Demography India
Vol.23, Nos. 1 A 2 (1994), pp. 149-155

S. Mitra*

The Models of Mortality Rates

'THE earliest experiment to develop mathematical models for life table functions took place more than a century and a half ago when Gompertz (1825) suggested an exponential function of ageBC* to approximate the force of mortality \i(x) for ages beyond adolescence. More so then, than it is now, the force of mortality, very high at birth, declined relatively rapidly to reach the minimum value somewhere around age ten and kept on increasing thereafter without any further change in direction. It is well known that the pattern of mortality so described has been found everywhere at all times. Improvement in levels of mortality has merely pushed the curve, which is concave upward, near the age axis. Be that as it may, the quality of fit by the Gompertz function above the age of minimum mortality left plenty of room for improvement and it was Makeham (1860) who suggested a modification by adding a constant *A* to the Gompertz exponential function. Surely it was not enough, and later Perks (1932) suggested a five parameter model for the force of mortality given by

$$\mu(x) = \frac{A + BC^{x}}{KC^{-x} + 1 + DC^{x}} \tag{1}$$

which not only produced a better fit but also covered the life span in its entirety. Recently, the model was tried on Canadian life tables (Keyfitz 1991) covering the period 1921-81 at five year intervals. The goodness of fit measured by the root mean square of error to *l(x)* turned out to be excellent ranging from a minimum of .0016 to a maximum of .0058.

Of late, in depth studies of the distribution of mortality rates have convinced some model builders that a comprehensive model of mortality encompassing the entire age range is no longer adequate due to changes in structures of causes of death. In modern times degenerative and man made diseases have taken precedence over the infectious, contagious and parasitic diseases which used to be major killers in the not so distant past. Omran (1971) noted three stages of mortality transition in the history of mankind. The first stage characterized by pestilence and famine was followed by the second stage of receding pandemics. The third stage of degenerative and man made diseases perhaps began in the middle of this century in the industrialized nations. Olshansky and Ault (1986) hint at the fourth stage which is like the third but characterized by the appearance of degenerative and man made diseases much later in life than they did before. The proposition seemingly

---

* Department of Sociology, Emory University, Atlanta, GA 30322.

gathered strong support from the fact that the distribution of the mortality rates in many countries at present manifest a hump in the middle ages, a notable departure from the traditional (/-shaped pattern.

Consistent with this finding Heligman and Pollard (1980) suggested the treatment of mortality function in terms of three separate but additive components. Of these three, the first deals specifically with the infant and childhood mortality while its contribution rapidly declines with advancing age. The second describes the youth and adult mortality especially the hump in the middle ages and shows steep decline as the curve is extended on both sides of the hump. The third and the final component is the good old Gompertz function that takes care of mortality at older ages. The end result was a eight parameter model estimator of the ratio of *q(x)* and \-*q (x)* given by

$$\frac{q(x)}{1-q(x)} = A^{(x+B)^{C}} + De^{-E(\ln x - \ln F)^{2}} + GH^{x} \tag{2}$$

Three alternative forms of (2) were proposed by replacing the left hand side of (2) by *q(x)* and dividing the last term of the right hand side by 1 + GH^x in the first and by 1 + KGH^x in the second. The last two had nine parameters, which did not produce any significant improvement over the second. In the third, the parameter *K* was used as an exponent of *x* in H^x in both the numerator and the denom inator. Keyfitz *(ibid.)* fitted the first alternative model and found that it was somewhat better than Perks for the first six Canadian life tables covering the period 1921-46 while the opposite was true for the remaining seven covering the period 1951-81. Although, the eight parameter model was the best on an average, the parameters, as Keyfitz *(ibid.)* has noted, show no predictable pattern of association with the level of mortality.

Bah and Rajulton (1991) also fitted the first alternative model on the Canadian life tables covering the period 1951-86 at five year intervals separately for males and females. It is apparent that Keyfitz used the life tables combined for the two sexes since he did not mention anything about sex in his analysis. What is interesting however is that the estimates of the parameters generated by these two independent studies are not as consistent as one would expect them to be. The problem can, at least in part, be traced to the differences in the computer programmes used in those two studies since convergence is always at stake when the number of parameters of a non linear model is that many. Recently, Levin and Mitra (1994) have dealt with this particular issue and have generated solutions by using the technique of visualization in conjunction with the Table Curve software. It remains to be seen if this latter technique can produce consistent results when applied on life tables which are either combined or specific by sex. Be that as it may, there seems no reason to doubt the existence of consistency between such estimates when they are properly found but the real problem in the application of this model may lie beyond the logistics of curve fitting to a discussion of which we now turn.

Reality of the Single Hump Model

According to Heligman and Pollard's model curves, the existence of a hump can be traced back to half a century ago (1946-48) among the males but not among the females.

Two other sets of life tables of later years (1960-62,1970-72) however, show humps for both males and females. It should be noted that the *q(x)* values or the probability of dying must be obtained by single year of age for the humps to be visible at several ages among which one near age 20 was most prominent It may also be pointed out that the hump became visible when the *q(x)* values were plotted against age on a semi logarithmic graph paper which greatly magnifies the smaller values of *q(x)*, especially those that are around the most prominent hump. As usual, the humps would be barely visible if the data are plotted on a standard graph paper.

Then one may ask questions about the quality of the data that were used to suggest the model. Can the data be regarded so absolute as to contain no statistical and/or non statistical errors? Take for example the three *q(x)* values at ages 18, 24 and 26 for Australian males according to the 1946-48 life tables which are .00155, .00160 and .00155 respectively (Heligman *ibid.).* As such these values are quite similar in magnitude. They are also the three smallest and others are all larger and range from .00163 to .00183 in the age interval 18-29. Is this enough of an evidence to conclude that the fluctuation of the *q(x)* values as revealed by these figures are so real that they validate the theory of a hump or a genuine change in the pattern of mortality from the earlier times?

Granting the full benefit of doubt let us ignore the possibility that the mortality data may contain non statistical errors which among others include incorrect reporting of age of death by the survivors of the deceased. But we may not similarly ignore the error that can be attributed to statistical or random causes. Otherwise we shall have to accept the age of 24 years as something so special that the probability of dying at that age has to be less than those at ages immediately before and after it.

There being no just cause to support that premise, even when the data set reveals such fluctuations, a test for random variation of the variable *q(x)* naturally suggests itself. In that endeavour we first compute the mean value of *q(x)* in the age interval of 18-29 which includes the so called hump somewhere near age 20. For the 1946-48 male life table the mean of these *q(x)s* turn out to be .00169, a value that happens to match the value of *q(x)* at age 21. In order to test the hypothesis that the variations of all the *q(x)s* in the selected interval around the central value of .00168 can be deemed as random, we compute the standard deviation of a typical *q(x).* For that we need the value of the denominator which was used for the computation of the associated age-specific mortality rate. Assuming that the size of the male population in (1946-48) at any one of the single years of age in the aforementioned interval was of the order of 100,000 and assuming further that the age-specific rates were three year averages, the denominator of a typical rate can be taken as around 300.000. In that event, the standard deviation of the Poisson variable $q(x)$ given by $\sqrt{q(x)/300000}$ works out as .000075. It can then be seen that an interval of twice the standard deviation on either side of the central value of .00169 include all the twelve $q(x)$s thus supporting the hypothesis of their random variations in that age interval. In light of this analysis one may conclude that the existence of a hump can be justified no more than say an increasing trend of $q(x)$ connecting the two extreme values at both ends of the interval.

Equally convincing case can be made for more than one hump on the basis of the observed $q(x)$ values for all the Australian male life tables used by Heligman and Pollard.

Thus for the 1946-48 data set there were two humps one at age 20 and the other at age 23. For the years 1960-62, they were also two in number to be found at ages 19 and 26. Finally for 1970-72, there was one at age 19 and a mildly prominent second hump at age 25.

When the one hump model was filled, the hump as defined by the parameter *F* was found at ages 20.41,20.66, and 20.03, respectively for those three life tables. However, the actual locations of the hump as revealed by the data sets and the curves of *q(x)* seem to differ from the estimated values. These differences cannot be explained away as random errors but they result from artifacts of the composite function, the first derivatives of which do not necessarily vanish at those ages.

It should be mentioned at this point that the present analysis based on the age interval 18-29 has been performed only for the sake of expediency. This is not to suggest that the humps are to be found only in that interval. In fact, at least two more humps have been revealed at various ages before age 18 and after age 29 by these data sets.

The pattern is not much different for the females except for the fact that in comparison with the males, the humps generated by the female data sets are much smaller in amplitude. In fact, for the years 1946-48 the model curve for the females shows no visible hump at all, although quite a few, including one even at the advanced age of 56 may be found in the data set.

It is clearly evident by now that any attempt to construct a mathematical model that on the one hand is capable of describing the *q(x)* function with its major ups and downs, and at the same time, has parameters of substantive significance on the other, is likely to seek a few compromises. This must be so regardless of the number of parameters the model is based upon. Accordingly, an acceptable model can be defined as one which can hold the deviations of most of the observed from the model values within the tolerance limits set by the standard rules of probability measures. Among all such acceptable models the one with the least number of meaningful parameters merits serious consideration of the model builders. From that point of view, Heligman and Pollards' eight or nine parameter model, in spite of its demonstrated ability to capture the largest of the humps, does not seem to qualify as the best of the mortality models. For one thing, the prominence of the largest hump is primarily due to its getting blown up on the logarithmic scale which surely overemphasizes its true import. It seems that a comparison of say, the life expectancies, generated by this and other models with the actual values, is in order which may reveal differences of inconsequential magnitude. Second, allowing that many parameters in a model of mortality is more like an overkill and even then they bear little or no relationship with the overall level of mortality as has been found in the Canadian examples (Keyfitz *ibid.',* Bah and Rajullon *ibid.).* Therefore, a simpler and a smoother model is still to be preferred if it can reproduce the life table functions with acceptable levels of accuracy while its parameters, fewer in number, have substantive meanings and arc related with levels of mortality.

A Mathematical Model of Survivorship Function

The search for such a simpler and smoother function led to the development of a model of the survivorship function *l(x)* (Mitra 1994; Mitra and Denny 1994) as

$$In\ (-lnl(x)) = A + mlnx-nln(\alpha -x) \qquad (3)$$

where x is age and α is the life span, the three parameters, m, n, and A are all positive and need to be estimated for any given life table. As such, the model has four parameters but the number is reduced to three when a is arbitrarily set at 100 or at some value nearby. Again, by assigning an appropriate value to infant mortality and requiring the model to reproduce it, the number of parameters can be further reduced to only two. The model fitted on a number of life tables covering a wide range of life expectancy produced encouraging results.

An interesting feature of this two parameter model is that it can generate numerous life tables not only for any given value of infant mortality but also for any combination of values of the other two parameters. Both of them are meaningfully related with levels of mortality routinely measured by life expectancy. Thus the model has the ability to adjust itself with structural differences in the patterns of mortality in different regions of the world which earlier led to the construction of the regional model life tables (Coale and Demeny 1983).

Other Statistical Models

The first such model was developed by the United Nations (1956) following the discovery of high correlations between successive $q(x)$ values. Consequently, the expected $q(x)$ values can be successively generated from a given value of $q(0)$ which in turn could be used to construct model life tables. Noting the presence of structural differences in the pattern of variation in mortality rates by age, Coale and Demeny *(ibid.)* applied the same principle to generate regional model life tables.

Using the data set of the United Nations, Ledcrman and Breas (1968) found that the correlation matrix generated by the entire set of $q(x)$s of both sexes taken together, lent itself to the analysis by the method of principal components. They found five significant factors explaining more than 90 percent of the variation of the $q(x)$s. Sex was identified as one of the five factors. Even then, for the analysis of mortality, five factors are not few although Lederman and Breas did a good job in labelling the factors appropriately.

One of the reasons why there was that many principal components is of course the fact that as a function of age $q(x)$ changes direction at least once. The connection may now be made between the large number of factors and the large number of parameters in the mathematical models of Perks *(ibid.),* Hcligman and Pollard *(ibid.)* all involving the probability of dying $q(x)$. Naturally, in view of (3), similar connection may be expected Ua'.ween the mathematical and the statistical models involving the survivorship function $l(x)$.

In this context, we may recall the simple bivariate linear regression of the logit functions of $l(x)$ of two life tables developed by Brass (1968). Since the coefficient of determination of his model was found to be very high for any pair of life tables, Mitra and Levin (1993) subjected a large number of life tables complied by Keyfitz *(ibid.)* to the principal components analysis. They first proved that the correlation between the logit (/(x)) and logit (/(y)) for any two ages x and y decreases as the differences x-y increases in absolute value. The subsequent analysis done separately for the two sexes revealed only two factors explaining 98 percent of the total variation. The factor scores measuring two independent and orthogonal dimensions of mortality turned out to be excellent predictors of the survivorship function in general and the life expectancy in particular.

A recent investigation of Brass's logit model (Mitra 1994) revealed its failure to meet an important boundary condition. Forcing the boundary condition on the linear logit model converted it to a linear model of the reciprocal of *l(x)* of the two life tables for which the slope and the intercept coefficients are both positive and add up to one. The model has therefore one independent parameter and that turned out to be too few since the coefficient of determination for this new transformation of *l(x)* was found to lie in the neighbourhood of .9 which was certainly high but not high enough. However, the coefficient of determination shot up to little less than 100 percent when the reciprocals of *l(x)* of two life tables are entered as the two independent variables in a multiple regression set up. A two-parameter statistical model was thus produced as

$$\frac{1}{l\,(x)} = a\left(\frac{1}{l_s\,(x)} - 1\right) + b\left(\frac{1}{l_t(x) - 1} - 1\right)$$

where $l_s(x)$ and $l_t(x)$ are the survivorship functions of two standard life tables appropriately chosen and $l(x)$ stands for the same of any given life table. Equation (4) so written satisfies the boundary condition at $x = 0$ where the parameters *a* and *b* can be estimated by standard statistical procedure.

The same equation (4) can also be looked upon as a factor analytic model in which *a* and *b* are like factor scores and their coefficients are like factor weights. Accordingly, one can expect that a principal components analysis of the reciprocal of $l(x)$ will produce no more than two factors with eigenvalues greater than one. And indeed it does (Mitra and Levin 1995), when the components are extracted from the same set of life tables (Keyfitz *ibid.*) mentioned earlier. As before, 98 percent of the variation are explained by the two factors giving rise to the factor scores for any life table as two independent measures of the level of mortality.

Further Considerations and Conclusion

Equation (4) also leads to some postulates about the nature of the mathematical function that can describe the *l(x)* function or its reciprocal. That is to say, if (4) holds for any set of at least three life tables, then the reciprocal of any *l(x)* must be expressible as

$$\frac{1}{l(x)} = 1 + P_1 f_1(x, m_1, n_1, \ldots) + P_2 f_2(x, m_2, n_2, \ldots) + \ldots \tag{5}$$

where $m_1, n_1, \ldots$, etc.; $m_2, n_2, \ldots$ etc., must be constants for all life tables while $P_1, P_2, \ldots$, etc., are constants specific for a given life table. This is so because if *ms* and *ns* are not constants then (4) cannot be made to hold for all *x*. Observe also that (5) must have at least two *f* functions, otherwise either $l_s(x)$ or $l_t(x)$ but not both would be sufficient to describe $l(x)$. The just requirement that a mathematical model of the survivorship function be consistent with the factor analytic model seems to provide an important guideline for the formulation of the constituents of the models of life table function. Hopefully, future investigation will lead to the development of a model of $l(x)$ function that will meet the conditions laid down by equation (5). In this connection it may be interesting to note that the principal components analysis of $q(x)$ performed earlier by Lederman and Breas revealed four factors besides sex. Heligman and Pollard models has three additive components but none of the parameters is constant. Future research may show that it would become consistent with (5) when another term is added to their model equation.

References

Bah, Sulaiman M. and Fernando, Rajulton, 1991, Has Canadian mortality entered the fourth stage of the epidemiologic transition? *Canadian Studies in Population,* 18(2) : 18-41.

Brass, W., *et al.,* 1968, *The Demography of Tropical Africa.* Princeton, NJ: Princeton University Press. Coalc, A. and Dcmcny, P., 1983, *Regional Model Life Tables and Stable Populations.* New York : Academic Press. Gompertz, B., 1825, On the nature of the function expressive of the law of human mortality; and on a new mode of determining the value of life contingencies. *Philosophical Transactions,* R.S. 115 (27): 513-585. Heligman, L. and Pollard, J. II., 1980, The age pattern of mortality. *Journal of the Institute of Actuaries, Part 1,* 107 (434): 49-80.

Keyfitz, N., 1991, *Population.* San Francisco : W. H. Freeman and Company. Lcdermann, S. and Rrcas, J., 1959, Les dimensions de la mortalite. *Population,* 14:637-682. Levin, Martin L. and Mitra, S., 1994, Research methods: Visualization in the numeric solution of complex nonlinear equations in mathematical demography. Forthcoming in *Social Science Computer Review.* Makeham, W. M., 1860, On the law of mortality. *Journal of the Institute of Actuaries,* 13: 325-358. Mitra, S., 1994, Boundary condition sets limits on Brass's model life table parameters, presented at the 1994 Joint Statistical Meeting in Toronto. Forthcoming in *Canadian Studies in Population.* Mitra, S. and Denny, C., 1994, On the application of a model of mortality. *Canadian Studies in Population, 2*1(2): 117-132. Mitra, S. and Levin, Martin L., 1993, Principal components analysis of the logits of the survivorship function, presented at the 1993 meetings of the International Union for the Scientific Study of Population in Montreal, Canada. Forthcoming in *Janasamkhya.* Mitra, S. and Levin, Martin L., 1995, Modeling the reciprocal of the survivorship function, to be presented at the 1995 meeting of the Population Association of America in San Francisco. Olshansky, S. J. and Ault, A. B., 1986, The fourth stage of the epidemiologic transition: The age of delayed degenerative diseases. *Milbank Memorial Fund Quarterly,* 64: 355-391. Omran, A. R., 1971, The epidemiologic transition: A theory of epidemiology of population change. *Milbank Memorial Fund Quarterly,* 49:507-537. Perks, W., 1932, On tome experiments in the graduation of mortality stati sties. *Journal of the Institute of Actuaries,* 63(12): 26-40. United Nations, 1956, *Manual III: Methods for Population Projection by Sex and Age.* Sales No. 56, Xffl, 3.