## *Chapter I*

# DEMOGRAPHIC MODELS

### A. GENERAL BACKGROUND

Demographic models are an attempt to represent demographic processes in the form of a mathematical function or set of functions relating two or more measurable demographic variables. The primary purpose of modelling is simplification, to reduce a confusing mass of numbers to a few, intelligible basic parameters, or to make possible an approximate representation of reality without its complexity.

Because all demographic models attempt to represent reality, they are based to a greater or lesser extent on actual data. Yet, according to their degree of dependence upon observed data, two broad categories can be distinguished. On the one hand, there are models that can be derived solely from a set of simple assumptions or postulates. An example is the stable population model which arises from assuming that fertility and mortality have remained constant for a sufficiently long period. The proof of the convergence of almost any initial age distribution to a stable state after being subject for a long time to constant fertility and mortality can be carried out mathematically without recourse to any type of observed data.[1] Yet, a model of this sort is powerful only to the extent to which it reflects an actual process. Therefore, even though the stable population model is not derived from actual observations, only the good fit it provides to the age structures of populations that might reasonably be considered "stable" (i.e., having been subject to unchanging fertility and mortality for many years) establishes its value.

On the other hand, there are models that could not be derived at all if suitable data were not available. In this category are most of the model life tables that have been proposed. These models arise from the systematic analysis of observed mortality patterns and from the discovery and exploitation of common patterns present in them.

In general, most models fall between these two extremes; that is, their basis is neither purely empirical nor purely theoretical. In fact, some have evolved from a purely empirical to a purely theoretical foundation. An example of this type of evolution is the nuptiality model first presented by Coale.[2] It arose from the analysis of marriage rates among selected populations. Coale discovered that by suitably changing the spread, the location and the area under the nuptiality schedules of these populations, they could all be made to conform to the same standard shape. Further analysis showed that this shape could be closely approximated by a probability density function corresponding to the sum of a normal and several exponential random variables.

In probability theory, an exponentially distributed random variable is frequently used as a model of waiting-time periods. Hence, the density that approximates the standard nuptiality schedule suggests an appealing description of the process by which marriage takes place: women enter the "marriage market" (they become socially acceptable candidates for marriage) according to a normal distribution; and, once they reach this state, their actual time of entrance into marriage is just the sum of several waiting times (the time to find the right partner, the time to arrange a wedding etc.).

The description of this particular process of model creation seems interesting because it clearly illustrates how profitable the interaction between fact and theory may be. Besides, this model is especially appealing because it achieves an almost perfect amalgamation of reality and mathematical theory. Not only does the final equation arrived at describe the observed data with high accuracy, but once discovered, its interpretation appears most reasonable. The model proposed serves at the same time as a summary and as an explanation of reality. Ideally, every model should fulfil these two purposes.

Unfortunately, the modelling of other demographic phenomena has not been equally successful. Yet, undoubtedly, the aim of researchers in this field is to arrive at models that, while being as economical as possible in the number of parameters they incorporate, are also flexible enough to approximate all the relevant variations observable in real data, and whose form and parameters have meaningful interpretations in reality.

The next sections describe several useful demographic models in the areas of mortality, nuptiality, fertility and population age structure. Attention is focused on those models which have been relevant in the development of indirect estimation techniques, mainly because of the important part such models have played in allowing the simulation of demographic data. These simulated data have been used as a basis for the investigation of the relations existing between, on the one hand, quantities that are measured without difficulty and, on the other hand, more useful demographic parameters whose values are not easily determined directly.

---

[1] Alfred J. Lotka and F. R. Sharpe, "A problem in age-distribution", *Philosophical Magazine*, vol. 21, No. 124 (April 1911), pp. 435-438; and Ansley J. Coale, *The Growth and Structure of Human Populations: A Mathematical Investigation* (Princeton, New Jersey, Princeton University Press, 1972).

[2] Ansley J. Coale, "Age patterns of marriage", *Population Studies*, vol. XXV, No. 2 (July 1971), pp. 193-214.

## B. MORTALITY MODELS: MODEL LIFE TABLES

A life table provides a summary description of the effects of age-specific mortality rates upon a birth cohort. The very earliest demographic models attempted to describe in mathematical form the variations of mortality with age, particularly the increase in the risk of dying after childhood. Attempts to describe by a single mathematical function mortality experience throughout life have found it difficult to reproduce the characteristic U or J shape of mortality rates by age. This difficulty led to a new approach in creating mortality models or model life tables. Instead of trying to relate the risk of dying solely to age, risks at a given age were related to the risks observed at other ages or to risks observed in other populations at similar ages. Because the relations explored have, in general, shed no light on a plausible theoretical interpretation of how the process of mortality occurs, most of the model life tables existing to date depend heavily upon empirical data. At least four systems of model life tables have been developed on the principle of narrowing the choice of a life table to those deemed feasible on the basis of examination of mortality risks calculated for actual populations. These systems vary in the range of human experience they encompass, so that one may be more appropriate than another for a particular case. Each of them is described in detail below.

### 1. *United Nations model life tables*

The first set of model life tables was developed by the Population Division of the United Nations Secretariat during the 1950s.[3] This set, subsequently published in a revised form,[4] is based on a collection of 158 observed life tables for each sex. The model tables were constructed by assuming that the value of each ${}_5q_x$ (the probability of dying between age $x$ and $x+5$ in a life table) is a quadratic function of the previous $q$ value, namely, ${}_5q_{x-5}$ (except for the first two age groups, ${}_1q_0$ and ${}_4q_1$, all the other groups considered are five years in length). This assumption implies that the knowledge of only one mortality parameter (${}_1q_0$ or an equivalent "level" that indexes the ${}_1q_0$ values used) would determine a complete life table. The United Nations model life tables are thus a one-parameter system.

Since the coefficients of the quadratic equations relating each ${}_5q_x$ value to its predecessor were not known a priori, they had to be estimated on the basis of observed data. Regression was used to estimate these coefficients from the 158 mortality schedules available for each sex. Once they were estimated, calculation of the actual model life tables was straightforward: a convenient value of ${}_1q_0$ would be chosen arbitrarily; it would then be substituted in the equation relating ${}_1q_0$ to ${}_4q_1$ so that a value of ${}_4q_1$ would be obtained which, in turn, would be used to generate a value of ${}_5q_5$ through the equation relating ${}_5q_5$ to ${}_4q_1$, and so on.

[3] *Age and Sex Patterns of Mortality: Model Life Tables for Under-Developed Countries* (United Nations publication, Sales No. 55.XIII.9).

[4] *Manual III: Methods for Population Projections by Sex and Age* (United Nations publication, Sales No.56.XIII.3).

There are inherent disadvantages in the "chaining" technique used in calculating the United Nations model tables, especially when, as in this case, the quadratic equations relating one parameter to the next are not exact. Although the regression method of fitting equations to data does make some allowance for the existence of errors, it assumes that the distribution of these errors is known and that their mean is zero. The presence of systematic errors with non-zero mean (biases) may severely affect the estimates yielded by fitted regression equations, and their adverse effects are very likely to be augmented by the chaining technique described above whereby errors in one estimate would only contribute to accentuated errors in the next.

Furthermore, the 158 tables used as the data base from which the regression coefficients of the fitted equations were estimated were not all of the same quality. Because many tables from developing countries were used in the data base, mortality data with numerous deficiencies were included; and since large areas of the world did not possess any mortality statistics, life tables for those areas could not be included in the data base.

Owing to these shortcomings and to the fact that a one-parameter system lacks the flexibility required to approximate adequately the variety of cases encountered in reality, the United Nations life tables were soon superseded by other model sets which, though based on a similar approach, tried to avoid the pitfalls encountered by the creators of the first system. However, the United Nations set established the usefulness of a model life-table system.

### 2. *Coale and Demeny regional model life tables*

The Coale and Demeny[5] regional model life tables were published in 1966. They were derived from a set of 192 life tables by sex recorded for actual populations. That set included life tables from several time periods (39 of them relate to the period before 1900 and 69 refer to the period after the Second World War) and was heavily weighted towards Western experiences. Europe, Northern America, Australia and New Zealand contributed a total of 176 tables. The remaining 16 tables were from Africa and Asia: 3 from Israel; 6 from Japan; 3 from Taiwan Province; and 4 from the white populations of South Africa. The 192 life tables were chosen for inclusion from an original set of 326 tables. Life tables in which age patterns exhibited large deviations from the norm were excluded. All of the 192 life tables selected were derived from the registration data and from the complete enumeration of the populations to which they refer. Most tables covered entire countries, but a few representing the mortality experience of subregions also were included, especially when that experience showed distinctive characteristics that persisted over time.

A preliminary analysis of the tables revealed that four different mortality patterns could be distinguished

[5] Ansley J. Coale and Paul Demeny, *Regional Model Life Tables and Stable Populations* (Princeton, New Jersey, Princeton University Press, 1966).

among them. Those patterns were labelled "North", "South", "East" and "West" because of the predominance of European countries belonging to the various regions in each category. Hence, the adjective "regional" was applied to the entire set.

The countries whose life tables underlie each of the patterns identified and the outstanding characteristics of the patterns are discussed below:

(a) *East model.* The life tables underlying the East model come mainly from Austria, Germany (before 1900), the Federal Republic of Germany (after the Second World War), and northern and central Italy, although some from Czechoslovakia and Poland are also included. When the pattern of these tables is compared with the "standard pattern" (that exhibited by the majority of tables), their deviations from the standard follow a U shape, revealing their relatively high mortality rates in infancy and at older ages (over age 50). A total of 31 tables was used to estimate this model. The life expectancy in those tables ranges from a low of 36.6 years for Bavaria in 1878 to a high of 72.3 years for Czechoslovakia in 1958;

(b) *North model.* The observed life tables underlying the North model come from Iceland (1941-1950), Norway (1856-1880 and 1946-1955) and Sweden (1851-1890). Nine tables were used to derive this pattern of mortality, characterized by comparatively low infant mortality coupled with relatively high child mortality and by mortality rates above age 50 that fall increasingly below those of the standard. The populations displaying this mortality pattern were very probably subject to endemic tuberculosis (positive deviations from the standard pattern in the middle age range, from age 10 to 40, suggest this fact). Therefore, this model is recommended as an adequate representation of mortality in populations where the incidence of this disease is high. Life expectancy in these tables ranges from 44.4 years (Sweden, 1851-1860) to 74.7 years (Norway, 1951-1955);

(c) *South model.* The South model is based on life tables for Spain, Portugal, Italy, southern Italy and the region of Sicily, covering a period from 1876 to 1957. The levels of life expectancy range from 35.7 years (Spain, 1900) to 68.8 years (southern Italy, 1954-1957). A total of 22 tables was used in deriving this model. Their mortality pattern is characterized by high mortality under age 5, low mortality from about age 40 to age 60, and high mortality over age 65 in relation to the standard;

(d) *West model.* The West model is based on the residual tables, that is, those not used in the derivation of the other regional sets. Their mortality patterns do not deviate systematically from the standard pattern obtained when all the available life tables are put together; and, in this sense, they are closer to the standard than those on which other regional sets are based. Furthermore, because this model is derived from the largest number and broadest variety of cases, it is believed to represent the most general mortality pattern. For this reason, the West model is often recommended as a first choice to represent mortality in countries where lack of evidence prevents a more appropriate choice of model. It is of interest that the age pattern exhibited by the West model is very similar to that of the earlier United Nations life tables. Life expectancy in these tables ranges from 38.6 years (Taiwan Province, 1921) to 75.2 years (Sweden, 1959).

Having identified each of the four patterns present in the observed life tables, the coefficients of linear equations relating the values of ${}_nq_x$ to $e_{10}$, the expectation of life at age 10, and those relating the values of $\log_{10}({}_nq_x)$ to $e_{10}$ were estimated by using least-squares regression. From the equations thus established, it was simple to derive a complete set of ${}_nq_x$ values, and, therefore, a model life table, from any given value of $e_{10}$. The exact equation used (whether on ${}_nq_x$ or on $\log_{10}({}_nq_x)$) for each section of the age range changed according to some simple criteria;[6] but, essentially, the models derived from these equations depend, within each region, upon only one parameter, namely, $e_{10}$. Twenty-four values of $e_{10}$ for females were selected so that they would produce $e_0$ values ranging from 20 to 77.5 years, increasing in steps of 2.5 years.

The female $e_{10}$ values were then used to estimate $e_{10}$ values for males on the basis of the mortality differentials by sex present in the actual data. In this way, pairs of life tables (one for each sex) were generated for each level of mortality considered. For simplicity, the pair of life tables with a female $e_0$ of 20 was identified as level 1 and that with a female $e_0$ of 77.5 became level 24. An explanation of how the different functions that constitute a life table were calculated from the ${}_nq_x$ values can be found in the life tables themselves.[7]

In order to give the user a better idea of the relationships between the four different mortality patterns embodied by these models, figures 1-4 show plots of the proportional deviations of the ${}_nq_x$ values of models North, South and East from those of West. In all cases, the ${}_nq_x$ values compared refer to female level 9 (with an $e_0$ of 40 years). The exact function plotted in figures 1-4 is

$$Z(x) = \left[{}_nq_x / {}_nq_x^W\right] - 1.0$$

where the index $W$ indicates model West.

As figure 1 shows, model North is characterized by a somewhat lower infant mortality than that of West, by significantly higher child mortality (over the age range from 1 to 10) and by adult mortality (over the range from 20 to 50) that, although lower than that of West, is higher than the values associated with either South or East (see figure 4). It is also worth noting that at older ages (over 65) North is the model with the lowest mortality.

Model South (shown in figure 2) exhibits higher child mortality than West (over the range from 0 to 5), the

[6] *Ibid.*

[7] *Ibid.*

Figure 1. Relative deviations of North model values for the probability of dying, ${}_nq_x$, from those of the West model for females, level 9

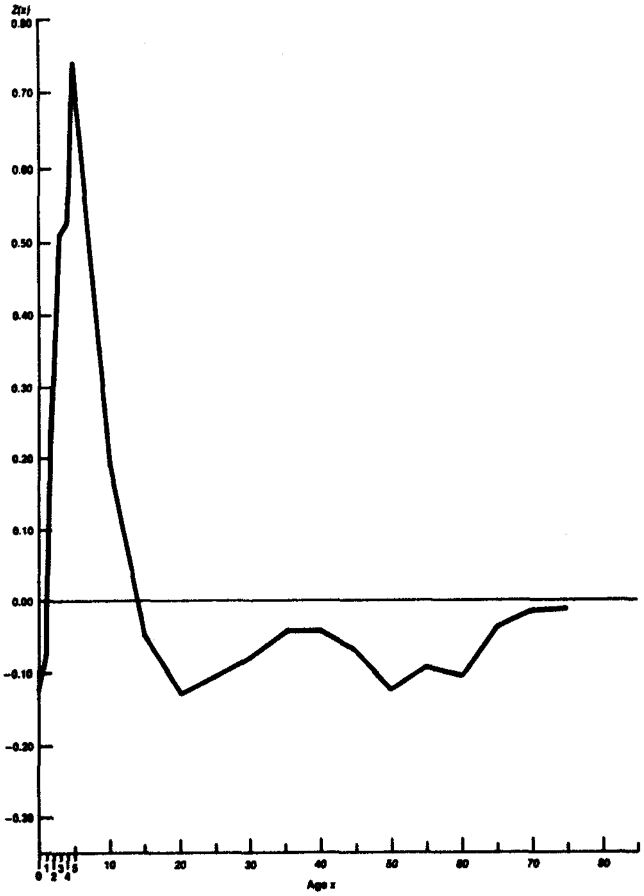
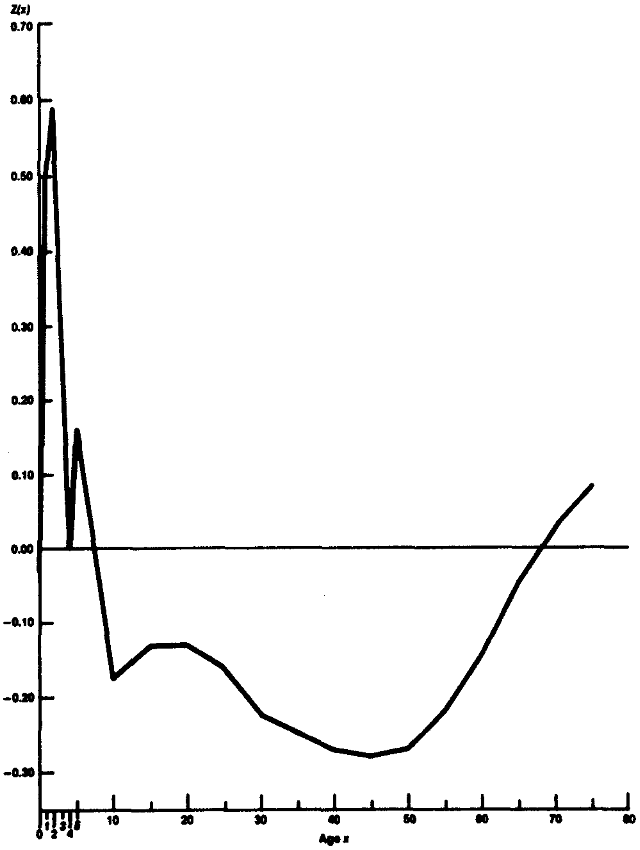



Figure 2. Relative deviations of South model values for the probability of dying, ${}_nq_x$, from those of the West model for females, level 9




lowest adult mortality among all models (see figure 4) and somewhat high mortality over age 65.

Model East (shown in figure 3) displays a relatively high infant mortality, coupled with the lowest child mortality in all models (see figure 4) and an adult mortality that falls between that of South and North. It is also characterized by having the highest mortality at older ages (over age 65).

Note should be taken that, as shown in figure 4, the model having the highest probabilities of dying at adult ages (20-50 years) is the West model (although when mortality is very low, the North model shows higher adult mortality); and only the East model is lower than the West model in terms of child mortality (probabilities of dying between ages 3 and 10).

At the time of its creation, this set of model life tables was probably the most general and flexible model of mortality available, and its widespread use has helped to demonstrate both its strengths and weaknesses. Among the latter, the most important is that, even though more varied than other sets of tables, the patterns embodied by the regional model life tables do not cover the entire range of human experience. For example, Demeny and Shorter[8] found that none of the four model patterns contained in the model life tables adequately reflects the Turkish experience. In general, because the data upon which the regional model life tables are based include very few examples from developing countries, it is possible that the patterns of mortality characteristic of these countries may be different from those represented by the models (see subsection B.5 for a discussion of the new United Nations model life tables).

However, owing to their long-standing use and acceptance, the Coale-Demeny life tables have become a necessary tool in indirect estimation; and they are consistently used in the worked examples presented in this *Manual*.

When employing these models, a problem that always arises is the selection of the mortality pattern that best represents the mortality prevalent in the country or region being studied. As described above, the regional model life tables contain four families or patterns of mortality. The most striking differences between these families are well known; and because the tables themselves are available, they can be investigated further by any analyst. Of course, if there are reasonable estimates of the mortality pattern for a given country, the best model can be selected by comparing the observed pattern to those embodied by the model tables. But when almost no reliable information on the incidence of mortality by age is available, the analyst can do little more than guess which pattern would be most appropriate.

[8] Paul Demeny and Frederic C. Shorter, *Estimating Turkish Mortality, Fertility and Age Structure* (Istanbul, Istanbul University, Statistics Institute, 1968).

**Figure 3. Relative deviations of East model values for the probability of dying, $_nq_x$, from those of the West model for females, level 9**

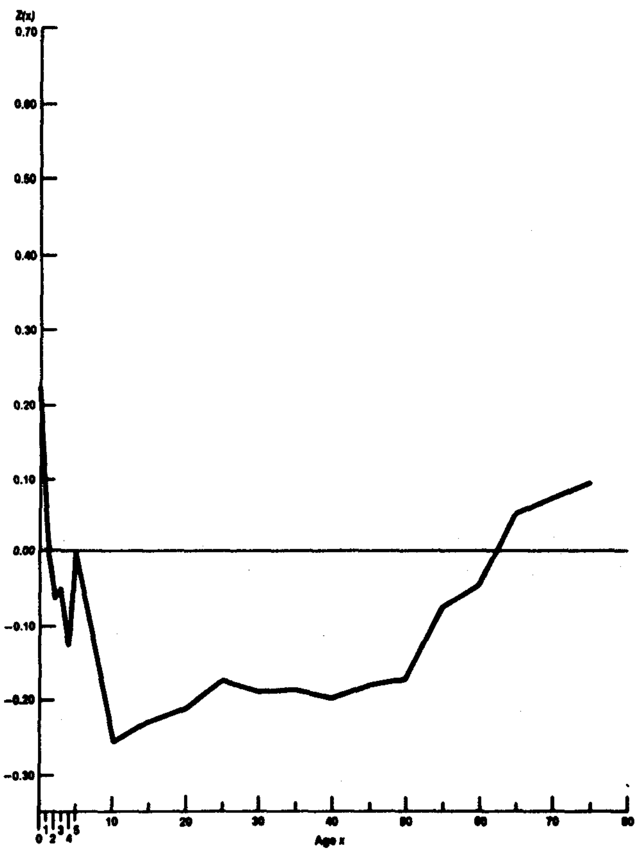
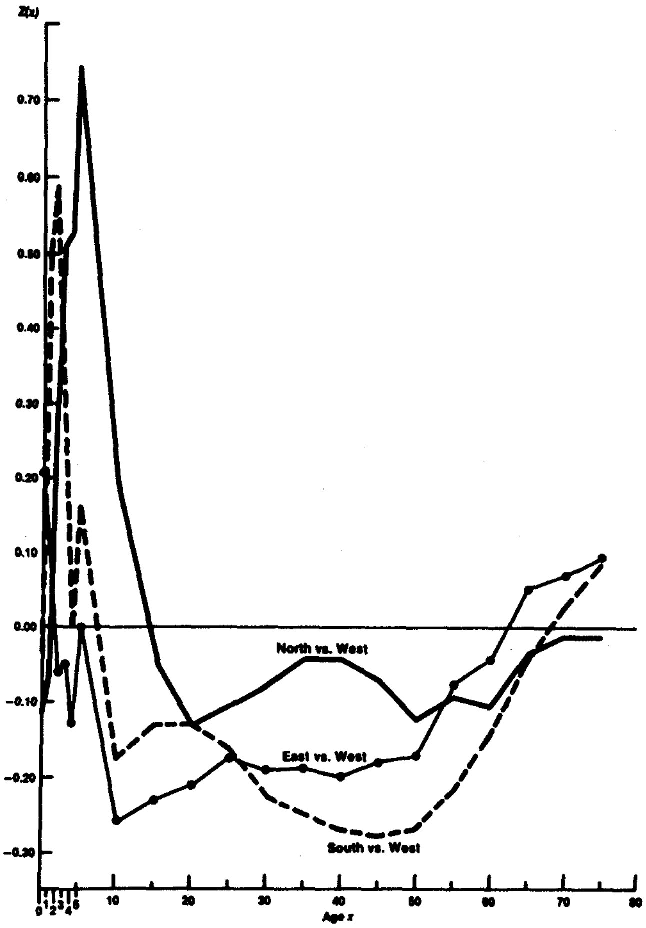



**Figure 4. Relative deviations of the North, South and East values for the probability of dying, $_nq_x$, with respect to those of the West model for females, level 9**




In situations where all extra information is lacking, the use of the West family is recommended because of the relatively wider data base from which it was derived. As for the "extra information" that would permit a more enlightened selection, it can vary considerably in type and quality, ranging, for example, from estimates of age-specific mortality rates derived from vital registration to knowledge of some fairly general facts, such as the extent to which breast-feeding is practised or the probable incidence of tuberculosis. When a set of observed age-specific mortality rates is available (preferably a set adjusted according to the methods described in chapter V), a model mortality pattern may be chosen by comparing the observed $_5m_x$ values to those corresponding to model life tables belonging to different families and whose levels (i.e., their expectation of life) roughly correspond to those of the observed rates. This comparison may be carried out by dividing the observed $_5m_x$ values by those of the model. Exact agreement between the two values would yield a quotient value of one. Therefore, the family whose quotient values are, on the average, closest to one could be selected as an appropriate representation of observed mortality. A slightly more sophisticated procedure would be to select the model set that minimizes the sum of the squared deviations from 1.0 of the ratios of observed $_5m_x$ over model $_5M_x$ values.

The selection of an adequate family is especially important when the indirect estimation of infant and child mortality is to be attempted. According to the description given above of the four mortality patterns contained in the regional model life tables, it is evident that they differ most markedly in their values at early ages and in the relation between infant ($_1q_0$) and child ($_4q_1$) mortality. It follows that quite different child mortality estimates may be obtained from the same information according to which family is selected as representative. Furthermore, in this case, reasonably sound external evidence that would permit a solid selection is very hard to obtain, mainly because infant deaths are very often grossly underreported. Currently, in the absence of adequate empirical data for selecting a family of model life tables, only the few general guide-lines given below may be proposed to narrow the possibilities and lead to a reasonable choice:

(*a*) In a population where breast-feeding is common practice and where weaning occurs at a relatively late age (12 months or over), one may reasonably expect child mortality ($_4q_1$) to be relatively higher than infant mortality ($_1q_0$) since breast-feeding may successfully prevent deaths due to malnutrition and infectious diseases among young infants. When weaning takes place, however, the child is less protected from these perils and is more likely to die. In these cases, mortality in childhood is likely to be well represented by the North family. Yet, it cannot be inferred from these observations that the North family will also provide an appropriate mortality model for other sections of the age range. More information on the incidence of mortality in adulthood is necessary to establish this fact;

(*b*) In some populations today, breast-feeding has been abandoned by a high proportion of the female population; and, from a very early age, infants are fed unsterilized and often inadequate rations of "milk formula". When this practice is adopted by women living in relatively unhealthy conditions, a higher degree of malnutrition and an increase of the incidence of infectious diseases among infants are observed. Under these conditions, a pattern of mortality similar to that of the South family may be a good representation of mortality in childhood;

(*c*) Early weaning may not be the only cause of malnutrition which results in a high child-to-infant mortality ratio. In some populations, breast-feeding is nearly universal but nutritional levels are low and both infant and child mortality are high. For such least developed countries, either the South or the North families of model life tables may be the most appropriate;

(*d*) If it is known that infant mortality is very high in relation to child mortality because of the prevalence of neonatal tetanus or some other cause, the East family may best reflect the actual age pattern of mortality;

(*e*) In the absence of data adequate to determine the most suitable family of model life tables to use for a particular country, one may select the same family as that employed for a neighbouring country with similar cultural and socio-economic characteristics;

(*f*) If little is known about the population under study, the West model is recommended, simply on the grounds of generality.

From these remarks, it is clear that the knowledge about mortality patterns is still fairly limited and that, certainly, better information concerning the mortality experience of populations in developing countries is needed to assess the adequacy of the models now available.

### 3. *Ledermann's system of model life tables*

Ledermann and Breas[9] used factor analysis to identify the most important variables or factors explaining the variation among a set of 154 observed life tables. The data base was nearly identical to that used for the earlier United Nations tables (see subsection B.1) and therefore had the same advantages and shortcomings.

Five factors were found to explain most of the variability among the observed tables. The first and largest is associated with a general mortality level; the second refers to the relation between child and adult mortality; the third is related to the pattern of mortality at older ages, while the fourth is related to the pattern of mortality under age 5; and lastly, the fifth reflects the differential between male and female mortality in the age range from 5 to 70.

At a later date, Ledermann[10] developed a series of one-parameter and two-parameter model life tables based on the regression analysis of the 154 actual life tables used in his first study of mortality patterns. The model life tables were obtained by estimating the probabilities of dying between ages $x$ and $x+5$, ${}_5q_x$, for males, females and both sexes combined, through logarithmic regression equations of the following type:

$$\ln {}_5q_x = a_0(x) + a_1(x) \ln Q \qquad \text{(B.1)}$$

for the one-parameter models, and

$$\ln {}_5q_x = b_0(x) + b_1(x) \ln Q_1 + b_2(x) \ln Q_2 \qquad \text{(B.2)}$$

for the two-parameter models, where $Q$, $Q_1$ and $Q_2$ are the independent variables used in each case, and $a_i(x)$ and $b_i(x)$ represent the estimated regression coefficients for the age group from $x$ to $x+5$.

Ledermann's models form a flexible system. While the early United Nations and the four Coale-Demeny regression models are based on just one independent variable (${}_1q_0$ and $e_{10}$, respectively), Ledermann estimated different sets of regression coefficients for equations (B.1) and (B.2), each based on a different independent variable or pair of variables. In the case of the one-parameter models, seven independent variables were used, namely: $e_0$, ${}_1q_0$, ${}_5q_0$, ${}_{15}q_0$, ${}_{20}q_{30}$, ${}_{20}q_{45}$ and $m_{50+}$ (the central mortality rate for ages 50 and over). The two-parameter models were obtained by using the following pairs of independent variables: ${}_5q_0$ and ${}_{20}q_{45}$; ${}_{15}q_0$ and ${}_{20}q_{30}$; and ${}_{15}q_0$ and $m_{50+}$. Every parameter refers to both sexes, except for ${}_{20}q_{30}$, which refers only to females. The use of different independent variables to generate each set of model life tables makes it easier for the user to avoid the bias introduced when a model life table is identified by way of an observed value that is not the independent variable used to generate the model. For example, this type of bias is introduced when, in the Coale-Demeny system, a life table is identified on the basis of the observed $l_2$ value rather than on the basis of the observed $e_{10}$. However, even though the Ledermann set does provide a wider variety of entry values that minimize the bias in identifying an appropriate model, in practice most of these values are not easily estimated with an adequate degree of accuracy for developing countries, so that the introduction of some bias cannot be avoided.

The Ledermann models also incorporate a feature absent in other tables. They provide not only the estimated values of the probabilities of dying but a measure of the dispersion of the observed values around the estimated values ($2\sigma_x$, where $\sigma_x$ is the standard error of the ${}_5q_x$ values estimated through a regression equa-

[9] Sully Ledermann and Jean Breas, "Les dimensions de la mortalité", *Population* (Paris), vol. 14, No. 4 (October-December 1959), pp. 637-682.

[10] Sully Ledermann, *Nouvelles tables-types de mortalité*, Travaux et Documents, Cahier No. 53 (Paris, Institut national d'études démographiques, 1969).

tion). Obviously, this measure refers only to the particular life tables from which the regression coefficients were calculated, and the former do not necessarily cover all possible situations. Nevertheless, the measures of dispersion presented do indicate the possible magnitude of the discrepancies between estimated and actual values.

In addition, the Ledermann tables reflect the sex differentials in age patterns of mortality and the way in which these differentials vary with respect to the overall level of mortality in actual life tables. Thus, for example, the effects of maternal mortality at high mortality levels are translated into a marked excess of female mortality in the early reproductive ages, but such excess disappears in tables of the same model corresponding to lower mortality levels.

This characteristic, however, may be a potential shortcoming of the system, because even though regression coefficients are given for the calculation of separate life tables for each sex, the independent variables used refer, with only one exception, to parameters obtained from data on both sexes combined. Thus, the user is forced to accept the relationships between male and female mortality that the model embodies, relationships that may not always be satisfactory. For instance, it is almost impossible to estimate from Ledermann's models a life table in which the male expectation of life exceeds that for females. When little is known about the sex differentials prevailing in a population, it is highly desirable to analyse data for each sex separately. On these grounds, the Ledermann tables are of limited value for the study of such populations. It may also be noted that, for applications to developing countries, the Ledermann system is not easy to use, as its independent variables, or points of entry into the tables, cannot be readily estimated by the indirect techniques currently available.

### 4. *Brass logit life-table system*

The main shortcoming of the model life-table systems described above is their dependence upon the type of data that generated them. The rather restricted data base used for this purpose and the fact that the model systems themselves consist of only a finite number of cases which cannot be expected to represent all possible human experience make them less than ideal. Another type of model is needed. Naturally, this model should adequately reflect the patterns found in empirical mortality data. However, it should not be constrained to represent exclusively the patterns these data embody for, as pointed out earlier, the true mortality experience of many populations has not yet been ascertained with any degree of accuracy, and it may or may not strictly conform to patterns observed in countries where accurate measurement has been possible.

A model that provides a greater degree of flexibility is that proposed by Brass and colleagues,[11] better known as the "logit system". Brass attempted to relate mathematically two different life tables. He discovered that a certain transformation of the probabilities of survival to age $x$ ($l(x)$ values in life-table terms) made the relationship between corresponding probabilities for different life tables approximately linear. In other words, if one lets $\lambda(l(x))$ represent some transformation of the value $l(x)$, for empirical data, the linear relationship:

$$\lambda(l^*(x)) = \alpha + \beta\lambda(l(x)) \qquad \text{(B.3)}$$

where $l^*(x)$ and $l(x)$ are two different life tables, and $\alpha$ and $\beta$ are constants, is approximately true for all vaues of $x$ if $\lambda$ is defined specifically as

$$\lambda(l(x)) = \text{logit}\,(1.0 - l(x)) =$$
$$0.5 \ln ((1.0 - l(x))/l(x)). \qquad \text{(B.4)}$$

Those familiar with the logit as defined in statistics will notice that $\lambda$ is just a special case of this function, being calculated for the probabilistic complement of $l(x)$ rather than for $l(x)$ itself, as would be usual practice in statistics where the logit of a probability $p$ is:

$$\text{logit}\,(p) = 0.5 \ln (p/(1.0 - p)). \qquad \text{(B.5)}$$

Assuming that, for any pair of life tables, values of $\alpha$ and $\beta$ can be found such that equation (B.3) is satisfied, it can be proved that the $\lambda$ transformation of any life table can be expressed as a linear function of the $\lambda$ transformation of some "standard" life table. That is, if equation (B.3) holds for every pair of life tables, all life tables can be generated from a single life table by changing the pairs of $(\alpha, \beta)$ values used. In fact, this proposition is not strictly true because the assumption made, namely, that (B.3) represents an exact relationship between life tables, is not entirely true. Equation (B.3) is only approximately satisfied by pairs of actual life tables, but the approximation is close enough to warrant the use of this relation to study and fit observed mortality schedules.

Before describing how equation (B.3) is used to generate model life tables, a word about the meaning of the parameters $\alpha$ and $\beta$ is in order. Consider the life tables $l^*(x)$ that can be generated by selecting a specific life table $l(x)$ and calculating $\lambda(l^*(x))$ for different values of $\alpha$ and $\beta$. If $\beta$ were to remain constant and equal to one, for instance, different values of $\alpha$ would produce life tables $l^*(x)$ whose shapes would essentially be the same as the $l(x)$ table used to generate them, but whose overall levels would change (see figure 5). If, on the other hand, $\alpha$ remains fixed and $\beta$ is allowed to vary, the resulting $l^*(x)$ life tables will no longer display the same shape as $l(x)$. All of the $l^*(x)$ tables will intersect at a single point located somewhere in the central portion of the age range. Therefore, their probabilities of survival will be either lower at younger ages and higher at older ages or lower at younger ages and higher at the older than the standard survival probabilities $l(x)$ from which they are generated (see figure 6). Hence, a changing

[11] William Brass and others. *The Demography of Tropical Africa* (Princeton, New Jersey, Princeton University Press, 1968).

Figure 5. Life tables derived through the logit system, letting $\beta = 1.0$ and using the Brass general standard

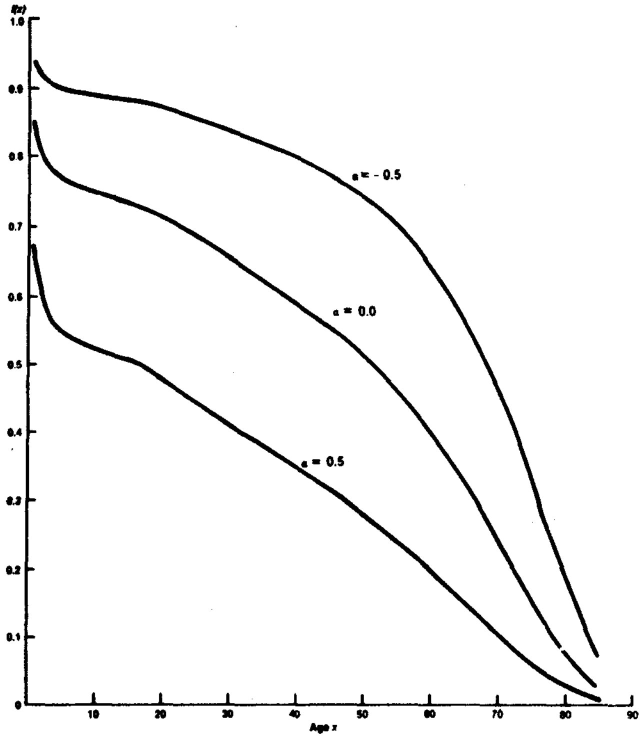
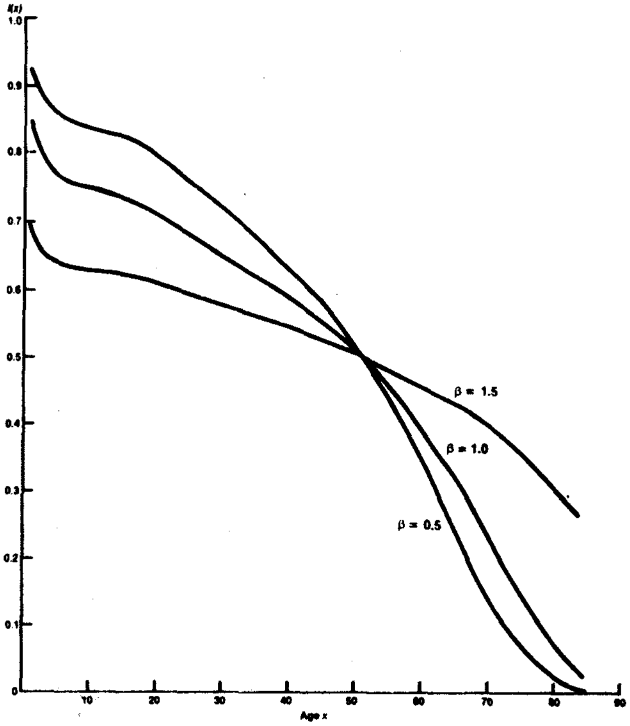


Figure 6. Life tables derived through the logit system, letting $\alpha = 0.0$ and using the Brass general standard



value of $\beta$ modifies the shape of the generated mortality schedule rather than its level. Naturally, simultaneous changes of $\alpha$ and $\beta$ will bring about changes in both the level and shape of the mortality schedule being generated.

From equations (B.3) and (B.4), the following expression can be derived:

$$l^*(x) = (1.0 + \exp(2\alpha + 2\beta\lambda(l(x))))^{-1} \qquad \text{(B.6)}$$

so that, for any set of $l(x)$ values defining a life table, another set $l^*(x)$ can be obtained by using any pair of $\alpha$ and $\beta$ values. (Note should be taken that at the end-points of the age range, where $l(x)$ is either 0 or 1, equation (B.6) cannot be used to calculate $l^*(x)$. Rather, $l^*(0)$ and $l^*(\omega)$ should be arbitrarily set equal to one and to zero, respectively.) Equation (B.6) can be used to generate model life tables simply by selecting an adequate standard. Potentially, any life table can be used as a standard, but for simulation and fitting purposes the standard proposed by Brass[12] is very frequently used. This "general" standard (presented in table 2) is different from the so-called "African" standard also proposed by Brass.[13] The latter is characterized by lower infant mortality and higher child mortality. In this *Manual*, only the Brass "general" standard is used.

Because of the mathematical simplicity of equations (B.4) and (B.6), the use of model life tables generated by means of the $\lambda$ transformation (also call the "logit" from here on) does not require that the resulting values be available in printed form. However, some such values have been printed. For example, Carrier and Hobcraft[14] produced a set of model tables from the African standard by fixing the value of $\beta$ at one. This set represents, therefore, a one-parameter system of model life tables.

The simple mathematical form of equation (B.6) also simplifies its use in computer applications. For this reason, life tables generated by the logit system are very often used for simulation purposes. Furthermore, the logit system is particularly appropriate for projecting mortality. If the past and current mortality schedules of a population are known, trends in the $\alpha$ and $\beta$ parameters can be determined by using the logit model life-table system to fit each mortality schedule, and with some caution these trends can be projected to generate estimates of future mortality.

In this *Manual*, the logit system is used to fit the adjusted mortality rates of a population and to synthesize independent estimates of child and adult mortality into coherent mortality schedules.

### 5. *United Nations model life tables for developing countries*

As more data of better quality have become available

[12] See K. Hill and J. Trussell, "Further developments in indirect mortality estimation", *Population Studies*, vol. XXXI, No. 2 (July 1977), pp. 313-333.

[13] W. Brass and others, *op. cit.*

[14] Norman H. Carrier and John Hobcraft, *Demographic Estimation for Developing Societies: A Manual of Techniques for the Detection and Reduction of Errors in Demographic Data* (London, London School of Economics, Population Investigation Committee, 1971).

TABLE 2. LOGIT VALUES FOR THE BRASS GENERAL STANDARD LIFE TABLE

| Age x (1) | Logit value λ(l(x)) (2) | Age x (3) | Logit value λ(l(x)) (4) | Age x (5) | Logit value λ(l(x)) (6) | Age x (7) | Logit value λ(l(x)) (8) | Age x (9) | Logit value λ(l(x)) (10) |
|---|---|---|---|---|---|---|---|---|---|
| 0.......... | — | 20.......... | −0.4551 | 40.......... | −0.1816 | 60.......... | 0.2100 | 80.......... | 1.2375 |
| 1.......... | −0.8670 | 21.......... | −0.4401 | 41.......... | −0.1674 | 61.......... | 0.2394 | 81.......... | 1.3296 |
| 2.......... | −0.7152 | 22.......... | −0.4248 | 42.......... | −0.1530 | 62.......... | 0.2701 | 82.......... | 1.4284 |
| 3.......... | −0.6552 | 23.......... | −0.4103 | 43.......... | −0.1381 | 63.......... | 0.3204 | 83.......... | 1.5346 |
| 4.......... | −0.6219 | 24.......... | −0.3963 | 44.......... | −0.1229 | 64.......... | 0.3364 | 84.......... | 1.6489 |
| 5.......... | −0.6015 | 25.......... | −0.3829 | 45.......... | −0.1073 | 65.......... | 0.3721 | 85.......... | 1.7722 |
| 6.......... | −0.5879 | 26.......... | −0.3686 | 46.......... | −0.0911 | 66.......... | 0.4097 | 86.......... | 1.9053 |
| 7.......... | −0.5766 | 27.......... | −0.3549 | 47.......... | −0.0745 | 67.......... | 0.4494 | 87.......... | 2.0493 |
| 8.......... | −0.5666 | 28.......... | −0.3413 | 48.......... | −0.0574 | 68.......... | 0.4912 | 88.......... | 2.2051 |
| 9.......... | −0.5578 | 29.......... | −0.3280 | 49.......... | −0.0396 | 69.......... | 0.5353 | 89.......... | 2.3740 |
| 10.......... | −0.5498 | 30.......... | −0.3150 | 50.......... | −0.0212 | 70.......... | 0.5818 | 90.......... | 2.5573 |
| 11.......... | −0.5431 | 31.......... | −0.3020 | 51.......... | −0.0021 | 71.......... | 0.6311 | 91.......... | 2.7564 |
| 12.......... | −0.5365 | 32.......... | −0.2889 | 52.......... | 0.0177 | 72.......... | 0.6832 | 92.......... | 2.9727 |
| 13.......... | −0.5296 | 33.......... | −0.2759 | 53.......... | 0.0383 | 73.......... | 0.7385 | 93.......... | 3.2079 |
| 14.......... | −0.5220 | 34.......... | −0.2627 | 54.......... | 0.0598 | 74.......... | 0.7971 | 94.......... | 3.4639 |
| 15.......... | −0.5131 | 35.......... | −0.2496 | 55.......... | 0.0821 | 75.......... | 0.8593 | 95.......... | 3.7424 |
| 16.......... | −0.5043 | 36.......... | −0.2364 | 56.......... | 0.1055 | 76.......... | 0.9255 | 96.......... | 4.0456 |
| 17.......... | −0.4941 | 37.......... | −0.2230 | 57.......... | 0.1299 | 77.......... | 0.9960 | 97.......... | 4.3758 |
| 18.......... | −0.4824 | 38.......... | −0.2094 | 58.......... | 0.1554 | 78.......... | 1.0712 | 98.......... | 4.7353 |
| 19.......... | −0.4694 | 39.......... | −0.1956 | 59.......... | 0.1821 | 79.......... | 1.1516 | 99.......... | 5.1270 |

for less developed countries it has become evident that the age patterns of mortality exhibited by their populations often differ from those recorded in the developed countries during the period 1850-1960, and consequently from those embodied in the models described above. For this reason, the Population Division of the Department of International Economic and Social Affairs of the United Nations Secretariat prepared and recently published a set of model life tables based on data from developing countries.[15]

Although the availability and reliability of data from the less developed regions have increased remarkably since the publication of the early United Nations life tables in 1955 and the Coale-Demeny tables in 1966, such data remain essentially poor. As a result, careful evaluation, selection and adjustment procedures were used to construct the data base upon which the new United Nations models were constructed. This data base consists of 36 life tables by sex (72 in total), covering a wide range of mortality levels (for 10 life tables, $e_0$ is below 50 years; and for another 10, $e_0$ is 70 years or greater). Geographically, 16 pairs of male/female life tables were obtained from 10 countries of Latin America, 19 from 11 countries of Asia, and one from Africa. The meagre contribution by Africa reflects both a shortcoming and a strength of this system—a shortcoming in that the lack of data from Africa casts doubts about the use of the resulting models to represent the experience of this major area, a strength in that the rejection of whatever data were available for sub-Saharan Africa, usually of extremely low quality, is to some extent a validation of the evaluation procedures used to select the data base and a partial assurance that the empirical tables underlying these models are of generally high quality so that the models themselves reflect distinctive patterns of mortality in developing countries rather than typical patterns of data flaws.

The new United Nations model life tables are similar to the Coale and Demeny set in that distinct patterns of age-specific mortality schedules have been identified and are published in detail. In addition, the new models incorporate a greater degree of in-built flexibility, allowing the user to construct mortality patterns different from those actually published. In this sense, they are more akin to the logit system proposed by Brass.

Four distinct patterns of mortality were identified on the basis of the data available. Because of the predominance of these patterns in certain geographical regions, they are identified in regional terms as the "Latin American", the "Chilean", the "South Asian" and the "Far Eastern" patterns.[16] A fifth pattern, called the "general" pattern, was constructed as the overall average of those listed above.

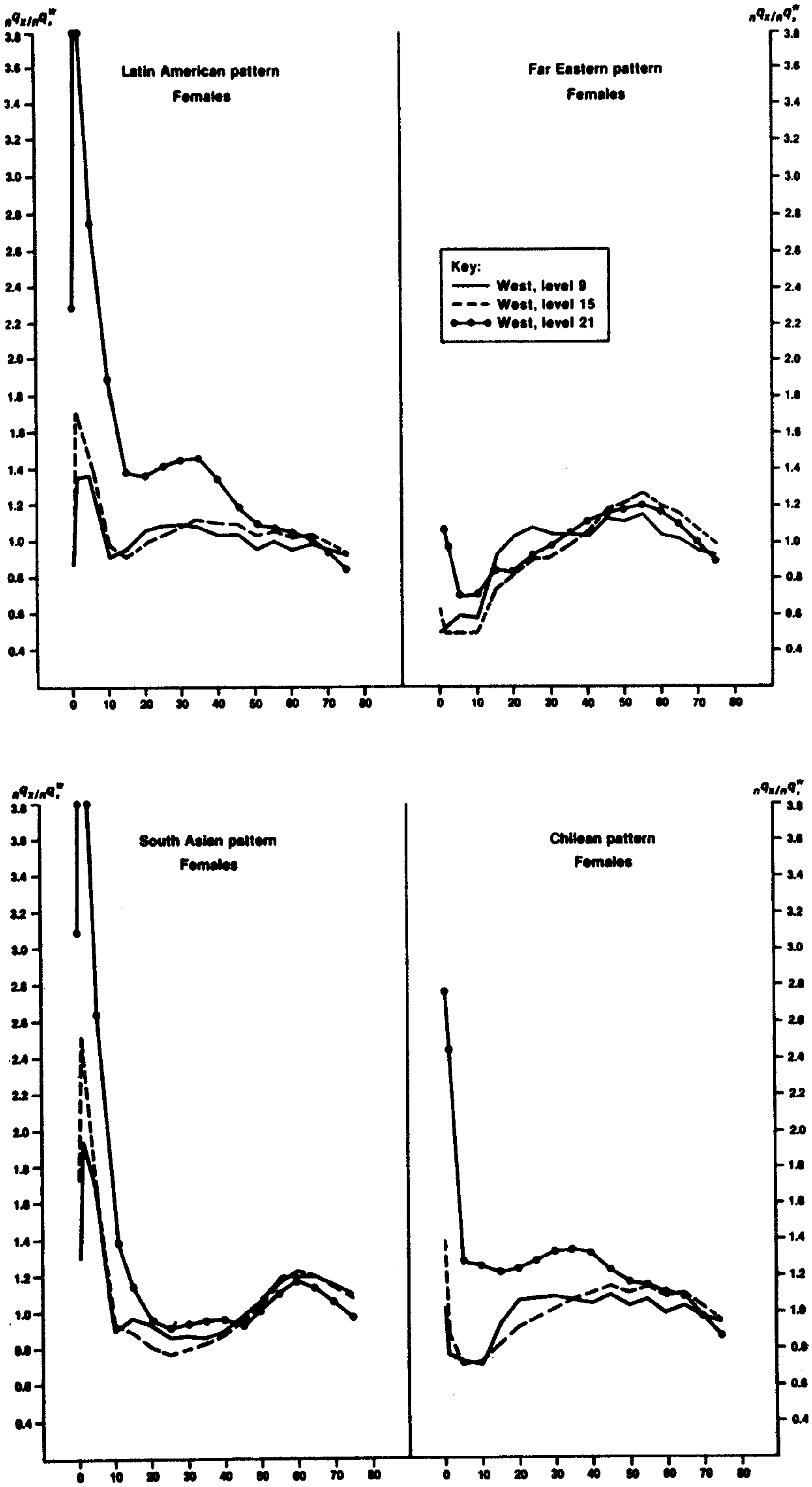
In order to illustrate the differences between the Coale and Demeny models and the United Nations patterns, figure 7 shows plots of the ratios ${}_nq_x / {}_nq_x^W$, where the superscript $W$ indicates the West model in the Coale-Demeny set. Comparisons are made between ${}_nq_x$ values corresponding to life tables with the same life expectancy at age 10. Levels 9, 15 and 21 of the Coale-Demeny tables for females are used as denominators. These comparisons highlight the most distinctive characteristics of the new United Nations patterns, which are discussed below.

The Latin American model, when compared with the West family of the Coale and Demeny models, exhibits high mortality during the infant and childhood years (due mainly to excess diarrhoeal and parasitic diseases), and again during the young adult ages (largely due to

[15] *Model Life Tables for Developing Countries* (United Nations publication, Sales No. E.81.XIII.7).

[16] The labelling used for these patterns does not relate to the geographical divisions established by the Population Division of the Department of International Economic and Social Affairs of the United Nations Secretariat.

**Figure 7. Comparison of the mortality patterns in the United Nations model life tables for developing countries with the West family of the Coale and Demeny set**




*Note:* The labelling used for these patterns does not relate to the geographical divisions established by the Population Division of the Department of International Economic and Social Affairs of the United Nations Secretariat.

accidents). It also exhibits relatively low levels at older ages, apparently because of comparatively low death rates due to cardio-vascular diseases.

The Chilean pattern is characterized mainly by extremely high infant mortality in relation to both the West family and its own child mortality. This excessive infant mortality appears to be due mainly to deaths from respiratory diseases and may also be related to early weaning.

The South Asian pattern in typified by extremely high mortality under age 15 and relatively high mortality at older ages (over age 55 approximately). Correspondingly, mortality during the prime adult ages is relatively low. Data about causes of death are scarce in this region; however, those gathered by the International Diarrhoeal Disease Research Centre in Matlab, Bangladesh, and by the Indian Model Registration Project reveal high rates of diarrhoeal and parasitic diseases at younger ages and high mortality from diarrhoeal and respiratory diseases at older ages.

The Far Eastern pattern exhibits very high death rates at the older ages compared with those at younger ages. There is some evidence that this distinctive pattern may be due to a past history of tuberculosis.

The general pattern (not shown), which can be considered an average of the four regional patterns described above, is very similar to the West family of the Coale-Demeny set.

As mentioned earlier, these tables combine the advantages of the Coale and Demeny regional system and its detailed publication format with the type of flexibility inherent in the Brass logit system. Such characteristics were achieved by using principal component analysis to construct each model, after preliminary clustering of the data had been carried out. Clustering allowed the identification of the four distinctive patterns described above. Within each cluster equations of the form

$$\text{logit}({}_nq_x) = U^c_{0x} + \sum_{i=1}^{k} a_i U^c_{ix} \tag{B.7}$$

were fitted, where ${}_nq_x$ is the observed probability of dying between ages $x$ and $x+n$; $U^c_{0x}$ is the overall average (in logit terms) for cluster $c$; $U^c_{ix}$ represents the characteristic deviation of the observed from the average; and the coefficient $a_i$ indicates the size of such deviations. Because the fitting procedure used identifies the $U_{ix}$ values as the principal components of the observed ${}_nq_x$ vectors (with $x$ ranging from 0 to $\omega$), the $U_{0x}$ vector can be interpreted as a measure of the average mortality model within each cluster, while $U_{1x}$ may be interpreted as a measure of the typical deviations from that average as mortality levels change. Deviations from the overall average not due purely to changes in level are embodied by the second and third principal components, $U_{2x}$ and $U_{3x}$, respectively. Hence, in constructing the one-parameter models printed in the tables, equation (B.7) with $k=1$ was used, but setting $k=2$ or $k=3$ and selecting $a_i$ values judiciously allows the user to derive mortality schedules whose pattern deviates from that of the printed tables. In this way, the flexibility of the models is enhanced.

In spite of their qualities, the United Nations life tables for developing countries are not used in this *Manual*, mainly because they were not yet available at the time of its preparation. In addition, because some of the methods described in the remainder of this volume have been developed on the basis of the Coale-Demeny models, it would be unfair to judge their performance on the basis of new models. The development of variations of these methods specifically designed for use with the new United Nations models is currently being undertaken and will soon add to the tools available for the analysis of demographic data in the developing countries.

## C. Model stable populations

The concept of a stable population was first formulated by Lotka,[17] who proved that almost any population that is subject to constant fertility and mortality for a sufficiently long time acquires ultimately an unchanging age distribution which is characteristic of the prevailing fertility and mortality rates, and which is independent of the initial age distribution. He called the end-product of such constant fertility and mortality conditions a stable population and established that the stable age distribution has the following form:

$$c(x) = b\,\exp(-rx)l(x) \tag{C.1}$$

where $c(x)$ is the infinitesimal proportion of the stable population at exact age $x$ ($c(x)$ is actually a density function), $b$ is the constant birth rate, $r$ is the constant rate of natural increase and $l(x)$ is the probability of surviving from birth to age $x$ (the usual life-table function).

Using equation (C.1) and recalling that an integral can be interpreted as a sum of infinitesimals, one can deduce that the proportion under age $y$, $C(y)$, is

$$C(y) = \int_0^y b\,\exp(-rx)l(x)dx \tag{C.2}$$

and, since for the highest age possible, denoted by $\omega$, $C(\omega)$ must equal one (that is, the portion of the population under the highest age attainable must be the entire population), it follows that

$$\int_0^\omega b\,\exp(-rx)l(x)dx = 1.0 \tag{C.3}$$

or

$$b = \left[\int_0^\omega \exp(-rx)l(x)dx\right]^{-1} \tag{C.4}$$

Equations (C.2) and (C.4) allow the computation of the age distribution of a stable population whenever its mortality schedule ($l(x)$) and its growth rate ($r$) are known.

[17] A. J. Lotka and F. R. Sharpe, *loc. cit.*

Model stable populations arise from the use of model mortality schedules or life tables to generate, through equations (C.2) and (C.4), stable age distributions for selected values of $r$. Among the five types of model life tables discussed above, at least four have been used to generate model stable populations. The Coale-Demeny set has probably been the most widely used for the purpose of demographic estimation (cf. chapter VII). Four families of model stable populations are included in this set: one for each of the patterns of mortality it contains. Stable age distributions are printed for each sex separately, for values of $r$ ranging from $-0.01$ to 0.05 (increasing in steps of 0.005 at a time), for values of the gross reproduction rate ranging from 1.0 to 6.0 (increasing in steps of 0.25) and for the 24 mortality levels of each family. Each age structure is accompanied by a series of other parameter values corresponding to the stable population it represents. These values include the death and birth rates and the gross reproduction rate. The degree of detail in which this set is tabulated makes it rather simple to use.

Another set of stable populations was published by Carrier and Hobcraft,[18] who computed stable age distributions based on life tables calculated by using the logit system with the African standard. Two types of life tables are presented: one in which $\beta$ is held constant with a value of one, and another where both $\alpha$ and $\beta$ are allowed to change. This set of stable populations is far less detailed than the Coale-Demeny set, but its use for estimation purposes may be required in situations where none of the Coale-Demeny mortality patterns is judged to approximate adequately the mortality experienced by an actual population.

The United Nations also published a set of model stable populations that can be used for demographic estimation, especially in cases where the other sets are not able to provide an acceptable fit.[19] Lastly, the United Nations has recently published a new set of model stable populations corresponding to its new model life tables.[20] Stable populations are presented for each of the five United Nations mortality patterns, for growth rates from 0.0 to 4.0 per cent by 0.5 per cent increments and for life expectancies at birth from 35 to 75 years by one-year increments. Intrinsic birth and death rates are also presented.

## D. Nuptiality models

As mentioned in section A of this chapter, while studying first-marriage frequencies (the number of first marriages taking place in a short age interval divided by the number of persons in that interval) in different female populations (or, more precisely, in different female cohorts), Coale[21] discovered that they could all be made to conform to a common standard pattern. The observed curves of first-marriage frequencies by age differed from one another only in the age at which marriage began, the rate at which marriages took place and the ultimate proportion that ever married. Therefore, a transformation of their origin, and of the horizontal and vertical scales, was all that was required to make them conform to a standard. Once a common pattern was discovered among the transformed distributions, a standard was calculated on the basis of period data from Sweden (1865-1869). The availability of this standard permitted the calculation of an empirical risk function $R(x)$, whose values quantify the risk of marrying for the first time at each age. By trial and error, Coale discovered that a good fit to $R(x)$ was obtained by the double exponential function

$$R_s(x) = 0.174 \exp(-4.411 \exp(-0.309x)). \quad \text{(D.1)}$$

It was later proved[22] that this standard risk function is closely approximated by the density function associated with an infinite sum of independent, exponentially distributed random variables. The exact form of this standard density function, $g_s(x)$, is

$$g_s(x) = 0.1946 \exp(-0.174(x-6.06) - \exp(-0.288(x-6.06))). \quad \text{(D.2)}$$

The fit of this function to the empirical standard is just as good as that of a density corresponding to the sum of a normally distributed random variable and several independent, exponential delays, and equation (D.2) was adopted as a model because it is easier to handle mathematically.

If the effect of differential mortality and migration by marital status on the proportion of women who have ever been married is neglected, the existence of a standard curve of first-marriage frequencies implies the existence of a standard curve describing the proportions ever married by age for any given cohort. The shape of this curve is standard but, naturally, there are differences in the beginning age at marriage (the age at which marriages begin taking place among members of a cohort), in the pace at which the curve rises and in the ultimate proportion getting married (the proportion ever married by the age at which first-marriage rates have fallen essentially to zero). If one denotes by $G_s(x)$ the standard proportion ever married $x$ years after marriages begin, then $G(a)$, the proportion married by age $a$ in some true cohort, can be expressed as

$$G(a) = \Theta G_s((a-a_0)/\gamma) \quad \text{(D.3)}$$

where $\Theta$ is the ultimate proportion ever married, $a_0$ is the age at which first marriages begin and $\gamma$ is the scale

[18] N. H. Carrier and J. Hobcraft, *op. cit.*

[19] *The Concept of a Stable Population: Application to the Study of Populations of Countries with Incomplete Demographic Statistics,* Population Studies No. 39 (United Nations publication, Sales No. 65.XIII.3).

[20] *Stable Populations Corresponding to the New United Nations Model Life Tables for Developing Countries* (ST/ESA/SER.R/44).

[21] A. J. Coale, "Age patterns of marriage".

[22] Ansley J. Coale and Donald R. McNeil, "The distribution by age of the frequency of first marriage in a female cohort", *Journal of the American Statistical Association,* vol. 67, No. 340 (December 1972), pp. 743-749.

factor expressing the number of years of nuptiality in the given population which are equivalent to one year in the standard population.

Since, mathematically, $G_s(x)$ is just the integral of $g_s(x)$,

$$G_s(x)=\int_0^x g_s(y)dy, \qquad \text{(D.4)}$$

the density $g(a)$ associated with the cohort described above can be obtained from equations (D.3) and (D.4) by making a change of variable in the latter and substituting in the former. The result implies that $g(a)$ has the form

$$g(a)=(0.1946\Theta/\gamma)\exp[(-0.174/\gamma)(a-a_0-6.06\gamma) - \exp[(-0.288/\gamma)(a-a_0-6.06\gamma)]]. \qquad \text{(D.5)}$$

No analytical experssion for $G(a)$ has been found, but its value for any age $a$ can be calculated by the numerical integration of $g(a)$, since

$$G(a)=\int_{a_0}^a g(y)dy. \qquad \text{(D.6)}$$

The problem of fitting a model of the type defined by equation (D.5) to an actual population consists of identifying values of the parameters $\Theta, a_0$ and $\gamma$ that adequately reflect the experience of the population in question. Approximate values for these parameters may be estimated from the knowledge of the proportions single classified by age. Normally, the proportion single in age group 50-54 may be considered an estimate of $U(\omega)$, the proportion who never marry. Therefore, the proportion who will eventually marry, $\Theta$, may be estimated by:

$$\Theta=1-U(\omega). \qquad \text{(D.7)}$$

Furthermore, it is known that the mean of the fitted first-marriage schedule is $a_0+11.37\gamma$. This mean, also known as the "singulate mean age of marriage", *SMAM*, can be estimated from the proportions single classified by age by a simple procedure first proposed by Hajnal[23] and described in detail in annex I. Since

$$SMAM=a_0+11.37\gamma \qquad \text{(D.8)}$$

then

$$\gamma=(SMAM-a_0)/11.37 \qquad \text{(D.9)}$$

so that if both *SMAM* and $a_0$ are known, $\gamma$ can be easily estimated. The estimation of $a_0$ is usually carried out in a very rough way. It is assumed that marriages in most developing countries begin early; and unless there is evidence to the contrary, $a_0$ is selected to be 13 or 14 years. Values of 12 or 15 may also be used. The exact value of $a_0$ is generally not crucial. Equation (D.9) can then be used to estimate $\gamma$. Once values for $\Theta, a_0$ and $\gamma$ have been estimated, Coale[24] provides tables that make the calculation of first-marriage frequencies and of the proportion ever married by age a simple matter.

To conclude, it may be mentioned that although, strictly speaking, model (D.5) only describes the first-marriage experience of a cohort, in practice, the proportions of women who have ever been married observed during a given period may also be closely approximated by this model, particularly if marriage patterns have remained constant, but also, and more surprisingly, in cases where marriage patterns have been changing.

## E. FERTILITY MODELS

### 1. *Coale and Trussell model*

Louis Henry[25] discovered that in populations where there is little or no voluntary control of fertility the age pattern of fertility within marriage is approximately constant. According to Henry, voluntary control is any behaviour affecting fertility that is modified as parity increases. He called "natural fertility", $h(x)$, the fertility observed in the absence of control and was able to infer its general pattern from the study of several populations where voluntary fertility control was presumed absent. In these populations, the level of natural fertility varied, but its age pattern remained the same. Henry attributed level variations between populations to differences in overall health, the practice of breast-feeding and any other physical or social factors that might affect the marital fertility experience of women irrespective of their parity.

In 1974, Coale and Trussell[26] proposed a model that, by generalizing the pattern of natural fertility, was able to represent the fertility experience of populations where voluntary fertility control was exercised. This model is based on the following assumption: marital fertility either follows natural fertility (if deliberate birth control is not practised); or it departs from natural fertility in a way that increases with age according to a typical pattern. Therefore, if one denotes by $\phi(x)$ marital fertility at age $x$ and by $h(x)$ natural fertility at the same age, in a population where fertility is controlled voluntarily,

$$\phi(x)=Mh(x)\delta(x) \qquad \text{(E.1)}$$

where $M$ is a parameter indicating the level of natural fertility that the population would experience in the absence of all voluntary control and $\delta(x)$ is a function of age indicating the typical pattern of departure from natural fertility when voluntary control is exercised.

By examining the function $\delta(x)$ calculated for several

23 John Hajnal, "Age at marriage and proportions marrying", *Population Studies*, vol. VII, No. 2 (November 1953), pp. 111-136.

24 A. J. Coale, "Age patterns of marriage"

25 Louis Henry, "Some data on natural fertility", *Eugenics Quarterly*, vol. VIII, No. 2 (June 1961), pp. 81-91.

26 Ansley J. Coale and T. James Trussell, "Model fertility schedules: variations in the age structure of childbearing in human populations", *Population Index*, vol. 40, No. 2 (April 1974), pp. 185-258.

populations, Coale and Trussell[27] found that it could be represented by

$$\delta(x) = \exp(m\,\nu(x)) \qquad \text{(E.2)}$$

where the function of $\nu(x)$ was very nearly the same for different populations and the parameter $m$ changed from population to population. They interpreted these results as meaning that $\nu(x)$ represents the typical pattern of deviation from natural fertility when deliberate birth control is practised, while $m$ measures the degree to which this control is practised. The final model of marital fertility arrived at is appealing both mathematically and theoretically. It can easily be derived from equations (E.1) and (E.2) and has the form

$$\phi(i) = Mh(i)\exp(m\,\nu(i)) \qquad \text{(E.3)}$$

where the index $i$ is used in place of age $x$ to indicate that, in general, only data referring to five-year age groups are used. Values of the two functions $h(i)$ and $\nu(i)$ have been estimated, and are shown in table 3.

TABLE 3. STANDARD PATTERN OF NATURAL FERTILITY AND OF DEVIATIONS FROM NATURAL FERTILITY, BY AGE GROUP, FOR THE COALE AND TRUSSELL FERTILITY MODEL

| *Age group (1)* | *Index i (2)* | *Natural fertility h(i) (3)* | *Deviation pattern from natural fertility ν(i) (4)* |
|---|---|---|---|
| 15-19 ............. | 1 | 0.411 | 0.000 |
| 20-24 ............. | 2 | 0.460 | 0.000 |
| 25-29 ............. | 3 | 0.431 | −0.279 |
| 30-34 ............. | 4 | 0.395 | −0.667 |
| 35-39 ............. | 5 | 0.322 | −1.042 |
| 40-44 ............. | 6 | 0.167 | −1.414 |
| 45-49 ............. | 7 | 0.024 | −1.671 |

Therefore, this model can be fitted to any population whose marital fertility rates are known by just identifying the values of the parameters $M$ and $m$. Coale and Trussell[28] suggest two possible ways of estimating $M$ and $m$. According to the first and simplest procedure,

$$M = \phi(2)/h(2) \qquad \text{(E.4)}$$

and

$$m = 0.2\sum_{i=3}^{7} \ln[\phi(i)/Mh(i)]/\gamma(i). \qquad \text{(E.5)}$$

That is, the level of marital fertility $M$ is defined by the relation between the observed marital and natural fertility rates at ages 20-24, at which ages voluntary control of fertility is deemed to have no effect on pattern; while $m$ is just the mean of the $m(i)$ values that can be estimated directly from the observed $\phi(i)$ once $M$ is known.

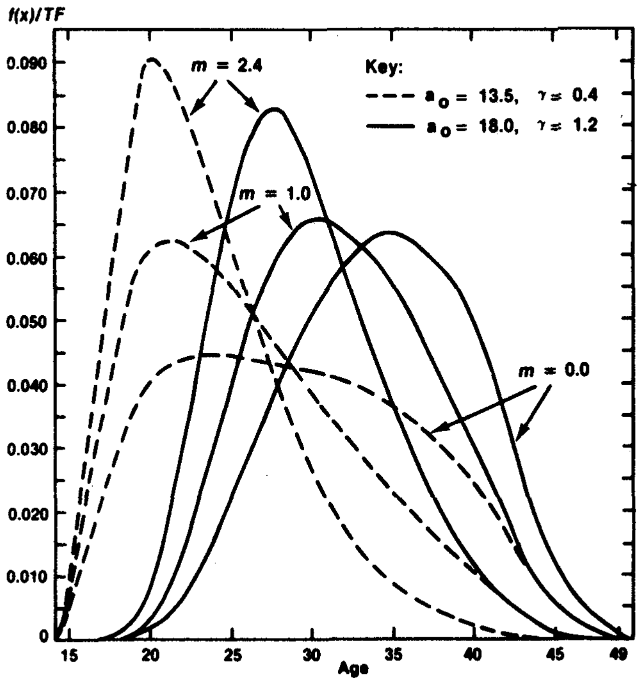
The second approach makes use of the following re-expression of (E.3):

$$\ln(\phi(i)/h(i)) = \ln(M) + m\,\nu(i) \qquad \text{(E.6)}$$

which clearly shows that the quantities $\ln(\phi(i)/h(i))$ and $\nu(i)$ are linearly related. Therefore, $\ln(M)$ and $m$ can be estimated by identifying the line that best fits the observed $[\ln(\phi(i)/h(i)), \nu(i)]$ points. Coale and Trussell[29] recommend the use of the least-squares method to fit this line and suggest that only the points corresponding to age groups from 20-24 to 40-44 ($i$ ranging from 2 to 6) should be considered.

Note that the values of the standard function $\nu(i)$ indicating the pattern of deviations from natural fertility (cf. table 3) are all either negative or zero. Furthermore, the absolute value of $\nu(i)$ increases as age increases. Therefore, a positive value of $m$ indicates that in the population being studied marital fertility falls increasingly short of natural fertility as age increases (see figure 8). A negative value of $m$ would indicate, on the contrary, that marital fertility is increasingly higher than natural fertility as age advances. If $m = 0$, natural fertility and the marital fertility under observation follow the same pattern.

**Figure 8. Coale and Trussell fertility schedules: combinations of early marriage and various degrees of fertility control; and late marriage with the same degrees of fertility control**




Model (E.3) can be used not only to investigate the pattern of marital fertility that a population experiences but to generate model marital fertility schedules that can be very useful for simulation purposes. Yet, since in

[27] *Ibid.*

[28] *Ibid.*; and Ansley J. Coale and T. James Trussell, "Technical note: finding the two parameters that specify a model schedule of marital fertility", *Population Index*, vol. 44, No. 2 (April 1978), pp. 202-213.

[29] A. J. Coale and T. J. Trussell, "Technical note: finding the two parameters that specify a model schedule of marital fertility".

many instances the function of interest is not marital fertility but overall fertility, the marital fertility model (E.3) may be combined with the nuptiality model described earlier (in section D) so that age-specific fertility, $f(x)$, can be obtained as

$$f(x) = G(x)\phi(x) \qquad \text{(E.7)}$$

where $G(x)$ is the proportion married by age $x$ defined by equations (D.5) and (D.6).

Model (E.7) is a five-parameter model of overall fertility. Its parameters are: $\Theta$, the ultimate proportion that ever marries; $a_0$, the age at which marriage begins; $\gamma$, the pace at which marriage takes place; $M$, the overall level of marital fertility; and $m$, the degree of departure from natural fertility. Since $\Theta$ and $M$ appear only as constant multipliers in $G(x)$ and $\gamma(x)$, respectively, they determine the level of $f(x)$ and not its shape. The latter aspect is influenced only by the values of the three other parameters present. Coale and Trussell[30] constructed a set of model patterns of age-specific fertility $f(x)$ by evaluating equation (E.7) for different values of $a_0$, $\gamma$ and $m$. The model fertility schedules generated in this way fit a wide range of observed fertility experience and permit the investigation of extreme patterns which have never or only rarely been accurately measured in practice.

2. *Brass relational Gompertz fertility model*

Brass[31] has sought to reduce the number of parameters determining the shape of age-specific fertility from the three required by the Coale-Trussell models to two by postulating, once more, a relational scheme between a "standard" fertility schedule and any other schedule. Specifically, denoting by $F(x)$ cumulated fertility up to age $x$ and by $TF$ total fertility, the ratio $F(x)/TF$, the proportion of total fertility experienced up to age $x$, is assumed to follow a Gompertz distribution function, whose form is

$$F(x)/TF = \exp(A\exp(Bx)) \qquad \text{(E.8)}$$

where $A$ and $B$ are constants, $A < 0$. This expression can be reduced to a linear function of $x$ by taking logarithms (ln) twice. The two steps necessary to carry out this transformation are

$$\ln(F(x)/TF) = A\exp(Bx) \qquad \text{(E.9)}$$

and

$$\ln(-\ln(F(x)/TF)) = \ln(-A) + Bx. \qquad \text{(E.10)}$$

A minus sign must be introduced when transforming equation (E.9) into (E.10) because the quantity $F(x)/TF$ is smaller than one; hence, $\ln(F(x)/TF)$ is negative and the logarithm of a negative number is not defined. To simplify notation, one may denote the $\ln(-\ln(F(x)/TF))$ transformation of $F(x)$ by $\eta(F(x))$. Then equation (E.10) becomes

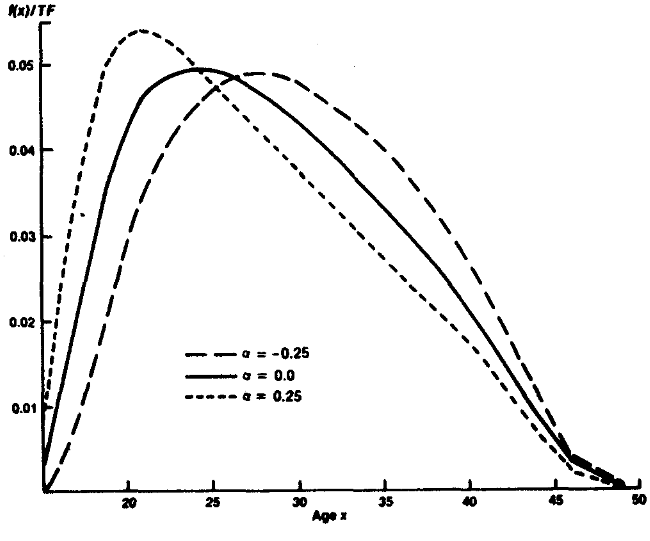
$$\eta(F(x)) = \ln(-A) + Bx. \qquad \text{(E.11)}$$

This model, in which $\eta(F(x))$ is linear function of age, approximates the observed $F(x)/TF$ ratios fairly well over the central ages of childbearing, but its fit deteriorates at the extremes. Brass discovered that a better fit can be obtained by substituting for the age variable $x$ a function of $x$ that can be interpreted as an $\eta$ transformation of a specific, standard fertility schedule. According to this finding, the relation expressed in equation (E.11) can be transformed into

$$\eta(F(x)) = \alpha + \beta\eta(F_s(x)). \qquad \text{(E.12)}$$

That is, the $\eta$ transformation of the observed fertility schedule is a linear function of the $\eta$ transformation of the standard. The parallel with the Brass logit life-table system is obvious; in both cases, a transformation that tends to linearize the distribution under consideration is used to relate any observed schedule to a standard pattern by the use of two constants. In the case of model life tables, the two parameters can be interpreted as determining the general level of mortality and the relationship between mortality early in life and that late in life. In the case of the relational fertility model, $\alpha$ in equation (E.12) can be taken as determining the age location of the fertility schedule or, more specifically, the age by which half of the total childbearing has occurred, while $\beta$ may be interpreted as determining the spread or degree of concentration of the schedule. (To see the effects that changes in $\alpha$ and $\beta$ have on the shape of fertility schedules, refer to figures 9 and 10.)

**Figure 9. Fertility schedules generated through the Gompertz relational model with $\beta = 1.0$**




[30] A. J. Coale and T. J. Trussell, "Model fertility schedules: variations in the age structure of child bearing in human populations".

[31] William Brass, "The relational Gompertz model of fertility by age of woman", London School of Hygiene and Tropical Medicine, 1978 (mimeographed).

Brass derived an appropriate standard from the Coale-Trussell model schedules. Values of its $\eta$ transformation for each age within the childbearing span are shown in table 4. The Brass model is definitely easier to use than the one developed by Coale and Trussell, and may prove very useful for simulation and projection purposes. However, its development is still fairly recent and experience with its use is limited.

**Figure 10. Fertility schedules generated through the Gompertz relational model with $\alpha = 0.0$**

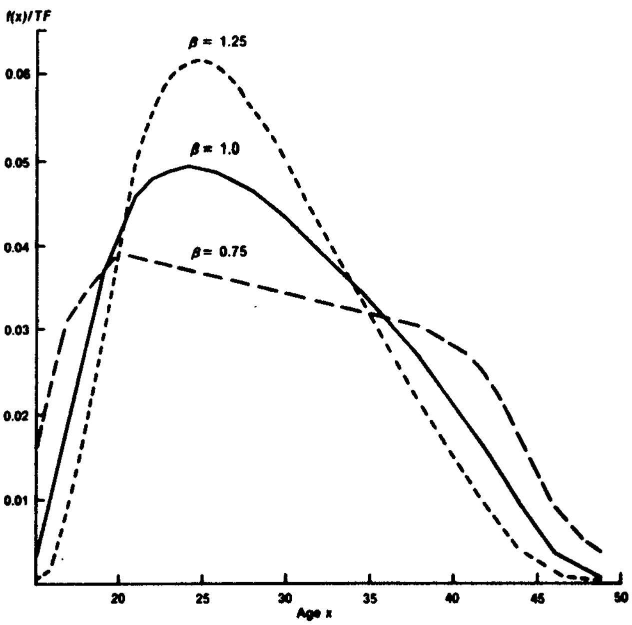



TABLE 4. VALUES OF THE $\eta$ TRANSFORMATION OF A STANDARD FERTILITY SCHEDULE, $\eta(F(x))$

| *Age* *x* *(1)* | *η Transformation* *η(F(x))* *(2)* | *Age* *x* *(3)* | *η Transformation* *η(F(x))* *(4)* |
|---|---|---|---|
| 11............. | 3.18852 | 31............. | −0.84272 |
| 12............. | 2.70008 | 32............. | −0.99014 |
| 13............. | 2.37295 | 33............. | −1.14407 |
| 14............. | 2.07262 | 34............. | −1.30627 |
| 15............. | 1.77306 | 35............. | −1.47872 |
| 16............. | 1.49286 | 36............. | −1.66426 |
| 17............. | 1.25061 | 37............. | −1.86597 |
| 18............. | 1.04479 | 38............. | −2.08894 |
| 19............. | 0.85927 | 39............. | −2.33192 |
| 20............. | 0.69130 | 40............. | −2.62602 |
| 21............. | 0.53325 | 41............. | −2.95500 |
| 22............. | 0.38524 | 42............. | −3.32873 |
| 23............. | 0.24423 | 43............. | −3.75984 |
| 24............. | 0.10783 | 44............. | −4.25499 |
| 25............. | −0.02564 | 45............. | −4.80970 |
| 26............. | −0.15853 | 46............. | −5.41311 |
| 27............. | −0.29147 | 47............. | −6.12864 |
| 28............. | −0.42515 | 48............. | −7.07022 |
| 29............. | −0.56101 | 49............. | −8.64839 |
| 30............. | −0.70000 | | |

## ADDITIONAL REFERENCES

Coale, Ansley J. *and* T. James Trussell. Erratum. *Population index* (Princeton, New Jersey) 41(4):572, October 1975.

Page, H. J. *and* G. Wunsch. Parental survival data: some results of the application of Ledermann's model life tables. *Population studies* (London) 30(1):59-76, March 1976.

Trussell, T. James. Selected applications of model fertility schedules. Unpublished doctoral dissertation. Princeton, New Jersey, Princeton University, 1974.