Beard, R.E. 1971. Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. In: Biological Aspects of Demography (ed. W. Brass). London: Taylor and Francis

SOME ASPECTS OF THEORIES OF MORTALITY, CAUSE OF DEATH ANALYSIS, FORECASTING AND STOCHASTIC PROCESSES

R. E. BEARD

Pearl Assurance Co., London

WHEN asked to prepare a contribution for the Society for the Study of Human Biology symposium I realized that there was little new material I could add to my previous work on mortality but it did seem that an outline of the development of my studies over the years might be of interest and value to other workers in the same field, particularly as other commitments have severely curtailed my opportunities for further research.

2. It should perhaps be noted that my interest in mortality arose from the studies needed for qualification as an actuary, but I was fortunate in having as a colleague at the Pearl Assurance Company, Wilfred Perks, whose original contributions to actuarial science were recently marked by the award of a gold medal by the Institute of Actuaries. Over many years we have had discussions and arguments which have been very productive of novel aspects on the subject of mortality and thus provided some of the stimuli for subsequent developments.

3. The actuary's main interest in mortality is that it is one of the essential factors in the calculus of life contingencies which underlies the proper financial management of life offices and pension funds. It is a very practical interest in that the contracts extend over many years into the future and a clear understanding of the magnitude and directions of variations in mortality is necessary if premiums and benefits are to be properly determined and if the periodic valuations are to be soundly based.

4. As events have turned out, for many years the trend in mortality until recently has been a fairly rapid decline at the younger ages and a much slower reduction at the older ages. This has been fortunate for life assurance business since an experienced mortality lighter than that assumed in the premium basis, operates in favour of the business. For annuity and pension business the opposite is however true and one of the difficult problems facing the actuary is the extent to which (and when) mortality will decline at the older ages in the future. Various techniques have been used to allow for future improvement; some are built on judgement and financial arrangements within the companies (or funds) and some are based on methods of extrapolation based on careful analysis of past trends. Among the latter may be mentioned the cohort analysis method used by Derrick (1927) and the rather sophisticated technique used by Prawitz (1954) and based on analysis of deaths by causes.

5. The calculus of life contingencies is based on the life table; the theoretical concept of a life table is, of course, the distribution by age at death of a group of births (or entrants at some other age). For calculations relating to human lives it is impracticable to build a life table by observation of the survivors of a group of births and tables are therefore constructed from the rates of mortality observed during a relatively short period of time. In using a table constructed in this way the actuary will have regard to its special characteristics and make due allowance for them. If circumstances indicate that "cohort" mortality is more appropriate, such tables will be constructed.

6. The necessity of basing calculations on rates of mortality and the practical problems of computation having regard to the absence of calculating machines meant that attention was focused on q_x (or μ_x) and early attempts to derive "laws of mortality" were based on the observed shape of these functions. The first mathematical formula to provide a reasonable approximation to adult rates of mortality was that proposed by Benjamin Gompertz in 1825 and a great deal of subsequent work has been based on the concept that age specific mortality rates increase in geometric progression. In 1867 William Makeham put forward his modification to the Gompertz form which materially improved the agreement at the younger ages for assured life data; the formula was still unbounded at the upper limit.

7. For a number of reasons the Makeham form has provided a

very useful basis for actuarial work but it became clear in the early part of this century that it was not universally applicable, particularly at the older ages where accumulating data suggested a slowing down of the rate of increase with age. In 1932 Wilfred Perks put forward further modifications of the Gompertz form and extensive calculations (Beard, 1951a) have shown that these provide very reasonable mathematical expressions of rates of mortality for a wide range of experiences.

8. The simplest member of the Perks's family is the logistic and the relationship between the three named curves can be briefly expressed as:

$$\mu_x = Bc^x \qquad \text{(Gompertz)},$$
$$\mu_x = A + Bc^x \qquad \text{(Makeham)},$$
$$\mu_x = A + Bc^x/(1 + Dc^x) \qquad \text{(Perks)}.$$

These formulae are, of course, applicable to adult mortality only.

9. If the deaths in a life table are looked upon as a frequency distribution then μ_x is the ratio of the ordinate at age x to the tail area above age x. For a very wide range of frequency distributions this function is of sigmoid form and it is not apparent whether the satisfactory representation of μ_x by a logistic (Perks) curve is because the formula has a theoretical significance or because it provides a good approximation to the particular function of a family of frequency curves which can be used to represent the distribution of deaths by age. Furthermore, apart from a few elementary cases, the ratio of the ordinate to the tail area is a complicated mathematical function. In seeking to describe this function (i.e. μ_x) by a simple mathematical form it may be that attention is being directed towards the wrong function and more progress in the understanding of mortality might be achieved by studying the curve of deaths (d_x or μl_x).

10. Karl Pearson (1897) did, of course, render a description of the curve of deaths as the sum of five different frequency curves and other writers, e.g. Phillips (1954), have endeavoured to encourage study of this function. In his 1932 paper Perks suggested that it might be worth while studying $d \log \mu l_x/dx$, i.e. the function which has the same mathematical relationship to μl_x as μ_x has to l_x and Ogborn (1953) did some related work but systematic studies are lacking.

11. The observed feature that a wide range of adult human

mortality tables could be expressed by the logistic curve prompted me to consider models in which the population was assumed to be heterogeneous and my first efforts were directed to deterministic forms in which the population was assumed to be stratified with common mortality rates. If μ_k^s is the force of mortality at time ($\equiv$ age) k for the group with "longevity" factor s and $\phi(s)/ds$ is the proportion of the initial population with factor s, then the mortality rate at time k for the whole population is:

$$\mu_k = \frac{\int \phi(s)\mu_k^s \exp\left(-\int_0^k \mu_t^s\, dt\right) ds}{\int \phi(s) \exp\left(-\int_0^k \mu_t^s\, dt\right) ds},$$

where the integrals are taken over the whole range of s.

12. If it be assumed that $\mu_k^s = \alpha + \beta s \exp(\lambda k)$, i.e. a Makeham form, and that $\phi(s) = \kappa s^p \exp(-\gamma s)$ $(0 \leqslant s < \infty)$, i.e. a Gamma distribution, then

$$\mu_k = \alpha + \frac{(p+1)\beta \exp(\lambda k)}{(\gamma\lambda - \beta) + \beta \exp(\lambda k)},$$

which is a logistic form (Beard, 1959).

13. Whilst the foregoing provides an interesting mathematical description of a process which leads to a logistic form it suffers from the disadvantage that it assumes that the mortality of the various strata is Makeham in form for which experimental verification seems very difficult. It is also a deterministic model and implies that each individual in the population has a unique "longevity" tag; whilst studies of heredity have shown correlations, the model does not seem to lend itself to development.

14. Concurrently with these ideas I was also thinking about the so-called "shot" models. In these it is assumed that individuals accumulate "shots" from random firings and are assumed to be dead when the total reaches a given figure. Forward models of this type did not lead to numerical results which fitted the facts. There are however "backward" models in which shots are lost and death is assumed to occur when the starting stock has been reduced to a given level and I found that this type of model did lead to sensible numerical results.

15. For example if it be assumed that l_t^α are the number of persons alive at time t with α units remaining and that the chance of losing a unit in the next interval is pdt times the number of units remaining then we can write:

$$\frac{dl_t^\alpha}{dt} = -p\alpha l_t^\alpha + p(\alpha+1) l_t^{\alpha+1}.$$

The solution to this equation is $l_t^\alpha = \binom{r}{\alpha} \exp(-p\alpha t)[1 - \exp(-pt)]^{r-\alpha}$, where r is the initial stock of units. If the initial population is so distributed that the proportion with stock r is $D/(1+D)^{r-\alpha+1}$, then it can be shown that the overall mortality will be:

$$\mu_t = \frac{p\alpha D \exp(pt)}{1 + D \exp(pt)},$$

i.e. a logistic form (Beard, 1964).

16. The deaths at time t for a given stratum are $p\alpha l_t^\alpha$ and the force of mortality is:

$$\mu_t^\alpha = \frac{p\alpha l_t^\alpha}{\sum_\alpha l_t^\alpha} \times \frac{p\alpha l_t^\alpha}{\int_t^\infty p\alpha l_s^\alpha\, ds}.$$

For the form l_t^α derived in paragraph 15 these expressions cannot be evaluated in simple terms although it may be noted that l_t^α can be expressed as a Beta ordinate (Pearson Type I) with a logarithmic transformation. For this curve μ_t^α will have a sigmoid form. This is not without interest as it suggests that the curve of deaths could be fitted with a Pearson type function subject to a change of variable.

17. One series of experiments I made was to use the Γ function for actuarial calculations (Beard, 1950, 1952) and some earlier unpublished work on fitting Pearson curves to the curve of deaths with a transformed variable gave encouraging results, but were not followed up.

18. From the foregoing two distinct models have been found which lead to the logistic form for the force of mortality, the first is a stratified population with the strata subject to Makeham mortality and the second is a different stratification but with mortality dependent on the state of "deterioration" reached at time ($\equiv$ age) t by a purely random process. These two models are reminiscent of the accident models in which the negative binomial distribution,

which describes so well the accident distribution, can be shown to arise from a number of different underlying models. In the mortality case the logistic may well be playing a similar role to the negative binomial.

19. The important point is, however, the fact that the purely random process model, provided the backward form is used, does lead to logistic values of the force of mortality which are actually found from statistical observations. The model implies that individuals start their life with a total quantity of "units" which diminishes throughout life on a probability basis, i.e. the chance that a unit is lost in a given interval is solely governed by the quantity remaining at the time. The backward model does not give a Gompertz form as a solution (except as a limiting form) but since the essential difference between the logistic and Gompertz forms is at the highest ages it is necessary to consider the data in this region before ruling out the logistic model.

20. The Gompertz model implies that μ_x continues to increase indefinitely with x whereas the logistic implies a limiting value. So far as human data are concerned there seems little firm data to work on. If a population is considered there is plenty of evidence that the increase in the rate of mortality slackens off at the older ages but this could be due, of course, to selective processes in which weaker members are the first to die (i.e. the first stratified model). The slackening of the increase in μ_x in the second model comes from the reduction in "resistance" by the random process until most survivors are within one unit of death. My own preference is to accept that there is an upper limit to μ_x; the evidence from actuarial sources of "damaged" lives (Beard, 1951b) is also interesting as the numerical values derived from select tables are in line with those derived by fitting a logistic formula to the adult age range.

21. However, the evidence may be studied from another angle. If the curve of deaths is regarded as the critical function, the discrimination between a logistic and a Gompertz form will depend on the few survivors at the tail of the curve. About 1% of births survive beyond age 90 or so and it would require a great deal of data to sort out the difference between the logistic and Gompertz distribution for the tail. Too close attention to the rates in this region may obscure the reliability of the data in terms of the overall observations.

22. So far, of course, the foregoing has been concerned with the

mortality observed from groups of individuals and the suggestion reached is that the mortality process is best defined in terms of deterioration of the individuals within the group by a random process. However, we know that individuals vary in regard to their expected longevity and susceptibility to different causes of death. To extend the model means that the notion of homogeneity must be dropped and allowance made for variations between individuals.

23. Before generalizing in this way I would refer to some ideas I tried in seeking a link between cellular processes and deterioration processes (Beard, 1960/1961a, 1961b). These were not particularly profitable but they were formative, particularly the statistical work on causes of death.

24. The animal organism is a highly complex and interlocking assembly of sub-systems, each of which is ultimately resolvable in biochemical terms. In principle it would seem reasonable to assume that the process of living is ultimately expressible in physical terms, although any relationship found would necessarily be complicated. Mortality is one extreme of the process of living and the study of mortality can be regarded as a limiting case of the study of the living organism. Although the various sub-systems are interlocked the feature that the behaviour of the aggregate can be described as a purely random process is a temptation to postulate that the sub-systems could also be so described.

25. Instead of attempting to analyse the various sub-systems a statistical approach would be to group the various causes of death according to their observed distributions and examine the resulting groups to see whether the particular mortality patterns did lead to different stochastic models. Following on this line of thought, my first experiment was to take the deaths by causes in England and Wales for 1958 (Registrar General) and group those causes which showed similar distributions of death by age. It was appreciated that this was a purely statistical exercise as the distribution of deaths is that arising from the particular population distribution at the time and that the underlying mortality rates were dependent rates.

26. The Pearson moment coefficients were calculated from these distributions (males and females separately) and attention was restricted to causes of death which might reasonably be regarded as constitutional in nature. The $\sqrt{\beta_1}$, β_2 values were then plotted on a chart which I had calculated to delineate the areas where the logistic fell. With one exception (cancer of the breast, female, which appeared

to be a "mixture") all the points fell within the logistic area, although, of course, the parameters differed from cause to cause.

27. Incidentally the $\sqrt{\beta_1}, \beta_2$ chart had been calculated to facilitate studies on mortality generally. There are growing numbers of series of observations on various animals and it seemed that the easiest method of obtaining a quick view of the form of the mortality was to calculate the moments of the distribution by ages at death and enter these on the chart. This technique handles the statistical significance problems arising from the limited number of observations in such series which tend to be obscured if attempts are made to estimate rates of mortality directly from the data.

28. The proper next stage in experimentation would have been to calculate for each of the (grouped) causes of death, independent rates of mortality and from them independent curves of death. This I did not do although some limited calculations were made to confirm that it was reasonable to expect the logistic to hold in the independent case. At this stage my attention was also drawn to the work by Prawitz (1954) who had used the idea of calculating independent rates by causes for use in forecasting mortality. More recent work in this field has been done by Chiang (1960).

29. However, at this time (1961) the controversy about smoking and lung cancer was becoming very active and it seemed worth while looking at the population data to see if the ideas had any application in this field. My idea was that if the increase in lung cancer was primarily associated with cigarette smoking then the mortality process might be expressed as a random process. This would also provide a difficult case on which to test the techniques.

30. The results have been described elsewhere (Beard, 1963) and will not be repeated here. The first job was to collate the relevant statistics and for this purpose the General Register Office Studies on Medical and Population Subject No. 13, provided the information in a convenient form. Consideration of the variations in the (independent) central rates of mortality for cancer of the lung showed that it was possible to find a reasonable representation of the data over the years 1913 to 1958 and for ages 27 to 82 by a product formula of the type:

$$\mu_x{}^T = k\phi(T-x)\chi(T)f(x),$$

where $\phi(T-x)$ is a function dependent on year of birth, $\chi(T)$ a function dependent on the calendar year of experience and $f(x)$ a function solely dependent on age. This, of course, is a descriptive formula.

31. However, if the hypothesis of a purely physical cause for this mortality held then it would be reasonable to expect the $f(x)$ function to be similar for males and females. After some experiments it was, in fact, found possible to find a set of values in which the k factors and the $f(x)$ functions were similar for males and females, provided the ϕ factor was related to the proportion of smokers in the generation and the χ factor related to the quantity of cigarettes smoked in the various calendar years, but with a time lag of about 15 years.

32. The critics can, of course, argue that this descriptive model still only reflects statistical association but there is one other piece of evidence to support the hypothesis. If the mortality rates $f(x)$ have a purely physical basis then it would be reasonable to expect them to be described by a stochastic process of some type. First attempts at this were unsuccessful but a careful study of the basic data showed that the frequency distributions of the deaths from lung cancer had patterns very suggestive that they were composed of deaths with distributions of two different types. No information is available to classify the population deaths according to the type of cancer but by making use of some Norwegian data (Kreyberg, 1962) and splitting the deaths into two series it was found that the progression of the proportion of $f(x)$ which could be assumed to be associated with epidermoid and oat cell carcinomas could be represented by a purely stochastic process.

33. Although the analysis thus proved to be a long and indirect process it did lead to the conclusion that if the deaths could be properly classified the results supported the hypothesis that the lung cancer incidence was associated with the quantity of cigarettes smoked. The stochastic model which describes the process is a "backward" type, i.e. the individual starts with a "stock" of resistance which is reduced over time on a probability basis according to the quantity of cigarettes smoked. The justification is a numerical agreement with observed facts and a common model for both male and female lives.

34. The model so far as I have taken it is incomplete but unless some observed data become available in a suitable form and give separate figures for the two main types of lung cancer there seems little point in refining the model.

35. During the course of this analysis some interesting points fell

to be considered. One of these was the behaviour of the independent mortality rates at the advanced ages. The Registrar General's data is given in quinquennial age groups up to age 85, but thereafter all ages are grouped and it is impossible to determine the true trend beyond this age and thus to decide on the mathematical behaviour of $f(x)$. However, Dr. Benjamin was able to provide some data relating to centenarians and it is not without interest to note that among the 148 male and 1084 female deaths of centenarians from 1956 to 1960 there was only one death (F) attributed to cancer of the lung. It proved impossible to form an opinion regarding the limiting behaviour of $f(x)$.

36. Another aspect of the calculations which provided some interesting thoughts arose from the early calculations of the $\mu = \phi\chi f$ formula. The values of $\phi(T-x)$, i.e. the parameters associated with each year of birth showed a remarkable increase from zero in the early 1800's to a maximum in about 1910 for males and rather later for females. The progression was so regular that I was tempted to seek an explanation on genetic grounds but it soon appeared that no reasonable model could be created to deal with the rates of increase found. This feature seemed to rule out a number of the theories being advanced which sought to identify the increase in lung cancer deaths with factors other than cigarette smoking. It was a satisfactory analytical result to find that the increase could be quantitatively agreed with the proportions of smokers in the generations. Similarly the variations in the numerical values of $\chi(T)$ did tie in with the quantity of tobacco smoked.

37. If this model for lung cancer mortality is correct, then it becomes possible to talk about forecasting the deaths expected in the future from this cause. Once the ϕ, χ, f factors are available the death rates in future years can be calculated but only by making assumptions about the smoking habits of future generations or about the discovery and elimination of the harmful constituents of tobacco. Whether similar models can be found for other causes of death is not known. From my earlier work on the statistical distributions of deaths by causes it would not seem unreasonable to apply the $\phi\chi f$ technique and I would expect the resulting "pure" mortality curves to be logistic in form and capable of representation by stochastic processes. It is conceivable, though I admit unlikely, that physical features will be found to describe the processes but the effort might be worth while. In any case extrapolation based on the

separate cause of death analyses would seem to offer a more scientific basis of forecasting than a purely arithmetical method.

38. The $\phi\chi f$ technique can be looked upon as a crude separation into "nature" and "nurture" since the ϕ parameter is associated with year of birth and the χ parameter with year passed through. But although I have made some limited experiments with other causes of death these have not led to any suggestions that the concept is useful.

39. As a rather different type of analysis I have done some work on the mortality of mice by causes. The idea came after reading an article by Lindop and Rotblat (1961) on their experiments in which series of mice were exposed to different doses of radiation and records kept of their age at death and cause thereof. The age distribution of deaths by causes showed marked changes according to the dosage received but the statistical uncertainty arising from the limited numbers and the size of the "unknown cause" groups were too large for useful conclusions to be drawn.

40. There are a number of interesting features of mortality tables based upon the over-simplifying assumption that causes of death operate independently but this is hardly the occasion to extend this paper. Clearly to move from the assumption that a group of lives is homogeneous to an assumption of heterogeneity must give rise to assumptions as to the form of the heterogeneity. Equally to move from a deterministic to a stochastic model will complicate the analysis still further but without these extensions the study of mortality must remain largely descriptive statistics. So soon as these extensions are introduced the study of mortality becomes the scientific study of living processes.

The following bibliography is a shortened one appropriate to the subject matter of this paper and as such is a necessarily biassed selection of the literature. Beard (1959, 1964) gives a more extended series relevant to this paper, but a more systematic collation of relevant papers will be found in Strehler (1962).

References

BEARD, R. E. (1950) *Proc. Centenary Assembly Inst. Act.* **2**, 89.
BEARD, R. E. (1951a) *J. Inst. Actuaries* **77**, 382.
BEARD, R. E. (1951b) *J. Inst. Actuaries* **77**, 394.
BEARD, R. E. (1952) *J. Inst. Actuaries* **78**, 341.
BEARD, R. E. (1959) *Ciba Fdn Colloq. Ageing*, edited by Wolstenholme and O'Connor, **5**, 302, Appendix, J. & A. Churchill, London.

BEARD, R. E. (1960/1961a) *Trans. Assur. med. Soc.* p. 17.
BEARD, R. E. (1961b) *International Population Conference*, **1**, 611.
BEARD, R. E. (1963) *Proc. R. Soc.* B, **159**, 56.
BEARD, R. E. (1964) *Int. Congr. Actuaries* **3**, 463.
CHIANG, C. L. (1960) *Proc. Berkeley Symp. math. Statist. Probab.*
DERRICK, V. P. A. (1927) *J. Inst. Actuaries* **58**, 117.
GOMPERTZ, B. (1825) *Phil. Trans. R. Soc.* **115**, 513.
KREYBERG, L. (1962) *Histological Lung Cancer Types*, Norwegian University Press, Oslo.
LINDOP, P. and ROTBLAT, J. (1961) *Proc. R. Soc.* B, **154**, 332.
MAKEHAM, W. M. (1867) *J. Inst. Actuaries* **13**, 325.
OGBORN, M. E. (1953) *J. Inst. Actuaries* **79**, 170.
PERKS, W. (1932) *J. Inst. Actuaries* **63**, 12.
PEARSON, K. (1897) *The Chances of Death and other Studies*, Arnold.
PHILLIPS, W. (1954) *J. Inst. Actuaries* **80**, 289.
PRAWITZ, H. (1954) *Svenska Aktuarieforeningen Femtio Av.*
REGISTRAR GENERAL (1958) *Statist. Rev.*
STREHLER, B. L. (1962) *Time, Cells and Aging*, Academic Press, London and New York.

SOCIETY FOR THE STUDY OF HUMAN BIOLOGY

Although there are many scientific societies for the furtherance of the biological study of man as an individual, there has been no organization in Great Britain catering for those (such as physical anthropologists or human geneticists) concerned with the biology of human populations. The need for such an association was made clear at a Symposium at the Ciba Foundation in November 1957, on "The Scope of Physical Anthropology and Human Population Biology and their Place in Academic Studies". As a result the Society for the Study of Human Biology was founded on May 7th, 1958, at a meeting at the British Museum (Natural History).

The aims of the Society are to advance the study of the biology of human populations and of man as a species, in all its branches, particularly human variability, human genetics and evolution, human adaptability and ecology.

At present the Society holds two full-day meetings per year—a Symposium (usually in the autumn) on a particular theme with invited speakers, and a scientific meeting for proffered papers. The papers given at the Symposia are published and the monographs are available to members at reduced prices.

Persons are eligible for membership who work or who have worked in the field of human biology as defined in the aims of the Society. They must be proposed and seconded by members of the Society. The subscription is £3 ($7.00) per annum (this includes the Society's journal *Human Biology*) and there is no entrance fee.

Applications for membership should be made to Dr. W. A. Marshall, Hon. General Secretary, Institute of Child Health, Guilford St., London, W.C.1.

PUBLICATIONS OF THE SOCIETY

Symposia, Volume I, 1958: *The Scope of Physical Anthropology and Its Place in Academic Studies*, edited by D. F. ROBERTS and J. S. WEINER (out of print).
Symposia, Volume II, 1959: *Natural Selection in Human Populations*, edited by D. F. ROBERTS and G. A. HARRISON. Pergamon Press (£1).
Symposia, Volume III, 1960: *Human Growth*, edited by J. M. TANNER. Pergamon Press (members 10s. 6d.).
Symposia, Volume IV, 1961: *Genetical Variation in Human Populations*, edited by G. A. HARRISON. Pergamon Press (members £1).
Symposia, Volume V, 1963: *Dental Anthropology*, edited by D. R. BROTHWELL. Pergamon Press (members 25s.).
Symposia, Volume VI, 1964: *Teaching and Research in Human Biology*, edited by G. A. HARRISON. Pergamon Press (members 25s.).
Symposia, Volume VII, 1965: *Human Body Composition, Approaches and Applications*, edited by J. BROZEK. Pergamon Press (members £3).
Symposia, Volume VIII, 1968: *The Skeletal Biology of Earlier Human Populations*, edited by D. R. BROTHWELL. Pergamon Press (members £2.).
Symposia, Volume IX, 1969: *Human Ecology in the Tropics*, edited by J. P. GARLICK and R. W. J. KEAY. Pergamon Press (members 30s.).

SYMPOSIA OF THE
SOCIETY FOR THE STUDY OF HUMAN BIOLOGY,

Volume X

BIOLOGICAL ASPECTS OF DEMOGRAPHY

Edited by
W. BRASS

TAYLOR & FRANCIS LTD
London
1971