Beard, R.E. 1964 International Congress of Actuaries, vol.3, p.463-477

SOME OBSERVATIONS ON STOCHASTIC PROCESSES WITH PARTICULAR REFERENCE TO MORTALITY STUDIES

by

R. E. BEARD (*Great Britain*)

In his text-book on stochastic processes [1] Professor M. S. Bartlett writes "by a stochastic process we shall in the first place mean some possible actual, e.g. physical process in the real world that has some random or stochastic element involved in its structure". This definition shows the extremely wide field of application and it is therefore not surprising that the whole subject has shown very rapid growth in recent years.

2. Since the business of insurance is based upon the happening at random of events in the real world it is to be expected that the operations of insurance companies would provide many instances where the theory would be applicable. This is certainly true and one of the earliest developments in the theory was the collective theory of risk associated with the names of Dr. Filip Lundberg and Professor H. Cramér [2]. In 1940 a further contribution to the theory was provided by Dr. Ove Lundberg [3] for which data from sickness and accident insurance were used as the observational material. In more recent years the theory of stochastic processes has been found a natural calculus for the study of many problems arising from non-life insurance and an appreciable part of the activities of ASTIN has been devoted to this development [4].

3. The theory of stochastic processes is essentially a branch of probability theory and it is proper to ask why the early Scandinavian applications to life insurance problems were not followed up more actively by actuaries generally. So far as the United Kingdom is concerned the conditions in the market for life insurance and annuity business have been such that the stochastic variations arising from the mortality component have been small relative to the other elements entering into the actuarial management of the funds and accordingly there has been no pressing need for the precision offered by the collective risk theory in the measurement of the variations. In other markets, and in the non-life field in particular, the position is however different and the random variations can be of much greater relative significance. No doubt a general change in the attitude of fiscal authorities to the taxation treatment of "fluctuation" reserves calculated from risk theoretical considerations as has occurred in at least one country, would provide a substantial impetus to activities in this field. Equally, a wider study of the potential applications to questions such as the estimation of claim reserves would be of value in classes such as motor which are becoming of considerable concern to insurers throughout the world.

4. Although these questions are of great interest they are not being further considered in this paper which is concerned with some applications to the study of mortality. The approach is basically theoretical, in the sense that it is not concerned with day-to-day actuarial problems; the impetus has however sprung from a desire to learn more about the mortality process in the hope of providing knowledge of value in the reduction of mortality and for making reasoned opinions of future trends. It is implicit in the paper that the causes of deaths considered can be regarded as arising from some form of breakdown of the constitution of the individual as distinct from external influences such as infections.

5. It may be noted that in the application of risk theory to insurance portfolios the assumption is usually made that events in successive intervals are independent, although they may well arise from the same underlying probability distribution. The results over a particular period of time are then essentially dependent on the sampling distribution of the expected events arising during the time interval. In the following notes the underlying idea is that the individuals comprising the population undergo changes in successive time-intervals, i.e. they move from one state to another with a certain probability. Thus the distribution of the individuals according to their "state" at any moment is dependent on the state at some earlier moment and the law defining changes of state. The resultant functions, e.g. the survivors at a given future time, are expectations based on the particular random process operating.

6. The first mathematical formula to provide a reasonable approximation to observed rates of mortality was that proposed by Benjamin Gompertz in 1825 [5] and a great deal of later thinking has been based upon the concept that age specific mortality rates increase in geometrical progression as proposed by Gompertz. The modification proposed by William Makeham in 1867 [6] subsequently improved the agreement at the younger ages with statistics derived from data for insured lives, but still left the unbounded limit at the upper ages. In 1932 Wilfred Perks [7] proposed a further modification of the Gompertz form, and the family of curves derived by him provided a reasonable mathematical form for most of the mortality data to which it has been applied. The simplest form of the family is more generally known as the logistic curve and the relationship between the three named curves can be briefly expressed as

$$\begin{aligned} \mu_x &= Bc^x && \text{(Gompertz)} \\ &= A+Bc^x && \text{(Makeham)} \\ &= (A+Bc^x)\div(1+Dc^x) && \text{(Perks)} \end{aligned}$$

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

7. Now mathematical relationships of the type found are essentially descriptive formulas and although various attempts have been made to provide some physical interpretation of the formulas most of them have

been in terms of vague concepts which have not led to any scientific developments. It is reasonable to ask whether the emphasis on the rate of mortality which has naturally followed from the needs of the actuary in his problems and from the calculus developed to enable him to deal with them has not been a retarding factor in the study of mortality for its own sake. To see this we may look at the problem from a life table point of view; if the curve of deaths is considered as a frequency distribution then the force of mortality μ_x becomes the ratio of the ordinate at age x to the tail area above age x. For a very wide range of distributions this ratio is of sigmoid form and will also show a pronounced geometric variation over quite a wide range. In other words the observed mathematical forms may be little more than a result of the definition of μ_x and the success of the Perks (or logistic) curve arises from its flexibility in relation to the general shape of the curve of deaths. On this general reasoning therefore it would seem that the curve of deaths would be a more suitable function for study than the rate of mortality, particularly when consideration is being given to deaths from specific causes.

8. There are, of course, other reasons why the curve of deaths, or the distribution by age at death of a group of births, is a more natural function for theoretical studies than the rate of mortality, particularly if studies are being made of populations, such as of mice or flies, which can be observed throughout their lifetime in an environment which can be held reasonably constant [8]. It is a well-known feature that a group of lives of a given age will exhibit quite a wide range of apparent health, and the greater the age the wider the range found. This suggests that an appropriate model for the study of a group of lives from birth is to postulate a characteristic which measures "vitality" or "resistance to destruction", to assume that the survivors at any age can be classified according to this characteristic and to assume that the transition from one state of the characteristic to another occurs according to some probability process. It is assumed that death occurs when the characteristic reaches a prescribed value. This model is essentially a stochastic process and the techniques developed in recent years are immediately available for its study [9] [10].

9. As I have shown elsewhere, the logistic form for μ_x can be obtained as a result of a process defined by

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

where l_t^α represents the number of lives at age t exhibiting characteristic α. It is assumed that the probability that a unit of α is lost in the time interval dt is p and that death occurs when α reaches a certain minimum value. The solution of equation (1) depends on the initial conditions; if it is assumed that

$$\begin{aligned} l_0^\alpha &= 1 \quad \alpha = r \\ &= 0 \quad \alpha \neq r \end{aligned}$$

the solution is well known [11] as

$$l_t^\alpha = \binom{r}{\alpha} e^{-p\alpha t}(1-e^{-pt})^{r-\alpha}. \tag{2}$$

The interpretation of (2) is that the lives start at age 0 in an identical condition. If however the initial population is not homogeneous in this respect but instead the lives are so distributed that the proportion with value r is proportional to $D/(1+D)^{r-\alpha+1}$ then it can be shown [12] that

$$l_t^\alpha = \frac{D(1+D)^\alpha e^{pt}}{(1+De^{pt})^{\alpha+1}}$$

and
$$l_t = \frac{(1+D)^\alpha}{(1+De^{pt})^\alpha}$$

whence
$$\mu_t = \frac{p\alpha De^{pt}}{1+De^{pt}}.$$

10. It is obvious that many different formulas could be derived by making different assumptions regarding the transition arrangements from state α to state $\alpha+1$, the initial distribution at $t=0$ and the definition of death. For the present purposes it will however be sufficient to restrict consideration to a process of the type defined by equation (1). Furthermore it would be possible to extend the study to processes which are continuous in both the state (α) and the time (t) dimensions, as distinct from the discrete variation in the state dimension. This leads to models which are akin to diffusion process of physics but since it has been shown that such processes can be approximated by discontinuous processes [11] there is no advantage in following this development in this paper.

11. Before leaving this latter point it may however be noticed that in his "general theory of mortality", C. D. Rich [13] derived an equation of the form

$$\frac{\partial f(x, r)}{\partial x} = -\frac{\partial}{\partial r}\{r\lambda_x f(x, r)\}$$

where $f(x, r)$ represented the distribution of lives at time x exhibiting a state of health r and λ_x measured the force of deterioration. By suitable adjustment of the time scales this he transformed to

$$\frac{\partial g(y, r)}{\partial y} = -\frac{\partial}{\partial r}\{r\lambda g(y, r)\}.$$

In this model it is assumed that increases in r correspond to decreases in "health"; the model corresponding to equation (1) above would be defined by

$$\frac{\partial g(y, r)}{\partial y} = \frac{\partial}{\partial r}\{r\lambda g(y, v)\}$$

which is seen to be the form taken by (1) when the interval on the right-hand side is allowed to decrease to zero. However the resulting equation loses its stochastic nature and the model reduces to a deterministic form.

It is however not without interest to note that the "backward" version of Rich's equation leads to analytical forms for μ_x which are more immediately recognised as approximations to the observed facts.

12. Returning now to equation (1) it is pertinent to ask whether it is likely to be directly applicable to observed mortality data. The over-all mortality of a population is made up of deaths arising from many causes; it is reasonable to assume that since a group of births at a given time are genetically different (apart from the small proportion of identical twins) their reaction to their subsequent environment, even if this be identical, will be different and their resulting mortality will in various ways reflect their different genetic constitutions. If consideration is limited to the study of deaths from one particular cause it cannot be assumed that the mortality pattern is a reflection of the degenerative process leading to death without making some assumptions as to the susceptibility of the population to the particular cause. As a necessary consequence it can also be noted that comparison of death rates from particular causes in different populations may also give rise to erroneous impressions. The significance of these points can be easily illustrated by considering hypothetical groups of births for which only two causes of death are possible, one at a young age and the other at an advanced age. If some only of the group are subject to the "young age" cause of death then the over-all comparisons for this cause will be essentially a comparison of the different proportions susceptible.

13. If the assumption be made that causes of death are independent, an obvious oversimplification in general, but of possible accuracy in some cases and sufficiently so in many others, it is possible to resolve the difficulty presented in paragraph 12 in a fairly convenient manner by focusing attention on the curve of deaths and not on the rates of mortality [14]. Thus if it be assumed that the population at birth can be divided into two groups, the first of whom can die of any cause and the second of whom of any cause except say cause A, then all the "A" deaths must arise from the first group and it can be shown that if p_A is the proportion susceptible to death from cause A, μ_t^A the rate of mortality from cause A on its own and μ_t^a the rate of mortality from cause A related to all the survivors at age t then the relationship $p_A\mu_t^A l_t^A = \mu_t^a l_t^a$ holds. In other words, if the age specific independent rates for cause A are found from the data the "pure" specific rates μ_t^A can be deduced by a relatively long but straightforward process. Alternatively the curves of deaths are identical except for the multiplier p_A.

14. However, a still more convenient approach exists, although estimation difficulties are involved, by taking the logarithmic differential of the curve of deaths, i.e.

$$\frac{d}{dt}\log \mu_t^a l_t^a = \frac{1}{\mu_t^a}\frac{du_t^a}{dt} - \mu_t^a = \frac{1}{\mu_t^A}\frac{d\mu_t^A}{dt} - \mu_t^A.$$

If the data are sufficiently extensive for the differential of μ to be determined then the form of this function can be determined from the observed data without a great deal of intermediate calculations.

This clearly provides a very convenient function for considering whether the mortality can have arisen from a stochastic process of given form, since general shape considerations are often sufficient to rule out many of the possible processes which can be written down.

15. It will have been appreciated that the shift of emphasis from rates of mortality to the curve of deaths brings into natural relief the distinction between life tables based on the mortality experienced in a given calendar year (or group of years) and that based on the subsequent mortality of a group of lives born at the same time, i.e. a generation life table. The importance of keeping these concepts clearly in mind does not need stressing for actuaries, but for analysis purposes some further elaboration seems desirable. In order to find a "pure" mortality curve, i.e. one reflecting only a particular cause of death, it is necessary to eliminate as many significant external factors as possible. For reasons given above there is every reason to expect the generation to be a significant factor, i.e. classification should be made by year of birth. It is also well known that mortality is sensibly influenced by environmental conditions and climate is a very important factor in giving rise to variations, both in the form of short-term movements and probably in the long term as well. In other words, the year of experience is also a proper factor to be considered. Finally there is a highly significant variation with age (in terms of stochastic processes this arises of course as a consequence of the efflux of time). In general then it seems as though the rate of mortality should be expressed as

$$\mu_x^{\mathrm{T}} = k\phi(\mathrm{T}-x)\psi(\mathrm{T})f(x) \qquad (3)$$

and in the particular application discussed below it is assumed that the force of mortality is that for a particular cause operating on its own, i.e. independent of other causes.

16. If it is assumed that $\phi(\mathrm{T}-x)$ is constant, i.e. that year of birth is not a significant factor, then μ_x^{T} is dependent on attained age and on the year of experience. If $\psi(\mathrm{T})$ is assumed constant then mortality is dependent on attained age and year of birth, i.e. a generation (or cohort) form. Now, since there are only two independent variables involved, T and x, there is a formal relationship between the three functions; in fact the form

$$\mu_x^{\mathrm{T}} = k\{\phi(\mathrm{T}-x)\varepsilon^{\mathrm{T}-x}\}\{\psi(\mathrm{T})\varepsilon^{-\mathrm{T}}\}\{f(x)\varepsilon^{x}\}$$

is identically equivalent.

If a set of values of μ_x^{T} for a range of values of T and x is available it is formally possible to find a set of numerical values for the functions ϕ, ψ and f which can be regarded as "fitting" the observed values according to some criterion. By selecting various values of ε different sets of the functions can be obtained and the selection of any one particular set must be determined according to consideration of the nature of the functions ϕ and ψ. It should be noted that the form of the "true" mortality function is related to the functions ϕ and ψ and until these have been fixed the shape of $\{f(x)\varepsilon^x\}$ cannot be determined.

17. The function ϕ is likely to be expressible as a continuous function since genetic changes take place slowly and $f(x)$ is also likely to be continuous since the deterioration which it can be assumed gives rise to mortality will be a progressively increasing function with age. ψ on the other hand is likely to be an irregular function since it will reflect changes in environmental conditions, such as climatic conditions, and will also reflect changes in definition which will be of a discontinuous nature.

18. If it is found possible to represent μ_x^T by this triple product the way becomes clearer for a more reliable numerical model for the "true" mortality and attempts to determine whether it can be described by some form of stochastic process becomes more hopeful of success. In particular it would become possible to see whether the Gompertz law, which is widely adopted as a model, is really appropriate.

19. The techniques outlined in the foregoing have been used in experiments on the mortality from lung cancer in the U.K. and the remainder of this paper is concerned with the results of these experiments. A more detailed description of some earlier results has been given in reference 15.

20. The raw data for the experiments consist of the annual statistics by the Registrar-General and the convenient summary of the death rates grouped in quinary age groups and quinary calendar years provided in reference 16 has been used, supplemented by the corresponding data for the years 1956-60. It has been assumed that the tabulated rates can be taken as the force of mortality for lung cancer independent of other causes of death at the centre of the age and year groups. From this set of data, i.e. the force of mortality at quinquennial ages for the successive quinquennia from 1911-16 to 1956-60, three sets of relationships of the type

$$\frac{\mu_{x+5}^{T+5}}{\mu_x^T}\cdot\frac{\mu_x^{T-5}}{\mu_{x+5}^T}=\frac{\psi(T+5).\psi(T-5)}{\{\psi(T)\}^2}=\bar{\psi}(T)$$

were first calculated for the functions ϕ, ψ and f respectively. Numerical integration of the three functions of type $\bar{\psi}$ then provided values of each of the three functions ϕ, ψ and f in terms of two integration constants and after some further calculations a set of values was found which provided a reasonable approximation to the observed values of μ_x^T over the range of 50 calendar years and for ages 20-80. The calculations were made for both male and female lives and although the results are reasonably satisfactory a more systematic method is very desirable. The use of an electronic computer seems the only practical way and studies of this are being undertaken.

21. The work to this stage had provided numerical values of three functions ϕ, μ and f which when multiplied together provided a reasonable approximation to the observed values of μ_x^T i.e.

$$\mu_x^T = \phi(T-x)\psi(T)f(x)$$

and it was next necessary to find some criteria, for fixing the value of ε so that some meaning could be atttached to the function, i.e.

$$\mu_x^T = \{\phi(T-x)\varepsilon^{T-x}\}\{\psi(T)\varepsilon^{-T}\}\{f(x)\varepsilon^x\}.$$

22. There has been considerable evidence obtained from studies of selected groups that the observed increase in the mortality from lung cancer is associated with smoking, in particular the smoking of cigarettes. If this hypothesis were true this mortality model seemed worth while investigating in terms of the facts regarding smoking habits throughout the population. ϕ is defined as a function solely dependent on year of birth but it can also be looked upon as a measure of the proportion of births who will ultimately die from lung cancer. If there is a strong relationship between lung cancer mortality and smoking, then ϕ will be closely related to the proportion of smokers in the generation concerned. Thus we would have $\sum_x P_x^T\phi(T-x)$ as a measure of the number of smokers in the population at time T. Furthermore, $\psi(T)$ in addition to reflecting classification changes could also be looked upon as a measure of the relative amount smoked at time T so that $\sum_x P_x^T\phi(T-x\psi(T)$ would be a measure of the total relative cigarette consumption at time T.

23. Some observations of these two quantities are available for the U.K. [17] and it was found possible to obtain sets of values of ϕ, ψ and f for males and females which were not unreasonable in the light of the cigarette hypothesis provided a time lag of 10-15 years was used. The numerical values are given in the appendices and the expressions for the mortality rates were

$$\text{Males} \quad \mu_x^T = \cdot 00508\ \phi^M(T-x)\psi^M(T)f^M_{(x)}$$
$$\text{Females} \quad \mu_x^T = \cdot 00445\ \phi^F(T-x)\psi^F(T)f^F_{(x)}$$

where ϕ measures the proportion of smokers in the population and ψ is a relative measure of average amount smoked among smokers, with an origin about 1925.

The similarity of the numerical constants was sufficiently encouraging to justify regarding the models as confirming the smoking hypothesis and thus to regard the values of $f(x)$ as a reasonable approximation to the "true" incidence of mortality from lung cancer. It will be noted that the values of $f(x)$ for male and female lives are not significantly out of line with each other; in other words when allowance is made for the incidence of smokers and of relative consumption the incidence of mortality from lung cancer is similar in the two sexes.

24. It was now felt reasonable for some studies to be made on the form of the "pure" mortality $f(x)$ which, if the hypotheses regarding ϕ and ψ are true, may be looked upon as a measure among those who smoke cigarettes of the incidence of death from lung cancer, operating independently of other causes of death, and approximately standardised for the consumption of cigarettes. It was first noted that the data did not support the Gompertz form, the rate of increase of $f(x)$ falling with increasing age. It was also apparent that the Perks (or logistic) form also failed to fit the data. Accordingly consideration was given to some form of random process. The clinical evidence now accumulating in regard to the effect of cigarette smoking points to a progressive change in the bronchial linings until a stage is reached when the cancer develops [18]. This suggests the "backward" type of model, i.e.

$$\frac{dl_t^\alpha}{dt} = -p\alpha l_t^\alpha + p(\alpha+1)l_t^{\alpha+1} \tag{4}$$

where l_t^α can be looked upon as the proportion of persons at age t whose condition has deteriorated to α units from an initial state of r units; p represents the chance of losing a unit and is interpreted as being the effect of smoking on the bronchial tissues. It will be noted that if at some stage p were to become zero (equivalent to stopping smoking) the process would cease and further deterioration would stop. This conforms to the observed reduction in the incidence of lung cancer among persons who have stopped cigarette smoking.

25. The "deaths" at time t from process (4) are

$$\mu l_t = p\alpha \binom{r}{\alpha} e^{-p\alpha t}(1-e^{-pt})^{r-\alpha}$$

and the logarithmic differential of this is

$$-p\alpha + \frac{p(r-\alpha)}{e^{pt}-1} \tag{5}$$

This function is infinite at $t=0$ and decreases steadily with t, passes through zero and finally asymptotes to $-p\alpha$ when $t=\infty$ The values found from the numerical values of $f(x)$ increase from the youngest age (22) to a maximum at about age 35 and then decrease so that at first sight the model is unsuitable.

However, it has been recognised that lung cancers are of a number of different types and it has been suggested that it is the epidermoid and oat cell types which arise from smoking and that the other types are of different origin.

26. It is impossible to obtain information regarding the types of cancer for the data concerned but some fairly reliable observations regarding the proportions of the different types are available from a Norwegian study [19]. Clearly these data cannot be properly used on the U.K.

statistics without a comparison of the smoking habits of the two populations, but in default of better information it can be used as a crude approximation. The following table gives the proportion of epidermoid and oat cell carcinomas to the total lung cancers in different age groups for the Norwegian data (males and females combined):

Age-group	–29	30–39	40–49	50–59	60–69	70–
proportion	·00	·15	·66	·77	·74	·65

27. From a rough graphical graduation of these figures proportions appropriate to the values of $f(x)$ can be found and the following table gives the derived relative pure mortality rates for the epidermoid and oat cell carcinomas among male lives in the U.K.:

Age		Age	
17	·0	52	116·0
22	·0	57	258·0
27	·0	62	443·0
32	·2	67	641·0
37	2·7	72	777·0
42	14·0	77	925·0
47	48·0	82	937·0

As compared with the value of $f(x)$ this function shows high contact at the younger ages; the slope relation has an infinite value in the age 20-30 group and the general behaviour is very much that of formula (5). Calculation shows that the curve of deaths can be approximately represented by

$$\mu l_t = {\cdot}04\binom{8{\cdot}33}{7}e^{-{\cdot}04t}(1-e^{-{\cdot}03t})^7 \qquad \text{origin } t=27.$$

In other words the mortality process can be represented by

$$\frac{dl_t^\alpha}{dt} = -{\cdot}03\alpha l_t^\alpha + {\cdot}03(\alpha+1)l_t^{\alpha+1}$$

where α is initially 8·33 for all individuals and death is assumed to occur when it falls to 1·33, the process starting in the 20–30 age-group. It is not without interest to note that the ratio (·16) of the final value of α to the initial value falls within the limits proposed by Szilard [20] in a paper on a theory of ageing based on considerations of damage to chromosomes.

28. This representation of the mortality process as a pure death process is, of course, approximate only and further experiments are needed but the present data do not justify closer investigation. A difficult point arises from the interpretation to be attached to the transition probability p which is rather loosely dealt with in paragraph 24. In deriving the values

of the pure mortality rate $f(x)$ it was assumed in the model that a proportional relationship (with a time lag) existed between the mortality rate and the average amount of cigarettes smoked per smoker. If the transition probability p is assumed to be dependent on the quantity of tobacco smoked then a much more complicated analysis which allowed for a distribution of smokers by amount smoked rather than the averaging feature of the model would be required. It does not appear from the numerical values that such a model would lead to any different conclusions, but since these are based on the use of the Norwegian data for cancer types some doubt must remain until more adequate data become available or can be investigated.

29. The rather elaborate reduction of the data outlined in the foregoing analysis brings out the difficulties to be faced in trying to determine whether mortality can be expressed in terms of stochastic processes. It seems reasonable to conclude that the statistics of deaths from lung cancer in England and Wales are not incompatible with the hypothesis that the incidence of the particular type of lung cancer currently thought to be related to smoking habits, particularly of cigarettes, can be described as a stochastic process of pure death type. Whether other causes of death can be similarly analysed and stochastic processes found to describe the mortality process remains to be determined, but it is suggested that the extension of actuarial technique described in this paper does provide a possible means of study which may lead to useful additions to the knowledge of mortality.

REFERENCES

1. Bartlett, M. S., *An Introduction to Stochastic Processes*, Cambridge University Press, 1956.
2. Cramér, H., For a comprehensive bibliography see for example *Collective Risk Theory, a survey from the point of view of the Theory of Stochastic Processes*, Skandia Jubilee Volume, Stockholm 1955.
3. Lundberg, Ove, *On Random Processes and their Application to Sickness and Accident Statistics*, Uppsala, 1940.
4. ASTIN Bulletin, Vol. 1, Parts 2 and 5, for bibliographies (1959-1961).
5. Gompertz, B., *Phil. Trans.*, Vol. 115, 513 (1825).
6. Makeham, W. M., *J.I.A.*, Vol. 13, 325 (1867).
7. Perks, W., *J.I.A.*, Vol. 63, 12 (1932).
8. Phillips, W., A Basic Curve of Deaths, *J.I.A.*, Vol. 80, 289 (1954).
9. Sacher, G. A., and Trucco, E., "The Stochastic Theory of Mortality", *Ann. N.Y. Acad. Sc.*, Vol. 96, 4 (1962).
10. Armitage, P., and Doll, R., *Stochastic Models for Carcinogenesis*, Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol. IV, 19 (Ed. Jerzy Neyman).
11. Bharucha Reid, A. T., *Elements of the Theory of Markov Processes and their Applications*, McGraw Hill, 1960.
12. Beard, R. E., *Ciba Foundation Colloquim on Ageing*, Vol. 5, Appendix, J. & A. Churchill, London 1959 (Eds. Wolstenholme and O'Connor).

13. Rich, C. D., *J.I.A.*, Vol. 70, 314 (1939).

14. Prawitz, H., "Investigation of Mortality from Different Causes of Death as a basis for forecasting the future trend", *Svenska Aktuarieforeningen Femtio Ar.* (1954).

15. Beard, R. E., (Royal Society Meeting on Demography, to be published? June 1963).

16. General Register Office, Studies on Medical and Population Subjects No. 13 (1957).

17. Todd, G. F. (Ed.), Tobacco Manufacturers Standing Committee, Research Papers No. 1, 3rd Edition (1962).

18. *Lancet*, The Cigarette as Co-carcinogen, p. 85, Jan. 13 (1962).

19. Kreyberg, L., *Histological Lung Cancer Types*, Norwegian University Press, Oslo (1962).

20. Szilard, L., *Proceedings Nat. Academy Sciences*, Vol. 45, 30-45, Jan. (1959).

APPENDIX I

x	$f(x)$		T	$\psi(T)$		$(T-x)$	$\phi(T-x)$	
	Males	Females		Males	Females		Males	Females
17	·6		1913	1·09	1·72	1836	·006	·005
22	1·0	1·0	18	1·00	1·14	41	·008	·005
27	1·9	2·2	23	·98	·99	46	·012	·007
32	4·0	5·3	8	1·04	·93	51	·018	·012
37	10·8	11·9	33	1·46	1·13	56	·025	·019
42	28·1	28·2	8	1·68	1·15	61	·036	·028
47	68·1	58·0	43	1·72	1·02	66	·056	·042
52	155·	112·	8	1·94	1·07	71	·082	·056
57	322·	218·	53	2·02	1·01	76	·131	·077
62	574·	376·	8	2·00	·90	81	·210	·103
67	892·	580·				86	·324	·129
72	1160·	850·				91	·420	·173
77	1492·	1180·				96	·595	·215
82	1645·	1460·				1901	·734	·286
						6	·855	·363
						11	·910	·450
						16	·905	·528
						21	·822	·538
						26	·760	·613
						31	·727	·628
						36	·570	·785
						41	·286	·980

$${}^{M}\mu_x^T = \cdot 00508\ \phi^M(T-x)\psi^M(T)f^M(x)\text{(per 1,000)}.$$

$${}^{F}\mu_x^T = \cdot 00445\ \phi^F(T-n)\psi^F(T)f_T(x)\text{(per 1,000)}.$$

Appendix II

Cancer of the lung and pleura—England and Wales
Age specific death rates per million living per year

Age	1916-1920		1926-1930		1936-1940		1946-1950		1956-1960	
	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.
MALES										
15–19	2	2	2	2	2	3	3	4	1	2
20–24	3	3	4	4	6	8	8	8	5	6
25–29	4	4	7	7	14	15	18	16	11	14
30–34	6	6	11	12	30	29	36	35	35	31
35–39	11	11	22	24	68	68	94	96	91	90
40–44	15	19	52	48	149	143	236	237	259	258
45–49	25	28	76	75	274	244	544	493	590	630
50–54	40	44	112	106	431	429	954	910	1,265	1,350
55–59	55	59	148	138	586	577	1,350	1,330	2,325	2,400
60–64	76	73	181	170	646	655	1,717	1,830	3,305	3,470
65–69	86	81	169	168	636	624	1,763	1,850	3,920	3,800
70–74	71	69	158	152	533	554	1,400	1,590	3,895	3,820
75–79	52	63	133	145	464	458	1,085	1,210	3,395	3,180
80–84	41	49	94	102	324	351	765	910	2,240	2,190
FEMALES										
20–24	1	1	1	1	3	3	3	3	3	3
25–29	2	2	2	3	5	5	6	6	4	6
30–34	3	3	3	5	8	10	12	13	15	13
35–39	5	6	9	8	16	17	24	26	35	27
40–44	10	11	16	15	32	31	48	50	60	60
45–49	14	16	23	25	49	51	73	79	104	104
50–54	22	24	32	36	78	74	117	115	170	164
55–59	34	31	58	51	107	113	169	180	241	249
60–64	36	36	61	66	153	149	221	231	330	325
65–69	34	35	69	68	179	166	302	295	383	401
70–74	38	30	74	68	192	183	316	311	447	439
75–79	29	30	73	59	184	157	309	315	489	485
80–84	31	37	49	42	152	142	280	292	451	450

Some observations on stochastic processes with particular reference to mortality studies

by

R. E. Beard (*Great Britain*)

(*Summary*)

The theory of stochastic processes has proved to be a branch of probability theory for which useful applications have been found in practically all branches of scientific activity. Although one of the earlier applications of the theory was concerned with insurance, namely the theory of risk

associated with the names of Filip Lundberg and Harald Cramér, and the activities of the ASTIN actuaries are concerned with various applications to non-life insurance, there has been little attention paid by actuaries to the possible application to the more traditional field of mortality studies.

The paper outlines some of the considerations which arise from attempts to apply the theory of stochastic processes to mortality data and summarises some experiments made by the author in analysing mortality from lung cancer.

It is shown that the statistics of deaths from lung cancer in England and Wales during the past 50 years are not incompatible with the hypothesis that the incidence of the particular type of lung cancer currently thought to be related to smoking habits can be described as a stochastic process of pure death type. The analysis also leads to a model in which the incidence for males and females is approximately the same when allowance is made for different smoking habits of the sexes.

Observations sur les processus stochastiques, compte tenu notamment des études de mortalité

par

R. E. Beard (*Grande Bretagne*)

(*Résumé*)

La théorie des processus stochastiques, branche de la théorie des probabilités, s'est avérée utilement applicable dans presque tous les domaines d'activité scientifique. L'une de ses premières applications intéressait les assurances: c'est la théorie du risque, qui évoque les noms de Filip Lundberg et de Harald Cramér. D'autre part, les travaux des actuaires de l'ASTIN sont consacrés aux applications de cette théorie en divers domaines d'assurances autres que l'assurance-vie. Mais les actuaires ne se sont guère préoccupés jusqu'ici d'étudier les applications de la théorie dans le domaine plus traditionnel des études de mortalité.

L'auteur définit brièvement diverses questions que suscitent certaines tentatives d'appliquer la théorie des processus stochastiques aux données relatives à la mortalité; il résume quelques-unes de ses expériences personnelles, tendant à analyser la mortalité du cancer pulmonaire.

Il démontre que les statistiques des décès attribuables au cancer du poumon en Angleterre et au Pays de Galles pour les 50 dernières années ne s'opposent pas à l'hypothèse suivant laquelle la morbidité d'une certaine forme du cancer pulmonaire, couramment attribuée à l'usage du tabac, peut être définie comme un processus stochastique du type de la mortalité pure et simple. L'analyse aboutit en outre à un modèle dans lequel la morbidité de la maladie chez les hommes et les femmes est à peu près la même, compte tenu de la différence entre les habitudes des deux sexes en matière d'usage du tabac.