JIA 70 (1939) 0314-0379

A GENERAL THEORY OF MORTALITY

By C. D. RICH, M.A., F.I.A.
Assistant Actuary of the Provident Mutual Life Assurance Association

[Submitted to the Institute, 24 April 1939]

THE PROBLEM

If a pebble on a steep hill is disturbed from a state of rest, it will probably roll down the hill gathering speed as it goes, and, if the hill ends in a precipice, the pebble will, unless it is checked in its course, eventually fall over the edge. So long as the pebble is rolling down the hill it is within sight of an observer on the hill, but when it falls over the precipice it passes out of his sight; its disappearance occurs suddenly. There is no violent change in the pebble's speed or direction as it passes over the edge, so the suddenness of its disappearance is not due to any alteration in its motion but to the fact that it has rolled down beyond the level of the edge of the hill.

Thus it is that death occurs. Suppose a man becomes ill, gets worse and dies. His death is instantaneous, but the cause of his death—deterioration of health—may have been progressing for some time. Death takes place because his health has deteriorated beyond a certain limit. If by some means it were possible to represent the man's state of health throughout his lifetime in the form of a graph, we should have a record of all the illnesses he had undergone and of his eventual death. Each illness would be indicated by a fall in the graph, the extent of the fall depending on the seriousness of the illness, and each subsequent recovery by a rise; death would be indicated by a fall to an extreme lower limit. Looked at in this way, death is only a single phase in a process continuing throughout life.

Reverting to our illustration, the observer could, by counting the number of pebbles that fell over the precipice, calculate the frequency with which the falls occurred. But he would probably learn much more about the phenomenon if he were to investigate what caused them to fall over. Actuaries are accustomed to measure mortality by counting the number of deaths among a group of observed lives, but is it not possible that we can learn more about it by studying deterioration of health? Suppose for instance that we were able to keep continuous health records of a large number of persons up to the time of death, is it not conceivable that they would teach us something new about mortality?

The practical difficulties in the way of carrying this out are that we have no convenient method of measuring health, and that even if we had it would be extremely difficult to collect the required data. We have of course data showing the mortality of persons who at some stated time satisfied a certain test of health. Thus we have mortality tables based on life assurance statistics, where at entry the lives were all of first-class health, and we have tables based on statistics of disabled lives, where at entry the lives all suffered from certain impairments. But none of these tables (whether in aggregate or select form) shows how the healths of the lives under observation varied after entry.

Although the practical difficulties in keeping continuous health records of a number of observed lives may be insurmountable, there is no reason why we should not speculate on the way in which in theory this could be done, and the purpose of this paper is to investigate whether anything can be learnt thereby. It will be necessary to devise a system for recording health observations, and we will see to what conclusions this leads us. To begin with, we must decide precisely what we are to mean by the expression "health", and in this connexion two hypotheses are formulated which it is believed will fit in with pre-conceived ideas. It will be found that a theory of mortality can be built up from these hypotheses, and it is hoped that the results so obtained may help to throw fresh light on the subjects of mortality and "selection".

Scientific accuracy is not claimed for the various physical analogies which are drawn in the paper; these are solely for the purpose of illustration. For convenience the greater part of the mathematical analysis has been segregated and formed into an Appendix.

THE DEFINITION OF HEALTH

The word "health" is used in this paper in a sense which means more than physical health and is intended to include consideration of all those factors such as habits, occupation, residence and environment which may affect the prospects of longevity.

It is often the practice of actuaries to divide lives according to their state of health into two classes: "select" lives, namely lives which are of first-class health and would be accepted for assurance by a life office at its ordinary rates of premium, and "damaged" lives, namely lives whose health is below this standard. It will probably be agreed that neither of these two classes is homogeneous; some select lives may be of better health than others, and damaged lives may vary from the person who only just fails to be ranked as first class to the person lying on his death-bed. In fact if the two classes are combined and the lives arranged in order of health, they will exhibit continuous variation from the best to the worst. The point of division between the select and the damaged lives is an arbitrary one and indeed its position is not fixed, for opinion as to what is meant by a first-class life may differ slightly as between one life office and another.

To repeat a simile made by H. E. Raynes (*J.I.A.* Vol. LXVIII, p. 256), 'the conception of dividing lives into two classes, the sheep and the goats, is fictitious and bears no relation to the facts; it is more probable that a whole Noah's Ark of animals exists'. If these words are interpreted literally, which of course was not Mr Raynes's intention, they mean that there are a limited number of states of health, all distinct from each other. It is preferable to assume that there are an infinite number of possible states of health forming a one-dimensional continuum; health may vary by infinitely small degrees between upper and lower limits. This then is our first hypothesis concerning health.

The range of possible states of health may be likened to the range of possible temperatures of water. Water can be of any temperature between freezing point and boiling point; health can be of any state between a limit of best health and a limit of worst health.

THE MEASUREMENT OF HEALTH

Temperature can be measured according to a fixed scale, and the temperature of any given article can be definitely determined. State of health, on the other hand, is largely a matter of opinion; this is partly because we are usually not fully aware of all the special factors which should be taken into account in assessing any particular life—medical science for instance is not far enough advanced to enable us to diagnose physical health infallibly—and partly because we have not sufficient experience to be able to deduce completely accurate conclusions from such information as is in our possession. Nevertheless, there must be a definite correct opinion regarding the state of health of any life, unable though we may be to arrive at it exactly, and we shall therefore proceed on this assumption.

To measure temperature it is necessary, first, to have an instrument which will indicate changes in temperature and, secondly, to calibrate the instrument so that it can be read. We will in our imagination devise a similar method for measuring health, and as an instrument for the purpose we will suppose we have an ideal referee or arbiter of health who possesses full knowledge of all the factors likely to affect the longevity of any life under consideration and is able to sum them up and arrive in his mind at a precise conclusion. Just as an uncalibrated thermometer can be used for purposes of comparison—it will tell us which is the hotter of two different temperatures, though it will not tell us by how much—so we will suppose that our referee records his opinion by comparing one life with another, and that given two lives of the same age he is able to say which is in the better state of health. He will be able to judge between two lives suffering from defects in health of such diverse characters that they may in the ordinary way appear difficult to compare. We may, if we like, assume that the person with the longer estimated expectation of life will be deemed to be in the better state of health, but this is not essential. It is necessary only that the referee shall adopt some consistent method.

We have referred to two lives of the same age; we shall obviously meet with difficulties if we attempt to compare lives of different ages, and there is no need to do this.

Given now any number of lives all of the same age, it will be possible by comparing them one with another to arrange them according to their states of health in order from the best to the worst. We may assume that we are never likely to find two lives of exactly the same state of health, for our referee will be able to judge with such accuracy that he will always find some slight superiority in one or other of the lives in the same way that a delicately adjusted thermometer can detect differences in temperature that would escape a cruder instrument.

We are now in a position to devise a health scale, or, in other words, to calibrate our instrument. To every possible state of health there must correspond some figure or "health rating" on the scale, and it will be convenient if we arrange that the better the state of health the lower the rating, the scale running from 0 to 1. Thus a life in perfect health, if such existed, would have the rating 0, while a life on the point of death would have the rating 1.

Supposing we have under observation a large group of lives all the same age, let us take a random sample and arrange the lives forming this sample in order of health. To the best life we allot a rating slightly greater than 0, to the worst life a rating slightly less than 1, and to the others we allot intermediate ratings, taken in order of health and distributed throughout the scale; the method of doing this is to a considerable extent arbitrary. Any other life of the group can then be given a rating by reference to and comparison with the lives to which ratings are already allotted; to do this we proceed to find the two such lives whose states of health are the next better and the next worse than the life in question and the latter is allotted a rating between those of the two so found. No life would ever be allotted the rating 0 or 1, for however good (or bad) the state of health of any particular life, it might always be possible to find one still better (or worse).

Such a method of rating is similar to what is known as the "decimal system" of indexing, used for instance in libraries, whereby the index can be extended indefinitely and an appropriate number in any desired position can always be found for any addition to the index that may be required.

The health ratings can be fixed in this way for every age, subject to the proviso that it is necessary to arrange for the scale to progress smoothly with change in age. In other words, if at age x a particular life is of rating r, then at age $x+\delta x$, where δx is small, he must be of rating $r+\delta r$, where δr is also small. The most straightforward way of effecting this would be to fix the ratings for each age in turn, commencing at the youngest age at which the lives come under observation and proceeding up to the oldest age.

The conditions to be observed in framing our scale are thus:

(1) At any age the ratings must all be between 0 and 1,

(2) Of two lives of the same age, the one in the better state of health must have the lower rating,

(3) For a small increase in age, the rating of any life must not vary by more than a small amount.

The only difficulty that could arise would be if the state of health of any life were to change so suddenly as to make it impossible to satisfy condition (3). It will be seen however from the next section of the paper that a sudden change of this kind cannot occur under our conception of health, and that if δx is made small enough, δr will also be small.

These conditions are sufficient for the theory developed in this paper. If however it were possible to make practical use of such a scale, it would also be advisable to arrange it so that the lives are at all ages well distributed throughout the range of ratings, and never "bunched up" in any one portion of it.

It will be clear that in framing the scale we shall have wide choice of method both in fixing the ratings at the youngest age and also in linking up the ratings between one age and another. As regards the former, this can be illustrated by a simple analogy. Suppose that we have a number of cubes of all sizes up to 1 cubic inch in volume, and that it is desired to devise a scale running from 0 to 1 by which we can record their sizes. Among the many ways of doing this are three obvious ones, namely, to record the size of a cube according to (A) the length of an edge, (B) the area of a face, (C) the volume. Thus a cube whose edge is of length half-an-inch would be rated at ·5 under scale A, ·25 under scale B, and ·125 under scale C. Diagram I shows how the scales would compare; the dotted lines join corresponding ratings on the three scales.

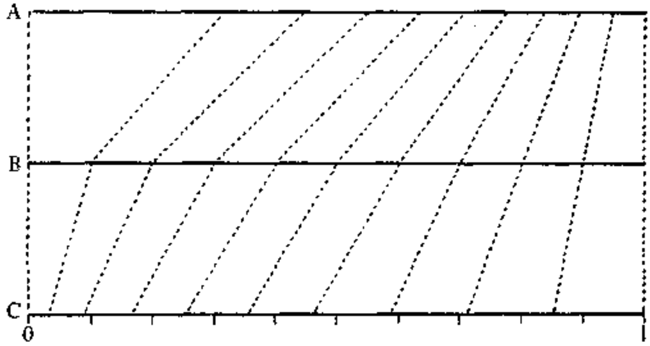


Diagram I. Comparison between different scales.

If we have two different health scales, then at any particular age any state of health will be represented, say, by rating r under one scale and by rating r' under the other. We can plot the corresponding values of r and r' in a graph; the resulting curve will represent the relationship between the two scales, and will take a form such as that shown in Diagram II. This curve will run from the origin to the point (1, 1), and, since an increase in r necessarily means an increase in r', the slope of the curve will always be positive.

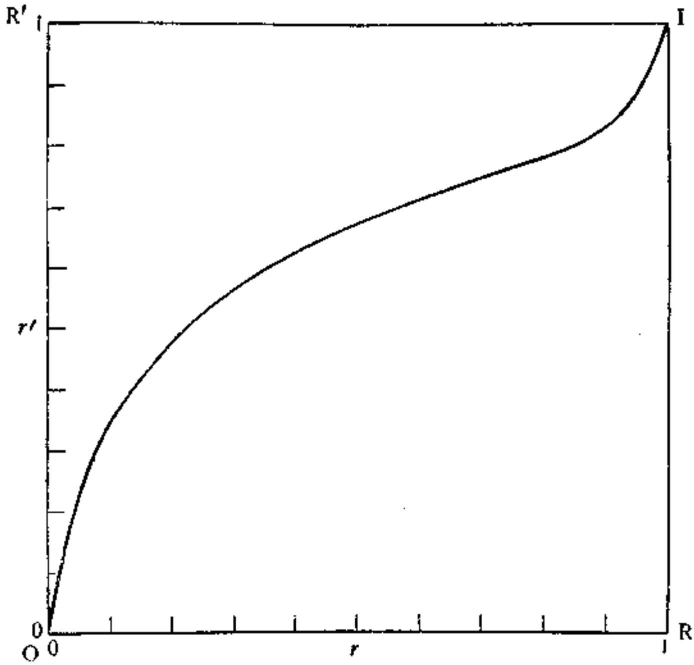


Diagram II. Relationship curve between two scales.

VARIATION IN HEALTH

Under a health scale such as has been described, the rating of any person will vary with the lapse of time corresponding to changes in his health due to illnesses which he may experience and special risks which he may undergo, all of which will, temporarily or permanently, affect his prospects of longevity. (It may also happen that on account of the way in which the scale has been framed the rating will vary without any change in his health; this means that the scale would probably not be very suitable for practical purposes.) The degree of variation in the rating when a person becomes ill will of course depend on the seriousness of the illness, and when he recovers on the completeness of the recovery. Even slight illnesses, such as common colds, may increase the possibility of more serious ailments developing, and will therefore be reflected by commensurate variations in the rating.

Our second hypothesis concerning health is that, though variations in rating may sometimes take place quickly, they are continuous and cannot take place instantaneously; thus a rating cannot change from one value to another without passing through all intermediate values, and a life cannot die without the rating attaining the value 1 at the instant of death. This is similar to saying that water cannot change in temperature from 40° C. to 60° C. instantaneously but only by passing through all temperatures between 40° C. and 60° C., and that water of temperature 40° C. cannot turn into steam until it has reached a temperature of 100° C.

This hypothesis may seem difficult to reconcile with the known occurrence of "sudden" deaths and accidents. Take, for example, the case of a person in good health who is walking in a street when a piece of stonework breaks off from a neighbouring building and falls on him causing his immediate death. It may appear that there is no interval of time in which the health rating could pass from its original value through all higher values up to 1. Nevertheless, on consideration it will be realised that however quickly death may take place, it is never entirely instantaneous; there will be a small fraction of time between the moment the person is hit and the moment he dies. Furthermore, he will commence to be in danger of his life when the piece of masonry starts to fall off the building; our imaginary referee would observe this and from the moment there was a possibility of the person being hit, his rating would commence to rise. Going further back, the referee would have been aware for some time that the building was in a dangerous condition, and the rating therefore would have begun to rise as soon as the person was likely to walk along the street in question. This implies that the ratings of all persons likely to use the street would rise temporarily while the danger existed.

No event, however suddenly it may appear to happen, is without its cause, and since the state of health of any life is assumed to include consideration of all factors which may affect the prospects of longevity, it is seen that our hypothesis will fit in satisfactorily with cases of supposed sudden death or accident. It also follows that the ratings of all lives may rise or fall quite frequently owing to possible dangers which may be incurred and of which the lives themselves may often be unaware.

THE HEALTH FREQUENCY CURVE

If we keep under observation a group of lives all of the same age, then at any moment of time the numbers of lives corresponding to different health ratings will form a frequency distribution, which we shall call the "health frequency distribution". The corresponding curve will be called the "health frequency curve". As time elapses the ratings of some lives may rise, of others fall, and some lives may die, with the result that the distribution will be undergoing continual change or "movement". The nature of this movement will now be considered. We will use x to denote time or age (which are synonymous) reckoned from the time the lives enter into observation, r to denote rating, and $f(x, r)$ to denote the height of the health frequency curve at time x for rating r. The area enclosed by the frequency curve will represent the size of the group; thus $\int_0^1 f(0, r)\, dr$ will be the number of lives entering into observation at time 0, and $\int_0^1 f(x, r)\, dr$ will be the number of survivors at time x.

In the short interval of time from x to $x+\delta x$ a certain number of lives will have ratings which pass through the value r, either passing upwards from a value lower than r to one higher than r or downwards from a value higher than r to one lower than r. The ratings of these lives will pass through the value r with speeds which may be fast or slow, but since by hypothesis all variations in rating are continuous, they will in a short interval of time change only by small amounts and can be regarded as having had values in the neighbourhood of r at time x. The numbers of the lives whose ratings pass upwards or downwards through the value r will therefore depend on the height of the health frequency curve in this neighbourhood, and the ratios which these numbers reckoned per unit of time bear to $f(x, r)$ we shall call the "forces of deterioration and recuperation", and we shall denote them by $\phi_1(x, r)$ and $\phi_2(x, r)$ respectively. The forces of deterioration and recuperation so defined will apply only to the group of lives under consideration, and will not necessarily be the same for a different group of lives.

We shall use the expression "flow through rating r" to mean the number of lives whose ratings pass through the value r. Thus the upward and downward flows through rating r in the interval from x to $x+\delta x$ will be $\phi_1(x, r)f(x, r)\,\delta x$ and $\phi_2(x, r)f(x, r)\,\delta x$ respectively.

The difference between the forces of deterioration and recuperation, i.e. the ratio which the net upward flow through rating r per unit of time bears to $f(x, r)$, will be called the "net force of deterioration" and will be denoted by $\phi(x, r)$. Thus

$$\phi(x, r) = \phi_1(x, r) - \phi_2(x, r).$$

The net force of deterioration will in general be positive. It may happen for some values of x and r to be negative, but for any given value of r it cannot remain negative indefinitely, for if so the net upward flow through rating r would always be negative, and the number of lives of rating lower than r would increase continually, which is impossible since the lives must die sooner or later. If our scale for measuring health is well devised, the net force of deterioration will always be positive.

Since there can be no flow in either direction through rating 0, the forces of deterioration and recuperation must be nil for this rating; and since there can be no downward flow through rating 1, the force of recuperation must here also be nil. On account of temporary illnesses and risks, variations in ratings of all lives may be frequent; the upward and downward flows may therefore be of considerable magnitude through all ratings except those near 0 or 1, and it is likely that, except for extreme values of the rating, the forces of deterioration and recuperation will be large compared with the net force of deterioration. We can thus obtain a rough idea of the forms taken by these forces, which will be somewhat as shown in Diagram III.

At time x, the upward flow through rating 1 per unit of time will be $\phi(x, 1)f(x, 1)$. This will be the rate at which deaths occur, and this rate is therefore dependent on the height of the health frequency curve for rating 1 and on the net force of deterioration for the same rating. Dividing by the number of persons alive at time x, we obtain the following expression for the force of mortality:

$$\mu_x = \frac{\phi(x, 1) f(x, 1)}{\int_0^1 f(x, r)\, dr}.$$

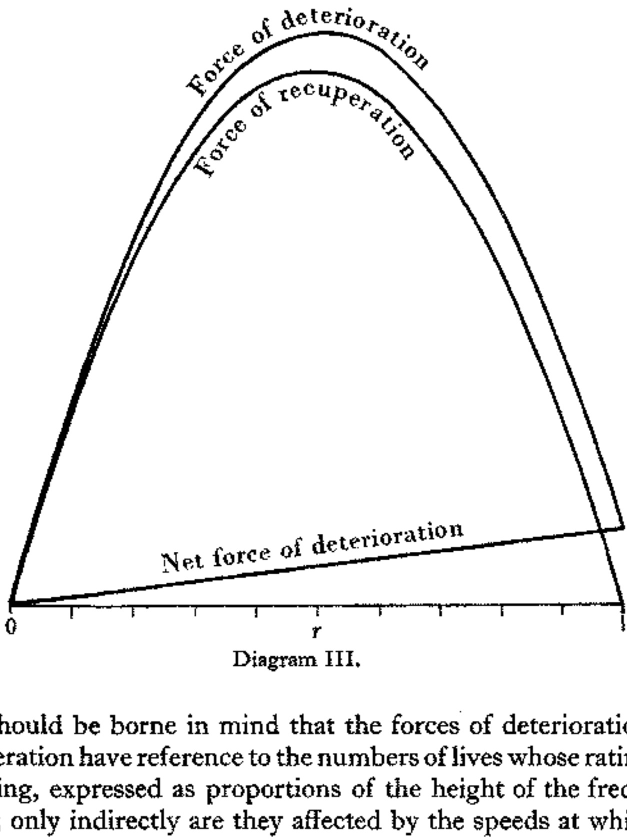


Diagram III.

It should be borne in mind that the forces of deterioration and recuperation have reference to the numbers of lives whose ratings are changing, expressed as proportions of the height of the frequency curve; only indirectly are they affected by the speeds at which the ratings of the individual lives are changing. The net force of deterioration may be looked upon as the function determining the general tendency of health to deteriorate, leading to eventual death. If the force of recuperation were always nil, the ratings of all lives would rise continually, some perhaps more rapidly than others, but on average according to this general tendency. The magnitude of the force of recuperation determines the extent by which there will be deviations from the general tendency, these deviations taking place equally in both directions.

We have now carried out our investigation to a point at which we can form a picture of how the mortality of a group of lives is connected with state of health. The force of mortality at any time is dependent both on the health frequency distribution and on the net force of deterioration for rating 1 at that time. Given the net force of deterioration for all values of x and r, and given also the health frequency distribution at time 0, then we have sufficient data to determine the health frequency distribution at any subsequent time (see Appendix, § 1). In its general form, however, this problem is not easy to solve, and our next step is to find whether it can be simplified.

THE EQUAL-DISTRIBUTION SCALE

The health frequency distribution at any time will of course depend on the scale adopted for measuring health, and this can be chosen in many ways. Reverting to the analogy of the cubes, Diagram IV shows the three different forms the frequency curve of a number of cubes might take if the sizes of the cubes are recorded according to the three scales previously mentioned. It is supposed that the cubes are equally distributed from 0 to 1 if measured under scale B, i.e. according to the area of a face. If measured under scale A the frequency of distribution will be heavier for the higher ratings, and if under scale C it will be heavier for the lower ratings, being in fact infinite for rating 0. (In dealing with health frequency distributions, an infinite or nil frequency for any rating is undesirable, and we should therefore endeavour to frame our scale so as to avoid this.)

Let us now arrange our scale for measuring health so that the lives in the group under consideration will at all times be equally distributed over the range of ratings from 0 to 1. To make the necessary transformation from a scale already fixed, we find the 99 ordinates (percentiles) which will divide the area enclosed by the health frequency curve under this scale into 100 equal areas. The state of health represented by the first of such ordinates under the old scale will then be represented by ·01 under the new scale, that by the second ordinate by ·02, and so on. In general the state of health represented by any rating under the old scale through which the ordinate would cut off a proportion r of the total area enclosed by the curve will be represented by the rating r under the new scale. This is done for all ages. It will easily be seen that our new scale satisfies the three conditions mentioned on p. 318. It is of course necessary to assume that we are dealing with a large group of lives, otherwise the occurrence of a single death would cause a sudden jump in the ratings of all the lives.

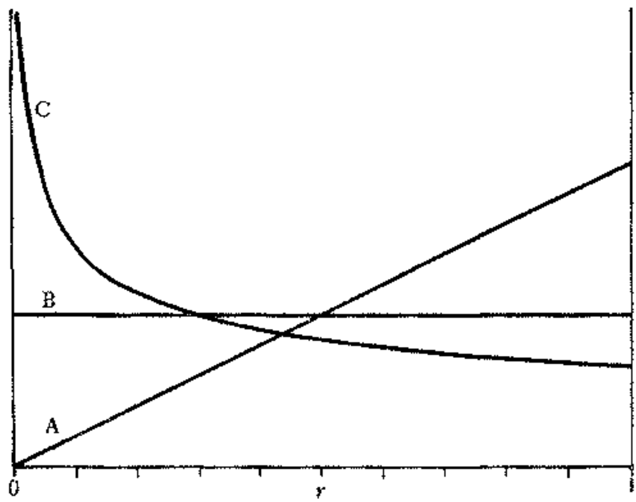


Diagram IV. Comparison between distributions under different scales.

A scale framed in this way we shall call the "equal-distribution scale". Such a scale leads to a simple form for the net force of deterioration, as will now be shown (see Diagram V). The frequency curve will be a straight line, the height of which at time x we shall denote by $f(x)$; $f(x)$ will be equal to the number of survivors at time x and will decrease as time elapses and as the lives gradually die. In a short interval of time δx, the height of the frequency curve will fall from $f(x)$ to $f(x+\delta x)$, and the number of lives of rating less than r will decrease by $r\{f(x)-f(x+\delta x)\}$. But this is the same as the net number of lives whose ratings have passed from a value less than r to a value greater than r in time δx, i.e. to $\phi(x, r) f(x)\,\delta x$. Hence $\phi(x, r) = r\,\frac{1}{f(x)}\,\frac{f(x) - f(x+\delta x)}{\delta x}$,

which in the limit becomes $r\left\{-\frac{d}{dx}\log f(x)\right\}$.

We can therefore write $\phi(x, r) = r\lambda_x$, where λ_x is a function independent of the rating, and which we shall call the "factor of deterioration". The factor of deterioration is always positive.

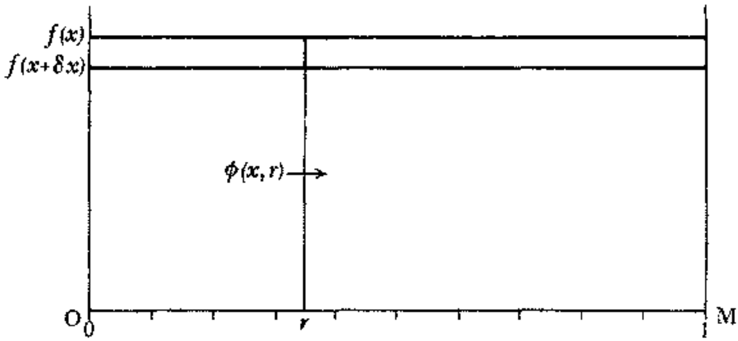


Diagram V.

The net force of deterioration will thus at any given time vary directly with the rating, it will be 0 for rating 0 and λ_x for rating 1. Its graph will take the form of a straight line (and has in fact been so drawn in Diagram III). At time x, the net upward flow through rating r per unit of time will be $\phi(x, r) f(x) = r\lambda_x f(x)$, the rate at which deaths occur will be $\lambda_x f(x)$, and the force of mortality will be $\mu_x = \lambda_x f(x) \div f(x) = \lambda_x$, i.e. the force of mortality is the same as the factor of deterioration.

It is interesting to observe that the flow in the frequency distribution under the equal-distribution scale is similar to the flow of water when allowed to run out of a rectangular tank. Suppose for instance that Diagram V represents a section of such a tank, there being an outlet hole at M. As the water runs out of the tank, its level will fall and there will be a flow of water from left to right across the tank. It is easily seen that at any given time the rate of flow through any cross-section of the tank is proportionate to the distance of this cross-section from the origin. If $f(x)$ is the level of the water at time x, and $\lambda_x f(x)$ the rate at which the water runs out at M, then the rate of flow through a cross-section distant r from the origin will be $r\lambda_x f(x)$. Now suppose further that the water in the tank is stirred round so that currents are set up in it, water moving both from left to right and from right to left at varying speeds, but the net flow remaining unaltered; we then have an almost exact picture of the flow in the frequency distribution.

The only difference is that in the case of the health frequency distribution λ_x is an unknown variable, whereas in the case of the water in the tank λ_x obeys a prescribed rule. The rate at which the water runs out, $\lambda_x f(x)$, depends on the pressure at the bottom of the tank, i.e. on the height of the water, and if we assume that the rate is directly proportional to the height, then λ_x is constant.

This gives rise to the idea of altering the measure of time in such a fashion as to replace λ_x for the health frequency distribution by a constant λ. Let us introduce a variable measure of time by which the length of a unit of time will be proportionate to the factor of deterioration; reckoned by such a measure the factor of deterioration will be constant. If y represents the time measured in this way, then $\lambda\delta y = \lambda_x \delta x$, i.e.

$$y = \frac{1}{\lambda}\int_0^x \lambda_x \, dx.$$

The artifice of a variable measure of time will be found considerably to simplify some of our work. The value λ assigned to the factor of deterioration by the variable measure is, of course, arbitrary. Ordinarily λ_x (which under the equal-distribution scale is the same as the force of mortality) will increase with x, so that the length of a unit of time by the variable measure will decrease as time elapses.

The function $g(y, r)$ will be used to denote the health frequency distribution of a group of lives when the variable measure of time is used, so that $f(x, r) \equiv g(y, r)$.

NATURAL SCALES

Any health scale under which the net force of deterioration of a group of lives takes the form $r\lambda_x$ will be called a "natural" scale for that group. The equal-distribution scale is a natural scale; the existence of other natural scales is proved later. In what follows it is assumed that the initial health frequency distribution of the group forms a continuous curve.

The particular advantage in the use of a natural scale is that it enables us easily to find the health frequency distribution at any time x if we are given the distribution at time 0, by means of the fundamental relationship (see Appendix, § 2):

$$f(x, r) = e^{-\int_0^x \lambda_x dx} f\left(0, re^{-\int_0^x \lambda_x dx}\right).$$

If time is reckoned by the variable measure this relationship becomes

$$g(y, r) = e^{-\lambda y} g(0, re^{-\lambda y}),$$

and means that the height of the frequency curve at time y for rating r is $e^{-\lambda y}$ times its height at time 0 for rating $re^{-\lambda y}$. In other words, if the frequency curve at time 0 be reduced in height or "compressed" in the ratio $e^{-\lambda y} : 1$, and then "stretched" in the direction away from rating 0 towards rating 1 so that distances from rating 0 are increased in the ratio $1 : e^{-\lambda y}$, we shall obtain the frequency curve at time y. $e^{-\lambda y}$ is a positive proper fraction which decreases as y increases, so the longer the time that has elapsed the greater the degrees of compression and stretching. This compressing and stretching effect may be called the "receding" movement of the distribution; it is due to the fact that the ratings of the lives forming the group are on the whole receding to higher values. The receding movement is illustrated in Diagram VI which shows an imaginary frequency curve AB at time 0, and the consequential frequency curves A_1B_1, A_2B_2, A_3B_3, etc., at successive equal intervals of time reckoned by the variable measure.

The point P corresponding to rating r on the initial frequency curve will recede to the successive positions P_1, P_2, P_3, etc., and it will reach rating 1 at time y given by $e^{-\lambda y} = r$, i.e. $y = -\dfrac{\log r}{\lambda}$. It will be seen therefore that the configuration of the curve at time 0 between ratings 0 and $e^{-\lambda y}$ is sufficient to determine its complete configuration at time y; that part of the curve at time 0 above rating $e^{-\lambda y}$ has no effect after time y. When the curve AB is compressed and stretched in the manner described, part of it will extend beyond rating 1, and this part has of course no real meaning. For the sake of interest however it is shown in Diagram VI, the true frequency curve together with its continuation beyond rating 1 taking the successive positions $A_1B_1b_1$, $A_2B_2b_2$, $A_3B_3b_3$, etc. Since the ratio in

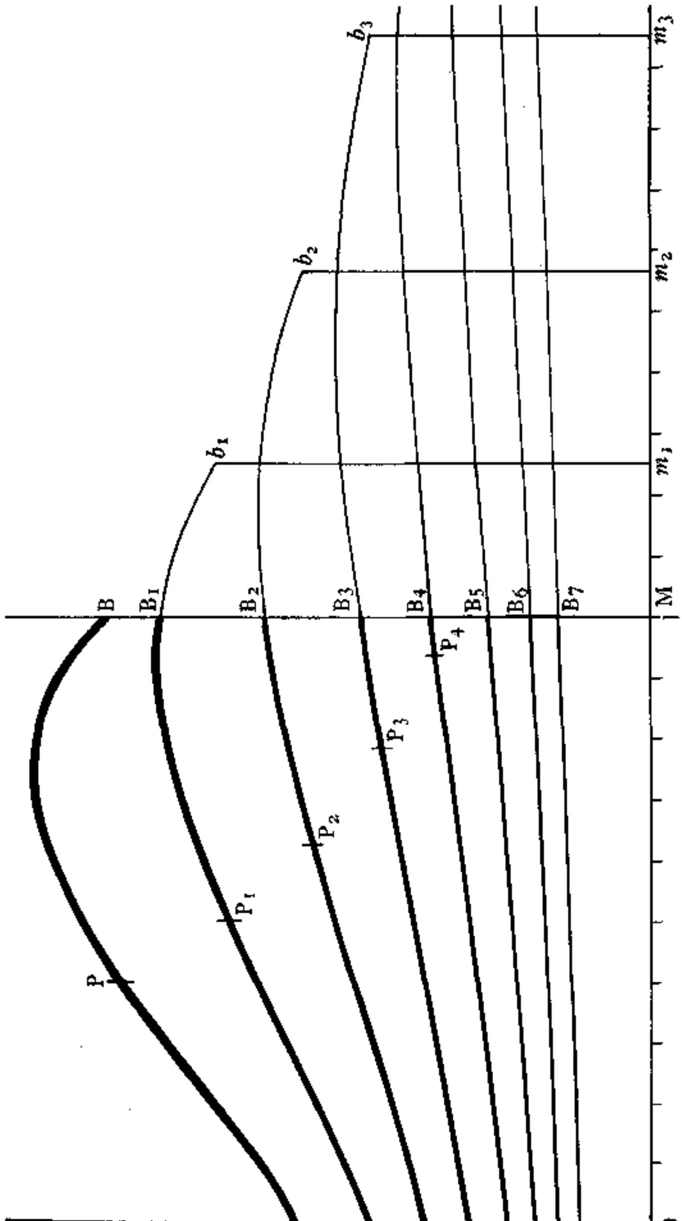


Diagram VI. Movement of the distribution under a natural scale.

which the frequency curve is compressed is the reciprocal of the ratio in which it is stretched, it is easily seen that the areas denoted in Diagram VI by ABMO, $A_1b_1m_1O$, $A_2b_2m_2O$, $A_3b_3m_3O$, etc. are all equal. Now the areas A_1B_1MO, A_2B_2MO, A_3B_3MO, etc. represent the numbers of survivors, and it follows that the areas $B_1b_1m_1M$, $B_2b_2m_2M$, $B_3b_3m_3M$, etc. represent the numbers of lives which have died.

The continual change in the configuration of the frequency curve is similar to that which we may suppose to take place in the surface of a viscous fluid which is allowed to run out of a rectangular tank having an outlet at one end of its base. As the amount of fluid in the tank diminishes, its general level will fall, the fluid will flow towards the end of the tank containing the outlet, and any irregularities in its surface will be stretched out in the same direction.

As will be seen from Diagram VI, the effect of the receding movement is that with the lapse of time the frequency curve is gradually flattened out. The survivors of the group of lives under consideration are therefore becoming more equally distributed over the range of ratings from 0 to 1, and we may say that the group is tending towards a state of equal distribution. (This does not apply if $f(0, 0)$ is zero or infinite.) Generally, however, the state of equal distribution is never actually attained.

Under a natural scale, the force of mortality will by the formula on p. 324 be

$$\mu_x = \frac{\lambda_x f(x, 1)}{\int_0^1 f(x, r)\, dr},$$

$$= \lambda_x \frac{f\left(0, e^{-\int_0^x \lambda_x dx}\right)}{\int_0^1 f\left(0, re^{-\int_0^x \lambda_x dx}\right) dr}.$$

Thus when x becomes large, $\frac{\mu_x}{\lambda_x}$ will tend to the limit $\frac{f(0, 0)}{\int_0^1 f(0, 0)\, dr}$,

which (unless $f(0, 0)$ is zero or infinite) is equal to unity, i.e. the force of mortality tends to become the same as the factor of deterioration.

The rate at which deaths will occur at time x is

$$\lambda_x f(x, 1) = \lambda_x e^{-\int_0^x \lambda_x dx} f(0, e^{-\int_0^x \lambda_x dx}).$$

The height of the frequency curve at time 0 for any rating is therefore directly connected with the rate at which deaths will take place at a definite time in the future, and the frequency curve at time 0 may thus be regarded as a distorted "curve of deaths".

The relationship between two different natural scales is investigated analytically in the Appendix, §§ 3–7. The existence of natural scales other than the equal-distribution scale is proved in § 4, where it is shown that, commencing at time 0 with any system of health rating we like (which means in effect that the initial health frequency distribution is at our choice), the system of rating at all subsequent times can be arranged so as to give a scale which is a natural scale. Every natural scale will have its own appropriate factor of deterioration λ_x, and its own appropriate variable measure of time by which the factor of deterioration will be a constant λ.

Since under every natural scale, the group of lives will tend with the lapse of time to a state of equal distribution, it follows that all natural scales tend to become the same, or in other words that the differences between the ratings of any life reckoned under the various natural scales will tend to disappear. Furthermore the factor of deterioration will tend to become the same on all natural scales, i.e. equal to the force of mortality. (These remarks do not apply to a scale under which $f(0, 0)$ is zero or infinite.)

THE NATURAL LAW OF MORTALITY

We have seen that for a given group of lives each natural scale has its own appropriate factor of deterioration λ_x. It is shown in the Appendix, § 7, that, given μ_x for all ages, we can, by suitable choice of natural scale (or, in other words, of the initial health frequency distribution) make λ_x take what form we please.

The formula
$$\mu_x = \lambda_x \frac{f(0, e^{-\int_0^x \lambda_x dx})}{\int_0^1 f(0, re^{-\int_0^x \lambda_x dx})\, dr}$$

(see p. 331) may be regarded as the natural law of mortality. The force of mortality of any group of lives can in theory, subject to its health frequency distribution forming a continuous curve, be expressed in this form, the function λ_x being at our choice, and the function $f(0, r)$ being the appropriate frequency distribution at time 0. Though the formula represents a general law it is not suggested that it is necessarily one convenient for curve-fitting; in fact as it stands it is probably of too general a character for this purpose. It seems doubtful however whether the existence can be proved of a general law of mortality of less wide scope, such as for instance the representation of the force of mortality by a formula containing a limited number of constants.

If for λ_x we choose a function whose graph is a perfectly smooth curve (e.g. a function of the Gompertz type bc^x, or of the Makeham type $a+bc^x$), then any irregularities in μ_x will result in corresponding irregularities in $f(0, r)$. Looking at the matter another way, λ_x can be regarded as an underlying true force of mortality, the actual force of mortality of the group being dependent both on λ_x and on the health frequency distribution. In a case where the group is equally distributed over the range of ratings from 0 to 1, μ_x will be always equal to λ_x. It should be noted (see Appendix, §7) that $f(0, 0)$ is zero, finite, or infinite according as $\underset{x\to\infty}{\mathrm{Lt}} \dfrac{\lambda_x}{\mu_x}$ is less than, equal to, or greater than unity.

The following special applications of the above formula for the force of mortality may be of interest. If λ_x be made constant and equal to k, the formula becomes

$$\mu_x = k\frac{f(0, e^{-kx})}{\int_0^1 f(0, re^{-kx})\,dr}.$$

If $f(0, r)$ can be expressed as a power series in the form

$$a+br+cr^2+\dots,$$

we have
$$\mu_x = k\frac{a+be^{-kx}+ce^{-2kx}+\dots}{a+\frac{1}{2}be^{-kx}+\frac{1}{3}ce^{-2kx}+\dots},$$
$$= \mathrm{A}+\mathrm{B}e^{-kx}+\mathrm{C}e^{-2kx}+\dots,\text{ say,}$$

which may be compared with Makeham's Law and the Multiple Geometric Law.

In the particular case when $\lambda_x = k$ and $f(0, r) = a + hr^n$,

$$\mu_x = k \frac{a + he^{-nkx}}{a + \frac{1}{n+1} he^{-nkx}},$$

$$= \frac{A + Bc^x}{1 + Dc^x}, \text{ say,}$$

which is the formula used so successfully by W. Perks (*J.I.A.* Vol. LXIII, pp. 12 *et seq.*). In our formula, however, the constants A, B, D, c are not entirely independent; they are connected by a certain relationship which is not found to be satisfied by any of the various sets of values given to the constants by Mr Perks in his applications of the formula.

It is interesting to observe that if λ_x is constant, an expression for the value of an assurance $\bar{A}_x$ can be found as follows. The rate at which deaths occur at age $x+t$ will be (see p. 332) $ke^{-k\overline{x+t}} f(0, e^{-k\overline{x+t}})$;

hence $$\bar{A}_x = \frac{k}{l_x} \int_0^\infty v^t e^{-k\overline{x+t}} f(0, e^{-k\overline{x+t}})\, dt.$$

Writing $e^{-k\overline{x+t}} = s$, this becomes

$$\bar{A}_x = \frac{1}{v^x l_x} \int_0^{e^{-kx}} s^{\delta/k} f(0, s)\, ds.$$

A similar expression can be obtained for $\bar{A}^{1}_{x:\overline{n}|}$, the lower limit of the integral being $e^{-k\overline{x+n}}$. These results might be of use in a case where $f(0, r)$ is of such a form (e.g. $a + hr^n$) that the integrand is an integrable function.

DISCONTINUITIES IN THE HEALTH FREQUENCY DISTRIBUTION

It has up to now been assumed that the initial health frequency distribution forms a continuous curve. Let us consider what happens if the curve is discontinuous. Diagram VII shows a curve APQB with a single discontinuity, the height of the curve suddenly falling at rating r from PN to QN.

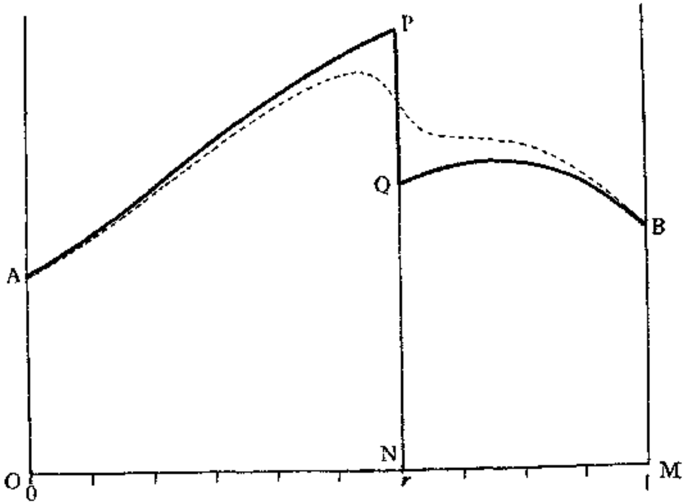


Diagram VII. A discontinuous distribution.

It is apparent that the initial upward flow through rating r will be at the rate $\phi_1(0, r).PN$, and that the initial downward flow will be at the rate $\phi_2(0, r).QN$. Hence the net upward flow through rating r will be at the rate

$$\phi_1(0, r).PN - \phi_2(0, r).QN$$

$$= \{\phi_1(0, r) - \phi_2(0, r)\}.PN + \phi_2(0, r).\{PN - QN\}$$

$$= \phi(0, r).PN + \phi_2(0, r).PQ \qquad \text{......(i)}$$

and

$$= \{\phi_1(0, r) - \phi_2(0, r)\}.QN + \phi_1(0, r).\{PN - QN\}$$

$$= \phi(0, r).QN + \phi_1(0, r).PQ. \qquad \text{......(ii)}$$

Considering now that section of the distribution which has ratings lower than r, it follows from (i) that, besides the normal upward flow $\phi(0, r).PN$ which would take place out of the section if the curve were continuous, there will be an additional upward flow $\phi_2(0, r).PQ$. Similarly, considering that section of the distribution which has ratings higher than r, it follows from (ii) that there will be an additional upward flow into the section of $\phi_1(0, r).PQ$. Except at rating r the upward flow will initially be everywhere normal, i.e. the same as if the curve were continuous throughout. The normal flow considered by itself will cause the configuration of the distribution to change with the receding movement already described. The effect of the additional flow through the discontinuity will be to "spread" the discontinuity, so that the frequency curve will be discontinuous not only at rating r but over a range of ratings in the neighbourhood and on both sides of r. Wherever there is a discontinuity, and so long as it persists, there will be an additional flow of the type described; eventually therefore the discontinuities will be smoothed out, and the curve become continuous throughout.

It is convenient to consider the movement of the distribution as being made up firstly of a "smoothing-out" process due to the additional flows where there are discontinuities in the frequency curve, and secondly of the normal receding movement such as would take place if the curve were continuous. The smoothing-out process considered by itself will result in the frequency curve assuming a configuration such as that indicated by the dotted line in Diagram VII. We have seen that $\phi_1(x, r)$ and $\phi_2(x, r)$ are probably large compared with $\phi(x, r)$; hence it is likely that the smoothing-out process will take place rapidly compared with the receding movement. The movement of the frequency distribution can thus be regarded as commencing with an upward "surge" through rating r; this surge will quickly die away leaving the normal receding movement.

If the discontinuity is in the opposite direction, i.e. if $PN < QN$, the effect is similar except that the surge will be downwards instead of upwards.

In practice the health frequency distribution of a group of lives can never form a continuous curve, for the lives being finite in number the curve really consists of a number of isolated points. Instead of comparing the group to fluid in a tank we should more accurately compare it, say, to sand in a box, each grain of sand representing a separate life. Provided however that the group is large, and so long as there is nowhere a sudden change in the distribution, the discontinuities in the frequency curve will on the whole balance each other and the fundamental relationship for $f(x, r)$ given on p. 329 will be approximately correct. There is nevertheless one exception of importance. If for any rating r the height of the frequency curve is small, then however great the total size of the group, there will be only few lives with ratings in the neighbourhood of r, so that at and near r there will be discontinuities of significance. This is illustrated in Diagram VIII where APM is the frequency curve, there being no lives of rating higher than r represented by the point P. The number of lives of rating slightly lower than r will be small, and there will be discontinuities near P which will result in an additional upward flow. The frequency curve will in a short space of time, in so far as it is legitimate to represent it

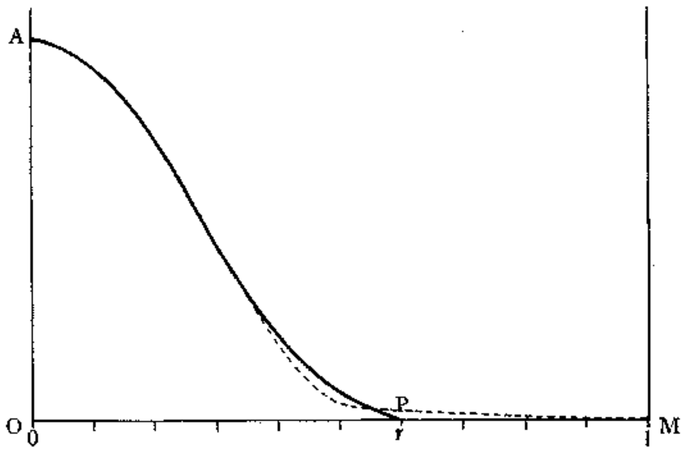


Diagram VIII. Special type of discontinuity.

by a continuous line, assume a configuration such as that indicated by the dotted line. This explains why for such a group of lives the force of mortality may attain a positive value almost immediately, and not, as would otherwise appear, remain at zero until by the normal receding movement the point initially represented by P will have reached rating 1.

To investigate fully the manner in which a discontinuous health frequency distribution will change with the lapse of time, it would probably be necessary to consider the speeds at which the ratings of individual lives are rising or falling. In fact the ideal way of dealing with the problem would be to approach it with the notion of speeds rather than of flows. But we should then have three variables to deal with—time, rating and speed—and the introduction of a third variable seems to make the problem too involved to enable any concrete results to be obtained.

HOMOGENEITY OF THE HEALTH FREQUENCY DISTRIBUTION

So far we have been considering a single group of lives, and the question arises as to whether any other group of lives would under the same health scale be subject to the same forces of deterioration and recuperation. It would be very convenient if this were so, because mortality could then be interpreted in terms of the health frequency distribution, any differences between the rates of mortality of various groups of lives being ascribed to differences in their distributions. Such an interpretation would be legitimate if lives of the same age and rating were all exactly similar to each other, but there is no justification for assuming this to be the case. For instance, lives of one type may have high ratings because they are suffering from a chronic ailment which will almost certainly reduce longevity; lives of another type may have similar high ratings because they are suffering from an acute ailment which will probably result either in death at an early date or in complete recovery. We cannot assume that both types of lives will be subject to the same forces of deterioration and recuperation. A group may be regarded as made up of lives of various types, and the forces of deterioration and recuperation for the group as the averages of those for the individual types. Two different groups may not contain the various types in the same proportions and hence will not be subject to identical forces.

The difficulty will be overcome if by some means the health scale can be so framed that lives of the same age and rating will always be subject to the same forces of deterioration and recuperation; in such a case the health frequency distribution could be said to be "homogeneous". How far this can be achieved is discussed below.

Given any group of lives, let us regard it as made up of a number of sub-groups, where each such sub-group is homogeneous, i.e. consists of lives all of the same type, lives being defined to be of the same type if they are subject to the same forces of deterioration and recuperation. The forces of deterioration and recuperation for each sub-group will in general be different from those for the whole group, and in fact the equal-distribution scale of the group may not be a natural scale for each of the different sub-groups. Now as has been explained (see p. 332), we can by suitable choice of natural scale make the factor of deterioration take what form we please; suppose therefore that for each sub-group we adopt a natural scale such that the corresponding factor of deterioration for the sub-group is the same as that for the whole group under its equal-distribution scale. By this device the method of measuring health is modified; the rating of any life is determined first by deciding to which sub-group the life belongs, and then using the appropriate scale. For the group as a whole the factor of deterioration will be unaltered, and it follows that its distribution will be unaltered, since a change in distribution necessarily means a change in the factor of deterioration. The result is therefore that at any particular age the ratings of some lives may be increased and of others decreased, but the effect on the whole group will be nil.

Thus, provided that it is legitimate to assume that the group of lives is constituted of a number of homogeneous sub-groups, we have shown that the health scale can be so framed that all the lives will be subject to the same factor of deterioration. The scale can then be applied to any other group consisting of lives of the same types but in different proportions, so enabling its health frequency distribution to be determined for purpose of comparison; or it can, in a similar way, be applied to a section of the original group, and this we shall make use of when we consider the question of "selection". It will be obvious that there are objections to using the same scale for groups which are constituted of entirely different types of lives, and generally speaking it would only be practicable to compare the health frequency distributions of groups of lives which are living contemporaneously and under somewhat similar conditions. Thus for instance we might compare the distributions of different sections of the same population; for such a purpose it would be convenient to adopt the equal-distribution scale of the population.

It will be observed that we have not entirely succeeded in making the health frequency distribution homogeneous. Different lives may be subject to different forces of deterioration and recuperation; it is only the difference between these forces, i.e. the net force of deterioration, which has been made common to all the lives. It will also be realised that our scale no longer depends solely on an arbitrary referee; we have found a more scientific criterion. The efficacy of the latter is, however, dependent on our having full knowledge of the net force of deterioration for the group of lives and for each sub-group at all ages, since without such knowledge it is not possible to change the scale for each sub-group in the manner described.

It may be objected that it is incorrect to regard the group of lives as being constituted of homogeneous sub-groups. No two lives are exactly alike, and there may therefore be as many types as there are lives; if a sub-group can consist of one life only, its net force of deterioration becomes meaningless, and our argument breaks down. Nevertheless, it is not unreasonable to suppose that if the group were sufficiently large it could be divided into sub-groups in each of which the lives are of type so similar that there would be little if any error in treating them as all subject to the same forces. In this way we can justify the assumption for a large group, and, since the whole theory of this paper rests on the basis that the numbers of lives dealt with are large enough to allow of the operation of the "law of averages", it follows that it is in general legitimate, by suitably framing the scale, to regard the group as made up of lives all subject to the same net force of deterioration.

The question as to how the "expectation of life" depends upon the rating may be of interest, and is discussed in the Appendix, § 8.

SELECTION

If from a group of lives all of the same age we exclude those lives whose rating exceeds some critical value r, we shall obtain what may be called a "perfectly select" group. The equal-distribution scale of the original group (which may be called the "control" group) can conveniently be used for both groups, it being assumed as explained above that the select group will be subject to the same factor of deterioration as the control group. The initial health frequency curves of the two groups will under this scale be as shown in Diagram IX, AC representing that of the control group, and APQM that of the select group. Let us consider the nature of the movement that will with the lapse of time take place in the distribution of the select group.

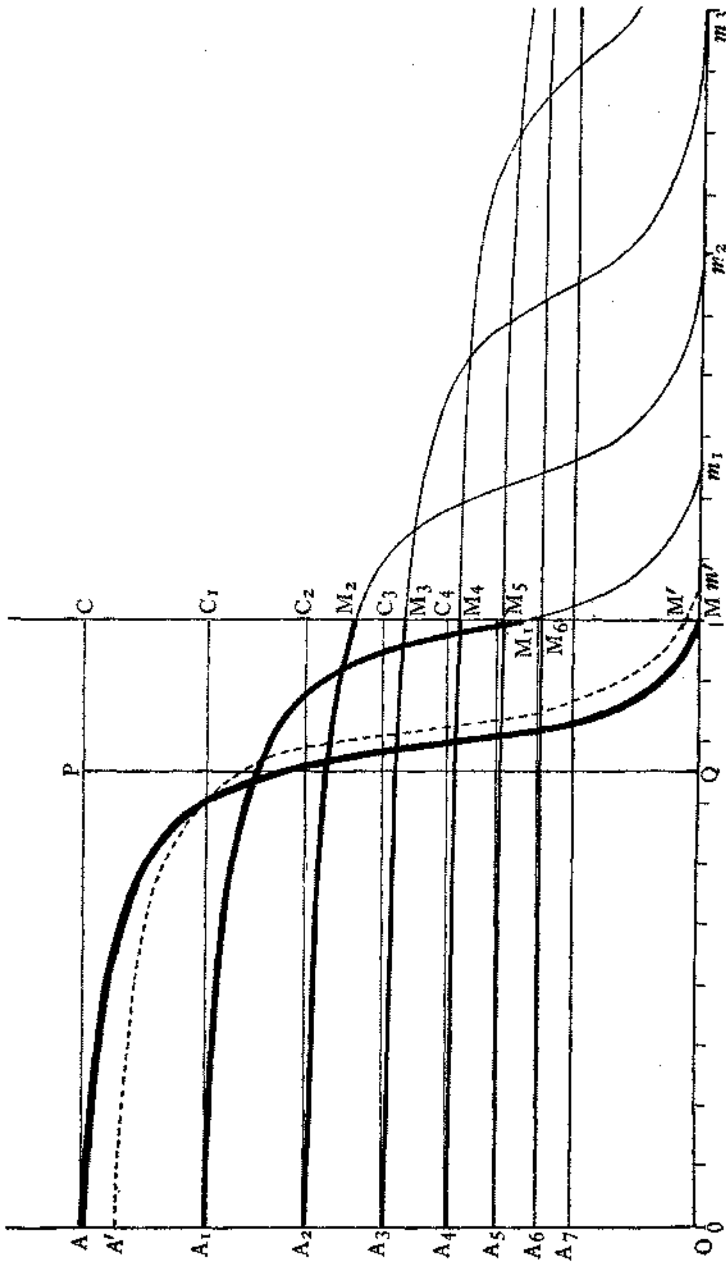


Diagram IX. Movement of the distribution of a select group.

The distribution has initially a discontinuity at rating *r*, so there will be, to commence with, an upward surge through this rating, as the result of which (and ignoring the normal receding movement) the distribution will in a short space of time be represented by a frequency curve such as that indicated by the curved line AM in Diagram IX. The frequency curve at any subsequent time will be that which would be arrived at by the normal receding movement assuming the curved line AM to be the frequency curve at time 0. The curves A_1M_1, A_2M_2, A_3M_3, etc., in Diagram IX thus represent the frequency curve at successive equal intervals of time (reckoned by the variable measure), being deduced from the curve AM by a similar method to that adopted in the construction of Diagram VI. A_1C_1, A_2C_2, A_3C_3, etc., correspondingly represent the frequency curve of the control group.

Broadly speaking, the movement of the distribution of a select group can be divided into two stages: the first, during which the smoothing-out process takes place, and the second, after the smoothing-out process has concluded. If the surge were to happen instantaneously, the frequency curve would at once move from the position APQM to the position of the curved line AM. Actually it must occupy some space of time, during which the normal receding movement will operate, so at the end of the first stage the frequency curve will possibly be as indicated by the dotted line A′M′. It is probable that though the initial surge occurs quickly, it never entirely dies away; but it may be assumed that when the second stage is reached the remaining effects of the surge are negligible, and the receding movement then alone operates.

The movement can be compared to the breaking of a wave on the seashore. When a wave breaks, the surface of the water becomes for a short space of time discontinuous and surges forward; it then smooths out and the edge of the water travels steadily up the beach.

As explained on p. 337, the configuration of the curve AM in Diagram IX can probably be found exactly only by taking account of the speeds with which the ratings of individual lives are rising or falling, but knowledge of its general shape can be obtained by the following argument.

The normal receding movement is determined by the net force of deterioration; so to consider the smoothing-out process by itself, it is necessary to ignore this force, i.e. we must assume it is everywhere nil. The forces of deterioration and recuperation must therefore be taken as equal to each other, being nil for rating 0, then rising to a maximum and falling again to nil for rating 1. Applying this assumption to the control group, the health frequency distribution will remain stationary since the net force of deterioration is nil; but there will nevertheless be continual movement in the distribution, some lives having ratings which are rising, and other lives ratings which are falling, the flows being equal in the two directions. If we take any two values between 0 and 1, then, however far apart these values may be, there will always be some lives with ratings which in a given space of time pass from one to the other. Turning now to the select group, this forms part of the control group, and the number of lives whose ratings pass from a value less than the critical value r to one greater than r is, to commence with, the same as in the control group, but of course there will be no lives whose ratings pass from a value greater than r to one less than r. Thus the frequency of distribution of the group will decrease for ratings below r, and increase for ratings above r. The extents of these decreases and increases will depend on the magnitude of the forces of deterioration and recuperation; i.e. they will be nil for ratings 0 and 1, and will grow greater as we move away from these extreme ratings towards the centre of the scale. Furthermore, since small changes in the ratings of lives take place considerably more frequently than large changes, and since small changes other than those from values slightly lower than r to values slightly higher than r will, to commence with, counterbalance each other, it follows that the greatest alterations in the distribution will be in the neighbourhood of rating r.

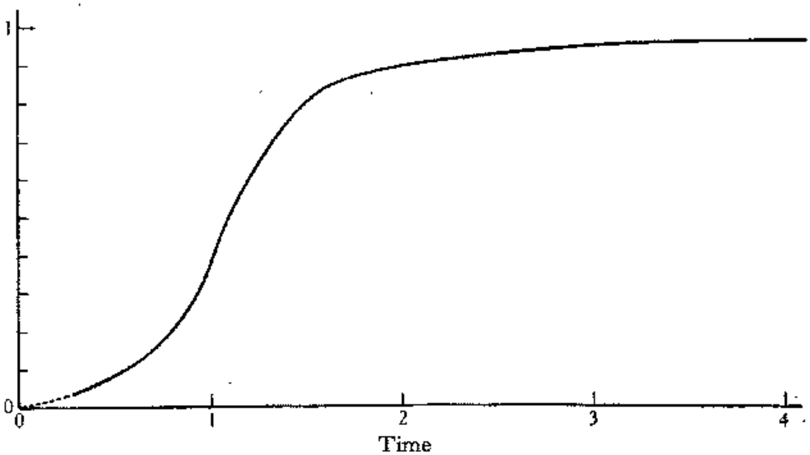
The curve AM in Diagram IX has been drawn so as to fulfil the above conditions and to enclose the same area as the original frequency curve APQM. The only matter in doubt is the degree of curvature which it should be given; this we are unable to determine, and it must be borne in mind therefore that any conclusions which we reach regarding the effects of "selection" will be qualitative only, and not quantitative.

The curves A_1M_1, A_2M_2, A_3M_3, etc., in Diagram IX have been continued beyond rating 1 to m_1, m_2, m_3, etc. (cf. Diagram VI), the areas thus cut off beyond this rating representing the numbers of deaths that have occurred. The force of mortality of the select group at any time can be deduced from the diagram; for instance when the frequency curve has reached the position A_tM_t, the force of mortality will be

$$\lambda \frac{M_tM}{\text{Area } A_tM_t \text{ MO}}.$$

The ratio which the force of mortality of the select group bears to that of the control group will thus be

$$\frac{M_tM}{\text{Area } A_tM_t \text{ MO}}.$$



(The unit of time corresponds to that adopted in Diagram IX, i.e. time t is the time when the frequency curve is in the position denoted by A_tM_t in Diagram IX.)

Diagram X. Ratio of force of mortality of select group to that of control group.

By plotting this ratio in the form of a graph, we can show how the mortality of the select group runs into that of the control group. This has been done in Diagram X, and it will be seen that the ratio increases at a rate first slow, then more rapid, slowing up again as it approaches the value unity. Eventually the difference between the ratio and unity becomes of negligible magnitude.

The following points call for special remark:

(1) What precisely happens during the short interval of time while the initial surge is taking place may be open to doubt. At

time 0 the force of mortality must be nil, since there are no lives of rating 1; it must almost immediately attain a positive value, but when the surge is over it is still small. Hence we may reasonably suppose that it increases steadily during the interval.

(2) We know the general shape but not the exact configuration of the curve AM in Diagram IX, and the same therefore applies to the graph in Diagram X. In particular we cannot tell what length of time is represented by the initial portion of the graph in which the ratio of the force of mortality of the select group to that of the control group is increasing slowly, nor what length of time must elapse before the ratio can for practical purposes be taken as unity. The former may be a matter of days or of years; it is only by comparison with observed facts that we can say it is as a rule likely to be nearer the first of these than the second. The value of the critical rating r is of course of predominating importance with regard to both Diagrams IX and X.

(3) The variable measure of time is used in Diagram X, and if the force of mortality of the control group is increasing, the length of a unit of time is becoming shorter. Hence by the true measure of time the mortality of the select group will approach that of the control group increasingly more rapidly than as indicated in the diagram. This will generally be most significant at the older ages at entry, but in any case the general shape of the graph is not likely to be altered by the transformation from the variable to the true measure of time.

SELECT AND ULTIMATE MORTALITY

It is proposed now to consider what light the foregoing analysis can throw on the practical aspect of select mortality. The arguments used below can be applied with appropriate modifications to any mortality experience where the phenomenon of selection occurs, but to help fix our ideas we will suppose we are dealing with the experience of lives assured at ordinary rates under ordinary (as distinct from industrial) policies.

It is necessary first to decide exactly what are meant by select mortality and ultimate mortality. Lives accepted for assurance are selected by the life offices out of those who apply for it by excluding those not in first-class health. Confining our attention to lives born at the same time and applying for assurance at the same age, we can regard the lives accepted for assurance as a select group, the lives applying for assurance being the corresponding control group; and the ratio between the forces of mortality of the two groups will, if represented graphically, be as depicted in Diagram X. It is not unreasonable to suppose that the force of mortality of the control group will increase steadily with age; the force of mortality of the select group will therefore commence at zero and increase, first slowly, then more rapidly, then slowly again as it approaches that of the control group. We will define the mortality of the lives accepted for assurance as ultimate on and after the time when its difference from the mortality of the lives who applied for assurance becomes negligible; before this time the mortality is select, depending of course on the duration since entry into assurance.

The lives selected for assurance will not as a rule form a *perfectly* select group, for the method of selection employed by the life offices is not such as to fix a definite point of division in the health scale between the ratings of the lives accepted and of those rejected; for instance medical science may not be sufficiently accurate to prevent an occasional bad life from being admitted or good life from being excluded, also the standard may vary slightly between one medical examiner and another or between one life office and another. It is probable therefore that the health frequency curve of the accepted lives is similar to that of a perfectly select group after the initial surge has partly taken place, such for instance as is represented by curve (i) in Diagram XI. Thus the select mortality, at any rate for short durations, may be somewhat different from what it would be if the lives formed a perfectly select group, but nevertheless the differences will not be such as to alter the general form of the curve in Diagram X.

According to our definition, lives on entering into assurance experience select mortality until such time as the effects of the initial selection exercised by the life offices have disappeared. After this time they experience ultimate mortality. The select mortality will depend on the age at entry, and so also, as far as we can see at present, will the ultimate mortality. The latter point requires further investigation.

Let us consider two groups of lives, U and S, consisting of those lives which were accepted for assurance at ages u and s respectively ($u<s$), all the lives having been born at the same time; and let us suppose that when the lives of group U have reached age s they are experiencing ultimate mortality. Now if the lives who presented themselves for assurance at age s were exactly similar to the survivors at age s of group U, the relationship between the forces of mortality of the two groups from age s onwards would be the same as that described above in the case of the select and control groups. Represented under the equal-distribution scale of group U, the health frequency distribution of group S would at entry be like curve (i) in Diagram XI; and when the mortality of group S becomes ultimate it would be identical with that of group U. This however involves the assumption of an ideal state of affairs which cannot be justified for two reasons.

In the first place, that section of the population consisting of lives aged s and living in circumstances such that they are prospective applicants for assurance may not necessarily be exactly similar to the survivors of the corresponding section of the population $s-u$ years ago then aged u. There may have been a variation in the class of persons from whom assured lives are drawn. This feature was discussed by W. P. Elderton and H. J. P. Oakley (*J.I.A.* Vol. LIV, pp. 43 *et seq.*) and called by them "class selection". In an extreme case of such variation it might happen that the equal-distribution scale of group U would not be a natural scale for group S, but we need not pursue this eventuality.

Secondly, even if there has been no change in the class of persons from whom assured lives are drawn, it does not follow that those actually applying for assurance will form a representative sample. It is possible for instance that there may be some discrimination against the life offices, the proportion of lives presenting themselves increasing as the health rating rises. This would apply to both groups U and S, but by the time the lives of group U have reached age s the effects of this discrimination will have, in part if not altogether, disappeared. The initial health frequency distribution of group S would therefore under the equal-distribution scale of group U, be somewhat as represented by curve (ii) in Diagram XI (though perhaps not so exaggerated as there drawn). In the case of annuitant lives, on the other hand, the discrimination would be in the opposite direction, and the initial health frequency distribution

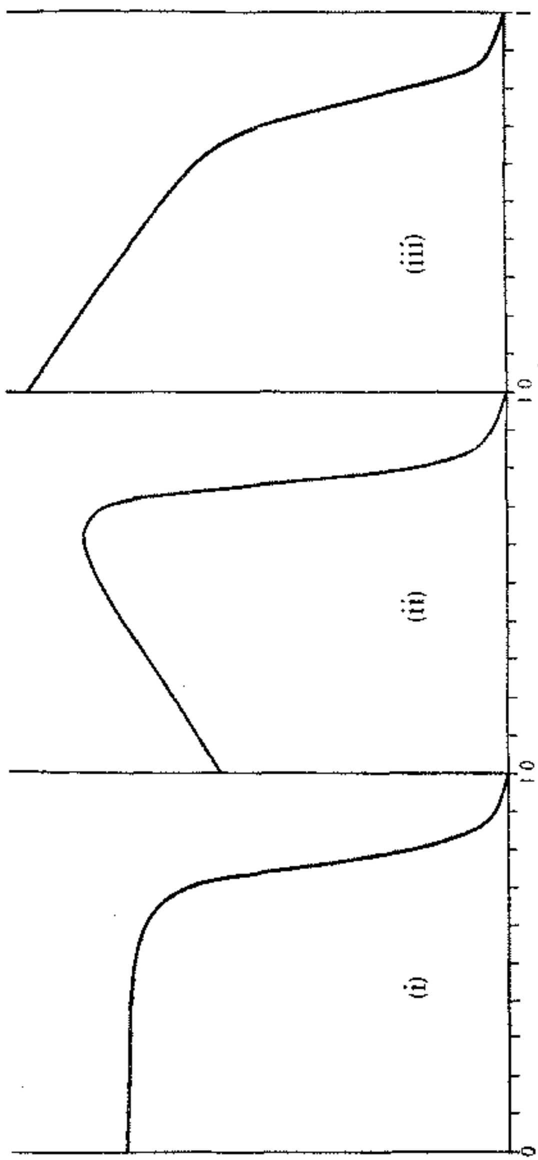


Diagram XI. Different types of select distributions.

as shown in curve (iii). For annuitant lives it is even possible that the general downward slope of the frequency distribution may be as important a feature as its tailing-off at the critical rating.

What now are the effects if the initial health frequency curve of group S is of a kind different from the ideal curve (i)? Though the select mortality of group S may be somewhat altered, the general shape of its graph will remain unchanged provided the distribution contains few or no lives of rating higher than the critical value r. The ultimate mortality of group S will, however, not be identical with that of group U, though it will tend to become the same with the lapse of time (for, save in the exceptional case where the distribution is zero or infinite for rating 0, the force of mortality of any group of lives will, as we have seen, tend to become the same as the factor of deterioration). It may possibly happen that the difference between the ultimate forces of mortality of the two groups will always be practically negligible, but such a result cannot be expected as a matter of course.

In the case of a frequency distribution such as that represented by curve (ii), the mortality of group S will before the end of the select period become greater than that of group U, and will then approach it from above instead of from below. This may perhaps explain why for assured lives there is sometimes not much difficulty in making the select rates of mortality run into a common ultimate table irrespective of the age at entry. On the other hand, in the case of a frequency distribution such as that represented by curve (iii), the mortality of group S will always be below that of group U; and this possibly accounts, at any rate to some extent, for the feature encountered in connection with the $a(f)$ and $a(m)$ experiences.

It is hardly necessary to draw attention to the fact that the above discussion has been confined to lives all born at the same time and entering into assurance at two ages u and s only. In making a mortality investigation all ages at entry have to be considered, and furthermore, if as often happens the experience is that of a number of years, the "exposed to risk" at any age may be spread over several years of birth. These factors introduce further complications, though not any fresh principles.

The duration of the select period may vary with the age at entry, and is primarily dependent on the value of the critical rating above which the medical selection exercised by the life offices comes into operation. It is therefore interesting to note that in practice this duration is sometimes found to be independent of the age at entry. May the explanation be that the distance we can see into the future as regards the probable health of any life is approximately the same for all ages? If so, then the critical rating is automatically fixed for each age at entry so as to make the duration at which the effects of the initial selection by the life offices disappear independent of the age. A similar argument would apply to annuitant lives.

CONCLUSION

The question naturally arises as to whether the theory of mortality formulated in this paper is capable of independent confirmation. The theory is not one which readily lends itself to verification; and, since it is entirely built up from two fundamental hypotheses regarding health, the necessity for such verification does not arise if these hypotheses are accepted. The observed mortality of a group of lives may enable us to make deductions about its health frequency distribution, but generally speaking the correctness of these deductions cannot easily be proved or disproved.

The nature of the mortality of a select group of lives during the select period gives rise to interesting speculations. The force of mortality should, according to our theory, commence at zero and increase at a rate first slow, then more rapid, eventually slowing up again. Most published select tables of mortality exhibit only the last two of these three stages. This may be explained by the fact that the first stage is of duration probably less than a year, so that the period of a year usually adopted as the unit of time in mortality tables is too long to enable it to be brought to light. To investigate this point the mortality at short durations would have to be examined in detail. There are certain obvious difficulties in making such an examination. For instance, the "census" method of tabulating the data might, at any rate in the form with which we are familiar, be unsuitable. Furthermore, the experience would have to be a large one, as the rate of mortality will at short durations be small. It would be necessary to take special care to tabulate for every life the correct date of entry into observation; for instance "dated back" contracts would need special treatment. In the case of assured lives it is arguable that the date of entry into observation should be the date on which the evidence of health is obtained and not the date on which the assurance is completed.

Another matter which might be usefully investigated would be to find what proportion of lives is rejected at each age by life offices for assurance at ordinary rates. Given such data, it might be possible to judge whether they are under our theory compatible with the observed duration of the select period.

APPENDIX

Determination of the health frequency distribution

§ 1. The purpose of the following is, given the net force of deterioration and given also the frequency distribution at time 0, to find the frequency distribution at any subsequent time (i.e. given $\phi(x, r)$ and $f(0, r)$, it is required to find $f(x, r)$).

Consider the section of the frequency distribution which at time x lies between ratings r and $r+\delta r$ (see Diagram XII). As time elapses some lives will have ratings which enter this section and others ratings which leave it; this will happen whenever the rating of a life passes through the value r or $r+\delta r$. In time δx the net upward flow into the section through rating r will be

$$\phi(x, r) f(x, r)\, \delta x,$$

and the net upward flow out of the section through rating $r+\delta r$ will be

$$\phi(x, r+\delta r) f(x, r+\delta r)\, \delta x.$$

Thus the net increase in the section will be

$$\phi(x, r) f(x, r)\, \delta x - \phi(x, r+\delta r) f(x, r+\delta r)\, \delta x.$$

But the number of lives whose ratings are within the section at time x is $f(x, r)\, \delta r$, and at time $x+\delta x$ is $f(x+\delta x, r)\, \delta r$. Hence

$$f(x+\delta x, r)\delta r - f(x, r)\delta r = \phi(x, r) f(x, r)\delta x - \phi(x, r+\delta r) f(x, r+\delta r)\delta x,$$

$$\therefore \frac{f(x+\delta x, r) - f(x, r)}{\delta x} = \frac{\phi(x, r) f(x, r) - \phi(x, r+\delta r) f(x, r+\delta r)}{\delta r},$$

or, proceeding to the limits,

$$\frac{\partial}{\partial x} f(x, r) = -\frac{\partial}{\partial r}\{\phi(x, r) f(x, r)\}. \qquad \ldots\ldots(1)$$

This is a partial differential equation; to solve our problem we should have to find the function $f(x, r)$ which satisfies it and which becomes the same as the given $f(0, r)$ when x is put equal to 0.

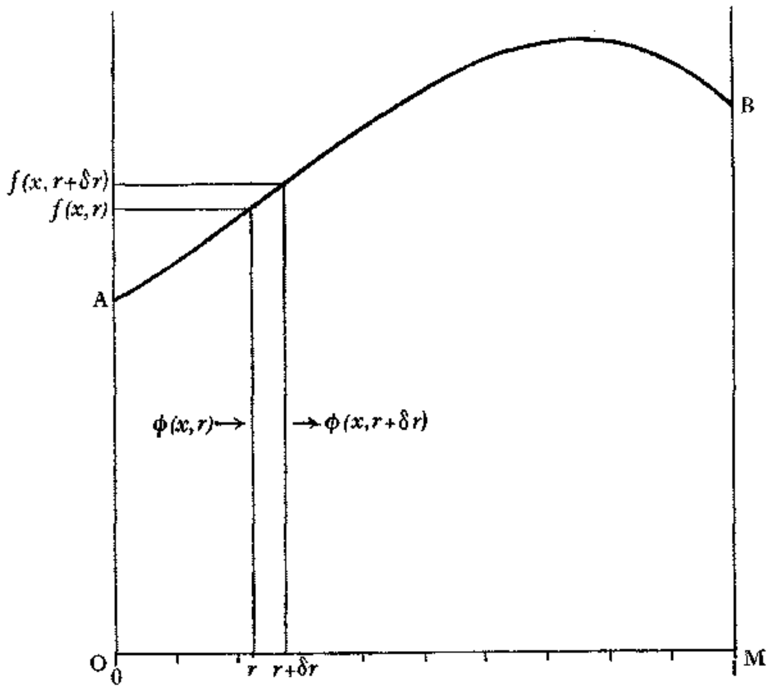


Diagram XII.

In arriving at equation (1) it has been assumed that the given function $\phi(x, r)$ is continuous with respect to both variables and the given function $f(0, r)$ continuous with respect to r, and it holds good therefore only under these conditions, a point which is of importance.*

* Strictly speaking, it has also been assumed that the functions are differentiable. It is taken for granted here and elsewhere in the Appendix that the functions dealt with are reasonably "well-behaved". Reference will not therefore be made to matters, such as uniform continuity, which might from a rigorous mathematical standpoint require consideration.

§ 2. Under a natural scale, $\phi(x, r) = r\lambda_x$. Making this substitution in (1), we have

$$\frac{\partial}{\partial x} f(x, r) = -\frac{\partial}{\partial r}\{r\lambda_x f(x, r)\}, \qquad \ldots\ldots(2)$$

or, altering the measure of time so as to make the factor of deterioration constant (see p. 328),

$$\frac{\partial}{\partial y} g(y, r) = -\frac{\partial}{\partial r}\{r\lambda g(y, r)\}, \qquad \ldots\ldots(3)$$

i.e.
$$\frac{\partial}{\partial y} g(y, r) + r\lambda \frac{\partial}{\partial r} g(y, r) = -\lambda g(y, r). \qquad \ldots\ldots(4)$$

The solution is now seen to be

$$g(y, r) = e^{-\lambda y} g(0, re^{-\lambda y}), \qquad \ldots\ldots(5)$$

since this will be found on substitution to satisfy (4), and it takes the required form when y is put equal to 0.* Reverting to the true measure of time,

$$f(x, r) = e^{-\int_0^x \lambda_x\,dx} f\left(0, re^{-\int_0^x \lambda_x\,dx}\right). \qquad \ldots\ldots(6)$$

This result holds good provided $f(0, r)$ is continuous with respect to r.

Relationship between two natural scales

§ 3. Suppose we have two natural scales, A and B. All functions reckoned according to scale B will be distinguished by means of dashes. Thus the net force of deterioration under scale A will be $\phi(x, r)$ and under scale B will be $\phi'(x, r')$.

* To arrive at the solution from (3), let $G(y, r) = \int_0^r g(y, r)\,dr$. Integrating (3) with respect to r between the limits 0 and r,

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

To satisfy this equation, it can be seen that $G(y, r)$ must be a function of $re^{-\lambda y}$, and in order that it may take the required form when y is put equal to 0 it must be $G(0, re^{-\lambda y})$. Hence, differentiating with respect to r, we obtain the solution for $g(y, r)$.

At any given time the state of health which is denoted by rating r under scale A will be denoted by rating r' under scale B. The curve showing the relationship between r and r' at any time (as in Diagram II) will be called the "relationship curve" between the two scales; this curve will vary with the time x. We can regard r' as a function of x and r, and when necessary will write it as $r'(x, r)$.

At time x the number of lives whose ratings under scale A are between r and $r+\delta r$ must be the same as the number of lives whose ratings under scale B are between r' and $r'+\delta r'$,

$$\therefore\ f(x, r)\,\delta r = f'(x, r')\,\delta r'(x, r),$$

or, proceeding to the limit,

$$f(x, r) = f'(x, r')\frac{\partial}{\partial r}r'(x, r) \qquad \ldots\ldots(7)$$

The function $r'(x, r)$ will fulfil the following conditions (see also p. 320):

(i) $r'(x, 0) = 0$ and $r'(x, 1) = 1$,

(ii) $r'(x, r)$ is continuous with regard to both variables,

(iii) $\dfrac{\partial}{\partial r}r'(x, r)$ is always positive.

Conditions (i) and (ii) are obvious, and condition (iii) follows from (7) above.

Should either $f(x, r)$ or $f'(x, r')$ be for any values of x and r zero or infinite, $\dfrac{\partial}{\partial r}r'(x, r)$ will by (7) also be zero or infinite. This is an eventuality to which we shall give special consideration later; in §§ 4–6 it is assumed that $f(x, r)$, $f'(x, r')$ and $\dfrac{\partial}{\partial r}r'(x, r)$ are never zero or infinite.

It is assumed throughout the paper that the factor of deterioration is always positive, never for finite values of x attaining the values zero or infinity, and that it does not tend to zero as x becomes large. The object of this assumption is to avoid having to deal with special cases which would rarely arise in practice. For the equal-distribution scale the assumption can be fully justified, since under this scale the factor of deterioration is equal to the force of mortality, and it is reasonable to suppose that at no age is death either impossible or a certainty and that at old ages the chance of death does not continually decrease.

§ 4. The purpose of the following is, given the factor of deterioration under scale A, and given also the relationship curve between scales A and B at time 0, to find the relationship curve at any subsequent time (i.e. given λ_x and $r'(0, r)$, it is required to find $r'(x, r)$).

At time x, the number of lives whose ratings under scale A are lower than r must be the same as the number of lives whose ratings under scale B are lower than $r'(x, r)$. Also, at time $x+\delta x$, the number of lives whose ratings under scale A are lower than r must be the same as the number of lives whose ratings under scale B are lower than $r'(x+\delta x, r)$, i.e. the same as the number of lives whose ratings are lower than $r'(x, r)$ together with the number of lives whose ratings are between $r'(x, r)$ and $r'(x+\delta x, r)$. Hence the net upward flow in time δx through rating r under scale A must be the same as the net upward flow in time δx through rating $r'(x, r)$ under scale B less the number of lives whose ratings are between $r'(x, r)$ and $r'(x+\delta x, r)$.

$$\therefore\ \phi(x, r) f(x, r)\,\delta x$$

$$= \phi'(x, r') f'(x, r')\,\delta x - \{r'(x+\delta x, r) - r'(x, r)\} f'(x, r'),$$

$$\therefore\ \phi(x, r) f(x, r) = \{\phi'(x, r') - \frac{\partial}{\partial x} r'(x, r)\} f'(x, r'),$$

$\therefore$ using (7),

$$\phi(x, r)\frac{\partial}{\partial r} r'(x, r) = \phi'(x, r') - \frac{\partial}{\partial x} r'(x, r),$$

$$\therefore\ r\lambda_x \frac{\partial}{\partial r} r'(x, r) = r'(x, r)\,\lambda'_x - \frac{\partial}{\partial x} r'(x, r). \qquad \text{......(8)}$$

In this result put $r = 1$:

$$\lambda_x \left[\frac{\partial}{\partial r} r'(x, r)\right]_{r=1} = \lambda'_x - \frac{\partial}{\partial x} r'(x, 1),$$

$$= \lambda'_x, \qquad \text{......(9)}$$

since $r'(x, 1) = 1$ and is independent of x. Substituting this value for λ'_x in (8):

$$r\lambda_x \frac{\partial}{\partial r} r'(x, r) = r'(x, r)\,\lambda_x \left[\frac{\partial}{\partial r} r'(x, r)\right]_{r=1} - \frac{\partial}{\partial x} r'(x, r),$$

or $$r\lambda_x \frac{\partial}{\partial r} r'(x, r) + \frac{\partial}{\partial x} r'(x, r) = r'(x, r)\,\lambda_x \left[\frac{\partial}{\partial r} r'(x, r)\right]_{r=1},$$

which can also be written

$$r\lambda_x \frac{\partial}{\partial r} \log r'(x, r) + \frac{\partial}{\partial x} \log r'(x, r) = \lambda_x \left[\frac{\partial}{\partial r} r'(x, r)\right]_{r=1}. \quad \ldots(10)$$

We must now find the function $r'(x, r)$ which satisfies this differential equation and which becomes the same as the given $r'(0, r)$ when x is put equal to 0. Altering the measure of time so as to make the factor of deterioration under scale A constant, the equation becomes

$$r\lambda \frac{\partial}{\partial r} \log r'(y, r) + \frac{\partial}{\partial y} \log r'(y, r) = \lambda \left[\frac{\partial}{\partial r} r'(y, r)\right]_{r=1}. \quad \ldots(11)$$

The required solution is

$$r'(y, r) = r'(0, re^{-\lambda y}) \div r'(0, e^{-\lambda y}), \qquad \ldots\ldots(12)$$

since this will be found on substitution to satisfy (11), and it takes the required form when y is put equal to 0. Reverting to the true measure of time,

$$r'(x, r) = r'\left(0, re^{-\int_0^x \lambda_x dx}\right) \div r'\left(0, e^{-\int_0^x \lambda_x dx}\right). \quad \ldots\ldots(13)$$

It follows from this that, if we are given any natural scale (A), it is possible to find another natural scale (B), such that the relationship curve between the two scales at time 0 may have any given form (subject of course to its fulfilling the conditions mentioned in § 3). Since the equal-distribution scale is known to be a natural scale, this result proves the existence of other natural scales.

§ 5. In order to understand the relation between r and r' expressed by (13), it is of help to consider the relationship curve between the two scales at time 0; this will take a form such as in

Diagram XIII. Along the axis OR, OX is taken equal to $e^{-\int_0^x \lambda_x dx}$ (which is of course less than 1) and OX_1 to $r.OX$. The ordinates through X and X_1 meet the curve at P and P_1, and the corresponding abscissae meet the axis OR′ at Y and Y_1. The points P and P_1 thus represent ratings under scale A of $e^{-\int_0^x \lambda_x dx}$ and $re^{-\int_0^x \lambda_x dx}$, and under scale B of $r'(0, e^{-\int_0^x \lambda_x dx})$ and $r'(0, re^{-\int_0^x \lambda_x dx})$ respectively. Hence by (13),

$$r'(x, r) = \frac{P_1X_1}{PX}.$$

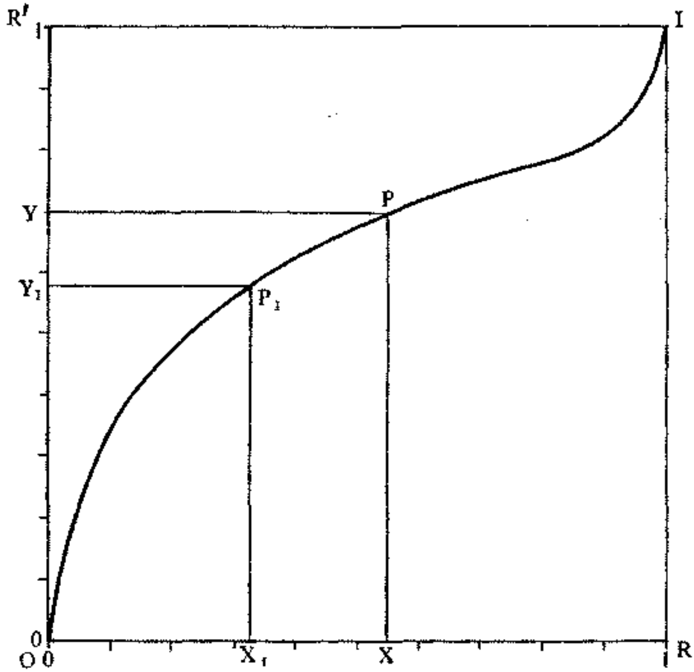


Diagram XIII.

For any given value of x, r may vary between 0 and 1, so that P_1 may be any point between O and P; any such position of P_1 will give us a pair of corresponding ratings at time x as follows:

$$r = \frac{P_1Y_1}{PY}, \ r' = \frac{P_1X_1}{PX}.$$

If now we suppose the rectangle OXPY enlarged by increasing all ordinates in the ratio IR : PX $(= 1 : e^{-\int_0^x \lambda_x\,dx})$, and all abscissae in the ratio IR′ : PY $(= 1 : r'(0, e^{-\int_0^x \lambda_x\,dx}))$, it will then occupy the position of the square ORIR′. The curve OP will be enlarged so as to run from O to I and will become the relationship curve at time x.

As x increases, $e^{-\int_0^x \lambda_x\,dx}$ tends towards zero, so that P approaches O and the extent by which the curve OP must be enlarged becomes greater. In the limit when x becomes infinitely large, the relationship curve will become a straight line from O to I, and the ratings under the two scales will be identical (see also p. 332).

It may be of assistance to imagine that the square ORIR′ is a shallow box, that the curve OI is a piece of elastic fixed at O and passing out of the box through a hole at I, and that the box is filled with some substance which hinders the free motion of the elastic. It can be further imagined that if the piece of elastic is gradually pulled out of the box through the hole at I, its shape inside the box will pass through the same changes as the relationship curve.

By (9), $\frac{\lambda'_x}{\lambda_x} = \left[\frac{\partial}{\partial r} r'(x, r)\right]_{r=1}$, which is the slope at I of the relationship curve at time x. But this is $\frac{\mathrm{PY}}{\mathrm{PX}}$ times the slope at P of the relationship curve at time 0. Hence the ratio at any time between the factors of deterioration under the two scales can easily be found from the relationship curve at time 0 as follows:

$$\frac{\lambda'_x}{\lambda_x} = \left[\frac{r}{r'(0, r)}\frac{\partial}{\partial r} r'(0, r)\right]_{r=e^{-\int_0^x \lambda_x\,dx}}. \qquad \ldots\ldots(14)$$

As x increases, the relationship curve tends to the form of a straight line, so that $\frac{\lambda'_x}{\lambda_x}$ tends to unity.

§ 6. The point P in Diagram XIII represents a rating at time 0 of $e^{-\int_0^x \lambda_x\,dx}$ under scale A. It will now be shown that it represents a rating at time 0 of $e^{-\int_0^x \lambda'_x\,dx}$ under scale B (i.e. that

$$r'(0, e^{-\int_0^x \lambda_x\,dx}) = e^{-\int_0^x \lambda'_x\,dx}).$$

By differentiation,

$$\frac{\partial}{\partial x} r'\left(0, e^{-\int_0^x \lambda_x dx}\right) = -\lambda_x e^{-\int_0^x \lambda_x dx} \left[\frac{\partial}{\partial s} r'(0, s)\right]_{s=e^{-\int_0^x \lambda_x dx}}. \quad \ldots(15)$$

Also, from (13),

$$\frac{\partial}{\partial r} r'(x, r) = \frac{\partial}{\partial r} \frac{r'\left(0, re^{-\int_0^x \lambda_x dx}\right)}{r'\left(0, e^{-\int_0^x \lambda_x dx}\right)},$$

$$= e^{-\int_0^x \lambda_x dx} \frac{\left[\frac{\partial}{\partial s} r'(0, s)\right]_{s=re^{-\int_0^x \lambda_x dx}}}{r'\left(0, e^{-\int_0^x \lambda_x dx}\right)}.$$

Putting $r = 1$,

$$\left[\frac{\partial}{\partial r} r'(x, r)\right]_{r=1} = e^{-\int_0^x \lambda_x dx} \frac{\left[\frac{\partial}{\partial s} r'(0, s)\right]_{s=e^{-\int_0^x \lambda_x dx}}}{r'\left(0, e^{-\int_0^x \lambda_x dx}\right)},$$

by (15)

$$= -\frac{1}{\lambda_x} \frac{\frac{\partial}{\partial x} r'\left(0, e^{-\int_0^x \lambda_x dx}\right)}{r'\left(0, e^{-\int_0^x \lambda_x dx}\right)},$$

$$= -\frac{1}{\lambda_x} \frac{\partial}{\partial x} \log r'\left(0, e^{-\int_0^x \lambda_x dx}\right).$$

But, by (9),

$$\left[\frac{\partial}{\partial r} r'(x, r)\right]_{r=1} = \frac{\lambda'_x}{\lambda_x};$$

hence

$$\frac{\partial}{\partial x} \log r'\left(0, e^{-\int_0^x \lambda_x dx}\right) = -\lambda'_x,$$

$$\therefore \log r'\left(0, e^{-\int_0^x \lambda_x dx}\right) = -\int_0^x \lambda'_x \, dx,$$

$$\therefore r'\left(0, e^{-\int_0^x \lambda_x dx}\right) = e^{-\int_0^x \lambda'_x dx}. \qquad \ldots\ldots(16)$$

This result means that if we are given λ_x and λ'_x for all values of x, we can find the relationship curve between the two scales at time 0. Thus $r = e^{-\int_0^x \lambda_x dx}$ and $r' = e^{-\int_0^x \lambda'_x dx}$ represent corresponding ratings under the two scales at time 0; by varying the value of x we can obtain a number of such corresponding pairs of ratings, and these can be plotted in the form of a graph to give the relationship curve.

§ 7. Suppose now that scale A is the equal-distribution scale of the group of lives under consideration, and that the factor of deterioration λ_x is known for all values of x. Let λ'_x be any positive function which satisfies the assumption mentioned at the end of § 3 with regard to zero and infinite values. Then § 6 shows that we can find another scale (B) such that the factor of deterioration under it will be λ'_x. In other words, by suitable choice of scale we can make the factor of deterioration take what form we please.

It is necessary to show that the relationship curve between scales A and B at time x fulfils the conditions stated in § 3. From the nature of the way in which the relationship curve at time x is connected with that at time o, it is clear that it will be sufficient to consider the latter curve only, i.e. the curve defined by $r = e^{-\int_0^x \lambda_x dx}$, $r' = e^{-\int_0^x \lambda'_x dx}$. Putting x equal to o and ∞, it is seen that condition (i) is satisfied; also since a small change in x results in small changes in r and r', both in the same direction, it follows that conditions (ii) and (iii) are satisfied.

The possibility of zero or infinite values of $\frac{\partial}{\partial r} r'(x, r)$ which was referred to in § 3 needs special consideration; for this purpose it will again be sufficient to investigate the relationship curve at time o only. Equation (14) shows that for any value of r other than o, $\frac{\partial}{\partial r} r'(0, r)$ can be zero or infinite only if either λ_x or λ'_x is zero or infinite, which is contrary to hypothesis (see end of § 3); hence for such values of r the possibility does not arise. The position at the origin is not however the same, for here

$$\left[\frac{\partial}{\partial r} r'(0, r)\right]_{r=0} = \underset{r\to 0}{\mathrm{Lt}} \frac{r'(0, r)}{r},$$

$$= \underset{x\to\infty}{\mathrm{Lt}} \frac{e^{-\int_0^x \lambda'_x dx}}{e^{-\int_0^x \lambda_x dx}},$$

$$= \underset{x\to\infty}{\mathrm{Lt}}\ e^{\int_0^x \lambda'_x \left(\frac{\lambda_x}{\lambda'_x} - 1\right) dx}, \qquad \ldots\ldots(17)$$

$$= \underset{x\to\infty}{\mathrm{Lt}}\ e^{-\int_0^x \lambda_x \left(\frac{\lambda'_x}{\lambda_x} - 1\right) dx}. \qquad \ldots\ldots(18)$$

(17) shows that if $\frac{\lambda'_x}{\lambda_x}$ tends to a limit less than unity (including zero) as x becomes large, then $\frac{\partial r'}{\partial r}$ will be infinite at the origin; and (18) shows that if $\frac{\lambda'_x}{\lambda_x}$ tends to a limit greater than unity (including infinity) as x becomes large, then $\frac{\partial r'}{\partial r}$ will be zero at the origin. It has already been shown (see § 5) that if $\frac{\partial r'}{\partial r}$ is not zero or infinite at the origin, then $\frac{\lambda'_x}{\lambda_x}$ tends to unity as x becomes large.

The preceding paragraph is of general application. In the particular case where scale A is the equal-distribution scale, $f(0, 0)$ cannot be zero or infinite, and it follows from (7) that, if $\frac{\partial r'}{\partial r}$ is zero or infinite at the origin, then $f'(0, 0)$ will be infinite or zero respectively. There are therefore three possibilities:

(i) $\underset{x\to\infty}{\mathrm{Lt}} \frac{\lambda'_x}{\lambda_x} < 1$, $\left[\frac{\partial}{\partial r} r'(0, r)\right]_{r=0}$ is infinite, $f'(0, 0)$ is zero;

(ii) $\underset{x\to\infty}{\mathrm{Lt}} \frac{\lambda'_x}{\lambda_x} = 1$, $\left[\frac{\partial}{\partial r} r'(0, r)\right]_{r=0}$ is finite, $f'(0, 0)$ is finite;

(iii) $\underset{x\to\infty}{\mathrm{Lt}} \frac{\lambda'_x}{\lambda_x} > 1$, $\left[\frac{\partial}{\partial r} r'(0, r)\right]_{r=0}$ is zero, $f'(0, 0)$ is infinite.

λ_x, the factor of deterioration under the equal-distribution scale, is the same as the force of mortality μ_x. Hence $f'(0, 0)$ will be zero, finite, or infinite according as $\underset{x\to\infty}{\mathrm{Lt}} \frac{\lambda'_x}{\mu_x}$ is less than, equal to, or greater than unity.

Expectation of life

§ 8. Out of a group of lives whose frequency distribution at time 0 is $g(0, r)$, the number of survivors at time y is

$$\int_0^1 g(y, r)\, dr = \int_0^1 e^{-\lambda y} g(0, re^{-\lambda y})\, dr.$$

The total of the lifetimes reckoned from time 0 of the lives forming the group is therefore

$$\int_0^\infty \left\{\int_0^1 e^{-\lambda y} g\,(0, re^{-\lambda y})\,dr\right\} dy,$$

$$= \int_0^1 \left\{\int_0^\infty e^{-\lambda y} g\,(0, re^{-\lambda y})\,dy\right\} dr.$$

Writing $s = re^{-\lambda y}$, so that $\dfrac{ds}{dy} = -\lambda re^{-\lambda y}$, the above becomes

$$\int_0^1 \left\{\int_0^r \frac{1}{\lambda r} g\,(0, s)\,ds\right\} dr,$$

reversing the order of integration,

$$= \int_0^1 \left\{\int_s^1 \frac{1}{\lambda r} g\,(0, s)\,dr\right\} ds,$$

$$= \int_0^1 -\frac{\log s}{\lambda} g\,(0, s)\,ds.$$

This result holds whatever the form of $g\,(0, r)$, provided it is continuous and such that the movement of the distribution obeys the fundamental relationship (5). The result is the same as would be obtained if a life of rating r at time 0 were assumed to have an expectation of life of $-\dfrac{\log r}{\lambda}$. It seems therefore that we may regard $-\dfrac{\log r}{\lambda}$ as equivalent to the expectation of life for rating r; this applies whatever the age, the expectation being reckoned by the variable measure of time under which the length of a unit of time will generally decrease as the age increases. On this basis the expectation will vary from infinity in the case of an imaginary "perfect" life of rating 0 to nil in the case of a life of rating 1.

It was mentioned on p. 329 that under the receding movement of the frequency distribution the point on the frequency curve corresponding to rating r at time 0 will reach rating 1 at time $-\dfrac{\log r}{\lambda}$. Considering therefore the group as a whole, the position is the same as if the ratings of all lives were to increase regularly with the lapse of time, a life of rating r at time 0 attaining rating 1 at time $-\dfrac{\log r}{\lambda}$.

We have not proved that every life of rating r has an expectation equal to $-\frac{\log r}{\lambda}$, and in fact it is doubtful whether this is so. To find the average length of lifetime of a number of lives all of exact rating r would involve dealing with a group of discontinuous distribution, and the methods we have been using would not apply. Furthermore, lives of the same rating are not all exactly similar, for their ratings may be moving in different directions and with different speeds. It is likely that the expectation of life depends not only on the rating but also on other factors, and it is only when considered as part of a continuous distribution that a life of rating r can be treated as having an expectation of $-\frac{\log r}{\lambda}$.

The question may be raised as to whether it is not possible to assume that the ratings are in the first place fixed so as to be dependent on the estimated expectation of life. We may, if we like, make this assumption, but unless our referee is endowed with prophetic powers, which is not intended, the true expectation of life will not necessarily be the same as the estimated expectation. As an alternative it could be assumed that the ratings are fixed so as to be dependent on the estimated value of a continuous annuity at some given rate of interest payable throughout life; this might not give the same result, but would be a more reasonable assumption if it is thought that the referee can predict the immediate future more accurately than the distant future. In fact, there are unlimited numbers of ways in which the ratings may be fixed, and there is no need to tie ourselves down to any particular one. Furthermore, any method which may in the first place be adopted will lose some of its significance if the health scale is afterwards changed, as explained on pp. 338–40, so as to make the health frequency distribution of the group homogeneous.

Acknowledgement is due to Mr W. Perks for the use, firstly, of the phrase "natural scale" which appeared, though with not precisely the same meaning, in his paper "On some Experiments in the Graduation of Mortality Statistics" (*J.I.A.* Vol. LXIII, pp. 12 *et seq.*) and, secondly, of the illustration involving cubes of different sizes which is taken from the same paper.

ABSTRACT OF THE DISCUSSION

Mr W. G. Bailey, in opening the discussion, said that everyone was familiar with the expression that "So-and-so was going downhill", but those who had considered the operation of mortality from the point of view of deterioration must have found difficulty in fitting into the picture those lives who died from accident or some other cause when they were to all intents and purposes in excellent health. The author's health frequency distribution purported to classify lives according to their state of health, but the author had insisted that every life of rating r must pass through every intermediate rating before passing out through the final ordinate at death, and to meet the difficulty mentioned he had introduced a "universal selector" who was empowered to allot to the lives in question special ratings higher than those to which they would be entitled on health alone. Unfortunately, that was an integral part of the theory, because unless those lives were granted such special ratings the flow through rating r would not depend on the height of the health frequency curve at r alone, so that the conclusions at the bottom of p. 322 and the top of p. 323 would be invalidated. On that account the distribution was not, he thought, entitled to the description of "health frequency distribution".

Furthermore it seemed that the author was toying with the question of determinism. He appeared to regard a probability as arising simply on account of heterogeneity of data, i.e. if the data could be subdivided sufficiently a probability would be turned into a certainty. Though that was denied in the last paragraph of the Appendix, colour having been lent to the denial by the substitution for the "universal selector" of the net rate of deterioration, the objection still remained, because the rate of deterioration of those lives who died from "force of impact", as it might be called, was artificially rated down. The special treatment which those lives received suggested that the health frequency distribution was related more to the curve of deaths than to a classification by state of health, which was confirmed by the expression for the expectation of life on p. 362 and by the equation at the top of p. 332.

In effect the author had provided a draughtsman's pantograph. If one pointer of the apparatus were run over the μ_x curve, the health frequency distribution would be produced and *vice versa*. If it were possible, therefore, by some means to arrive at the health frequency distribution direct from the data, the curve of deaths could be obtained therefrom; but it seemed more probable that it would be necessary to rely upon the curve of deaths to give the health frequency distribution, which was in itself of comparatively little significance.

He was rather puzzled to know just what the author was seeking when he set out on his classification. If it was homogeneity of data, then homogeneity in respect of what? It was obviously not homogeneity in respect of "health"—at any rate not in respect of "health as we know it"—because of the treatment of sudden deaths. In the paper it was stated that homogeneity in respect of rate of deterioration or expectation of life had not been achieved. Possibly the author could say whether any practical result would have been forthcoming had he constructed his distribution according to the normal meaning of "state of health" and allowed exits from each grade by death as well as transference.

He was not sure that actuaries were not beating the air in their search for homogeneity, and were sometimes apt to forget that heterogeneity did not matter provided that the constitution of the data remained constant. It was only because the constitution was liable to vary that in constructing tables a distinction was made between black and white, male and female, ordinary and industrial, and other factors leading to discrete variation. How much further were they to go in that subdivision? They must go some distance, because they knew from the phenomenon of temporary selection that not only were their data not homogeneous but the constitution of the data varied. But were they justified in assuming, as was suggested on p. 340, that no two lives were exactly alike, or in assuming that for every age x, μ_x could take any value? They had no evidence beyond knowledge of the variation of μ_x with the duration of true selection, and they had no evidence that would provide them with a frequency distribution of lives according to their individual forces of mortality at age x. Even supposing that lives did differ, there was only sufficient evidence for two groups, select and non-select; and, grateful as they must be to the author for throwing light into yet another dark place, it seemed that more practical results would be obtained by exploring the possibilities of such a subdivision than by postulating infinite individual differences whose frequencies could not be measured.

Mr H. W. Haycocks said that the author had given a very excellent exercise in pure theory, having started off with certain premises, from which the rest of his argument logically followed. He had not asserted that his premises were empirically true, but the implication of the concluding section of the paper was that a theory must be so formulated that determinate propositions could be inferred from it by logical means alone, and, further, that among those propositions must be some capable of empirical corroboration or refutation—or, briefly, that a theory should be verifiable. Surely the most important test of a theory was its application to practice.

Mathematics was nothing more than an efficient method of reasoning; the conclusions always asserted either the whole or a part of the initial premises, although in another language. Fontenelle had said: "Mathematicians are like lovers. Grant a mathematician the slightest principle and he will draw therefrom a consequence which you must also grant him, and therefrom another consequence, and so on." It followed, then, that in order to appraise the author's theory, it was necessary first of all to examine carefully his initial hypotheses. They were asked to imagine a super-referee who had full knowledge of all the factors affecting the longevity of any life, and who could sum them up and arrive at precise conclusions. Thence they derived their initial health frequency distribution and the net force of deterioration. It should be noted, however, that the referee was not endowed with any prophetic powers; it appeared therefore that the frequency distribution could be traced from the past to the present, but that, as with present-day mortality tables, the future was speculative. Apparently the referee knew which buildings were just on the verge of collapse, but could not predict wars, epidemics or earthquakes.

The idea of a scale ranging from 0 to 1 drew attention to the theory of probability where there was a similar scale, the precise meaning of which had given rise to much controversy. The classical theory regarded probability as a degree of belief in a proposition capable of quantitative measurement. In the case of mathematical probability, the implied probabilities were deduced analytically from given data; there was no question of postulating a referee with *a priori* knowledge of quantitative measurements of degrees of belief. In practice the problems for solution usually either were concerned with games of chance, involving the mathematical theory of arrangements, or depended on observed relative frequencies, from which the probabilities were inferred. That in fact was the basis of most of the practical work of actuaries; they inferred probabilities of death from observed relative frequencies. There was, however, no such simple way of inferring relative degrees of health. Medical science was too little advanced and the social background too unstable for any referee even to approximate to the author's force of deterioration or even his ideal health frequency distribution; and even if the medical officer were to classify a group of lives into five or six broad groups from the health point of view, it would be possible to do no more than fit a graduation curve to the result and say that according to the medical officer's opinion that curve represented the health distribution of a particular group of lives at a particular time. He was doubtful whether any further observations would lead to a satisfactory and intelligible function for the net force of deterioration.

With regard to the possibility of verifying some of the author's propositions deduced by his mathematical analysis, the reduction of the expression for the force of mortality to a Perks formula appeared to be the most hopeful case. But the reduction had been achieved by two very arbitrary assumptions, one of the expressions having been made constant, and the other having been given a Makeham form; and without some empirical evidence to justify such substitutions, it seemed that they constituted no evidence in favour of the theory. The author's remarks that the constants A, B, D and c were interdependent were interesting, and in an able investigation, *Mortality Variations in Sweden*, Cramér and Wold had arrived at a similar conclusion. It was impossible to regard either Makeham's formula or Perks formula as exact representations of the forces of mortality; they were merely convenient graduation formulae, useful short-hand descriptions of certain characteristics of the data under consideration. It was necessary to avoid giving any biological significance to them, such as that the constant c represented a physiological factor and A and B environmental factors. He was doubtful, therefore, whether anything of value could be gained by trying to reduce the author's expressions to either of those types of formulae.

It seemed that advances in actuarial science, particularly in the field of prediction, which was after all its most important function, would depend rather on investigations such as that of Cramér and Wold, or upon investigations which attempted to correlate the force of mortality with such factors as housing density, social class, and occupation. Since the author had postulated a being who had almost *a priori* knowledge of the law of mortality, it was not possible to expect that his theory would arrive at any practical results. Nevertheless, he (the speaker) attached some value to the theory, and considered that it was of use as an aid to thought and to the clarification of certain problems such as selection. He would however, for the sake of students, have preferred a rather simpler exposition, as, for example, one based entirely on geometrical curves; for such curves could give a very simple picture of intricate problems and had been used with great effect in other social sciences, particularly in pure economic theory. It was natural to regard health as being represented on a scale ranging from bad to excellent, and he thought that the paper must lead to speculations which might eventually arrive at an increase in knowledge of mortality.

Mr L. E. Coward criticized the author's assumption that health could be measured in a simple scale from 0 to 1. Many qualities, for example a mixed colour, could not be so measured. He asked whether in the author's view the function λ_x was more fundamental than μ_x, and whether it was a function which in ideal conditions could be represented in a simple form.

In the mathematical part of the paper there were two arbitrary assumptions which seemed to strike at the root of the whole theory. The first was that in choosing his "natural scale" the author had made his force of deterioration equal to $r\lambda_x$. That was purely arbitrary, and by making the force of deterioration equal to some other function of r a different formula for μ_x could be obtained—provided of course that the resulting differential equation could be solved. Thus there was an indefinite number of expressions for the force of mortality of equal generality with that given in the paper as "the natural law of mortality". As an experiment he had put the force of deterioration equal to λ_x alone, and found that

$$f(x, r) = f\left(0, r - \int^x \lambda_x dx\right)$$

the frequency distribution thus moving along unchanged in shape. There were thus no lives at the lowest ratings after a certain time, but that was not an insuperable difficulty because it merely signified that no life at age 80 could be as healthy as the best lives at age 40, for instance. Earlier in the paper it was assumed that health could be measured on a scale running from 0 to 1. By taking different limits for the scale, other expressions for μ_x could be found. If those two arbitrary assumptions were taken away there would appear to be very little left in the theory.

It was mentioned on p. 332 that the frequency curve at time 0 might be regarded as a distorted curve of deaths and therefore he wondered what would happen if it were made to coincide with the actual curve of deaths. That was perfectly feasible. It was necessary to measure health from 0 to ω and to make the frequency curve move back at uniform speed (death occurring at rating 0), and eventually the satisfactory conclusion was reached that $\mu_x = \mu_x$. For those reasons he thought that the paper could be likened to a scientific experiment, valuable in itself, but which had given a negative result.

Mr A. W. Joseph said that the author had placed in the hands of actuaries an instrument of great power for clarifying and explaining certain features of mortality. Though the paper as a whole appealed to him, certain parts of it raised doubts in his mind which he had not been able to dispel.

He agreed with what had been said about the difficulties of having a supernormal referee, and thought it was particularly unfortunate that such a conception had been introduced into the theory because the theory gained nothing thereby. In that respect he disagreed with the opener. The "universal selector" was not an integral part of the theory. As the author developed his theory he discarded all the abnormal powers of the referee. The further the referee could see into the future, the smaller the force of recuperation. Yet the author contemplated quite high forces of recuperation. No more than consistency was required of the referee for the theory to work. If the lives were placed in order a second time after a short interval many violent changes of order should not occur. That condition, coupled with the fact that deaths must be placed at the end of the line, certainly implied that the referee had good judgment of health. For instance the lives might be ordered according to alphabetical order of names; and that would be consistent to some extent. There would, however, be a sudden jump to the end of the line when a death took place, and a large number of such violent changes would not be in accordance with the criterion which should be applied to the referee. With good judgment of health there would not be violent changes. There was no need for the powers of the referee to be greater than those of some actual person or group of persons, which was fortunate, for otherwise the theory would have no practical value.

He found great difficulty in understanding the section of the paper headed "Homogeneity of the Health Frequency Distribution". The author tried to divide the group of lives into sub-groups consisting of lives all of the same type, lives being defined to be of the same type if they were subject to the same forces of deterioration and recuperation. If, however, the original scale were altered as described, and hence also the forces of deterioration and recuperation, would the lives still be homogeneous? There was another objection to the method used for separating the lives into groups, for consideration of Diagram III showed that the sub-groups would consist mainly of lives with ratings of, say, 0 to 0·1 and 0·9 to 1 for the first group, 0·1 to 0·2 and 0·8 to 0·9 for the second group, and so on. When the scale for each group was changed—or distorted would be a better description—to make the ratings extend from 0 to 1, and the sub-groups then united, the result would be to bring into conjunction lives most decidedly non-homogeneous, and the final order would be worse than the first. It was assumed that the resulting scale would be a natural one, which was true though not at all obvious, and incidentally it was an interesting mathematical problem to show that two groups of lives of different mortality and ordered by different natural scales were ordered in a natural scale on being combined only if the factors of deterioration were identical or if the equal-distribution scale were used for both groups.

He could see the author's reason for including the section he had criticized. The device of natural scales made possible the solution of a partial differential equation which otherwise might be insoluble. But the solution was not given without cost; it had to be paid for, and the payment was that the scale depended on the group of lives itself. That characteristic of the method made it difficult to compare two or more groups of lives. If, for example, two groups of lives were ordered by the same referee, could natural scales be chosen such that when the groups were combined the lives would appear in the same order in which the referee would have placed them if he had been given them all in the first place? Without having examined the question at all thoroughly, he felt fairly certain that the answer was in the negative. Yet something of that nature was required if the theory were to be of use as a means of comparing and contrasting the mortality of two groups of lives. He did not think the author had succeeded in his attempt to overcome that disadvantage of the theory.

On the other hand the behaviour of one group of lives only could be investigated from a qualitative, not a quantitative, point of view very successfully by the theory. For example, the discussion of selection was very good, and he admired the way in which it was shown from the shape of the frequency curves of Diagram XI that the mortality of a group of select lives might become greater than that of the ultimate and then approach the ultimate from above instead of below. In fact a simple digest of the paper, omitting all mathematics, giving the idea of ordering the lives and the device of an appropriate scale, showing the shape of the frequency curves when such a suitable scale was chosen and then proceeding to discuss selection, would be most helpful to students.

A point of interest was that the factor of deterioration for any group of lives could be made to take any required form by correct choice of the initial frequency distribution. In particular the factor could be made constant, and all the advantages of a constant factor of deterioration could thus be obtained without the scale of time being changed. The discussion of expectation of life at the end of the Appendix would still be applicable.

It was possible to simplify § 4 and § 6 of the Appendix, which, as they appeared in the paper, were rather difficult to follow, without the necessity of solving partial differential equations. The essential (and obvious) relation between r and $r'(x, r)$ was

$$\int_0^r f(x, r)\,dr = \int_0^{r'(x,r)} f'(x, r')\,dr'. \qquad \text{......(1)}$$

That relation was really more fundamental than the author's equation (7) which, however, followed immediately from it by differentiation with respect to r. In particular

$$\int_0^r f(0, r)\, dr = \int_0^{r'(0,r)} f'(0, r')\, dr'. \qquad \ldots\ldots(2)$$

Using S_x to denote the survivors of all the lives at time x,

$$S_x = \int_0^1 f(x, r)\, dr = \int_0^1 e^{-\int_0^x \lambda_x dx} f(0, re^{-\int_0^x \lambda_x dx})\, dr = \int_0^{e^{-\int_0^x \lambda_x dx}} f(0, s)\, ds,$$

where $$s = re^{-\int_0^x \lambda_x dx}.$$

Similarly $$S_x = \int_0^{e^{-\int_0^x \lambda'_x dx}} f'(0, s')\, ds'.$$

Therefore $$\int_0^{e^{-\int_0^x \lambda_x dx}} f(0, s)\, ds = \int_0^{e^{-\int_0^x \lambda'_x dx}} f'(0, s')\, ds'.$$

By comparison with (2) above, if $e^{-\int_0^x \lambda_x dx} = r$, then $e^{-\int_0^x \lambda'_x dx} = r'(0, r)$,

i.e. $$e^{-\int_0^x \lambda'_x dx} = r'(0, e^{-\int_0^x \lambda_x dx}),$$

which was the result obtained in § 6.

Substituting the values of $f(x, r)$ and $f'(x, r')$ in (1),

$$\int_0^r e^{-\int_0^x \lambda_x dx} f(0, re^{-\int_0^x \lambda_x dx})\, dr = \int_0^{r'(x,r)} e^{-\int_0^x \lambda'_x dx} f'(0, r'e^{-\int_0^x \lambda'_x dx})\, dr',$$

i.e. $$\int_0^{re^{-\int_0^x \lambda_x dx}} f(0, s)\, ds = \int_0^{r'(x,r)e^{-\int_0^x \lambda'_x dx}} f'(0, s')\, ds'.$$

Comparing with (2) above, if $re^{-\int_0^x \lambda_x dx} = R$, then

$$r'(x, r)\, e^{-\int_0^x \lambda'_x dx} = r'(0, R),$$

i.e.

$$r'(x, r) = r'(0, re^{-\int_0^x \lambda_x dx}) \div e^{-\int_0^x \lambda'_x dx} = r'(0, re^{-\int_0^x \lambda_x dx}) \div r'(0, e^{-\int_0^x \lambda_x dx}),$$

which was the result obtained in § 4.

Mr W. Perks referrred to the formula of which he had made certain use, and which as shown in the paper could be arrived at by the author's theory, subject to a certain relationship between the constants A, B, D and c. He pointed out, however, that the relationship would make it impossible for the formula to fit any usual form of mortality, whatever value was given to n.

The suggestion in the last paragraph of the Appendix that it would be reasonable to assume that the referee could predict the immediate future more accurately than the distant future led to the idea that the referee might be able to estimate in respect of each life the probability of death within one year, as an actuary had to do in his every-day work. Instead of one year a period of six months or three months might be taken, and eventually the point might be reached where it was assumed that the referee could estimate for each life the force of mortality. Under the author's theory, except for lives where health rating was unity, the force of mortality was nil; in the frequency distribution there was a general flow towards the right, caused by deterioration of health, death taking place on the extreme right of the distribution. If, on the other hand, it were assumed that there was such a thing as $\mu(x, r)$, the movement of the distribution could be interpreted on a different basis, namely with no flow but with deaths occurring at all ratings. The fall taking place in the frequency distribution for any rating would be due entirely to deaths at that rating, no lives moving to higher or lower ratings. The whole of the mathematics would be the same, as could be seen from equation (2) on p. 353. If the right-hand side were differentiated by parts, then dividing through by $f(x, r)$ the equation became $\mu(x, r)=\lambda_x\{r\omega(x, r)+1\}$, where the symbol $\omega(x, r)$ was used for the slope-relation of $f(x, r)$, in the r-dimension, in the same way that $-\mu(x, r)$ was used for the slope-relation of $f(x, r)$ in the x-dimension. The slope-relation in the x-dimension was thus the force of mortality on the basis of no flow. Corresponding to the author's analogy of the tank of water, the assumption that $\mu(x, r)$ was a force of mortality was the same as supposing that the tank had a perforated bottom, the water flowing out everywhere downwards but not sideways. A better analogy perhaps would be to suppose that heat was applied to the bottom of the tank so that the water evaporated into steam, which appeared to be particularly apt since death might be considered as a transition of life from one form to another.

It was interesting to consider certain special cases of the author's distributions from the point of view of $\mu(x, r)$. The author had dealt with the case of equal distribution; for such a distribution the slope-relation in the r-dimension was nil, i.e. $\omega(x, r)=0$. The author had shown that $\mu_x=\lambda_x$; and it could also be shown that $\mu(x, r)=\lambda_x$, i.e. that notwithstanding the health rating, the lives all had the same force of mortality. Another case was obtained by assuming that the slope-relation in the r-dimension was $1/r$, so that $f(0, r)=r$; the force of mortality for age x then became $2\lambda_x$ and $\mu(x, r)$ was also equal to $2\lambda_x$. If $\omega(x, r)=a/r$, then $f(0, r)=r^a$, and $\mu(x, r)=(a+1)\lambda_x$. An interesting case was obtained by assuming that $\omega(x, r)=1-1/r$, for it led to the result $\mu(x, r)=r\lambda_x$, i.e. the force of mortality was proportionate to the rating; it had, however, the disadvantage that $f(0, r)=e^r/r$, which became infinite when $r=0$.

It was clear that two distributions with the same values of λ_x could be combined in any proportions and that the combined distribution would conform to the author's conditions. If one of the distributions were assumed to be subject to flow and no mortality (except for rating unity) and the other to forces of mortality and no flow, the combined distribution would be subject to a mixture of flow and mortality. It was possible to go further and to write the differential equation in the form:

$$\mu(x, r) + m(x, r) = \lambda_x\{r\omega(x, r) + 1\},$$

where $m(x, r)$ was the net rate of flow in the r-dimension and was subject to the condition that $\int_0^1 m(x, r) f(x, r)\, dr = 0$. On the basis of an equal distribution, an obvious special case, fulfilling the conditions, was obtained by putting $\mu(x, r) = 2r\lambda_x$, whence $m(x, r) = (1 - 2r)\lambda_x$. Other special cases could be obtained by assuming alternative forms for $\mu(x, r)$ and then deducing the forms taken by $m(x, r)$ and $f(x, r)$.

The modification in the theory of the paper which he had suggested by the introduction of $\mu(x, r)$ provided a link with his own paper on Graduation (*J.I.A.* Vol. LXIII, p. 12), where he had discussed a similar problem under the heading "Some Speculations on the Theory of Mortality"; and he congratulated the author on solving the problem which some ten years ago he himself had given up.

Sir William Elderton remarked that the paper had impressed him as a refreshing piece of originality. He regretted that two of the previous speakers should have suggested that it was something that required making easy for "students", in the examination sense, to digest. He did not consider that the author need worry if he wrote an original paper and some students could not fully understand it, because some day perhaps they would. He had enjoyed the paper so much that he would not like the author or anyone else to think that such remarks as he might make were meant in any way to detract from it. His remarks were largely in the nature of questions, and probably the author had already thoroughly considered and formed the answers to them in his mind.

In connexion with the scale of health, he thought that the author's picturesque phrases had perhaps been misunderstood by some previous speakers; he had never understood the author to say that a scale could in the present stage of knowledge be formed, but merely that it was possible to imagine that one might be formed. He considered that it was easier to regard the health frequency distribution as tailing off to zero at the ends of the scale, rather than taking the form in which it was dealt with in the paper. Such a distribution, under which $f(0) = f(1) = 0$, immediately suggested the expansion of $(a + b)^n$ where $a + b = 1$, or (if a discrete series was unsuitable) a formula of the type $r^m (1 - r)^n$. In particular he thought that from the latter expression it might be possible to study the effect of variation in health or age, or by changing the values of m and n of variation in mortality. However, the way in which the particular thing was interpreted did not seem to him necessarily to matter; it was largely a question of the temperament of the person who interpreted it, and certainly the author had got a great deal of amusement out of his own interpretation. Incidentally, he was quite sure that the author had given considerable pleasure to those who had studied the paper, and certainly to himself.

With regard to the remark concerning the duration of the select period at the top of p. 350, he thought that it would be more correct to say that the duration was often assumed, for purpose of convenience, to be independent of the age at entry. There were certainly many experiences in which the duration varied; he instanced the O[NM], Japanese and Norwegian experiences, and said that there were other experiences where the period of selection was shorter for the younger than for the older age at entry. Concerning the reference at the foot of p. 350 to the applicability of the "census" method to determine the rate of mortality at short durations, he thought that it would be difficult to improve on a form of that method by which the deaths were tabulated for each month of duration over one, or two, calendar years and the number of insured for the same duration in the middle of each calendar year.

It was of interest to note that it was not necessarily the case that the health of the lives in a group would on the whole continually deteriorate. Suppose, for instance, that a country had passed through a serious epidemic, or had a population where the mortality was improving from, say, age 20 to age 35 (as did sometimes happen), then it might be that, under a reasonable health scale, the frequency curve was shifting not up the scale from 0 to 1, but down the scale from 1 to 0. In conclusion, he asked whether it was not possible to extend the author's method and wipe out age altogether. Under certain circumstances, for example, it was conceivable that a man aged 80 might have the same rating as another man aged 40; one such case might be if both men were dying from pneumonia, and there were many other cases as well. If, as he believed was permissible, a complete scale were assumed running from 0 to 1, every individual at every age would fall somewhere on that scale; and if the whole population were treated in one, disregarding age, a method on the lines of the author's might lead to interesting results.

Mr R. E. Beard said that he had been interested in the study of factors underlying mortality and that his researches had led to the differential equation arrived at in the paper. His starting point had been the following conception of mortality (for which he claimed no originality):

"Consider an individual as starting life with n units of resistance to destruction and let there be a natural time such that the chance of losing a unit in a short interval of this time is constant at all parts of the scale. The process of growing older is such that units are lost as time increases, the chance of a loss being as defined above, and an individual is to be considered dead when he has lost all n units."

From such assumptions a differential equation could be written down and various solutions could be obtained depending on the initial conditions. A finite number of condition groups would arise from the analysis, but if continuous variation were assumed the solution followed the lines developed by the author. If necessary a rate of annihilation could also be introduced to deal with accidental deaths.

As a result of his investigations he had been led to focus his attention on the curve of deaths, the difficulties encountered in dealing with which had been brought out in Mr Phillips's paper in *J.I.A.* Vol. LVI. He had not yet developed a sufficiently powerful technique for dealing satisfactorily with the problem and as a result the few arithmetical trials he had made had proved interesting rather than successful. A special case of the solution, which gave considerable promise, arose when the curve of deaths, related to the natural time, was the Incomplete Gamma Function. The natural time and age scales he had been led to consider as geometrically related, in which connexion it was interesting to note that the system of curves investigated by Mr Perks gave rise to expressions for l_x related to the Pearsonian system of frequency curves. However, it should be remembered that his own experiments were based on the curve of deaths and not on the l_x-curve and considerable differences in the value of the parameters were therefore to be expected.

Further evidence of a different nature was difficult to find, but in a book entitled *Biological Time* by Le Comte de Noüy, some investigations into the question of a natural time had been made, and the author had derived a relationship of the form $1/x$ where x was the age in years. It was, however, comparatively easy to fit a simple geometrical progression to the statistics and some progress might result from considerations on such lines.

Though his remarks might be considered rather irrelevant to the paper, they should be regarded as endeavours to obtain special cases from the general solution and, in his opinion, it was in that direction that the practical utility of the paper lay.

Mr Duncan C. Fraser remarked that the author's ideas were particularly interesting at the date of the discussion when the winter with its forces of deterioration was past, and they felt the forces of recuperation coming into full play. Ratings of members one month before and one month after Easter might be very different.

The early part of the paper might, he thought, be regarded as an exercise in the relations of forces analogous to the force of mortality. Instead of using the author's function $f(x, r)$ he had found it helpful to make use of the integral of that function. The integral represented the number living at age x with ratings not greater than r, and might be called $l(x, r)$. The function $f(x, r)$ was then the differential coefficient of $l(x, r)$ with reference to r (x remaining constant): and f/l represented the force of flow of l through r when r varied and x remained constant. Commencing with expressions for the force of flow of $l(x, r)$ with reference to x (r constant) and with reference to r (x constant), the equations connecting those with the author's functions would lead easily to the equal-distribution scale, to the natural scale, and to the fundamental differential equation on p. 353.

The solution of the differential equation depended in practice on familiarity with the differential coefficients of such functions as e^x and e^{e^x}, with the variations which were introduced by the introduction of constants and + or − signs.

The author had used a variable measure of time proportional to the factor of deterioration, which in certain circumstances was equal to the force of mortality. In that connexion it was interesting to notice that the reciprocal of μ was itself a measure of time, since its value at age x might be defined as the number of units of time in which the number living at the age of x would be exhausted if the daily deaths continued to be constant and the same as at the exact age x.

At the end of the paper a simple formula was given for the expectation of life, depending on the function λ, which in the equal-distribution scale became μ; and it would be useful if numerical illustrations could be given of the application of the formula. For a given age x there must be an average value of the rating r for which the expectation would be equal to $\mathring{e}_x$.

Diagram VI showed how lives passed through various ratings as time went on, and the essence of it appeared to be that the rating at a time x, was connected with the rating at a previous time 0. They had had discussions on the subject of the mortality of lives of the same generation, and the idea of "vintage years" of birth had been suggested. It seemed that the important fact was that the group of lives born in the same year passed through the same phase of the infantile epidemics, and that the mortality of the group in the adult years of life was influenced by the condition in which members of the group were left as a result of such epidemics. That seemed to fit in with the author's theory and he thought that there might be a subject there for practical investigation.

Mr R. D. Anderson, in closing the discussion, said that, to his mind, the principal point which emerged was that various speakers had made suggestions as to alterations which could be made in the assumptions made in the paper without having any particular effect on the result, so that the same conclusion was reached. It seemed to him that the health distribution was really an *ex post facto* distribution designed to throw up the deaths when required, whether they were deaths from illness or accident. In view of that the author could make his distribution change its shape at pleasure, and, as he had said, by a suitable choice of scale it was possible to make the factor of deterioration take any required form. There was no objection, of course, to making imaginary distributions, provided they were recognized throughout as mathematical artifices, like co-ordinate systems, and not later on given objective reality. Perhaps the author had sought a unified formula because he was a mathematician, and the following quotation from Eddington in that connexion seemed to be apposite:

"A unified theory does not necessarily mean a unified formula. The latter kind of unification is exemplified by the theory of the 'Generalized astronomical instrument' which combines in a single equation the theory of the altazimuth, meridian circle, prime vertical instrument, equatorial and almucantar. Such compression appeals more to the mathematician than to the physicist."

—to which, he thought, might be added "or to the actuary". In his opinion that was the objection to a general formula, which, by making suitable assumptions, could be reduced to Mr Perks' formula or to other formulae.

He thought that the conclusion to be drawn from the paper was that if it were desired to study mortality it would be necessary to study mortality and not health, and that what the author had done was not to explain mortality in terms of health but health in terms of mortality.

The President (Colonel H. J. P. Oakley), in proposing a vote of thanks to the author, said that there had been a well-sustained discussion, and he was sure that the members would wish him to congratulate the author on a most interesting and original contribution to the proceedings. It was not often that any reference was made to the title of a paper, but he remembered many years ago, when a paper was read on "Masculinity", a distinguished member of the profession questioned the title and asked "If masculinity why not femininity?" When he looked at the title chosen by the author and became absorbed in his unique presentation of the subject, he said to himself "Why 'Theory of Mortality'? Why not 'Practice of Living'?" He liked the simile in the author's opening paragraph wherein he likened life to a pebble rolling downhill until finally it disappeared over the precipice. It was a pretty picture, but where was the precipice? If that were known, then life would be a term certain, but the precipice itself was variable. For some few fortunate people it might never be, and life to them was a gentle roll all the way until they glided down to the shore at the edge of the great river.

There had seemed to be a danger at one point in the discussion of considering the matter too much from the point of view of the individual life. Actuaries were not concerned with individual lives except in underwriting, and then, having underwritten the life, they put it into a group, and it was with groups of lives that they were concerned.

Whether it would be possible to look for any practical development from the author's theory remained to be seen. It would, he was sure, start many actuaries on a new line of thought, and for that alone there was good reason to thank the author. It might be that in the future the paper would be referred to as an epoch-making one, giving fresh impetus to the study of mortality or the duration of human lives.

Mr C. D. Rich, in a preliminary reply, said that he felt in relation to his paper rather like the novelist who was described as writing a book not by inventing a plot, but by introducing characters and then waiting to see what happened to them. When he commenced his investigation he had no preconceived ideas as to the results he would obtain, and it was partly for that reason that he believed his conclusions were right.

Mr C. D. Rich has sent the following further reply to the discussion.

Before replying to the various points raised in the discussion, I would first like to express my thanks to Mr Anderson whose paper "Select

Mortality Tables with special reference to the A 1924–29 Mortality Tables" (*J.I.A.* Vol. LXVIII, p. 223) set me on the train of thought which led to the theory I have put forward. I felt that the assumption on which Mr Anderson's paper was based, namely that in a mixed body of lives the net rate of flow from the select group to the non-select group is proportionate to the size of the former, was open to objection; and the idea occurred to me therefore of investigating select mortality on the assumption of an infinite number of possible states of health. I found that before considering the mortality of a select group of lives it was necessary to deal with mortality generally, and some of the results I obtained in the course of the investigation seemed of sufficient interest to warrant my calling my paper "a general theory of mortality" rather than "a theory of select mortality". It is perhaps worth while remarking that though many criticisms were made regarding the paper, there was so far as I can see none relating specifically to that part of it which is concerned with select mortality.

A few speakers found fault with the paper on grounds which can be summed up in Mr Anderson's words: "what the author had done was not to explain mortality in terms of health, but health in terms of mortality". I think that this criticism is a little unfair, because it is impossible to explain the unknown except in terms of the known. It would be more accurate to say that I have shown that if health were measured it would be found to be closely related to mortality, which is of course what might be expected. As regards Mr Coward's question whether λ_x is more fundamental than μ_x, the answer is that probably both functions are determined by the same underlying clause.

In view of the remarks made by some speakers regarding the health scale, it may be desirable to state my reasons for having discarded the human referee and introduced the imaginary "ideal referee". The first, and more important one, is the same as Mr Haycocks' reason for condemning him, i.e. "medical science is too little advanced". Many of the arguments raised might equally well have been applied, before the discovery of the thermometer, to prove that there was no such thing as temperature. Sir William Elderton's answer to these criticisms was very much to the point. Incidentally it was my intention that the referee should be capable of foreseeing the possibility of war or epidemic just as well as other factors which might be of significance to the life under consideration. My other reason was in order to avoid instantaneous changes in rating due to unforeseen accidents resulting in death or injury. This reason, however, disappears if it be admitted that every accident can be foreseen by a human referee by some interval of time, however small. Mr Joseph realized this, and I am in general agreement with much of what he said on this matter. There is, too, another way of dealing with sudden deaths (which could be extended to cover sudden injuries) as Mr Perks remarked; he showed that the problem of a health frequency distribution with exits vertically as well as horizontally can be resolved into the simpler problem of one with exits horizontally only, thus answering the question raised by the opener of the discussion. I think, however, that the idea of immediate exits by death at each state of health leads to complications if the scale is changed, and I prefer to adhere to the assumption in the paper that death can only take place if the rating passes through the value 1.

The exposition in the paper that Mr Perks' formula is a special case of what I called the "natural law of mortality" was included as a matter of interest only. There was no intention, as Mr Haycocks thought, of trying to justify the formula otherwise than for curve-fitting purposes; I might observe, however, that to make λ_x constant is not an arbitrary assumption since it is shown in the Appendix, § 7, that this can be done by suitable choice of scale. I had not realized that, as Mr Perks pointed out, the relationship which must exist between the constants would make the formula unsuitable for any normal mortality table.

The purpose of the section of the paper entitled "Homogeneity of the Health Frequency Distribution" was to demonstrate that it is possible to frame the health scale so that all lives of the same age and rating may be treated as subject to the same net force of deterioration. There is, of course, no reason for this if a single group of lives is under consideration, but it is of fundamental importance if one group is to be compared with another, or part of a group with the whole, as is necessary when dealing with select mortality. Unless the factor of deterioration is the same for the select as for the control group, then I do not think that the conclusions in the paper regarding selection necessarily follow. Mr Joseph appears to have misunderstood the remark at the foot of p. 338: "lives being defined to be of the same type if they are subject to the same forces of deterioration and recuperation". The forces of deterioration and recuperation are functions of the age and rating, and it was meant that lives of the same sub-group would be subject to forces of the same magnitude if of the same rating, but not if of different ratings. My intention was that lives suffering from similar impairments to health (though not necessarily of the same degree), or subject to similar occupational risks, would be regarded as being of the same type. The necessity does not arise, therefore, of having specially to bring together lives from opposite ends of the scale, and there should not be any violent distortion of the scale for each sub-group.

Some of the criticisms raised in the discussion were met by other speakers, and there is no need for me to deal with them. The point regarding the "arbitrary assumptions" referred to by one speaker was answered by Mr Anderson when he remarked that alterations could be made in the assumptions without having any particular effect on the result. For instance, to measure by a natural scale rather than by some other scale is no more an arbitrary assumption than to measure a length in feet rather than in metres. The use of natural scales does however enable us to reach the results regarding selection which might not otherwise be obtainable. Another speaker would have preferred a simpler exposition of the theory based on curves, but such an exposition would be useless unless first justified by mathematical proof; this proof is given by the paper, and it is open to anyone to construct curves to deal with particular applications of the theory.

I was greatly interested in Mr Fraser's suggestion concerning $l(x, r)$, defined as $\int_0^r f(x, r)\, dr$. Mr Joseph's elegant proofs also in effect involve this function, and it seems that the mathematical work in the Appendix can be very much simplified by its use. Mr Fraser said that: "For a given age x there must be an average value of the rating r for which the expectation would be equal to $\mathring{e}_x$." If this average value is r' and the variable measure of time be used, the corresponding expectation is $-\frac{\log r'}{\lambda}$, which must be equal to

$$\mathring{e}_x = \int_0^1 -\frac{\log r}{\lambda} f(x, r)\, dr \div \int_0^1 f(x, r)\, dr.$$

This enables r' to be determined, and the answer will depend on the form of $f(x, r)$. In the case of equal distribution, $f(x, r) = 1$, and the answer is easily seen to be $r' = 1/e$.

I am greatly indebted to Sir William Elderton for his kind remarks, and can only regret that I did not have the benefit of his advice before writing the paper.

There are two further matters to which I would like to refer. The first is an omission in the paper; in the discussion of select mortality there is no reference to withdrawals. If there are withdrawals from the experience, then unless the lives which withdraw are exactly similar to those which remain under observation, the conclusions in the paper do not necessarily hold good. The other matter is with regard to the shape of the curve AM in Diagram IX, which represents the effect of the "surge" on the distribution of a select group of lives. It is stated in the paper that this curve can probably not be determined exactly except by taking account of the speeds with which the ratings of individual lives are changing. Nevertheless, I think that without this knowledge it may yet be possible to find a general formula for the curve. A likely method would be by means of Fourier's Series, which has the advantage that it can be used to represent both discontinuous and continuous functions as, for example, the distributions before and after the surge. My attempts in this direction have not met with any success, but I mention the idea in the hope that someone with more skill than myself may solve the problem. If a general formula of this kind could be found, it would provide a theoretical relationship between select and ultimate mortality, and might be of great value.