JIA 85 (1959) 0165-0210

A STATISTICAL REVIEW OF THE EVIDENCE FOR THE EXISTENCE OF TEMPORARY SELECTION

By HILARY L. SEAL
Lecturer, Department of Zoology, Yale University

[Submitted to the Institute, 26 January 1959]

...Mr Higham deduces from his facts the simple rule that the effect of selection lasts for half the length of time intervening between the age at entry and 80; but...I am...led to the conclusion that, much as Mr Higham's estimate of the duration of the effect of selection exceeds that of some authors, it still falls short of the truth at the middle and most important ages of life.

T. B. SPRAGUE, 1870

Owing, indeed, to the long period after entry during which the mortality rates are affected by the 'self selection' of the annuitants, it is not sufficiently accurate to assume that after a period of five, or even ten, years the annuity values on select lives can safely be based upon an aggregate table.

G. F. HARDY, 1903

The honour of the ancient authors, and indeed of all, remains untouched; since the comparison I challenge is not of wits or faculties, but of ways and methods, and the part I take upon myself is not that of a judge, but of a guide.

FRANCIS BACON, *Novum Organum* (1620)

INTRODUCTION

The theory of 'temporary selection' is concerned with the variation, for fixed x, of $q_{[x-t]+t}$ the observed rate of mortality at age x during the $t+1$th year after the issue of an assurance or annuity contract. The classical view is that—apart from chance variations—$q_{[x-t]+t}$ increases gradually with increasing t until the effects of selection have disappeared after which time $q_{[x-t]+t}$ is a constant depending on x only.

Various reasons have been suggested for the persistence of temporary selection in an observed series of values of $q_{[x-t]+t}$. The chief of these are:

(1) The continuing effects of an initial selection on the part of the assurance company or by the annuitant (Morgan, 1834);
(2) The gradual withdrawal from assurance of healthy lives (Higham, 1851); and
(3) Secular improvements in medical selection or in the self-selection of annuitants (Karup, 1903).

But although some of the earlier writers emphasized the long continuance of selective effects, practical considerations have nearly always limited the period of temporary selection to a relatively short term of years. Thus, for example, the first graduated 'ultimate' table, the $H^{M(5)}$, was prepared by omitting the first five* years of assurance. However, longer select periods have been used;

* More correctly, the first 4½.

for example, the seven-year period of Karup's graduation of the Gotha 1852–96 data, the eight-year period used in the nine Russian companies' tables, and the 10-year period of the $O^{[M]}$ table and of Stolz's graduation of the seventeen Swedish companies' data—after an original graduation based on an unlimited period of selection.

In order to provide an idea of the size of the differences between successive $q_{[x-t]+t}$ in some well-known *graduated* experiences Table A shows various sets of values of $10^2 q_{[x-t]+t}/q_{[x-w]+w}$ for illustrative values of t. Here w is (except for the Swedish experience) the value of t at and after which $q_{[x-t]+t}$ was assumed constant. The relatively wide range of values between the experiences is noteworthy. And whereas the x-wise trend is upward for the $O^{[am]}$, it is approximately constant for the $O^{[M]}$ and the Swedish table, and is mainly downward for the $H^{[M]}$, the Gotha, the $O^{[NM]}$ and the British Government annuitants experiences. *Note that the difference between* $q_{[x-3]+3}$ *and* $q_{[x-4]+4}$ *is generally of the order of* 5 % *at all ages, but that the difference between* $q_{[x-4]+4}$ *and* $q_{[x-7]+7}$ *depends very largely on the select period adopted in the graduation.*

Table A

Age x	Experience	w	Values of $10^2 q_{[x-t]+t}/q_{[x-w]+w}$			
			$t=0$	3	4	7
35	$H^{[M]}$	5	49	95	100	100
	Gotha	7	57	89	93	100
	Swedish*	20	33	69	73	81
	$O^{[NM]}$	5	60	97	100	100
	$O^{[M]}$	10	43	78	82	93
	$O^{[am]}$	5	45	87	97	100
55	$H^{[M]}$	5	47	92	97	100
	Gotha	7	49	82	85	100
	Swedish*	25	38	68	72	79
	Brit. Govt. annuitants	4	71	95	100	100
	$O^{[NM]}$	5	65	94	99	100
	$O^{[M]}$	10	50	77	82	93
	$O^{[am]}$	5	56	91	98	100
75	$H^{[M]}$	5	42	100	100	100
	Brit. Govt. annuitants	4	55	83	100	100
	$O^{[NM]}$	5	60	93	99	100
	$O^{[M]}$	10	55	77	82	94
	$O^{[am]}$	5	63	94	99	100

* Ratios of m not q.

The theory of temporary selection that has gained general recognition in the text-books thus envisages the initial rise of $q_{[x-t]+t}$ followed by constancy for values of $t \geqslant t_0$ (t_0 being 5, 10 or even, in the limit, ∞). But there are valid reasons why observed series of values of $q_{[x-t]+t}$ could decrease with increasing t. Two such are:

(*a*) Secular improvements in mortality rates (Höckner, 1934); and
(*b*) Changes in the class of lives under observation (Elderton, 1906).

For example, the changes envisaged in (*b*) would occur where a young life office with relatively heavy 'class mortality' has no data to contribute for values of t exceeding 15, say. Although the effects of (*a*) and (*b*) are perceived more clearly for the larger t-values these changes in the constituent data actually begin to affect the q-values from the commencement. A glance at Table 2 *post* reveals that several authors have been convinced of the reality of a downturn in the curves of $q_{[x-t]+t}$. A number of additional examples may be found in Abel (1914).

Further doubts about the essentially upward trend in the q-values are raised when careful consideration is given to the reasons (1)–(3) quoted earlier. One may suspect that the supposed selective effect of withdrawals in assurance experiences should have made the presence of temporary selection more marked in these experiences than in annuity data. Such does not appear to have been the case. One may wonder, too, whether the passage of time has led to a more stringent or a less stringent medical (or 'self') selection. Finally, one may ask whether—'class' differences apart—an individual who was passed as a normal life by the doctor three years ago should be any different from an individual of the same age whose assurance dates from one or more years earlier. In other words do 'symptoms of death' manifest themselves among otherwise insurable lives more than three years preceding death?

If then the arguments against a continually increasing—or, at least, a never-decreasing—series of values of $q_{[x-t]+t}$ are plausible, why have so many authors found strong evidences of the continuance of temporary selection for 5, 10 or even more years following the issue of an assurance or annuity contract? It is suggested that the answer lies in two analytical oversights made by nearly every author responsible for the detection of the continuance of temporary selection. Both of them are concerned with grouping, in the one case of ages and in the other of durations.

Consider first the quinquennial grouping of entry ages so commonly adopted in the analysis of select tables (e.g. Higham, 1851; Sprague, 1870; Hardy, 1903). The following figures (Table B) are based on a table prepared by Chatham (1891). The right-hand side of the table shows that for every 5-year upward movement in age at entry there is a 'regression' in the mean age of the group of between ·1 and ·2 of a year on the average. Now since the upward change in q_x for a 1-year change in x is of the order of 5–10% in the $H^{M(5)}$ experience we may argue that the regression in the mean age of successive, quinquennially centred, values of $q_{[x-t]+t}$ is of the order of $\frac{1}{2}$% to 2% of q_x (ultimate) for every 5-year upward movement in t. A comparison of $q_{[x-5]+5}$ and $q_{[x-15]+15}$ in this manner would thus create an upward difference of between 1% and 4% even if the mortality rates at the corresponding individual ages were sampled from identical universes.*

* The age-drift in $q_{[x-t]+t}$ calculated for individual entry ages is about one twenty-fifth of the regression in the quinquennially grouped case. This may be seen by comparing the (approximate) mean-age of entrants distributed over the age-range $(x-2\frac{1}{2},\ x+2\frac{1}{2})$, namely

$$\int_0^5 z^2\,dz \bigg/ \int_0^5 z\,dz\ -2\tfrac{1}{2}=\tfrac{5}{6},$$

with that of entrants between $x-\frac{1}{2}$ and $x+\frac{1}{2}$, namely

$$\int_2^3 z^2\,dz \bigg/ \int_2^3 z\,dz\ -2\tfrac{1}{2}=\tfrac{1}{30}.$$

Table B

x	Deviation from x of mean entry-age of quinquennial group centred at x			Δ		
	Gotha 1829–78	H^M	Mutual of N.Y.	Gotha 1829–78	H^M	Mutual of N.Y.
25	+·57	+·19	+·79	·73	·21	·79
30	−·16	−·02	·00	−·11	·08	·10
35	−·05	−·10	−·10	·11	·06	·07
40	−·16	−·16	−·17	·04	·01	·06
45	−·20	−·17	−·23	·01	·03	·07
50	−·21	−·20	−·30	·04	·03	·07
55	−·25	−·23	−·37	·43	·11	·22
60	−·68	−·34	−·59	Mean ·18	·08	·20

A different type of error is introduced by the grouping of several durations t. Consider the two probability distributions generated by the binomials

$$(p+q)^n \quad \text{and} \quad (p+q)^m \quad (p=1-q;\, n \geqslant m).$$

If q, the chance of 'success', is small the probability that the relative frequency of 'success' in the first case (s/n say) exceeds that in the second (r/m say) is not one-half but may exceed it quite substantially. For example,

$$\Pr\left(\frac{s}{n}>\frac{r}{m}\right)+\tfrac{1}{2}\Pr\left(\frac{s}{n}=\frac{r}{m}\right)=\sum_{r=0}^{m}\binom{m}{r}q^r p^{m-r}$$

$$\times\left[\sum_{s=[nr/m]+1}^{n}\binom{n}{s}q^s p^{n-s}+\begin{cases}\frac{1}{2}\dbinom{n}{nr/m}q^{nr/m}p^{n-nr/m} & (nr/m)\text{ integral}\\ 0 & (nr/m)\text{ not integral}\end{cases}\right]$$

$$=\begin{cases}\cdot 512, & m=500, & n=1000, & q=\cdot 01,\\ \cdot 509, & mq=10, & nq=20, & q\to 0,\\ \cdot 544, & m=100, & n=1000, & q=\cdot 01,\\ \cdot 518, & mq=10, & nq=100, & q\to 0,\\ \cdot 506, & mq=100, & nq=1000, & q\to 0.\end{cases}$$

If therefore an 'ultimate' value of q_x is compared with $q_{[x-t]+t}$ where, as is very likely, $E_x \geqslant E_{[x-t]+t}$ the chance that the former exceeds the latter is greater than ·5 even though the corresponding universal mortality rates are identical. The same situation occurs whenever a comparison is made between two values of $q_{[x-t]+t}$ based on different numbers of lives exposed to risk. Note that if there is an even chance of 'plus' or 'minus' in a series of 100 plus and minus signs the probability of 59 or more pluses is ·0443; the occurrence of 59 plus signs in such a series would thus be considered significant if an excess of pluses was expected *a priori*. However, if the chance of an individual plus sign is really ·54 the probability of 59 or more pluses in a series of 100 signs has become ·1835 which is far from being significant.

The foregoing doubts about the validity of arguments in favour of long periods of temporary selection, and the realization that natural methods of grouping ages and durations create an appearance of continuing selection where none exists, led the author to re-examine mortality data that had already been subjected to analysis by the actuarial masters of yesterday. The results were very surprising.

This paper is thus a description of the statistical analysis of every available published mortality experience of assured and annuitant *lives* from which the values of $q_{[x-t]+t}$ could be calculated for *individual* values of x and successive integral values of t. During the course of some years of search the author has located and analysed 38 such sets of mortality data. On the basis of a 'null hypothesis' that the 'underlying true values' of $q_{[x-3]+3}$ and $q_{[x-4]+4}$ are identical at each age x, two of these 38 experiences would be expected to be 'significant' at the 5 % level. In fact four of them proved significant, and this is not an unusual deviation from expectation on the basis of the null hypothesis. It can thus be argued that temporary selection has never lasted for more than 3 years in any of the experiences in which it could be accurately tested.

Although the statistical techniques described herein do not apply directly to mortality experiences based on 'selections' instead of lives, a dozen such sets of observations were tested in a similar manner. Among these 12 experiences two were found to be 'significant' so far as temporary selection between $t=3$ and $t=4$ is concerned. The net result of the analysis of these 50 sets of mortality data is that although temporary selection may have amounted to as much as 5 % of the mortality rate between the fourth and fifth contract year in one or two of the mortality experiences published prior to 1916, it cannot have exceeded 1 % or 2 % in many of the others. Nor is there any evidence that there was a recrudescence of selective effects after the point of time $t=5$.

It is thought that a description of the statistical methods used to establish these results will be of interest and will serve to confirm the practical actuary in his employment of ultimate mortality tables excluding only the first two or three contract years.

ARE MORTALITY RATES DISTRIBUTED BINOMIALLY?

A typical set of mortality observations consists of a series of pairs of numbers $y_{[x-t]+t}$ and $E_{[x-t]+t}$ for each integral age x and with t-values equal to 0, 1, 2, ... Here y stands for the deaths at age x last birthday in the $t+1$th contract year and E represents the corresponding exposed-to-risk. The ratio

$$q_{[x-t]+t}=\frac{y_{[x-t]+t}}{E_{[x-t]+t}} \quad (t=0, 1, 2, \ldots)$$

is called the observed mortality rate at age x in the $t+1$th contract year and we may regard it as a sample value of a parameter to which q would tend if E were to become infinitely large. The question is: If repeated samples of E observations were taken from a universe with a fixed parameter value how would the resulting q-values be distributed?

To start with, let us assume that all the E individuals (*not* policies) come under observation at exact age x and that they can all be traced until they attain age $x+1$ or die prior thereto. We now make a further assumption that we will later discard, but which will be convenient in clarifying our ideas. This is that all E individuals have the same probability, $1-\pi$ say,* of dying before age $x+1$. We then have the familiar text-book situation and it may be shown that the probability of y deaths is given by the coefficient of z^y in

$$(\pi+\overline{1-\pi}.z)^E. \tag{1}$$

The resulting probability distribution for $y=0, 1, 2, \ldots, E$ is known as the Binomial of Bernoulli.

Our first step towards generalizing this set-up is to drop the assumption that $1-\pi$ has the same value for all the E lives in each sample. Even though the E individuals have been chosen with an eye to their homogeneity with respect to sex, social class, state of health, occupational status, geographic location, etc., we are intuitively sure that the real risk of death is very serious for some of the E individuals (who have undetected neoplasmic growths, for example), and quite slight for others. This conception might be expressed mathematically by writing $\pi_j, j=1, 2, \ldots, E$, for the probability of survival of the jth individual in the sample.† It is well known that with this change in the mathematical formulation the probability of y deaths is given by the coefficient of z^y in

$$\prod_{j=1}^{E} (\pi_j+\overline{1-\pi_j}.z). \tag{2}$$

The resulting distribution (see, for example, Aitken, 1939) is called the Binomial of Poisson. It was indicated by Seal (1943) that if the range of variation in the π_j is small this distribution can be approximated by the Bernoulli Binomial with $\pi=E^{-1}\sum_{j=1}^{E}\pi_j$. More thorough investigations by Walsh (1952, 1956) give precision to this statement.

There is, however, something forced about this first generalization. It is implicit in the sampling theory leading to the Poisson Binomial that every sample of E contains precisely one individual with a survivorship probability of π_1, precisely one with a survivorship probability of π_2, and so on up to π_E.‡ This is not a realistic model of sampling E-tuples of individuals from a large group (universe) of apparently homogeneous, similarly circumstanced individuals. Recognizing that the probability of survivorship may vary quite widely in such a group it would be reasonable to write $f(\pi)\,d\pi$ for the probability that an individual chosen at random has a survivorship probability between π and $\pi+d\pi$. Although we have thus supposed that the distribution $f(\pi)$ is continuous this is not necessary to the discussion. However, it expresses mathematically the familiar statement that every man has his own mortality rate which is different from that of any other man in history! Finally let

$$\int_0^1 \pi f(\pi)\,d\pi=\bar{\pi}.$$

* In accordance with statistical usage we employ Greek letters for the universal parameters and Roman script for sampling values. Thus, the familiar actuarial notation ${}_tp_x$ is reserved for observational values of this function and their parametric counterparts are written ${}_t\pi_x$. There is one extension of this rule: the Greek letter θ is used for the expected number of deaths and the sampling values are written as y—the degenerate English form of the Old English runic letter þ, thorn.

† No confusion should exist between this π_j and π_x the probability of surviving from age x to $x+1$.

‡ Not all the π_j-values need to be disparate.

Suppose, now, that we repeatedly draw samples of E individuals from such a universe. The chance of the E-tuple of π-values specified by $\pi_j, j = 1, 2, \ldots, E$, is

$$\prod_{j=1}^{E} f(\pi_j)\, d\pi_j$$

and the probability of y deaths is then given by the coefficient of z^y in

$$\prod_{j=1}^{E} (\pi_j + \overline{1 - \pi_j}\,.\,z).$$

Hence, the probability of y deaths in general is given by the coefficient of z^y in

$$\int_0^1 d\pi_1 \int_0^1 d\pi_2 \ldots \int_0^1 d\pi_E \prod_{j=1}^{E} f(\pi_j)(\pi_j + \overline{1 - \pi_j}\,.\,z)$$

$$= \prod_{j=1}^{E} \int_0^1 f(\pi_j)(\pi_j + \overline{1 - \pi_j}\,.\,z)\, d\pi_j = \prod_{j=1}^{E} (\bar{\pi} + \overline{1 - \bar{\pi}}\,.\,z) = (\bar{\pi} + \overline{1 - \bar{\pi}}\,.\,z)^E. \quad (3)$$

The second stage of generalization of a naïve equi-probability of death for all individuals aged x has thus led us back to the Bernoulli Binomial distribution with $\pi = \bar{\pi} = \int_0^1 \pi f(\pi)\, d\pi$. This result is also due to Poisson (1837, pp. 65–6) and has been much discussed by writers on Lexis theory (for example, Bortkiewicz, 1917). As a corollary we may note that $f(\pi)$ may be different in each sample provided only that the mean π-value, namely $\bar{\pi}$, remains unchanged throughout the sampling. Furthermore, the procedure of generalization thus indicated allows us to extend the Binomial of Bernoulli involving a parameter π_x, the probability of surviving from age x to age $x+1$ based on a supposedly constant force of mortality throughout the year of age, to a new Bernoulli Binomial involving a parameter ${}_t\pi_x$, the probability of surviving from age x to age $x+t$ based on a variable force of mortality $\mu_{x+\tau}$, $0 \leqslant \tau < t$. This latter extension has been demonstrated directly—and unnecessarily—by recent writers (Seal, 1948; Nolfi, 1951).

Our second generalization must, however, be extended in two directions before it can be applied to the mortality observations commonly published by life offices individually or collectively. In the first place we notice that unless the observational period commences and ends when all the lives have attained an (assumed) exact integral age—as is the case with observational periods that open and close on policy anniversaries—a small proportion of the lives included at age x are not exposed from x to $x+1$, but from $x+a_j$ to $x+b_j$, say, $j = 1, 2, 3, \ldots, E$, where most of the $a_j = 0$ and most of the $b_j = 1$. This means that the product of E identical binomial terms of the type $\pi + \overline{1 - \pi}\,.\,z$ has to be replaced by

$$\prod_{j=1}^{E} \left({}_{b_j - a_j}\pi_{x+a_j} + \overline{1 - {}_{b_j - a_j}\pi_{x+a_j}}\,.\,z\right) \quad (4)$$

and the probability of y deaths is given by the coefficient of z^y in the expanded form of this product. The expected number of deaths on the basis of this (Poisson Binomial) distribution is

$$\sum_{j=1}^{E} \left(1 - {}_{b_j - a_j}\pi_{x+a_j}\right)$$

and if, following Cantelli (1914), we assume that

$$1 - {}_{b-a}\pi_{x+a} = (b-a)(1-\pi_x), \tag{5}$$

this expectation is equal to

$$(1-\pi_x)\sum_{j=1}^{E}(b_j-a_j),$$

namely $1-\pi_x$ times the exposed-to-risk at age x. The replacement of (4) by the binomial

$$(\pi_x + \overline{1-\pi_x}\,.\,z)^{\sum_{j=1}^{E}(b_j-a_j)}$$

thus leaves the mean unaltered but changes the variance from

$$\sum_{j=1}^{E} {}_{b_j-a_j}\pi_{x+a_j}(1-{}_{b_j-a_j}\pi_{x+a_j}) = (1-\pi_x)\sum_{j=1}^{E}(b_j-a_j)\{1-(b_j-a_j)(1-\pi_x)\}$$

to

$$(1-\pi_x)\,\pi_x\sum_{j=1}^{E}(b_j-a_j).$$

The latter is always less than the former unless all the $b_j - a_j = 1$ and the reduction is, in fact, equal to

$$(1-\pi_x)^2\sum_{j=1}^{E}(1-\overline{b_j-a_j})(b_j-a_j)$$

which, since $1-\pi_x$ is nearly always small, represents a relatively small error.

Broadly, then, we may conclude that the common use of $\sum_{j=1}^{E}(b_j-a_j)$ as the exposed-to-risk in the case where a few of the lives are exposed for only part of the year of age, only slightly distorts the accuracy of the Binomial of Bernoulli when it is assumed to represent the probability distribution of deaths.

Finally, we consider the situation common in assurance, as opposed to annuity, experiences where withdrawals as well as deaths are occurring among the exposed-to-risk. If all the E lives are observed from precise age x until the attainment of age $x+1$, or prior death or severance, it can be shown (Seal, 1948) that the Binomial form (1) is still valid provided $1-\pi$ is replaced by

$$\int_0^1 \exp\left[-\int_0^t (\mu_{x+\tau}+\nu_{x+\tau})\,d\tau\right]\mu_{x+t}\,dt,$$

where μ and ν are the forces of mortality and of withdrawal, respectively. However, the q_x of most assurance experiences is not an estimate of this probability—known in the English literature as the 'probability of death'—but of the so-called 'rate of mortality', $\exp\left[-\int_0^t \mu_{x+\tau}\,d\tau\right]$, which is stochastically independent of the force of withdrawal. The rate of mortality q_x differs from the unbiased estimate of the 'probability of death' by utilizing a *smaller* denominator which, because it involves the actual withdrawals, cannot be regarded as fixed in repeated sampling. The sampling distribution of q_x is thus no longer Binomial but is unlikely to differ from it by much. And if the probability distribution of y is being considered, the use of this reduced 'exposed-to-risk' will leave θ unchanged but will reduce the variance since the expected value of $1-q_x$ will be 'too small'.

We may thus summarize our conclusions as follows. The probability distribution of deaths at age x is quite generally Binomial. In particular the Binomial applies when each life's probability of survival is drawn from a hypothetical universe with mean $\bar{\pi}$. And when the lives are subject to observational 'cut off' during the year of age or may withdraw of their own free will the Binomial still appears reasonably apt except that the variance is then *slightly reduced.*

STATISTICAL TECHNIQUES USED

Assuming, then, that all the observed values of $q_{[x-t]+t}$ tabulated in our suitably chosen mortality experiences are sample values from one or another Binomial distribution, the statistical techniques to be used in looking for evidences of temporary selection are all standard.

Consider, first, the test of the hypothesis that a series of relative frequencies based on different numbers of experimental 'repetitions'—our 'exposed-to-risk'—are all sample values of one and the same unknown parametric probability value. The procedure is described in Ch. IV of Fisher (1954) and it is to be noted that arbitrary groups of these frequencies may be amalgamated and tested against one another (*vide* Ex. 15·1, *l.c.*).

Specifically, it is known that the variate

$$X^2 = \sum_{t=3}^{t_3} \frac{(y_{[x-t]+t} - E_{[x-t]+t}\,q_x)^2}{E_{[x-t]+t}\,q_x(1-q_x)}$$

$$\equiv \{q_x(1-q_x)\}^{-1}\left[\sum_{t=3}^{t_3} y_{[x-t]+t}\,q_{[x-t]+t} - y_x q_x\right],$$

where $y_x = \sum_{t=3}^{t_3} y_{[x-t]+t}$ with similar notations for E and q, is distributed approximately as χ^2 with $t_3 - 3$ degrees of freedom. The fit of the χ^2 distribution is, in general, better the larger the values of E, but a number of writers have shown that the approximation is still good in circumstances differing considerably from the optimum (Cochran, 1954). In what follows we require that $E_{[x-t]+t}q_x$ should be at least one unit and this is believed to be adequate.

Now suppose X^2 were calculated at a given age x for the value of t_3 that emerged from the 'unit rule' just stated and were found not significantly different from its expectation $t_3 - 3$. It might then be concluded that the hypothesis of equi-mean-probability of death at that value of x and for every available value of t had not been disproved. However, it is conceivable that all the negative deviations $y_{[x-t]+t} - E_{[x-t]+t}q_x$ were found at the lower values of t and all the positive deviations at the higher. In fact this situation is exactly what the protagonist of extended temporary selection is expecting and the non-significance of the value of X^2 would be put down to coincidence.

In order, therefore, to guard against this possibility we have divided the range of available t-values into *four* parts and have ignored the first of them in all the subsequent work. These four ranges of t-values are: 0–2, 3–4, 5–9 and 10 and up, respectively. At durations 0, 1 and 2 it was originally thought that temporary selection was so obviously existent that any statistical tests of it would be wasted effort. Since the classical view would be that the initial selective forces would then be most strongly evident at $t=3$ and $t=4$ this was considered a suitable low-mortality group of durations with which to compare the higher durations. Furthermore, durations 5–9 appeared to constitute a useful 'middle ground' on which opinion might be divided between the exponents of 'short' and 'long' enduring temporary selection. The final group, $t \geqslant 10$, was thought to constitute the most reasonable 'ultimate' part of many mortality experiences.

The computational scheme may thus be summarized as follows: Write $y_x^{(k)} = \sum_{t=t_{k-1}+1}^{t_k} y_{[x-t]+t}$ with similar notations for $E_x^{(k)}$ and $q_x^{(k)}$, where $k = 1, 2, 3$ and $t_0 = 2$, $t_1 = 4$, $t_2 = 9$ and t_3 is the largest available t-value at that age x. Then if the individual values of $q_{[x-t]+t}$, $t = 3, 4, 5, \ldots, t_3$, are sampling values of a single universal probability value the statistical criteria

$$X^2 = \sum_{k=1}^{3} \frac{(y_x^{(k)} - E_x^{(k)} q_x)^2}{E_x^{(k)} q_x(1 - q_x)}$$

$$\equiv \{q_x(1 - q_x)\}^{-1} \left[\sum_{k=1}^{3} y_x^{(k)} q_x^{(k)} - y_x q_x \right]$$

and $$X_k^2 = \{q_x(1 - q_x)\}^{-1} \left[\sum_{t=t_{k-1}+1}^{t_k} y_{[x-t]+t}\, q_{[x-t]+t} - y_x^{(k)} q_x^{(k)} \right]$$

are distributed approximately as χ^2 with 2 and $t_k - t_{k-1} - 1$ degrees of freedom, respectively. It follows that the sum

$$X_0^2 + \sum_{k=1}^{3} X_k^2 \equiv X^2$$

is distributed approximately as χ^2 with $2 + \sum_{k=1}^{3} (t_k - t_{k-1} - 1) = t_3 - 3$ degrees of freedom, as already stated.

In order to obviate any criticism that the computations included only favourable x and t values, a rigid rule was used in determining the inclusion or exclusion of duration t at any age x. Starting with duration 3 all *consecutive* E-values (and the corresponding deaths) were included so long as $E \geqslant q_x^{-1}$. If $E_{[x-3]+3} < q_x^{-1}$ the procedure was started at $t = 4$; and so on.* The result was a series of pairs of numbers $E_{[x-t]+t}$, $y_{[x-t]+t}$ at every individual age x and for consecutive values of t exceeding 2 and running up to, in some cases, 30 or 40.

However, there was one difficulty in applying this rule: The observed value of q_x could not be calculated until it was decided which durations t to include. This was resolved by choosing for q_x some previously published rate based on the observations being studied. In most cases this q_x was the ungraduated aggregate value but some sets of ungraduated five-year ultimate rates were used.† However, Karup's (1903) graduated seven-year ultimate rates were used in determining the critical E-values for the Gotha data and the classic 'slightly graduated' values of q_x designated as 'Kersseboom' and 'Deparcieux' were adopted for the corresponding data. Finally, the graduated quinquennial values of q_x reproduced on p. 164 of *Skand. Akt.* **2** (1919) were used for the Russian data.

* The only exception to this rule was where $E_{[x-3]+3} \geqslant q_x^{-1}$, $E_{[x-4]+4} < q_x^{-1}$, but $E_{[x-t]+t} \geqslant q_x^{-1}$ for a series of consecutive $t \geqslant 5$. This case occurred very infrequently.

† Namely, for the data designated in Table 1 as (14), (15), (16), (17), (l), (m) (these two latter being quinquennially centred values) (λ) and (μ).

The consequences of this subdivision of the range of t-values are that:

(1) X_1^2 tests whether $q_{[x-3]+3}$ and $q_{[x-4]+4}$ can be regarded as sample values from one and the same mean-probability of death;

(2) X_2^2 tests whether $q_{[x-5]+5}, q_{[x-6]+6}, \ldots, q_{[x-9]+9}$ are sample values based on their own mean-probability of death—which may be different from that of (1);

(3) X_3^2 makes a similar test for the observed values of $q_{[x-10]+10}, q_{[x-11]+11}, \ldots$; and

(4) X_0^2 tests the over-all group-values $q_x^{(1)}$, $q_x^{(2)}$ and $q_x^{(3)}$ for differences between the mean-probabilities of death hypothetically fixed for (1), (2) and (3).

Note that if the value of X_0^2 is significantly large the factor $\{q_x(1-q_x)\}^{-1}$ in X_k^2 must be changed to $\{q_x^{(k)}(1-q_x^{(k)})\}^{-1}$, since q_x is no longer a valid estimate of the probability of death in the group of durations designated by k.

It is mentioned, in passing, that the minor discrepancies from true Binomiality mentioned in the preceding section, owing either to 'cut-off' periods in the year of age or to the existence of withdrawals, both tend to *increase* the X^2-values we are calculating. They thus tend to promote significant differences between the q's where none exist.

The result of all the calculations described above is a set of up to four values of X^2 at each age x.* Now opinions differ whether the duration of temporary selection varies from age to age, some writers holding that it is longest at the young ages (Higham, 1851) and others arguing that medical selection is least efficacious at those ages (Macaulay, 1894; Abel, 1914). However, judging from the prevalence of mortality tables with uniform select periods at every age the weight of opinion has been against a period of temporary selection varying with age. In our investigations we have conformed to this view by obtaining an over-all picture of the existence of temporary selection in any given experience by adding the values of X_0^2, the values of X_1^2, the values of X_2^2, and the values of X_3^2, respectively, over all the ages included in the computations. As is well known each of these four results is distributed as χ^2 with a number of degrees of freedom obtained by summing all the individual numbers of degrees of freedom attached to the X^2-values involved. Table 1 at the end of this paper summarizes the results thus obtained but it is necessary to make some further general observations before they are reviewed in detail.

It will be noticed that by squaring a 'deviation' of the type $y_{[x-t]+t} - E_{[x-t]+t}q_x^{(k)}$ we are giving equal weight to positive and negative results. It may be objected that such a procedure is invalid since we 'know' that the trend in successive $q_{[x-t]+t}$ values is unlikely to be downward. In answer to this we observe that this supposed knowledge is derived from earlier analyses of precisely the same data that we are now investigating and it is, of course, an elementary error in statistical analysis to observe a 'trend' and then to 'test' it.

However, the writer is prepared to admit on *a priori* grounds that the obvious 'selection' at $t=0$ is likely to continue through $t=1$ and $t=2$ and that it may still be apparent at $t=3$ and $t=4$. Furthermore, if withdrawals are selective from a mortality standpoint, it may be conceded that their effect is becoming evident by $t=3$ in the case of assurance experiences. For these reasons the

* The presence of one instead of four values at a particular age is illustrated by the case where the criterion $E \geqslant q_x^{-1}$ resulted in the inclusion of only two durations $t=3$ and $t=4$.

judgment of the significance of $\sum_x X_1^2$ is replaced by that of the significance of $\sum_x X_1$, where the sign of X_1 at any age is that of $q_{[x-4]+4} - q_{[x-3]+3}$ and where we believe that positive deviations are more probable than are negative values. Since the variate X_1 is approximately Normal with zero mean* and unit variance, the sum $\sum_x X_1$ is essentially Normal with mean zero and variance m, where m is the number of ages involved. This means that the 5% limen of $\sum_x X_1/\sqrt{m}$ is 1·6449 and the occurrence of a value between this and 2·3263 is thus indicated by *s* in the appropriate column of Table 1.

'SELECTIONS' NOT 'LIVES' THE OBSERVATIONAL UNIT

From the standpoint of the Institute's members Tables 1A and 1B suffer from the fact that the most recent British data included therein are Finlaison's (1884) government annuitants. This is, of course, because no other data published in Britain since that date have excluded duplicates.

It is thus interesting to inquire what would be the effect of applying the X_j^2 ($j=0, 1, 2, 3$) and X_1 tests to experiences based on the 'selection' instead of the 'life' as the observational unit. In such data individuals effecting more than one policy at any given age are counted only as one exposure, usually for the period of the policy longest in force. However, if the same individual insures himself one, two, three,... years later he is counted again as a single individual at a later entry age and is again followed through until his death or earlier removal from observation. As a result any value of $q_{[x-t]+t}$ will be based on individual *lives* each counted once but $q_{[x-t_1]+t_1}$ and $q_{[x-t_2]+t_2}$, $t_1 \neq t_2$, may contain duplicate exposures of one and the same individual. This method of presenting data was adopted in the 60 British Offices', 1863–93, experience and was also used as an alternative in the Austrian, Hungarian, Swedish and Russian companies' data analysed in Table 1A. The Japanese offices also used it in their first mortality experience.

Without going into details here it can be shown that the effect of such a duplication of exposure in $q_{[x-3]+3}$ and $q_{[x-4]+4}$ is to cause X_1^2 to be distributed, not as χ^2, but as $\left(1 - \dfrac{2d}{y_{[x-3]+3} + y_{[x-4]+4}}\right)\chi^2$, where d is the number of deaths included both in $y_{[x-3]+3}$ and in $y_{[x-4]+4}$. If the proportion of lives thus duplicated is as high as 10% the expected value of X_1^2 is not unity but about ·9. Similar remarks hold for the comparisons within the group of durations $t=5$–9 and within the group $t=10$ and over. It would seem that, on the whole, the expected values of $\sum_x X_j^2$ ($j=1, 2, 3$) would be overestimated by some 5–10% if the χ^2 distribution were used as a significance test as in Table 1.

The situation is rather different when we consider X_0^2. Here the three component $q_x^{(k)}$-values are themselves based on duplicated individuals and their variances are thus increased (Seal, 1943). On the other hand, the tendency discussed in the preceding paragraph still holds for the sizes of the differences

* In general $E_{[x-3]+3}$ and $E_{[x-4]+4}$ are of the same order. If the former is substantially larger than the latter X_1 is more likely to be positive than negative. However X_1^2 is still closely approximated by a χ^2 distribution since the excess of positive X_1's is approximately balanced by the deficiency in negative X_1's.

occurring between the $q_x^{(k)}$ $(k = 1, 2, 3)$. These two opposing effects tend to cancel one another out although a precise result is difficult to formulate.

It was therefore with some diffidence that we turned to the British Offices', 1863–93, experience and made the analysis summarized in Table 1 C. It will be seen that the results* were supplemented by those based on the Japanese Offices' data.

THE CHANCE OF DETECTING MORTALITY DIFFERENCES

The X_j^2 $(j = 0, 1, 2, 3)$ and X_1 tests described above are known to be 'good' tests of the respective hypotheses and, in fact, they are close to the 'best' possible in the detection of differences in the q-values where they exist (see, for example, Hoel, 1945). Let us, therefore, consider the chances that these tests—or, more specifically, two of them—will detect differences of a given size.

Consider first X_1. Assume, for convenience, that $E_{[x-3]+3} \approx E_{[x-4]+4}$ and suppose that $1 - \pi_{[x-3]+3} = \kappa(1 - \pi_{[x-4]+4})$, where $\kappa \leqslant 1$, the equality sign occurring when the null hypothesis is true. If, then, $\pi_{[x-4]+4}$ is close to unity it may be shown (Patnaik, 1948) that X_1 is approximately Normal with unit variance about a mean $\frac{1}{2} \ln \kappa^{-1}\{\mathscr{E}y^{(1)}\}^{\frac{1}{2}}$. (This mean is zero, as it should be, when $\kappa = 1$.) Hence $\sum_x X_1$ is very nearly Normal with a mean of $\frac{1}{2} \ln \kappa^{-1} \sum_x \{\mathscr{E}y^{(1)}\}^{\frac{1}{2}}$ and a variance of m.

The probability that $\sum_x X_1/\sqrt{m}$ exceeds 1·6449 or, in other words, the probability of one or other of the letters *s*, *s.s* or *s.s.s* in the appropriate column of Table 1, is thus

$$Q(\xi) = \int_{\xi}^{\infty} e^{-\frac{1}{2}x^2}\, dx/\sqrt{(2\pi)}, \text{ where } \xi = 1{\cdot}6449 - \tfrac{1}{2} \ln \kappa^{-1} \sum_x \{\mathscr{E}y^{(1)}\}^{\frac{1}{2}}/\sqrt{m}$$

When $\kappa = 1$ this probability is ·05 independently of $y^{(1)}$ and m, but in other cases some approximation must be made to the values of $\{\mathscr{E}y^{(1)}\}^{\frac{1}{2}}$ involved.

The following table contains values of $Q(\xi)$ for three κ values and for various aggregate numbers of deaths where the square roots of each of these numbers have been equated, in turn, to $\sqrt{m}$ times the average value of $\{\mathscr{E}y^{(1)}\}^{\frac{1}{2}}$ supposedly encountered in the analysis of one or more experiences with approximately the same number of deaths at age x.

Table C. Values of $Q(\xi)$

Number of deaths	κ		
	·9	·95	·99
100	·132	·083	·055
500	·320	·142	·063
1,000	·508	·202	·069
2,000	·761	·309	·078
5,000	·981	·567	·099
10,000	·9999	·821	·127
30,000	$1 - {\cdot}7 \times 10^{-13}$	·997	·219

* The following four sets of data were omitted from the calculations: Females, Whole Life (Par and Non Par, New and Old).

It appears that if there are as many as 2000 deaths recorded at $t=3$ and 4 (as there were in the $H^{[M]}$ data) there are three chances in four of detecting an upward trend in q of 10 %, but it is more than two to one against detecting an upward difference of 5 %. On the other hand, with as few deaths as 1000 at those durations, there is an even chance against detecting such a 10 % upward trend, it is four to one against detecting a 5 % upward trend and fourteen to one against detecting a 1 % upward trend. Notice that even with 30,000 deaths it is still nearly four to one against detecting a 1 % upward trend in q between $t=3$ and 4. These figures must give considerable food for thought to those who believe that the detection of the continuance of temporary selection is a simple matter of graphical comparison.

Let us now extend this argument to determine the power of $\sum_x X_0^2$. We assume that $1-\pi_x^{(1)}=\kappa(1-\pi_x^{(2)})=\kappa(1-\pi_x^{(3)})$ and, for convenience,

$$E_x^{(1)}=\tfrac{1}{8}(E_x^{(2)}+E_x^{(3)}).$$

Here again the null hypothesis is based on $\kappa=1$ but the permissible alternatives are no longer restricted to $\kappa<1$; in fact $\kappa \lesseqgtr 1$.

Now X_0^2 is a χ^2-variate with two degrees of freedom. It is known that there is an unlimited number of ways of splitting X_0^2 into two component parts, namely $X_0^2=X_{01}^2+X_{02}^2$, where each of these parts is distributed independently as χ^2 with 1 degree of freedom when κ is unity. For our purpose it is desirable that X_{01}^2 should be unaffected by changes in the value of κ with the result that the effects of such changes are all reflected by variations in the value of X_{02}^2. This objective may be achieved by computing X_{01}^2 as the independence χ^2-variate for the 2×2 table

$y^{(2)}$	$E^{(2)}-y^{(2)}$	$E^{(2)}$
$y^{(3)}$	$E^{(3)}-y^{(3)}$	$E^{(3)}$
$y^{(2)}+y^{(3)}$	$E^{(2)}+E^{(3)}-y^{(2)}-y^{(3)}$	$E^{(2)}+E^{(3)}$

since X_{01}^2 then tests the numerical equivalence of $1-\pi^{(2)}$ and $1-\pi^{(3)}$ and this is assumed to hold true. The balance, X_{02}^2, is then obtainable as the independence χ^2-variate for the 2×2 table

$y^{(2)}+y^{(3)}$	$E^{(2)}+E^{(3)}-y^{(2)}-y^{(3)}$	$E^{(2)}+E^{(3)}$
$y^{(1)}$	$E^{(1)}-y^{(1)}$	$E^{(1)}$
y	$E-y$	E

(see, for example, Kimball, 1954) and will have a non-central χ^2 distribution so long as κ is not unity.

Analogously with X_1, X_{02} is approximately Normal with unit variance and mean $\sqrt{8/9}\ln\kappa^{-1}\{\mathscr{E}y\}^{\frac{1}{2}}$. Thus, finally, $\sum_x X_0^2=\sum_x X_{01}^2+\sum_x X_{02}^2$ is distributed as non-central χ^2 with $2m$ degrees of freedom and non-centrality parameter $\lambda=\frac{8}{81}(\ln\kappa)^2\sum_x \mathscr{E}y$.

The following Table D therefore shows the probabilities of exceeding the 5 % point of χ^2 with $2m$ degrees of freedom for various specified values of $2m$

and $\sum_x \mathscr{E}y$—the aggregate expected number of deaths. This table was computed using Abdel-Aty's (1954) approximation, namely, that the cube-root of non-central χ^2 is distributed Normally about a mean

$$(2m+\lambda)^{\frac{1}{3}}\left\{1-\frac{2}{9}\left(1+\frac{\lambda}{2m+\lambda}\right)\Big/(2m+\lambda)\right\}$$

with standard deviation

$$(2m+\lambda)^{\frac{1}{3}}\left\{\frac{2}{9}\left(1+\frac{\lambda}{2m+\lambda}\right)\Big/(2m+\lambda)\right\}^{\frac{1}{2}}.$$

Table D

$2m$	Aggregate number of deaths	κ		
		·9	·95	·99
90	5,000	·111	·062	·050
	20,000	·451	·107	·052
	50,000	·953	·244	·054
	100,000	1·000	·549	·059
600	75,000	·727	·141	·052
900	220,000	1·000	·375	·056
2,000	250,000	·992	·268	·054

Table D indicates that even with the enormous numbers of deaths assumed in the last lines of the table there is little chance of detecting a change of 1 % between $q_x^{(1)}$ and $q_x^{(2)}$, and that even a difference of 5 % could easily remain undetected.

After these preliminary theoretical and methodological considerations we may now turn to the actual mortality data and their analysis for continuing temporary selection.

EQUITABLE SOCIETY 1762–1828, HEALTHY MALES

As Buchanan (1927) observes, Morgan (1834) was the first to give numerical expression to the view that temporary selection was an important factor in assurance and annuity mortality. In the introduction to the mortality experience of the Equitable, 1762–1828, Morgan provided a table of $q_{[\bar{x}-t]+t}^{-1}$ for the individual entry ages 30(5)50; 33(5)48 and t the average of the quinquennial groups of durations 0–4, 5–9, 10–14, ... Although the effect of temporary selection was supposed to be clear from this table, if we ignore the eight values for $t=0$–4, only 15 of the 22 differences $\Delta_t q_{[\bar{x}-t]+t}^{-1}$ are negative and a further one is zero (to the number of significant figures shown). If it were pure chance whether $\Delta_t q_{[\bar{x}-t]+t}^{-1}$ should be positive or negative the probability of 15 or more negatives is ·0669 and the probability of 16 or more is ·0262.* Since at that time there could not have been valid theoretical grounds

* It is interesting to note that the supposedly extended period of selection in the $O^{[M]}$ (see Table 2 *post*) was based on Hardy's (1903) Table XVI showing values of the ungraduated $e_{[x-t]+t}$, $t=0(5)25$, with x the centre of quinquennial age groups extending from 20–24 to 85–89. In this Table the number of negative differences $\Delta_t e_{[x-t]+t}$, $t\geqslant 5$, is 22 whilst there is a single zero difference. There are 36 such differences in all. The chance of 22 or more negatives in a chance series of 36 'plus' and 'minus' signs is ·1215; the chance of 23 or more is ·0662.

for believing that the period of temporary selection exceeded 5 years neither of these results is significant at the 5 % level ($2\frac{1}{2}$ % in each tail of the binomial).

These considerations made it all the more interesting to review Morgan's (1834) data which were published in a suitable form for analysis. Unfortunately, these observations do not behave like random samples from universes with possibly differing mortality rates.

For example, using Chatham's (1891) method of calculating the exposed-to-risk,* the aggregate value of χ^2 for ages 26–81 and $t \geqslant 2\frac{1}{2}$ was 515·846 with 1157 degrees of freedom. The chance of a value as low as, or lower than, this on the (generalized) Binomial hypothesis is $\cdot 6 \times 10^{-57}$, or absolute impossibility (*vide* Borel, 1939). Expressed another way we may say that the mean variance at any age is of the form $\frac{4}{9}Epq$ instead of Epq. This suggested that some form of graduation of the deaths was adopted by Morgan before the data were published, but M. E. Ogborn has informed me that the manuscript book of deaths kept by the Equitable's then actuary definitely precludes this possibility.

Whatever the real reason for the strange behaviour of Morgan's figures, the q-values derived therefrom cannot be regarded as sample probabilities or treated as such. It is thus with regret that this important early set of data was omitted from the experiences analysed in Table 1.

DISCUSSION OF TABLE 1

A natural way of assimilating the results summarized in Table 1 is to concentrate first on the group of columns comparing the mortality rates at $t=3$ and $t=4$. The totals of the three subdivisions A, B and C of the Table may be interpreted as shown in the following table. Although on the small side none

Table E

Experiences	Number of deaths	Value of $\sum_x X_1$	Variance of $\sum_x X_1$	Approximate probability of a larger value of $\sum_x X_1$	
				On null hypothesis	On assumption that q at $t=3$ is 95 % of q at $t=4$
25 assurance	32,820	28·35	1055	·191	·9993
13 annuity	9,428	15·55	346	·202	·876
12 'selections'*	15,087	32·47	415	·055	·859

* Only 11 of these have X_1 values.

of the probabilities of the penultimate column gives grounds for abandoning the null hypothesis of equality of the universal values of $q_{[x-3]+3}$ and $q_{[x-4]+4}$ at all ages x. On the other hand, the last column indicates that the view that the former is only 95 % of the latter is barely tenable. This means that, in the experiences based on 'lives', if the effects of temporary selection still exist

* Elderton & Ogborn (1943) believe that the 'existings' should be added back into each of the exposures thus obtained. This would not affect the general magnitude of the χ^2-value quoted here.

between $t=3$ and $t=4$ they must affect the mortality rates on the average by less than 5%. It will be remembered that Table C *ante* indicates that we would have been lucky to discover a differential as small as 1% from the data of Table 1. We may note that Table E implies that there is no great difference between assurance and annuity experiences so far as temporary selection is concerned.

When we turn to the 49 individual values of $\sum_x X_1$ that were aggregated in Table E we find that 27 are positive and 22 are negative, an unsurprising result on the basis of the null hypothesis of no difference between the mortalities at $t=3$ and $t=4$. However, 6 of the 27 positive values of $\sum_x X_1$ are significant at the 5% or a lower limit instead of the 2·45 expected on the null hypothesis. This suggests that there may have been real upward trends in $q_{[x-t]+t}$ in about four of the experiences.* Although there are obvious statistical dangers in segregating a set of exceptional experiences it is of interest to consider the frequency distribution of the aggregate of the 284 individual X_1 values in the six experiences that were characterized as significant in Table 1. Table F provides the result. The observed frequencies when compared with those of the null hypothesis—a Normal distribution with zero mean and unit standard deviation, i.e. $N(0, 1)$—show a distinct shift to the positive side. However, when a Normal distribution is 'fitted' using the actual mean (obtained by adding the six values of $\sum_x X_1$ and dividing by 284) the observed and expected frequencies are in reasonable agreement.

Table F

Values of X_1	Observed frequencies	Expected on null hypothesis $N(0, 1)$	Expected on hypothesis $N(\cdot 3117, 1)$
$-\infty$ –(-1·960)	6	7·1	3·28
-1·960–(-1·645)	2	7·1	3·87
-1·645–(-1·282)	9	14·2	8·61
-1·282–($-$ ·524)	30	56·8	41·49
$-$ ·524–($-$ ·000)	47	56·8	49·97
·000– ·524	71	56·8	58·63
·524– 1·282	71	56·8	71·01
1·282– 1·645	23	14·2	21·23
1·645– 1·960	12	7·1	11·81
1·960– ∞	13	7·1	14·11
	284	284·0 $\chi^2=40{\cdot}912,\ \nu=9$	284·01 $\chi^2=9{\cdot}385,\ \nu=8$

* It may seem odd that 4 out of 6 significant results are from English and/or Scottish experiences. However, since 17 of the 49 experiences available at $t=3$ and 4 are English and/or Scottish the chance of 4 or more significant English and/or Scottish among 6 is:

$$\sum_{j=4}^{6} \binom{17}{j}\binom{32}{6-j} \Big/ \binom{49}{6} = {\cdot}0995.$$

It cannot be claimed that temporary selection continuing after $t=3$ is a national characteristic!

Now the total of the 284 values of $\sqrt{y_x^{(1)}}$ involved in the calculations of $\sum_x X_1$ for the experiences was 1,256. Using the arguments leading to Table C *ante* we may say that the average difference between $q_{[x-3]+3}$ and $q_{[x-4]+4}$ in some at least of these experiences is $1-\kappa$ times $q_{[x-4]+4}$, where κ is determined from

$$284 \times \cdot 3117 \approx \tfrac{1}{2} \ln \kappa^{-1} \, 1256 = 628 \ln \kappa^{-1}$$

i.e.

$$\kappa \approx \cdot 868.$$

The differential in these experiences thus appears to be of the order of 13 %, or much more than that usually assumed in a typical select mortality experience (see Table A *ante*).

We have thus concluded that in at least 43 of the 50 experiences temporary selection between $t=3$ and $t=5$ could only have produced differences in $q_{[x-t]+t}$ of the order of 1 % to 2 % if it had not in fact vanished by $t=3$. The existence of differences between the $q_{[x-t]+t}$ in the duration groups $t=5$–9 and $t=10$ and over would thus have to be explained by secular changes rather than by the continuance of the effects of initial selection. However, reference to the three totals for $\sum_x X_2^2$ and $\sum_x X_3^2$ in Table 1 is reassuring: none of these totals is significant, four of the six values being less than their expectations (namely, the corresponding degrees of freedom), and there are eight significant individual values of the criterion instead of the five expected.

Finally, we may turn to the results of the last set of columns of Table 1. The significant totals in two of the three sections of the table at first sight point to real differences in the values of $q_x^{(1)}$, $q_x^{(2)}$ and $q_x^{(3)}$. However, these large aggregate values of $\sum_x X_0^2$ are based in each case on a set of individual items which are undistinguished except for the presence of a few particularly large values. In fact if the nine significant values of $\sum_x X_0^2$ are replaced by their expectations the three aggregate results are all very close to the aggregate degrees of freedom corresponding thereto. Note, too, that only two of the six experiences in which $\sum_x X_1$ was significant at durations 3 and 4 now appear among the nine significant values of $\sum_x X_0^2$.* And the size of the differentials in $q_{[x-3\frac{1}{2}]+3\frac{1}{2}}$ and $q_{[x-7]+7}$ implied by these highly significant values of $\sum_x X_0^2$ (cf. Table D *ante*) suggests that some other explanation is to be sought than the continuance of temporary selection.

It has already been mentioned that the presence of duplicate policies might tend to increase the expected value of X_0^2. This would occur if the duplicates were relatively numerous, and reference to Table 1 C indicates that this may be an explanation of the highly significant values of $\sum_x X_0^2$ in three of the four experiences of Whole Life assurance mortality among British offices. The proportion of duplicates may be assumed to have been much less in the other experiences included in that portion of Table 1.

But this possibility helps to explain some of the other six significant values of $\sum_x X_0^2$ in Table 1 A and B. The larger the experience and the more offices contributing the harder it must have been to eliminate every one of the duplicate

* That these two should have been the $H^{[M]}$ and the $O^{[NM]}$, respectively, is an unfortunate coincidence for British actuaries.

policies. This would have been particularly true of policies effected many years ago and of those on the lives of women. It is perhaps no coincidence that the two significant values of $\sum_x X_0^2$ in Table 1 B both derive from large experiences of female lives.

It will be noticed that if the unremoved duplicate observations occur at the higher values of t (as we conjecture is the case) this will result in exceptionally large q-values at durations 10 and over whenever a duplicate set of policies becomes a claim. These large q-values would lead to improbably high values of X_0^2 at the ages at which these claims occur. In order to see whether this phenomenon occurred in the nine experiences with significant values of $\sum_x X_0^2$ in Table 1 we have enumerated in Table G the values of X_0^2 in excess of 4·605 (the 10% point of χ^2 with 2 degrees of freedom) in each of those experiences. Although not every frequency of excesses is significant by itself the over-all effect is overwhelmingly significant. We may therefore tentatively conclude in favour of our 'duplicate' explanation.

Table G

Experience	No. of values of X_0^2 with 2 D.F.	No. of values X_0^2 ($\nu=2$) in excess of 4·605	Probability that at least the foregoing number would have occurred by chance
Amicable	15	4	·0556
H^M	52	8	·1441
Conn. Mutual	39	9	·0131
Gotha	39	8	·0366
Brit. Govt. Females	42	7	·1214
French Females	45	12	·0012
	232	48	$< 10^{-5}$
B.O. Males, W.L., N.P., Old	36	8	·0235
B.O. Males, W.L., N.P., New	49	10	·0215
B.O. Males, W.L., P., New	50	19	10^{-7}
	135	37	$< 10^{-5}$

The foregoing review has been based on consideration of all 50 mortality experiences as a whole. But such an over-all picture was not of course available to the nineteenth-century actuaries who had become convinced of the reality of long-continuing temporary selection before the commencement of the last quarter of that century. This view was based on the analyses of the experiences of the Equitable (Morgan, 1834), the Seventeen Offices (Higham, 1851), the Scottish Amicable (Spens, 1863), and the Healthy male portion of the 10 English and 10 Scottish Offices (Sprague, 1870). Now we have already mentioned that the statistical techniques used at that time tended to exaggerate the effect of temporary selection whilst we are suspicious of the data published by Morgan. Furthermore, the Seventeen Offices' experience was based on policies instead of lives and it would seem that the skewness of the resulting distribution o 1q makes its mean a dubious statistic to use in the search for trends.

This leaves the $H^{[M]}$ experience as the *clef de voûte* on which all subsequent authors have leaned. As we have seen it was not until the turn of the century that another assurance experience was published in which strong selective effects could legitimately have been claimed up to the completion of five years from date of policy issue.* For these reasons we have thought it worth while to make an extended analysis of the $H^{[M]}$ data. The results are summarized in Appendix I and provide no grounds for believing that temporary selection in that experience continued beyond the point of time $t=4\frac{1}{2}$. In other words, a more critical attitude towards apparent trends in grouped mortality data would have prevented the preparation of so many mortality tables with unnecessarily (and incorrectly) long periods of selection (see Table 2).

In concluding this paper I would like to thank all those students and employees of Yale University and of Morss, Seal and Bret, New York, who so generously contributed their time and skills in making the tedious calculations summarized in Table 1. My thanks also go to Dr Stefan Vajda, who kindly read this paper in draft form and made helpful suggestions for its improvement.

APPENDIX I

The $H^{[M]}$ *data*

Two of the 50 lines of figures in Table 1 contain five s's, namely those lines summarizing the statistical tests made on the $H^{[M]}$ and the 'New' portion of the $O^{[NM]}$ data, respectively. Because the commonly accepted theory of temporary selection owed so much to Sprague's (1870) paper which was based solely on an analysis of the $H^{[M]}$ data, it was considered appropriate to make a more thorough investigation of the unusual results shown for those data in Table 1.

Since the comparisons between $q_{[x-3]+3}$ and $q_{[x-4]+4}$ indicated a continuance of temporary selection until at least the point of time $t=5$† it only remained to consider policy durations 5 and over. The values of X_2^2 were thus recomputed by replacing $q_x^{(2)}$ in numerator and denominator by the 'ultimate' (ungraduated) rate of mortality obtained by dividing the sum of the deaths at all the utilized durations $t=5$ and over by the corresponding exposed to risk. Entirely analogous X^2 values were then calculated at each age x for $t=10$–14, $t=15$–19, ..., $t=40$–44, respectively. Finally, an X^2-value analogous to X_0^2 of Table 1 was computed; whereas X_0^2 tests the differences between the three mean mortalities at durations 3 and 4, 5–9 and 10 and over, the new X^2 tests the differences between the mean mortalities at durations 5–9, 10–14, 15–19, ..., and 40–44, respectively. The results are given in Table 3 where it is to be noted that, except in the last column, the degrees of freedom corresponding to an X^2 value are only shown (in parentheses) when they are not equal to 4. Notice that all the values of X^2 at any age x are based on the 'ultimate' after 5 (i.e. $4\frac{1}{2}$) years mortality rate mentioned above.

* Although we found a similar continuing selection in the case of the male British Government annuitants, Finlaison (1884) chose a four-year period of selection for this experience.

† Or, more correctly, $t=4\frac{1}{2}$ because of the 'calendar year' method of observation.

Before drawing any conclusions about the continuance of temporary selection it is interesting to collect together into a frequency distribution the 282 values of X^2 (4 degrees of freedom) found in all the columns except the last. This distribution is shown in Table H, together with the theoretical distribution (χ^2 with 4 degrees of freedom) and the usual χ^2 test for goodness of fit. Although the fit is rather poor this is due entirely to an excess of very small values of X^2 and too few values of X^2 between 3·357 and 4·045. The former may just be 'bad luck', but it would seem that the scarcity of values near the mean could be due to the presence of undetected duplicate lives among the deaths. In the introduction to Brown (1869) it was mentioned that the elimination of duplicates would have been 'greatly facilitated' if the date of birth had been given on the data cards instead of the office age at entry. Reference was also made to the difficulty occasioned by 'noblemen' who sometimes used their (different) titles and sometimes their family names. It cannot be doubted that a proportion of simultaneously effected duplicate policies escaped detection and the occurrence of the consequent multiple claims has thus inflated the corresponding value of X^2.

Table H

Value of X^2	Observed frequency	Expected frequency	Contribution to χ^2
·000– ·297	7	2·82	6·196
·297– ·711	11	11·28	·007
·711– 1·064	16	14·10	·256
1·064– 1·649	34	28·20	1·193
1·649– 2·195	33	28·20	·817
2·195– 2·753	26	28·20	·172
2·753– 3·357	23	28·20	·959
3·357– 4·045	14	28·20	7·150
4·045– 4·878	33	28·20	·817
4·878– 5·989	23	28·20	·959
5·989– 7·779	21	28·20	1·838
7·779– 9·488	21	14·10	3·377
9·488–13·277	16	11·28	1·975
13·277– ∞	4	2·82	·494
	282	282·00	$\chi^2=26{\cdot}210$ $\nu=13$

Now although the largest value of X^2 is only slightly improbable* the contribution to χ^2 in Table H resulting from the excess of X^2 greater than 7·779 is 5·846 which is close to the 7·150 obtained from the lack of X^2-values near the mean. If some of these larger X^2-values were 'deflated' towards their mean the fit could be materially improved. There are thus grounds for believing that the poor 'fit' of the distribution in Table H may be due to the presence of undetected duplicates among the deaths. It may be mentioned that 12·3 of the 12·8 excess of actual X^2-values greater than 7·779 derives from durations 10–14 and 15–19; in fact there are 9·9 'too many' large X^2-values at durations 10–14 and 2·4 'too many' at $t=15$–19. These excesses account for the large $\sum_x X^2$-values shown for these durations at the foot of Table 3.

* The probability of at least one X^2 among 282 in excess of 21·504 is ·071.

Bearing in mind our doubts about the q-values at durations 10–19 let us form the differences $\Delta_k \bar{q}_x^{(k)}$ ($k=1, 2, \ldots, 7$), where $\bar{q}_x^{(k)}$ is the mean mortality of the group of durations $t=5k, 5k+1, \ldots, 5k+4$. There is, of course, a strong tendency for the exposed-to-risk to decrease at successive k-values but we may argue that the chance that the sign of any such difference is negative is unlikely to be more than ·55 given the underlying equality of the universal mortality rates at all these durations. The individual results are shown in Table 3 and these are summarized in Table I below. Only the result at $k=1$ is improbable on the null hypothesis and this would be rectified if some 4 or 5 of the supposedly inflated values of $\bar{q}_x^{(2)}$ were reduced sufficiently to change the corresponding values of $\Delta_k \bar{q}_x^{(1)}$ from positive to negative. However, the excess of positive values at $k=2$ is rather unexpected, since the supposed inflation at $t=10$–14 should tend to produce an excess of negative signs at this point. And although there may also be some inflated values of $\bar{q}_x^{(3)}$ to be corrected, this only pushes the argument one step further to $k=3$ where positive differences are still in excess. Nevertheless, these peculiarities are not conclusive arguments against the plausibility of the 'undetected duplicates' hypothesis. It may be mentioned in passing that Table I gives no support to Sprague's (1870) 'maximum mortality' theory which he developed on the basis of the $H^{[M]}$ data (cf. Table 2) and defended strenuously for many years (1896, 1904).

Table I

Sign of $\Delta_k \bar{q}_x^{(k)}$	Number of indicated signs at k equal to						
	1	2	3	4	5	6	7
Positive	40	28	24	19	17	9	9
Negative	15	23	22	22	17	18	9
Total	55	51	46	41	34	27	18

Table J

Sign of $\Delta_t q_{[x-t]+t}$	Number of indicated signs at t equal to								
	5	6	7	8	9	10	11	12	13
Positive	33	32	25	28	28	25	$29\frac{1}{2}$	$31\frac{1}{2}$	$28\frac{1}{2}$
Negative	24	24	30	27	27	29	$24\frac{1}{2}$	$22\frac{1}{2}$	$24\frac{1}{2}$
Total	57	56	55	55	55	54	54	54	53

One other piece of evidence may be offered in favour of the notion of inflated values at $t=10$–14 and against the continuance of temporary selection until a point of time between $t=5$ and $t=10$—the latter view being based on the figures of Table I. Consider the differences $\Delta_t q_{[x-t]+t}$ ($t=5, 6, \ldots, 13$) at each age x. Since the exposed to risk at successive durations are roughly equal the null hypothesis would imply that the sign of each such difference is almost equally likely to be plus or minus. If a few sets of undetected duplicate claims exist in the data the isolated excessively large $q_{[x-t]+t}$ would cause a sequence plus-minus or minus-plus among the differences $\Delta_t q_{[x-t]+t}$ at any age x. The essential equality in the number of plus and minus signs would thus be unaffected by individual inflated q-values. With this in mind we may review Table J which summarizes the results of an enumeration of the positive and negative values of $\Delta_t q_{[x-t]+t}$ $(t=5, 6, \ldots, 13)$ used in preparing Table 3. None of the nine sets of results is significant even at the 20% level. This suggests that the significant values of $\sum_x X^2$ at the foot of Table 3 are due to a disturbance such as undetected duplicates rather than any continuance or recrudescence of selective forces.

We may thus conclude that temporary selection has ceased to have any perceptible effect on the values of $q_{[x-t]+t}$ after the point of time $t=4\frac{1}{2}$.

APPENDIX II

Comparison of durations 2 and 3

Table 1 revealed a surprising lack of evidence for the existence of temporary selection after the third year of insurance. It is natural to ask what the situation would have been if the analysis had been commenced at $t = 2$ instead of $t = 3$.

It was observed from Table A that $q_{[x-3]+3}$ was about 95% of $q_{[x-4]+4}$ in some of the well-known graduated tables. Table AA shows the percentages that $q_{[x-2]+2}$ bears to $q_{[x-3]+3}$ at a number of ages in the officially graduated tables prepared from the data analysed in this paper.* Although there is considerable variability it could be stated that actuaries have accepted the view that, on the average, $q_{[x-2]+2}$ is a little more than 90% of $q_{[x-3]+3}$.

In order to test this conclusion the comparisons of the first part of Table 1 (namely, $t = 3$ *vs.* $t = 4$) were repeated on $t = 2$ and $t = 3$, using the 'unit criterion' described earlier to determine whether or not to include a particular age x. However, since there is no reason to suspect that the differences $q_{[x-3]+3} - q_{[x-2]+2}$ are negative, the criteria analogous to X_1^2 were not computed. Instead the analysis was restricted to the calculation of X_1' where

$$X_1' = \operatorname{sgn}(q_3 - q_2)\{q(1-q)\}^{-\frac{1}{2}}(y_2 q_2 + y_3 q_3 - yq)^{\frac{1}{2}},$$

$$y_2 \equiv y_{[x-2]+2},\ q_2 \equiv y_{[x-2]+2}/E_{[x-2]+2}, \text{ etc.},$$

$$y \equiv y_2 + y_3, \quad E \equiv E_2 + E_3, \quad q \equiv (y_2+y_3)/(E_2+E_3).$$

Actually these computations were made on an IBM 704† using an alternative form of X_1', namely

$$X_1' = (E_2 y_3 - E_3 y_2)(E_2+E_3)^{\frac{1}{2}} E_2^{-\frac{1}{2}} E_3^{-\frac{1}{2}} (E_2+E_3-y_2-y_3)^{-\frac{1}{2}} (y_2+y_3)^{-\frac{1}{2}}.$$

Table 1 *bis* summarizes the results.

We notice at once that the hypothesis that $q_{[x-2]+2}$ is equal to $q_{[x-3]+3}$ is unreasonable in each of the three sets of experiences. While a number of the individual values of $\Sigma X_1'$ are not significant—some are even negative—this is to be expected even when $q_{[x-2]+2}$ is only 90% of $q_{[x-3]+3}$ (see Table C).

* The anomalous results for the Austrian and Hungarian tables are due to the fact that separate Makeham graduations were made of the data at each duration t and no attempt was made to link up successive values of $q_{[x-t]+t}$.

† I am very grateful to Miss Myrna L. Knopf for her valuable help on this and the other calculations of Appendix II.

In order to summarize the behaviour of the 1743 individual values of X_1' calculated for Table 1 *bis* they were grouped into a frequency distribution as shown in Table FF. Since Table 1 *bis* had shown that a zero mean for X_1' was untenable it was considered of interest to test this distribution against the hypothesis N (·1320, 1), the mean of this Normal distribution being obtained as the average of the 1743 values of X_1'. In Table FF this would mean that the expectation of each cell is equal to 87·15, or one-twentieth of 1743. The resulting χ^2-goodness-of-fit test produces a value of 23·713 with 18 degrees of freedom, which is quite satisfactory.

The implication of Table FF is that the 50 sets of mortality data analysed in this paper are consonant with the view that there is a universal biometric constant characterizing the difference between mortality during the third and fourth years after policy purchase. And this holds true whether the experience refers to annuitants or to assured lives.

Table 1 *bis*. To test the significance of the differences $q_{[x-3]+3} - q_{[x-2]+2}$

Experience	Ages used	No. of deaths	$\sum_x X_1'$	d.f.
(1) Amicable	41½–68½*	31	− 5·92	16
(2) Scottish Amicable	35, 38–48, 53, 55, 57	48	7·10	15 *s*.
10 English and 10 Scottish:				
(3) Males	21–76	2,091	17·41	56 *s.s.*
(4) Females	25–74	335	−1·06	50
(5) Mutual of N.Y.	25–67	1,084	11·06	43 *s*.
(6) Mutual Benefit of N.J.	25–67	861	13·40	43 *s*.
23 German companies:				
(7) Males, medical	23–71	5,896	6·18	49
(8) Males, non-medical	20–74	3,181	7·67	55
(9) Females, medical	19–69	2,189	7·14	51
(10) Females, non-medical	18–72, 74	2,913	·93	56
(11) Connecticut Mutual	23–67	1,163	10·74	45
(12) Canada Life	24–60	270	− ·37	37
(13) Gotha	23–68	1,316	1·74	46
(14) Stuttgarter	24–62	1,236	6·52	39
28 Austrian companies:				
(15) Males, W.L.	23, 25–71	2,636	14·75	48 *s*.
(16) Males, E.A.	24–62	1,708	− ·34	39
(17) Females	24–67	786	5·94	44
18 Hungarian companies:				
(18) Males, W.L.	26–67	1,487	11·27	42 *s*.
(19) Males, E.A.	26–61	1,251	4·26	36
(20) Females	22–62	439	− 5·14	41
17 Swedish companies:				
(21) Males	19–62	722	11·11	44 *s*.
9 Russian companies:				
(22) Males, W.L., I	33–54, 58	68	− 7·40	23
(23) Males, W.L., II	29–62	270	− 6·84	34
(24) Males, W.L., III	28–69	749	4·55	42
(25) Males, E.A., III	25–62	963	− 4·21	38
(1)–(25) Totals		33,693	110·49	1032 *s.s.s.*

Table 1 *bis* (*continued*)

Experience	Ages used	No. of deaths	$\sum_x X_1'$	d.f.
Dutch provinces:				
(*a*) Males	6–12	17	− 1·12	7
(*b*) Females	6–12	21	2·05	7
(*c*) Monks of St Maur	22–27	16	2·58	6
English Tontine:				
(*d*) Males	7–9, 17–19, 21, 22	20	·00	8
(*e*) Females	6–7, 14–20, 22	27	− 1·49	10
British Govt.:				
(*f*) Males	44, 46–89	807	− 7·68	45
(*g*) Females	40–87	973	11·88	48 *s.*
15 American companies:				
(*h*) Males	56–85	310	12·32	30 *s.*
(*i*) Females	58–87	207	4·24	30
3 French companies:				
(*j*) Males	34–92†	3,242	5·25	55
(*k*) Females	39–92	3,515	1·34	54
5 Swiss companies:				
(*l*) Males	65–77	43	2·09	13
(*m*) Females	55–69, 73–80	130	11·79	23 *s.s.*
(*a*)–(*m*) Totals		9,328	43·25	336 *s.s.*
60 British:				
(α) Males, W.L., N.P., old	56, 60	5	·57	2
(β) Males, W.L., N.P., new	23–77	792	11·02	55
(γ) Males, W.L., P., old	23–69	517	4·31	47
(δ) Males, W.L., P., new	18–78	5,849	23·24	61 *s.s.*
(ζ) Males, E.A., new	22–55	773	4·02	34
42 British:				
(θ) Males, Annuitants, new	54–88	475	− ·37	35
(ι) Females, Annuitants, old	65–72, 75–80	41	− 1·22	14
(κ) Females, Annuitants, new	50–90	915	15·06	41 *s.s.*
3 Japanese Offices:				
(λ) Males	20–62	4,335	5·37	43
(μ) Females	20–62	1,403	14·34	43 *s.*
(α)–(μ) Totals		15,105	76·34	375 *s.s.s.*

* Excepting ages 42½, 45½, 46½, 49½–50½, 53½, 59½, 60½, 62½, 65½–67½.
† Excepting ages 35, 38, 41, 42.

Since the sum of the square roots of the numbers of deaths, y, involved in the computed values of X_1' is 8635, the value of this biometric constant may be estimated from the equation

$$1743 \times \cdot 1320 \approx \tfrac{1}{2}\ln \kappa^{-1} 8635$$

i.e. $$\kappa \approx \cdot 948.$$

In general, then, the mortality at any age x at duration $t = 2$ is 95% of that at the same age at duration $t = 3$. Once again we have found that the 'graduating actuary' consistently exaggerates the effects of temporary selection.

Table AA. Graduated values of $10^2 q_{[x-2]+2}/q_{[x-3]+3}$

Experience	Age x					
	25	35	45	55	65	75
$H^{[M]}$	91	92	91	91	97	100
Canada Life	87	89	88	—	—	—
Gotha, Neue Bankliste	95	93	94	94	93	—
Austrian:						
Males, W.L.	94	103	102	94	91	90
Males, E.A.	89	96	102	107	110	112
Females	108	104	96	84	73	65
Hungarian:						
Males, W.L.	101	98	94	90	87	—
Males, E.A.	91	96	101	105	107	—
Females	134	145	166	197	230	—
Swedish companies, Males*	91	91	92	92	93	—
British Govt.						
Males	—	—	98	99	99	99
Females	—	—	95	94	90	94
$O^{[M]}$	94	94	93	93	92	92
$O^{[NM]}$	94	93	92	91	91	90
$O^{[am]}$	82	84	86	89	90	91
$O^{[af]}$	83	83	84	85	88	90
Japanese, Males $J^{[M]}$	98	98	98	98	98	98

* Ratios of m not q.

Table FF

Values of X_1'	Observed frequencies	Values of X_1'	Observed frequencies
$-\infty$ –(−1·513)	83	·132– ·258	105
−1·513–(−1·150)	85	·258– ·385	84
−1·150–(− ·904)	91	·385– ·517	79
− ·904–(− ·710)	76	·517– ·656	71
− ·710–(− ·542)	82	·656– ·806	87
− ·542–(− ·392)	99	·806– ·974	93
− ·392–(− ·253)	76	·974–1·168	90
− ·253–(− ·121)	84	1·168–1·414	93
− ·121– ·006	115	1·414–1·777	89
·006– ·132	82	1·777 – ∞	79
			1743

Table 1. To test the significance of the differences at each attained age between the mortality rates at various contract durations

(*s.* = significant at the 5 % level but not at the 1 % level; *s.s.* = significant at the 1 % level but not at the ·1 % level; *s.s.s.* = significant at the ·1 % level.)

Experience	Reference	Observational year[1]	Ages used	Between the mortality at $t=3$ and $t=4$: No. of deaths	$\sum_x X_1^2$	d.f.	$\sum_x X_1$	Between the mortality at $t=5, 6, 7, 8$ and 9: No. of deaths	$\sum_x X_2^2$	d.f.	Between the mortality at $t=10, 11, 12, \ldots$: No. of deaths	$\sum_x X_3^2$	d.f.	Between the mean-mortality at $t=3$ and 4, at $t=5$–9 and at $t=10$ and over: No. of deaths	$\sum_x X_4^2$	d.f.	Computer
A. Assurance experiences																	
(1) Amicable Society, 1808–41, Whole Life	Galloway (1841)	P	41½–75½[2]	28	12·700*	19	−3·56*	148	62·314*	76	83	63·475*	48	259	56·482	39 *s.*	C.J.C.
(2) Scottish Amicable, 1826–60, non-hazardous males	Spens (1861)	P	35, 38–48, 53, 55, 57	54	13·673	15	−·01	55	21·770	21	—	—	—	109	11·617	15	C.J.C.
10 English and 10 Scottish Offices, –1863:	Brown (1869)	C															
(3) Healthy males			21–88	2,173	72·849*	58	15·42* *s.*	4,645	194·375*	223	11,077	1,291·376*	1,205 *s.*	17,895	165·428	114 *s.s.s.*	C.J.C.
(4) Healthy females			25–82, 84–86	307	31·161	50	−4·19	735	196·370	195	1,202	500·515	466	2,244	90·089	96	M.L.K.
(5) Mutual of New York, 1843–74, healthy lives	Bartlett (1875)	C	25–74, 77	1,085	42·102	44	4·43	1,572	154·743	177	1,294	564·476	477 *s.s.*	3,951	93,372	86	C.J.C.
(6) Mutual Benefit, New Jersey, 1845–79, healthy lives	Miller (1880)	C	25–78	902	51·055	44	5·96	1,901	151·856	182	2,603	791·841	714 *s.*	5,406	83·576	90	J.C.H.
23 German companies, –1875:	Zillmer *et al.* (1883)	B															
(7) Males, medically examined			23–85	5,528	46·300	49	1·42	10,818	165·227	189	13,474	1,056·112	1,023	29,820	105·741	94	J.U.T.
(8) Males, incomplete medical			20–79, 81–2	3,049	48·331	56	−4·94	6,377	220·870	225	4,736	413·620	374	14,162	86·826	108	D.T.
(9) Females, medically examined			19–72, 76–83[3]	2,059	44·792	51	9·46	3,312	222·377	194	1,796	356·885	386	7,167	117·029	95	J.U.T.
(10) Females, incomplete medical			18–82[4]	2,820	60·366	56	4·86	5,800	200·988	221	4,120	380·138	378	12,740	91·538	110	J.U.T.
(11) Connecticut Mutual, 1846–78, males	Wells (1884)	P	23–79	1,158	37·514*	46	·37*	2,677	167·264*	180	2,747	617·343*	620	6,582	123·757	91 *s.*	C.J.C.
(12) Canada Life, 1847–93, healthy males	Sanderson (1895)	P	24–64, 68, 71	255	36·372	38	−·85	616	136·420	142	925	388·252	370	1,796	75·797	71	J.R.N.
(13) Gotha, 1852–96, males, Whole Life	Karup (1903)	P	23–89	1,353	46·420*	47	−1·60*	3,816	190·068*	179	17,604	1,090·654*	1,128	22,773	150·121	91 *s.s.*	S.W.M.
(14) Stuttgarter, 1854–1901, males	Lohmüller (1907)	P	24–86	1,219	39·585	40	−9·59	3,242	164·109	163	12,336	1,131·174	1,084	16,797	101·722	90	C.J.C.
28 Austrian companies, 1876–1900:	Klang *et al.* (1907)	P															
(15) Males, Whole Life			23–87[5]	2,774	43·230	49	3·28	7,374	196·660	200	30,311	1,200·201	1,200	40,459	89·528	96	T.W.G.
(16) Males, Endowment Assurances			24–69, 73	1,686	42·829	40	16·65 *s.s.*	3,195	130·521	157	2,852	339·224	353	7,733	83·320	78	T.W.G.
(17) Females			24–88	734	48·489	45	−13·35	1,866	176·736	189	10,129	1,162·265	1,199	12,729	80·475	94	T.W.G.
18 Hungarian companies, 1876–1900:	Ormody *et al.* (1910)	P															
(18) Males, Whole Life			26–85	1,451	57·979	43	−8·78	3,671	197·527	175	12,479	992·672	948	17,601	89·914	87	T.W.G.
(19) Males, Endowment Assurances			26–69	1,163	31·251	38	3·18	2,196	139·959	139	1,910	277·371	269	5,269	56·874	70	J.U.T.
(20) Females			22–83	385	32·172	42	−2·22	967	179·734	168	3,772	864·492	839	5,124	69·821	84	T.W.G.
17 Swedish companies, 1895–1906, medically examined lives:	Jäderin *et al.* (1915)	P															
(21) Males			19–79	762	27·042	44	5·41	1,840	189·678	170	4,332	858·539	795	6,934	91·607	85	J.R.N.
9 Russian companies, 1835–1910:	Yastremski *et al.* (1916)	P															
(22) Males, W.L., Period I			33–66	69	24·106	25	−·58	210	106·157	95	367	190·203	210	646	52·270	47	J.U.T.
(23) Males, W.L., Period II			29–79	284	29·366	35	12·87 *s.*	756	193·002	140 *s.s.*	3,475	858·427	780 *s.*	4,515	57·082	70	J.U.T.
(24) Males, W.L., Period III			28–69, 74	736	48·038	42	1·28	1,693	123·056	151	2,169	344·530	334	4,598	58·994	76	J.U.T.
(25) Males, E.A., Period III			25–66	786	30·857	39	−6·57	1,386	161·505	143	1,069	226·124	239	3,241	60·846	70	D.H.S.
(1)–(25) Totals				32,820	998·579	1,055	28·35	70,869	4043·286	4,094	146,862	15,959·909	15,436	250,550	2143·832	2047	
B. Annuity experiences																	
Dutch provinces, 1613–:	Kersseboom (1742)	P															
(*a*) Males			6–58	22	6·739	8	·80	57	29·464	32	1,125	374·269	340	1,204	14·959	16	P.T.
(*b*) Females			6–67	29	8·824	8	−·02	64	37·607	32	1,140	411·170	412	1,233	20·067	16	P.T.
(*c*) Monks of St Maur, 1607–1745	Deparcieux (1746)	P	22–88	27	3·951	7	−·42	74	31·747	28	1,830	430·853	380	1,931	9·168	14	P.T.
English Tontine of 1789:	Seal (1949)	C															
(*d*) Males			7–9, 17–85[6]	35	7·858	10	5·18	74	26·389	41	1,729	886·137	900	1,838	13·905	22	J.M.W.
(*e*) Females			6–7, 13–87[7]	30	7·839	10	1·91	75	36·292	39	1,806	877·524	902	1,911	12·549	19	J.M.W.
British Government, 1808–75:	Finlaison (1884)	B															
(*f*) Males			44, 46–94, 97	837	44·128	45	11·08 *s.*	2,012	162·128	171	4,132	713·824	638 *s.*	6,981	100·053	85	A.F.A.
(*g*) Females			40–95, 97	1,086	41·387*	49	5·00*	2,916	184·424*	185	8,620	947·804*	942	12,622	126·706	93 *s.*	A.F.A.
15 American companies, –1892:	Weeks (1896)	P															
(*h*) Males			56–87, 89	305	24·450	31	−7·11	455	122·873	103	69	19·480	20	829	59·983	49	H.L.S.
(*i*) Females			58–59, 61–88	214	36·157	30	3·14	343	88·885	92	53	6·902	12	610	45·940	45	H.L.S.
3 French companies, 1820–98:	Duplaix *et al.* (1902)	C															
(*j*) Males			34–92[8]	3,185	62·089	55	2·77	6,631	180·595	187	9,365	764·657	694 *s.*	19,181	98·570	96	J.U.T.
(*k*) Females			39–92	3,469	49·302*	54	−1·58*	7,929	171·338*	190	14,515	771·390*	756	25,913	157·446	97 *s.s.s.*	J.U.T.
5 Swiss companies, –1899:	Kihm (1904)	B															
(*l*) Males			65–78	47	18·636	14	·73	70	31·986	30	1	—	—	118	25·957	16	H.L.S.
(*m*) Females			55–86	142	29·260	25	−5·93	194	76·106	70	55	22·307	21	391	31·115	34	M.L.K.
(*a*)–(*m*) Totals				9,428	340·620	346	15·55	20,894	1179·834	1,200	44,440	6,226·317	6,017	74,762	716·427	602 *s.s.s.*	
C. Experiences based on 'Selections' not 'Lives'																	
60 British Offices, 1863–93:		P															
(*α*) Males, W.L., Non-Par, Old	Higham & Low (1900)		33–92	75	20·464*	21	−6·44*	501	179·911*	171	10,092	1,610·359*	1,619	10,668	119·643	82 *s.s.*	C.J.C.
(*β*) Males, W.L., Non-Par, New			23–86	874	73·595*	56	19·71* *s.s.*	1,805	181·909*	214	3,358	726·654*	749	6,037	159·769	106 *s.s.s.*	C.J.C.
(*γ*) Males, W.L., Par, Old			23–96	795	49·336	50	12·80 *s.*	4,452	188·661	218	106,993	2,362·432	2,337	112,240	132·565	108	C.J.C.
(*δ*) Males, W.L., Par, New			18–89	5,880	57·797*	61	2·53*	15,165	277·988*	235 *s.*	35,620	958·398*	979	56,665	292·045	119 *s.s.s.*	C.J.C.
(*ε*) Males, E.A., Old	Manly & Deuchar (1900)		33, 35, 37–59	1	—	—	—	62	21·101	29	605	302·337	312	668	12·327	16	C.J.C.
(*ζ*) Males, E.A., New			22–60	732	28·603	36	−·29	1,385	146·900	138	1,612	335·112	335	3,729	76·928	68	C.J.C.
42 British Offices, 1863–93:																	
(*η*) Males, annuitants, old	Manly & Deuchar (1899)	P	66–90	27	4·715	7	−·87	146	73·961	63	591	279·794	258	764	34·053	31	M.L.K.
(*θ*) Males, annuitants, new			54–90	475	32·793	36	9·78	890	121·835	123	680	185·344	181	2,045	55·247	61	M.L.K.
(*ι*) Females, annuitants, old			60–89, 94, 96	82	23·020	18	−·01	454	101·222	102	2,634	505·132	526	3,170	43·931	48	M.L.K.
(*κ*) Females, annuitants, new			50–92	967	33·980	42	2·05	1,984	147·935	146	2,529	357·753	339	5,480	77·494	74	M.L.K.
3 Japanese offices, 1881–1905:	Ebihara *et al.* (1910)	P															
(*λ*) Males			20–74	3,945	47·470	44	2·79	6,717	190·453	180	3,332	277·595	287	13,994	93·936	90	J.U.T.
(*μ*) Females			20–73	1,234	59·412	44	−9·58	1,810	155·280	180	607	143·371	138	3,651	92·135	89	J.U.T.
(*α*)–(*μ*) Totals				15,087	431·135	415	32·47	35,371	1787·156	1,799	168,653	8,044·281	8,060	219,111	1190·071	892 *s.s.s.*	

* This $\sum_x X_k^2$ value is based on divisors $q_x^{(k)}(1-q_x^{(k)})$ instead of $q_x(1-q_x)$ at every age x.

1 P = contract years; C = calendar years; B = life years.
2 Excepting ages 42½, 45½, 46½, 50½, 53½, 62½, 66½, 67½, 71½, 73½.
3 Excepting ages 77, 79, 81, 82.
4 Excepting ages 78, 80.
5 Excepting ages 24, 86.
6 Excepting ages 34, 41, 46.
7 Excepting ages 25, 82, 86.
8 Excepting ages 35, 38, 41, 42.

[P.T.O

Table 1. To test the significance of the differences at each attained age between the mortality rates at various contract durations

(s. = significant at the 5% level but not at the 1% level; s.s. = significant at the 1% level but not at the ·1% level; s.s.s. = significant at the ·1% level.)

Experience	Reference	Observational year[1]	Ages used	Between the mortality at t=3 and t=4: No. of deaths	$\sum X_x^2$	d.f.	$\sum X_x$	Between the mortality at t=5, 6, 7, 8 and 9: No. of deaths	$\sum X_x^2$	d.f.	Between the mortality at t=10, 11, 12, …: No. of deaths	$\sum X_x^2$	d.f.	Between the mean-mortality at t=3 and 4, at t=5–9 and at t=10 and over: No. of deaths	$\sum X_x^2$	d.f.	Comparer
A. Assurance experiences																	
(1) Amicable Society, 1808–41, Whole Life	Galloway (1841)	P	41½–75½[2]	28	12·700*	19	−1·16*	148	62·314*	76	83	63·475*	48	250	56·482	39 s.	C.J.C.
(2) Scottish Amicable, 1826–60, non-hazardous males	Spens (1861)	P	34, 38–48, 53, 55, 57	54	13·673	15	−·02	35	21·770	21	—	—	—	109	11·617	15	C.J.C.
10 English and 10 Scottish Offices, –1863:	Brown (1869)	C															
(3) Healthy males			21–88	2,173	72·840*	58	13·42* s.	4,645	194·375*	203	11,077	1,291·376*	1,205 s.	17,895	165·428	114 s.s.s.	C.J.C.
(4) Healthy females			24–82, 84–86	307	31·161	50	−4·19	725	196·279	193	1,202	500·513	466	2,244	90·069	96	M.L.K.
(5) Mutual of New York, 1843–74, healthy lives	Bartlett (1875)	C	25–74, 77	1,085	40·102	44	4·43	1,374	154·742	177	1,094	564·476	477 s.s.	2,951	93·374	86	C.J.C.
(6) Mutual Benefit, New Jersey, 1845–79, healthy lives	Miller (1880)	C	25–78	902	51·055	44	3·96	1,901	151·836	182	2,603	791·641	714 s.	5,406	83·576	92	J.C.H.
23 German companies, –1875:	Zillmer et al. (1883)	B															
(7) Males, medically examined			23–85	1,328	46·900	49	1·42	10,818	163·227	189	13,474	1,056·212	1,023	29,820	107·741	94	J.U.T.
(8) Males, incomplete medical			20–79, 81–2	5,049	48·321	56	−4·94	6,377	220·870	225	4,736	413·640	374	14,162	86·646	108	D.T.
(9) Females, medically examined			19–79, 76–81[3]	2,059	44·794	51	9·46	3,312	212·377	194	1,796	356·885	386	7,167	117·089	91	J.U.T.
(10) Females, incomplete medical			18–82[4]	2,820	60·366	56	4·86	5,800	200·988	221	4,120	380·138	378	12,740	91·558	110	J.U.T.
(11) Connecticut Mutual, 1846–78, males	Wells (1884)	P	23–79	1,158	27·214*	46	·37*	2,677	167·264*	180	2,747	617·343*	600	6,582	125·757	91 s.	C.J.C.
(12) Canada Life, 1847–93, healthy males	Sanderson (1893)	P	24–64, 68, 72	255	36·372	38	−·85	616	136·400	144	925	368·452	370	1,796	75·797	71	J.R.N.
(13) Gotha, 1852–96, males, Whole Life	Karup (1903)	P	23–89	1,352	46·420*	47	−1·60*	3,816	190·068*	179	17,604	1,090·634*	1,128	22,772	130·121	91 s.s.	S.W.M.
(14) Stuttgarter, 1854–1901, males	Lokmüller (1907)	P	24–86	1,419	39·583	40	−9·39	2,242	164·109	163	12,236	1,131·174	1,084	16,797	101·782	90	C.J.C.
26 Austrian companies, 1876–1900:	Klang et al. (1907)	P															
(15) Males, Whole Life			23–87[5]	2,774	42·230	49	2·28	7,374	196·660	200	30,311	1,200·202	1,200	40,459	89·328	96	T.W.G.
(16) Males, Endowment Assurance			24–69, 72	1,086	40·829	40	16·65 s.s.	3,195	130·581	137	3,852	339·844	353	7,733	83·320	78	T.W.G.
(17) Females			24–88	724	48·480	41	−13·25	1,866	176·736	189	10,129	1,162·265	1,199	12,729	80·475	94	T.W.G.
18 Hungarian companies, 1876–1900:	Ormody et al. (1910)	P															
(18) Males, Whole Life			26–85	1,431	37·970	43	−8·78	3,671	197·327	175	12,479	992·672	948	17,601	87·914	87	T.W.G.
(19) Males, Endowment Assurance			26–69	1,163	31·251	38	1·18	2,196	139·959	139	1,910	277·371	289	5,269	58·874	70	J.U.T.
(20) Females			22–83	183	32·172	42	−2·22	967	179·734	168	3,778	864·492	899	5,184	69·841	84	T.W.G.
17 Swedish companies, 1855–1906, medically examined lives:	Jäderin et al. (1913)	P															
(21) Males			19–79	762	27·044	44	5·41	1,842	189·678	170	4,134	858·339	758	6,934	91·607	85	J.R.N.
9 Russian companies, 1835–1910:	Yastremski et al. (1916)	P															
(22) Males, W.L., Period I			33–60	69	24·106	25	−·58	210	106·137	93	367	190·203	210	646	52·276	47	J.U.T.
(23) Males, W.L., Period II			20–79	484	29·366	35	12·87 s.	736	199·009	140 s.s.	3,475	858·427	780 s.	4,515	57·082	70	J.U.T.
(24) Males, W.L., Period III			26–69, 74	736	46·056	42	1·18	1,693	125·036	131	2,169	344·330	334	4,598	58·904	76	J.U.T.
(25) Males, E.A., Period III			25–66	786	30·857	39	−6·57	1,986	161·505	143	1,069	220·124	239	3,841	60·646	70	D.H.S.
(1)–(25) Totals				32,820	998·579	1,055	28·32	70,869	4042·286	4,094	146,862	15,615·909	15,436	250,550	2143·852	2047	
B. Annuity experiences																	
Dutch provinces, 1615–:	Kersseboom (1742)	P															
(a) Males			6–98	22	6·735	8	·80	37	29·464	32	1,125	174·269	340	1,304	14·919	16	P.T.
(b) Females			6–67	29	8·684	8	−·02	64	37·607	32	1,140	411·170	412	1,232	20·067	16	P.T.
(c) Monks of St Maur, 1607–1745	Deparcieux (1746)	P	22–88	27	3·951	7	−·44	74	31·247	48	1,830	430·853	380	1,931	9·168	14	P.T.
English Tontine of 1789:	Seal (1949)	C															
(d) Males			7–9, 17–85[6]	35	7·658	10	5·18	74	16·389	41	1,729	886·137	900	1,838	13·905	24	J.M.W.
(e) Females			6–7, 13–87[7]	30	7·839	10	1·92	75	56·494	39	1,806	877·344	904	1,911	12·649	19	J.M.W.
British Government, 1808–75:	Finlaison (1884)	B															
(f) Males			44, 46–94, 97	837	44·128	43	11·06 s.	2,022	162·128	171	4,152	713·844	638 s.	6,981	110·053	85	A.F.A.
(g) Females			40–95, 97	1,086	41·987*	49	3·00*	2,326	184·424*	183	8,620	947·804*	942	12,604	126·706	99 s.	A.F.A.
15 American companies, –1892:	Weeks (1896)	P															
(h) Males			56–87, 89	305	24·450	31	−7·11	435	112·873	102	69	29·480	20	809	59·983	49	H.L.S.
(i) Females			58–59, 61–88	214	36·187	30	2·14	343	88·883	92	53	6·900	14	610	45·940	48	H.L.S.
3 French companies, 1819–98:	Duplaix et al. (1902)	C															
(j) Males			34–98[8]	3,185	62·089	55	4·77	6,631	180·595	187	9,263	764·637	694 s.	19,181	98·879	96	J.U.T.
(k) Females			39–94	2,460	49·202*	54	−1·38*	7,029	171·238*	190	14,515	771·390*	786	24,023	157·448	97 s.s.s.	J.U.T.
5 Swiss companies, –1899:	Kihm (1904)	B															
(l) Males			65–78	47	18·636	14	·73	70	21·986	30	1	—	—	118	45·957	16	H.L.S.
(m) Females			55–86	242	20·280	25	−5·93	194	76·106	70	55	22·307	21	391	32·215	34	M.L.K.
(a)–(m) Totals				9,498	340·680	346	13·35	20,894	1179·834	1,200	44,440	6,226·327	6,017	74,760	726·487	602 s.s.s.	
C. Experiences based on 'Selections' not 'Lives'																	
60 British Offices, 1863–93:		P															
(α) Males, W.L., Non-Par, Old	Higham & Low (1900)		33–92	75	20·464*	21	−6·44*	501	170·911*	171	10,092	1,610·359*	1,619	10,668	119·641	82 s.s.	C.J.C.
(β) Males, W.L., Non-Par, New			23–86	874	73·593*	56	19·71* s.s.	1,805	181·006*	214	2,358	726·634*	749	6,037	159·769	106 s.s.s.	C.J.C.
(γ) Males, W.L., Par, Old			23–96	795	49·330	50	12·60 s.	4,492	188·661	218	106,903	2,362·432	2,337	118,240	132·363	108	C.J.C.
(δ) Males, W.L., Par, New			18–89	5,880	57·707*	61	4·32*	15,163	277·988*	235 s.	35,880	958·398*	979	56,665	202·043	135 s.s.s.	C.J.C.
(ε) Males, E.A., Old	Manly & Deuchar (1900)		32, 33, 37–39	1	—	—	—	62	21·101	25	605	302·327	312	668	12·367	16	C.J.C.
(ζ) Males, E.A., New			42–80	734	28·609	36	−·29	1,383	146·900	138	1,618	335·114	335	2,729	76·928	68	C.J.C.
42 British Offices, 1863–93:																	
(η) Males, annuitants, old	Manly & Deuchar (1900)	P	66–90	27	4·715	7	−·87	146	73·961	63	591	279·794	258	764	24·953	21	M.L.K.
(θ) Males, annuitants, new			54–90	475	22·793	26	0·78	800	121·825	122	880	185·344	181	2,045	55·947	61	M.L.K.
(ι) Females, annuitants, old			60–89, 94, 96	82	23·020	18	−·01	454	101·212	108	2,634	309·132	306	3,170	43·931	48	M.L.K.
(κ) Females, annuitants, new			50–92	967	23·980	42	2·03	1,984	147·925	146	2,529	337·752	339	5,480	77·494	74	M.L.K.
3 Japanese offices, 1881–1915:	Shibata et al. (1916)	P															
(λ) Males			20–74	2,043	47·470	44	2·79	6,717	190·431	180	1,330	377·191	367	13,094	93·936	90	J.U.T.
(μ) Females			20–73	1,034	59·418	44	−9·58	1,810	155·280	180	807	143·371	138	3,031	94·135	89	J.U.T.
(α)–(μ) Totals				11,087	451·133	413	21·47	38,271	1787·136	1,799	168,653	8,044·481	8,060	219,111	1190·071	892 s.s.s.	

* This $\sum X_x^2$ value is based on division $q_x^{(0)}(1-q_x^{(0)})$ instead of $q_x(1-q_x)$ at every age x.

[1] P = contract years; C = calendar years; B = life years.
[2] Excepting ages 42½, 45½, 46½, 50½, 53½, 60½, 66½, 67½, 71½, 73½.
[3] Excepting ages 77, 79, 81, 82.
[4] Excepting ages 76, 80.
[5] Excepting ages 24, 86.
[6] Excepting ages 24, 41, 46.
[7] Excepting ages 45, 82, 86.
[8] Excepting ages 32, 38, 41, 42.

[P.T.O.

REFERENCES

ABDEL-ATY, S. H. (1954). Approximate formulae for the percentage points and the probability integral of the non-central χ^2 distribution. *Biometrika*, **41**, 538–40.

ABEL, A. (1914). Wirkungen der Auslese in der Versichertensterblichkeit der deutschen Lebensversicherung. *Schr. Zentralstelle gemein. dtsch. Sterblichkeits-Untersuchungen.*

ACKLAND, T. G. (1907). Notes on the British Offices Life Annuity Tables (1893). *T.F.A.* **3**, 285–349.

ACKLAND, T. G., HARDY, G. F., BAKER, H. J. & CHATHAM, J. (1903). An account of the principles and methods adopted in the compilation of the data, the graduation of the experience and the construction of deduced tables. *British Offices Life Tables* 1893. London.

AITKEN, A. C. (1939). *Statistical Mathematics.* Edinburgh.

BARTLETT, W. H. C. (1875). *Report of the Mortuary Experience of The Mutual Life Insurance Company of New York from* 1843 *to* 1874. New York.

BOREL, E. (1939). *Valeur pratique et Philosophie des Probabilités.* Paris.

BORTKIEWICZ, L. V. (1917). *Die Iterationen.* Berlin.

*BROWN, S. (1869). *The Mortality Experience of Life Assurance Companies, collected by the Institute of Actuaries.* London.

BUCHANAN, J. (1927). The theory of selection: its history and development. *T.F.A.* **11**, 43–104.

CANTELLI, F. (1914). Genesi e costruzione delle tavole di mutualità. *Boll. Notiz. Cred. Previd.*

CHATHAM, J. (1891). An examination of the published experience of several life offices and groups of life offices, etc. *J.I.A.* **29**, 81–184.

COCHRAN, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, **10**, 417–51.

DEPARCIEUX, A. (1746). *Essai sur les Probabilités de la Durée de la Vie Humaine.* Paris.

DUPLAIX, H., OLTRAMARE, F. & QUIQUET, A. (1902). *Tables de Mortalité* 1900 *des Rentiers et Assurés en Cas de Vie.* Paris.

EBIHARA, K. *et al.* (1910). *Japanese Three Offices Life Tables.* Tokyo.

ELDERTON, W. P. (1906). On a form of spurious selection which may arise when mortality tables are amalgamated. *J.I.A.* **40**, 221–46.

ELDERTON, W. P. & OGBORN, M. E. (1943). The mortality of adult males since the middle of the eighteenth century as shown by the experience of life assurance companies. *J. R. Statist. Soc.* **106**, 1–30.

ENGELBRECHT, G. (1905). Die Wirkung der Auslese auf die Sterblichkeit in den ersten Versicherungsjahren. *Z. ges. VersichWiss.* **5**, 66–79.

ENGELBRECHT, G. (1906). Der Einfluss der Versicherungsdauer auf die Sterblichkeit in der Lebensversicherung. *Z. ges. VersichWiss.* **6**, 108–52.

FINLAISON, A. J. (1884). Report on the mortality of government life annuitants. *Government Annuities Act,* 1882: *Copy of the Treasury Minute, dated* 7 *September* 1883. London.

FISHER, R. A. (1954). *Statistical Methods for Research Workers.* Edinburgh.

GALLOWAY, T. (1841). *Table of Mortality deduced from the Experience of the Amicable Society for a Perpetual Assurance Office, during a Period of* 33 *years, ending April* 5, 1841. London.

HARDY, G. F. (1904). The British Offices Life Tables, 1893. Memorandum on the graduation of the whole-life without-profit mortality table—male lives. *J.I.A.* **38**, 501–32.

HIGHAM, J. A. (1851). On the value of selection as exercised by the policy-holder against the company. *J.I.A.* **1**, 179–202.

*HIGHAM, C. D. & LOW, G. M. (1900). *Combined Experience of Assured Lives* (1863–1893) *Unadjusted Data*: *Whole-Life Assurance*: *Males.* London.

HÖCKNER, G. (1934). Einfach und doppelt abgestufte Sterbetafeln. Möglichkeit der Herabsetzung des Auslesezeitraumes. *Atti Dec. Congr. Int. Attu.* **2**, 201–10.

HOEL, P. G. (1945). Testing the homogeneity of Poisson frequencies. *Ann. Math. Statist.* **16**, 362–8.

JÄDERIN, E., FREDHOLM, I. *et al.* (1915). *Undersökning of Dödligheten enligt Erfarenheten hos Sjutton Svenska Lifförsäkringsbolag.* Stockholm.

KARUP, J. (1903). *Die Reform des Rechnungswesens der Gothaer Lebensversicherungsbank, a. G.* Jena.

KERSSEBOOM, W. (1742). *Derde Verhandeling over de Probable Meenigte des Volks in de Provintie van Hollandt en Westvrieslandt.* The Hague.

KIHM, C. (1904). Die Sterblichkeit der schweizerischen Rentner. *Proc. Fourth Int. Congr. Actu.* **1**, 214–338. New York.

KIMBALL, A. W. (1954). Short-cut formulas for the exact partition of χ^2 in contingency tables. *Biometrics*, **10**, 452–8.

KLANG, J. *et al.* (1907). *Absterbe-Ordnungen aus Beobachtungen an österreichischen Versicherten.* Vienna.

LOHMÜLLER, A. (1907). *Sterblichkeitsuntersuchungen auf Grund des Materials der Stuttgarter Lebensversicherungsbank, a. G.* 1854–1901. Jena.

MACAULAY, T. B. (1894). The influence of the age at entrance on the force of selection. *Trans. Actu. Soc. Amer.* **3**, 177–220.

*MANLY, H. W. & DEUCHAR, D. (1899). *Combined Experience of Life Annuitants* (1863–1893). *Unadjusted Data.* London.

*MANLY, H. W. & DEUCHAR, D. (1900). *Combined Experience of Assured Lives* (1863–1893). *Unadjusted Data: Endowment-Assurances and Minor Classes of Assurance: Male and Female.* London.

MILLER, B. J. (1880). *Mortuary Experience of The Mutual Benefit Life Insurance Company, Newark, N.J.*, 1845–1879. Newark.

MORGAN, A. (1834). *Tables showing the Total Number of Persons assured in The Equitable Society from its Commencement in September* 1762, *to January* 1, 1829, *etc.* London.

NOLFI, P. (1951). Betrachtungen über konsekutive Verteilungen. *Mitt. Verein. Schweiz. Versich. Mathr.* **51**, 53–62.

ORMODY, W. VON & ALTENBURGER, J. (1910). *Die Sterblichkeit der Ungarischen Versicherten.* Budapest.

PATNAIK, P. B. (1948). The power function of the test for the difference between two proportions in a 2 × 2 table. *Biometrika*, **35**, 157–75.

POISSON, S. D. (1837). *Recherches sur la Probabilité des Jugements en Matière criminelle et en Matière civile.* Paris.

SANDERSON, F. (1895). *Mortality Experience of the Canada Life Assurance Company from* 1847 *to* 1893. Hamilton.

SEAL, H. L. (1943). Tests of a mortality table graduation. *J.I.A.* **71**, 5–67.

SEAL, H. L. (1948). The probability of decrements from a population. A study in discrete random processes. *Skand. Akt.* **31**, 14–45.

SEAL, H. L. (1949). Mortality data and the Binomial probability law. *Skand. Akt.* **32**, 188–216.

SPENS, W. (1861). *Tables of the Mortality Experience of The Scottish Amicable Life Assurance Society from* 1826 *to* 1860. Glasgow.

SPENS, W. (1863). Observations on the Tables of the Mortality Experience of The Scottish Amicable Society. *J.I.A.* **10**, 61–82.

SPRAGUE, T. B. (1870). On the rate of mortality prevailing among assured lives, as influenced by the length of time for which they have been assured. *J.I.A.* **15**, 328–54.

SPRAGUE, T. B. (1884). Report of Mr Sprague, President of the Institute of Actuaries, on the conclusions of Mr Finlaison. *Government Annuities Act*, 1882: *Copy of the Treasury Minute, dated 7th September* 1883. London.

SPRAGUE, T. B. (1896). Maximum mortality percentages. *J.I.A.* **32**, 197–200.

SPRAGUE, T. B. (1904). Mr Moir's paper: maximum mortality periods. *T.F.A.* **1**, 185–6.

WALSH, J. E. (1952). Large sample validity of the binomial distribution for lives with unequal mortality rates. *Skand. Akt.* **35**, 11–15.

Walsh, J. E. (1956). Actuarial validity of the Binomial distribution for large numbers of lives with small mortality probabilities. *Skand. Akt.* **39**, 39–46.

Weeks, R. W. (1896). Mortality experience on annuitants (European and American) in American life insurance companies. *Trans. Actu. Soc. Amer.* **4**, 275–307.

Wells, D. H. (1884). *Mortality Experience of the Connecticut Mutual Life Insurance Company, of Hartford, Conn., from* 1846 *to* 1878. Hartford.

Yastremski, B. *et al.* (1916). *Tables de Mortalité Russes des* 9 *Compagnies d'Assurances.* Petrograd.

Zillmer, A. *et al.* (1883). *Deutsche Sterblichkeits-Tafeln aus den Erfahrungen von dreiundzwanzig Lebensversicherungs-Gesellschaften.* Berlin.

* In the above list of references, Institute and Faculty publications have been indexed under the names of the Presidents at the time, in conformity with the practice of Austro-Hungarian, French, German, Japanese, Russian, and Swedish actuarial associations.

ABSTRACT OF THE DISCUSSION

Dr P. G. Moore, in opening the discussion, said that the theory and practical applications associated with selection went back a long way and formed an accepted part of actuarial science. The Institute's official text-book *The Practice of Life Assurance* stated, in a definitive manner, that 'A select table should normally be used for the calculation of office premiums...'. In attempting to investigate the logic of following that cardinal principle of actuarial practice, there were two main avenues of approach, the financial and the theoretical.

First, in the financial approach it was admitted that selection existed but it was argued that its effect on premium rates was so small as to be negligible. It thus joined countless other features concerning an assured life which were regarded as being too small to be worth worrying about. Secondly, in the theoretical approach, efforts were made to discover whether selection existed or not, however small the actual effect might be, using the most refined statistical tools available.

The author had chosen the second line of approach, but before pursuing that line the speaker wished to spend a few moments on the first line, if only to put the subject of selection into perspective. Since the necessary adjustments were small but were applied so scientifically, and with such trouble, it was clear that they must have been meticulously worked out. The main factors considered in a premium were interest, mortality and age. The last two were, of course, interrelated and age was used merely to determine the point of entry into the mortality table as accurately as possible. The effect of selection on premium rates was illustrated in the tables he had circulated (reproduced below). It would be seen that for endowment assurances selection had little effect. Interest was in that case the dominant factor, selection being roughly equivalent to a change in interest rates ranging from ·01 % at the longer terms to ·1 % at the shorter terms. For whole life assurances selection was more important since mortality formed a larger element in the contract, but even there, over the age-range 25–55, selection was only equivalent to changes in age ranging from 5 to 7 months or in interest ranging from ·1 % to $\frac{1}{2}$ %. In the paper the first three years of selection were, in general, being omitted, and the middle column of the table of whole life rates demonstrated the somewhat artificial case where only the last two years of selection were allowed for. The premium rates showed differences from the ultimate which were minute except at the very advanced ages.

Turning to the consideration of selection from the theoretical angle, he said that the author's method consisted essentially in using the χ^2 test to see whether there were any differences in the mortality rates at different durations. The author had gathered together an impressive array of pre-1914 material to which he applied the χ^2 tests. There were three comments which the speaker desired to make on the form of the χ^2 test used.

First, the χ^2 test took no account of the sign of the deviation, observed minus expected, and hence the groups with excess deaths could occur all at low durations or all at high durations or spread over all durations, and yet give the same value of χ^2 in each case. The effect could be illustrated by a simple example taken from a completely different field. Suppose that in a factory making articles on a large scale a sample of 100 articles was taken from each hour's production. The numbers of defectives found in successive hours were

Endowment assurance annual premiums
'Hypothetical Select': 4-year select period*
$3\frac{1}{2}$ % values in roman type, $2\frac{1}{2}$ % *in italics*

Age	Duration	Select rate	Ultimate rate	% increase of ultimate over select rate
20	45	·01123	·01127	·3
		·01400	*·01403*	*·2*
30	35	·01712	·01718	·4
		·02031	*·02036*	*·3*
40	25	·02854	·02866	·4
		·03218	*·03229*	*·3*
50	15	·05577	·05622	·7
		·05996	*·06039*	*·7*

Whole life by annual premiums throughout life
$O^{[NM]}$ Table: 5-year select period
3 % values in roman type, $3\frac{1}{2}$ % *in italics*

Age	Select rate	Rate with last 2 years' selection only	Ultimate rate	% increase of ultimate over select rate
25	·01656	·01679	·01680	1·4
	·01550	*·01576*	*·01577*	*1·7*
35	·02212	·02245	·02246	1·5
	·02093	*·02127*	*·02129*	*1·7*
45	·03129	·03184	·03187	1·9
	·02997	*·03055*	*·03059*	*2·1*
55	·04697	·04815	·04826	2·7
	·04557	*·04679*	*·04690*	*2·9*

* The source and construction of this table are described by R. E. Beard in his paper to the Centenary Assembly of the Institute of Actuaries (*Proceedings*, **2**, 106).

4, 3, 5, 8, 7, 9. It was desired to test whether the proportion of defectives had remained constant or had changed over the day. A simple χ^2 test, for constant proportion of defectives, would take the expected defectives per hour as 6 (the average number) and lead to a χ^2 of 4·7 with 5 degrees of freedom. That χ^2 was not significant, the corresponding P being slightly over 0·5, but a study of the data indicated a rather different story. That was because the differences 'observed minus expected', although at first always negative, were later always positive. There was something to be gained from considering the order as well as the magnitude of the deviations. In fact the chance that the first three observations were all below expectation and the second three above, or *vice versa*, was only about ·1.

Secondly, there were two stages in the author's X^2 tests, the tests between groups, denoted by X_0^2, and within groups, denoted by X_1^2, X_2^2, X_3^2. On the hypothesis that was being tested, the main differences should occur between groups and that was the test which should be looked at first. If that test was significant, then it would be worth while examining the within-groups criteria X_1^2, etc.; but it seemed difficult to interpret the significance of those unless X_0^2 was itself significant.

Thirdly, in the formula for X^2 on page 173, q_x was defined as

$$\sum_{t=3}^{t_3} y_{[x-t]+t} \Big/ \sum_{t=3}^{t_3} E_{[x-t]+t}$$

and the expression given there for X^2 was algebraically exact. However, in applying the test to the data the q_x used was not as so defined but a q_x based on some previously published rate. The value of q_x used affected the value of X^2 and the corresponding degrees of freedom, and bearing in mind the small differences which it was desired to detect, it was difficult to be certain that X^2 was testing selection exclusively rather than the values of q_x used.

Table A gave some idea of the numerical magnitude of the differences that were being examined. It would seem that, for the whole data, the mortality rates in the three groups considered by the author would, according to the protagonists of selection, probably be in the approximate ratios 94:97:100. In the section entitled 'The chance of detecting mortality differences', the author investigated the power of his X_0^2 test, and it was noted that the order of magnitude of his parameter κ was 0·97. In the 25 assurance experiences of Table 1 there were roughly a quarter of a million deaths, or an average of 10,000 for each experience, and from Table D the probability of detecting such a difference was very small indeed. The value of ΣX_0^2 was significant in 4 (or 16 %) of the experiences instead of the expected $1\frac{1}{4}$. That did not seem to be in any way incompatible with a value of 0·97 for κ. It was, however, rather difficult to interpret some of the within-groups ΣX^2. For example, what was the interpretation of experiences such as number 23, where ΣX_0^2 was not significant but all three within-groups X^2 were significant? That in turn raised doubts as to whether, in data that went as far back at 1808, the individual deaths were independent and not due to epidemics or similar causes that would have the effect of invalidating the binomial form of variation.

Two further statistical points were worth mentioning. First, there was a proof, in the second section of the paper, of the generalized binomial. Most of that algebra could be short-circuited. If E_j represented the jth individual selected at random, then the probability that the individual lived to the next age was $\int_0^1 \pi f(\pi)\, d\pi = \bar{\pi}$ and the probability that he did not live was $1-\bar{\pi}$. The probability that out of E individuals independently selected at random exactly k lived through the ensuing year was

$$\binom{E}{k}\bar{\pi}^k(1-\bar{\pi})^{E-k}$$

and the distribution was binomial with parameters E and $\bar{\pi}$.

The second point referred to Table G. There nine experiences from Table 1 for which ΣX_0^2 was significant were separated out and the individual values of X_0^2 contributing to ΣX_0^2 were examined to see how many were significant. The conclusion reached was that since a very large number of values of X_0^2 were beyond the 10 % point of X^2 with 2 degrees of freedom, namely, 85 instead of the expected 37, there was strong evidence in favour of the duplicates explanation. But since only the values of ΣX_0^2 which were significant had been chosen for examination, it was to be expected that the individual values of X^2 would be more than usually significant. The point could be illustrated by a simple example.

Random samples of size four were drawn from a unit normal population, that is, a normal population with mean zero and standard deviation unity, by means of random sampling numbers. Using the 5 % level ($2\frac{1}{2}$ % at each end) the null hypothesis of zero mean was rejected if the sample mean fell outside the limits $\pm$·98. In the first 200 samples drawn exactly 10 fell outside those limits. Suppose that the 10 significant samples out of the 200 were further examined and each was subdivided into two samples of two—the order used for division being the order of drawing. The 5 % significance limits for the mean of a sample of 2 were $\pm$ 1·385 and of the 20 samples available, no fewer than 9, or 45 %, fell outside the 5 % limits. That demonstrated that the probabilities of individual portions making up the total being significant were altered if the total itself was significant.

It was clear that, on the basis of the tests used and the data available, the hypothesis of no selection for $t \geqslant 3$ could not be rejected. That did not necessarily imply that further data, or other statistical techniques which allowed more of the information to be utilized, would not give a different result. For example, it would be possible to fit to the values of q_x at any age x a regression of the form $q = a+bt$ and then to test whether or not b was equal to zero. Such a procedure would take account of the order of the groups as well as the magnitude of the variations. Until such alternatives had been explored the problem could not be regarded as finally resolved.

Mr M. E. Ogborn described an investigation by Mr R. H. Storr-Best and himself into the methods used by Arthur Morgan in the construction of the earliest Equitable mortality experience. [This is treated more fully in the note on page 300. Eds.]

With regard to the paper, he liked the quotations (which always appealed to him). They suggested that the author regarded himself as a guide rather than as a judge; but from his conclusions he might be regarded as being similar to a guide who conducted visitors round the National Gallery with a new X-ray machine which proved that all the masterpieces were fakes! The definition of selection included spurious selection, so the author undertook the onerous duty of proving that there was no evidence of selection worth talking about, whether genuine or spurious.

There was perhaps a small objection to be made to Table B in that it related to the ages at the time of entry. The figures with the minus signs would be numerically larger if the same groups were considered some years after entry. In relation to the early years of duration that would not matter; but after 15 or 20 years it might.

It was necessary to have regard not merely to the weight of the statistical evidence but also to the worth of the data on which the tests were made. In the case of the Equitable there had been no medical examination. How far the other experiences consisted of medically examined lives was not known. It was interesting that most evidence of selection appeared in the H^M, because that was a most heterogeneous experience. As the author pointed out, it was an experience of 10 English and 10 Scottish offices. Also it was the experience from the inception of the offices; since they began at various dates and grew at various rates, it would have been a very heterogeneous experience. The mortality at durations 40 and over could have had little relation to the data at the shortest durations. Again, circumstances had changed greatly since the period covered by the experience, by reason of the rapid progress in medical science and treatment. In his view selection should be of greater force than in those days for the reason that, while the advances in medicine had meant a lower mortality both for those who were in health and for those who were impaired, it had led to a greater differentiation between the two classes. The impairments of 150 years earlier were rather masked by the high rates of mortality from all causes. Especially since the last war, new methods had been discovered of keeping alive those who would formerly have died; although all might benefit from the lower rate of mortality, there should be a greater differentiation than in the past. He did not suggest that there should be a return to a long period of selection, but rather that it would be worth studying those impaired lives.

Mr E. Jones said it had been a courageous act on the part of the author to embark on a project which had involved such a large amount of research. He would have liked to go all the way with the author's conclusions, because he had always found it difficult to see, on general grounds, why selection should last more than three or four years. But while the author's methods of analysis and conclusions were on the whole convincing, the speaker was left with a big question mark in his mind on the subject of heterogeneity. It had cropped up time and time again in past papers. Higham's paper of 1851, which could be found in volume 1 of the *Journal*, expressly mentioned the fact that in the Seventeen Offices investigation, the experience of the Equitable, which had contributed a good proportion of the data, had been from the outset better than that of the other companies involved. Elderton's paper of 1906 (*J.I.A.* **40**) had shown that the combination of a table of heavy mortality and one of light mortality could exaggerate or understate the effects and duration of selection. Elderton had gone on to show that a good proportion of the selection actually shown by the $O^{[M]}$ table was due precisely to that factor—the amalgamation of 'old' and 'new' assurances, the old ones having had a heavier mortality—and he had come to the conclusion that a select period of seven years would have been more accurate than the 10 years actually adopted. The author had dealt with 'old' and 'new' assurances separately, but it seemed that each of those experiences was likely to have suffered from heterogeneity between offices, just as Higham had found in 1851. The speaker's doubts in that direction had been increased by reading again the reports and discussions on the A 1949–52 table, which had clearly shown how unreliable the information on selection really was. It would be recalled that it had not been possible to investigate more than the first five years and that the figures had shown no indication at all in that period of any tendency to run into an ultimate table. One of the main features of the A 1949–52 table had been the very wide variations in mortality between offices, and the proportions of data provided by heavy and light offices had varied greatly from one duration to another. It did not seem possible to draw reliable conclusions about the effects of selection in that period. It seemed probable that similar heterogeneity due to variations between offices existed in much of the data which the author had analysed. Changes in the overall levels of mortality during the period of experience and in the type of life were other well-known disturbing effects. The average length of the experiences which had been covered by the author was about 40 years, which was quite long enough for those factors to take effect. In statistical terms, it was doubtful whether the condition that $\bar{\pi}$ remained unchanged throughout the sampling was satisfied. He was left with the feeling that the data might have been given rather more sophisticated treatment than was justified. For the same reason he was doubtful whether the explanation of undetected duplicates in the analysis of the $H^{[M]}$ in Appendix I of the paper was the main reason. He was also puzzled by the tenor of the author's remarks (page 175) on the question of looking into variations of signs when dealing with squared deviations. He thought it was always necessary to consider that point, but there was no reference to the question of signs in the analysis of ΣX_2^2, etc. In the first section of the paper, the figures which gave the probability that $s/n > r/m$ were interesting and incontrovertible. They showed that there could be a risk of drawing wrong conclusions if mortality was compared at, say, two different durations at one particular age. But he would have thought that the risk of going wrong would be much smaller when dealing with a complete mortality table because different methods would be used. The figures in Tables C and D were particularly interesting because they showed what a formidable volume of homogeneous data was required to detect a modest degree of selection. That was one of the main reasons why the measurement of selection was still such a knotty problem. He had some reservations about the paper, but it was certainly an interesting and stimulating one which would give much food for thought to anyone who was concerned with analysing select mortality.

Professor G. A. Barnard (a visitor) expressed thanks on behalf of members of the Research section of the Royal Statistical Society for the invitation to be present at the meeting.

He desired to comment on the question of dispensing with algebra in the argument about $\bar{\pi}$ in the second section of the paper. It seemed that the algebra could in fact be dispensed with altogether if it was accepted that the laws of probability alone could not determine for or against the existence of 'second sight'. If two persons were examining samples from an urn, one of whom had second sight while the other had not, the appropriate probabilistic analysis for the former might involve a distribution with varying values of π; while the man without second sight would use the customary formula with $\bar{\pi}$. So long as attention was confined to the frequencies of black balls in the samples, the results of repeated sampling should be the same for both and that was what was expressed in the final formula on page 171.

In the first section of the paper it was interesting to see in another field the paradox which he had first met in connexion with industrial sampling, concerning the skewness of the binomial series. It was first emphasized by Colonel Simon in his book entitled *An Engineer's Manual of Statistical Methods*. He drew attention in that book to the fact that the small sample was better than the lot from which the sample was taken, much more often than it was worse than the lot; and he pointed out the misleading conclusions which might be drawn because of that fact. It was hard to believe that men with such heads for figures as actuaries would be seriously misled by such a phenomenon. The largest bias quoted by the author was ·544 and that occurred only when the expected frequency in the smaller sample was 1. Surely the least statistically-minded person would hesitate to draw conclusions from such a small number of occurrences.

An idea of the amount of that bias in moderate samples could be obtained by noticing that the smaller sample would show the smaller q when it had less than its due share of the deaths available. When the E's were, respectively, m and n, the smaller sample should get a fraction $\lambda = m/(m+n)$ of the total deaths, and the probability that it got less than that would be the probability of getting less than m occurrences in a binomial series $[\lambda+(1-\lambda)]^{m+n}$.

If λ were around $\cdot 1$ the Poisson distribution could be used, and applying the usual χ^2 approximation it was found that the bias was given by the normal probability corresponding to a standard deviate estimated by $\sqrt{(4r)}-\sqrt{(4r-1)}$, where r was the expected number of deaths in the smaller sample. That gave $\cdot 53$ instead of $\cdot 52$ for the case $mq = 10$, $nq = 100$, and $\cdot 51$ instead of $\cdot 506$ for the case $mq = 100$, $nq = 1000$, showing that the rather crude approximation exaggerated the effect. It did, however, suggest a quick way of seeing whether such possible bias was worth bothering about, and it would have been interesting to try it out on the figures analysed.

On the main contention of the paper, he was really out of his depth, because the hypothesis which the author set out to disprove was rather vague. The speaker was not altogether convinced that the partitioning of X^2 which the author adopted was really the more sensitive one. It would be interesting to see the two degrees of freedom in X_0^2 in Table 1 separated into one representing the difference between three and four years and another representing $t > 10$. In fact, he wondered whether that was not a question of estimation rather than a test of significance. It seemed that honour could be done to all parties—to ancient pioneers as well as to modern critics—by taking $q_{[x-t]+t} = q_x[1-(\cdot 2)^t]$. That would square with the figures which the author quoted. It could then be said that selection went on indefinitely up to age 90 and, at the same time, after three years it was so small as to be not worth bothering about.

Mr W. Perks recalled that there was a story which concluded with the lady concerned saying 'If I must ask silly questions I must expect to get silly answers'! As he did not agree with the author's answer, he suspected that the question he was asking in the paper was a silly one.

It was Jeffreys, he thought, who said that the trouble with the χ^2 test was that it asked all the questions at once and provided a single composite answer. That led the speaker to support the opener when he suggested that certain other questions ought to be asked, particularly questions relating to the slope of the regression lines in t.

He had no doubt, from his own observation, that certain mortality experiences showed duration-effects long after three years of duration and some did not. He found it silly, therefore, to ask the general question whether there was such a thing as temporary selection after three years, as if it were a biological constant. The author had referred to American experience and to the practice of analysing mortality for durations up to 15. The American 1946–49 mortality experience of assured lives showed very significant duration-differences all the way up to 15 years' duration. The C.M.I. 1949–52 experience showed significant duration-effects up to and beyond duration five years, after which the data were not analysed. Those effects must be recognized as real; what their causes were was another matter.

He did not suggest that the duration-effects after three years were very important in most life assurance activities, but the meeting was not discussing what was a suitable mortality table for life assurance work. It was discussing the differences in the mortality data. He had used the expression 'duration-effects', because that, in his view, was what the paper was about and not necessarily temporary selection, which was only one of the causes of duration-effects. If in a life assurance experience the duration-effect after three years was in fact nil, and if that experience was extended over a long period of time, during which mortality was improving, giving rise to positive spurious selection, and if there were also lapse selection, it would seem that the residual effect of the medical selection would have to be negative, and that he found quite unacceptable.

Further, evidence for long-duration temporary selection could be obtained from the statistics of substandard lives. It was known that the excess mortality of substandard assured lives went on for many years after entry. Their mortality as a whole was probably something like double the normal. Then there were the very bad lives that did not reach the proposal stage and the declined lives, both of which groups suffered very heavy mortality; it might be that for practical purposes those two groups could be assumed to die out within three or five years. But the plain fact was that substandard lives, subject to extra mortality, survived a much longer period. If from a group of lives there were removed a subgroup which was subject to heavy mortality for many years, the result would be positive selection for the residue. How much selection was another question. [See note on p. 206. Eds.].

For those reasons he supported the previous speaker in suggesting that the real problem before them was one of estimating the amount of the duration effects rather than of testing whether such effects existed.

He had a feeling that what had gone wrong in the paper was that the author had used a steam-hammer to crack a nut, and that the head of the hammer had a large mesh with the result that the nut which he had tried to crack had slipped through the mesh! A large part of the author's work was concerned with testing the difference between q_t and q_{t+1}—a small difference and one difficult to detect. The speaker agreed with Professor Barnard that it would be more appropriate to seek the difference between the mortality in the first group of durations 3 and 4 and the mortality in the last group of durations 10 and over. That led him to suggest that one of the biases in the X^2 test could be seen in ΣX_0^2. That was the test of the between-group differences. The mortality for the middle group, durations 5–9, might be expected to be round about the average for all durations 3 and over. The X^2 for that group should therefore be near its expected value. If, therefore, the three groups were added together, the differences, which would be apparent if the two extreme groups were taken separately, would be obscured.

In his view there were one or two other biases in the use of the X^2 test over the various experiences as well as over ages and durations. There was a combination of a number of small experiences in respect of some of which, with energy, patience and a knowledge of the original language, it would be discovered that nobody had ever expected them to show selection after three years. If a number of experiences of that kind were combined with a few others where selection was expected to last beyond three years, it would be found that the X^2 test would obscure the differences. Another bias would arise because equal weight was given to the large and to the small experiences.

Yet another bias was in using q_x instead of $q_x^{(\kappa)}$ in the divisors of the X^2 for many of the experiences. For example, in ΣX_1^2 for durations 3 and 4, if in fact there was selection so that $q_x^{(\kappa)}$ was 10 % or 15 % smaller than q_x, then in those cases where q_x had been used, the X^2 would be 10 % or 15 % too small, and so the test had a bias.

Many of the age and duration cells had small frequency contents. The test as to whether each particular age and duration should be used was that the expected deaths must be at least 1. If the q for duration 3 was in fact zero at every age, and if all the cells at duration 4 had no more than 2 expected deaths in them, then ΣX_1^2 could not exceed the number of degrees of freedom and it did not matter how many experiences were added together, a significant answer could not be obtained for ΣX_1^2. It was difficult to judge, but it looked as if for ΣX_1^2, that is, for durations 3 and 4, the average number of deaths in each cell was only about 15 over the whole of Table 1 A. If a few of the larger experiences were left out, the average was considerably less than 15, so it would appear that in the process of including a large number of cells of small content a bias was introduced. In fact, although it was in order to include one or two small cells in frequency distribution in the t dimension, he wondered whether it was sound to include a large number of small cells when summing over x and over a number of experiences.

Dr S. Vajda (a visitor) expressed his admiration for the great industry which had gone into the preparation of the paper. The author stated that he had analysed 38 sets of mortality data, and he pointed out that two of them would be expected to be significant at the 5 % level, but he found that in fact there were not two but four. He then stated that it was not an unusual deviation from expectation. It would be interesting to know what the words 'not an unusual deviation' meant. The speaker had made a few calculations and had found that in those circumstances a deviation of 4 or more might be expected with a probability of 12 %. That was, of course, more than 5 %, and it could be said that 4 was not significant; but it was not certain whether the significance of the second level should be tested at the same level as the original one. In any case, a specific figure of 12 % revealed more than the words 'usual' or 'unusual'.

At the foot of page 168 the author wrote: '...the occurrence of 59 plus signs in such a series would thus be considered significant if an excess of pluses was expected *a priori*'. The words *a priori* probably meant, in that connexion, under an alternative hypothesis. It was perhaps dangerous to call something *a priori* which was not in fact assumed to be *a priori*.

To turn to something more fundamental, he had enjoyed the privilege of being able to discuss matters with the author, and it was not surprising that he agreed with almost everything that the author said. However, one point remained on which he disagreed, and that was on page 175 where the author dealt with the question of elementary error. In the penultimate paragraph there were the words: '...it is, of course, an elementary error in statistical analysis to observe a "trend" and then to "test" it'. The speaker did not consider that that was an elementary error; it was, in his view, a natural thing to do. A series of figures might appear to show an increasing trend; if somebody said that they did not, and that it was simply due to chance, the thing to do would be to sit down and put the matter to the test. The observer expected to find that it was true but was prepared to be told by test that in fact it was not so. It was not suggested that that confirmed or contradicted the remarks which had already been made by the opener about that paragraph. The particular question was whether it was legitimate to look at something, suspect that something had happened and then apply a statistical test. He was not convinced that it was an elementary error.

Finally, it would be of interest to learn whether the author considered that what the speaker was about to say was nonsense, because if it were not, then the paper, in his view, was not of very much substance. He referred to the paragraph immediately below Table D. Did not that paragraph mean that the author had used a test which was not very sensitive? The speaker submitted, as a statistician, that there was possibly still a difference of 5 % but that the tests which had been applied could not detect it. He would leave it to the actuaries to decide whether such a difference should or should not be accepted as evidence for the existence of temporary selection.

Mr K. A. C. Wheeler was particularly interested in what Mr Ogborn had said with regard to the change in selection, if it existed, resulting from the improvements in medical treatment. He would have thought that with the continuous development of medical science selection would become more and more diffused and possibly spread over longer and longer terms. He realized that that was precisely the opposite to what Mr Ogborn had suggested, and perhaps he could be enlightened on the matter. Possibly the answer was that different forms of improvement would produce different effects on selection, but nevertheless it seemed, *a priori*, that selection would change in form and intensity as the possible causes varied.

Mr R. E. Beard, in closing the discussion, said that the actuarial history of temporary selection was built around three main topics. First, the practical needs of the actuary in the recognition and treatment of options; secondly, the practical aspects of devising mortality tables exhibiting selection and, thirdly, the theoretical formulation of the processes which led to the results obtained from statistical studies. Those three aspects were fundamentally different, and unless they were kept clearly in mind in discussing the subject, there was grave danger of developing pointless arguments.

The selection 'built in' to a mortality table might, or might not, reflect closely the variations shown in the crude data. Unless the philosophy underlying the construction of the table was known, it was dangerous to draw conclusions from the behaviour of the numerical values of the mortality table. Equally so the use of such a table in regard to option calculations could be misleading. On the theoretical aspects it was worth mentioning that even after some 200 years of collation of mortality statistics there was no real theory of mortality in the sense of a scientific law. The laws which had been propounded were essentially mathematical formulae for describing the results. Nevertheless, the three aspects were related and the paper provided an opportunity for linking them up. He proposed to return to that aspect later.

The author offered two quotations from earlier generations of actuaries and a third from a philosopher in support of his scientific approach to his subject. The last would serve as a reminder that the meeting was concerned with the facts behind the statistics and not the practical instruments which might be created from them. Had he wished, the author could have found many quotations echoing the thought that temporary selection lasted many years; few could be found supporting the opposite view.

That actuaries were troubled by the interpretation of statistics of mortality in select form was apparent from the discussions which took place at both the Institute and the Faculty on the A 1949–52 table. The Committee in its reply to the discussions stated that 'the length of time over which temporary initial selection persists is still a question on which the profession has little information'. The form of available data, the secular improvement in mortality, the changes in proportions of various classes of business and other factors all contributed to making the interpretation of statistical features a matter of considerable difficulty.

There were a number of clearly defined reasons why mortality data would exhibit the feature of temporary selection. First, the initial selection by which certain lives were eliminated from a population group (with a converse effect on disabled lives' experience); secondly, a continuing effect arising from selective withdrawals; thirdly, real effects arising from secular variations and, fourthly, spurious effects arising from the way in which the data were compiled. In the absence of adequate statistical analysis (and of mathematical theory) much confusion and argument had arisen from an inability to sort out and measure the true effects of those various components.

Recognizing in that lack of precision the scope for the application of modern techniques, the author had engaged on a monumental study which was a model of the way in which such techniques could be used. He had set up his hypothesis, designed experiments to test the hypothesis, specified his tests and interpreted the results. In that context the author's comments on page 169 were pertinent. The speaker would certainly be prepared to argue with him on the final sentence of the second paragraph on that page, since the question was surely that the tests failed to identify selection, not that it did not exist. But a careful rereading of the paper showed that the author was well aware of the point.

Notwithstanding the difficulties of interpretation of the observed statistics and the lack of a complete understanding of the components of selection, modern mortality tables had recognized the impracticability of long periods of selection. In some ways the paper might be looked upon as providing justification for the practical decision taken.

In searching for data upon which to test his hypothesis, the author immediately ran into difficulty because of the systematic bias introduced into parameters derived from grouped data by the distribution of business by ages and durations; hence no modern data were found. His hypothesis was based on an equality of mortality rates derived from varying amounts of data; by showing that the probability distribution of deaths at any age when lives were the unit of observation might be taken as binomial in form, with a slight safety margin, he was immediately able to use standard techniques for his tests, which were essentially tests of comparability of mean values. For most of his work he used tests based upon mean squared deviations corresponding to χ^2 tests, but in certain critical cases he used tests based upon mean deviations to avoid the danger of losing the significance of runs of similar signs. The speaker referred to his own use of mean deviation tests in connexion with graduation of mortality data (*J.I.A.* 77, 382).

With regard to the second type of error referred to on page 168, it might be pointed out that the lack of balance of numbers of positive and negative deviations arose when the distribution of deviations about the mean was skew in form. The greater number of positive deviations was balanced by a smaller number of negative deviations of large average amount (or *vice versa*), a warning that a sign test of deviations could be misleading if skew variation were involved.

From his carefully prescribed tests, the author concluded that his hypothesis was supported by all the data he could find which were in a suitable form to enable him to test his hypothesis. He could, of course, have developed his tests for application to data in which duplicates were known to exist, but the sharpness would have been greatly reduced because of the lack of real knowledge of the proportion and distribution of duplicates. A note in the *Journal* (**83**, 34) gave some information about the distribution of duplicates, but it was clear from the author's efficiency tests (Tables C and D) that there would be no point in developing such tests.

If the author's hypothesis were accepted, it was important to spend a few moments looking at the wider picture, because some of the uncertainties which were voiced in the Institute at the time of the A 1949–52 tables could be resolved thereby.

The essential features of the data were that (i) the select rates from the medical section seemed to reach a plateau at duration 3; (ii) the non-medical rates reached a plateau at duration 2, and (iii) in both sections there was a substantial gap to be spanned from the plateau to the ultimate mortality level. Those features were consistent with (*a*) a short duration of temporary initial selection, longer for medical than for non-medical data, and (*b*) the combined effect of many other factors which caused the mortality rates to increase slowly with duration, the latter including the bias arising from grouped data and the spurious selection arising from the distribution of business. Those concerned might be satisfied to accept the fact that temporary selection wore off in a few years, but a substantial area was still left where the composition of the mortality was unknown. If it were assumed that the factors other than initial temporary selection were not significant until after duration 3 or 4, then it would be reasonable to conclude that the true ultimate mortality level was somewhere near that shown at duration 3 or 4, and that the observed ultimate mortality at duration 5 and over arose from the many other factors. Since the difference between the two was of the order of 15% for the combined data, there was considerable scope for further research.

Among the suggestions put forward for explaining the long period of selection had been that of selective lapsing, in which the proportion of damaged lives tended to increase with duration. If that were the reason for most of the 'gap', then it could be concluded that the true ultimate mortality for a randomly selected group would be about that at duration 4, while the observed ultimate mortality was applicable to a group subject to selective decrements. The significance for the calculation of options was obvious.

With regard to the other aspects of selection, namely, the consideration of scientific or mathematical formulation of laws of mortality, there had been various attempts in that field, and he would refer only to those arising from the 'damaged lives' concept, originally due to Dr Sprague. Bulina, in a paper to the Tenth International Congress of Actuaries, showed that the reversed selection appearing in a disabled lives' table could be interpreted as due to a proportion of bad lives subjected to very heavy mortality rates which were independent of age. Perks pointed out in the discussion on the same paper that a similar concept could be appropriately applied to assured lives' select mortality. Anderson in his 1936 paper (*J.I.A.* **68**, 223) put forward a theory of select tables in which the damaged lives were subject to a mortality of about ·57, a value of the same order as that implied by Bulina's (and Perks's) work. Anderson showed that the difference between $q_{[x-4]+4}$ and q_x at age 60 was only about 4%, a difference which Dr Seal showed would probably be undetectable from observed data. Finally, he referred to the note by Carpmael in the *Journal* (**82**, 123) in which concepts similar to those of Anderson were applied to a Table following a logistic—or Perks—curve. It would be an interesting exercise to set up mathematical models based on those concepts and to develop statistical tests based on the determination of the parameters of the model from observed data. It was, however, an interesting line of thought that the duration of temporary selection derived from those 'damaged lives' models, whilst in theory limited only by the duration of lives, was probably undetectable beyond duration 4.

He referred also to a paper by Levinson entitled *A Theory of Mortality Classes* which was being submitted to the Society of Actuaries at its next meeting. It would not be right to discuss the paper prior to its presentation, but the underlying theory was pertinent to the discussion, particularly the author's comments on the duration of temporary initial selection and the influence of selective withdrawals.

It was clear from the discussion that there was plenty of scope for experimental work, and while the author had provided reasons for considering that temporary initial selection was of short duration, the speaker was of the opinion that there was a great deal which had not been explained. The author had stimulated a discussion which had been very interesting, and there was no doubt that future generations of students would find in the subject a large area of work still to be done, a number of critical questions still to be answered, and much scope for exercising their ingenuity in finding ways of answering those questions.

The President (Mr F. M. Redington), in proposing a vote of thanks to the author, said members would all be greatly indebted to him for a noteworthy paper which had produced a lively discussion. From the quotations at the beginning to the Appendices, each line of the paper had been marked with the author's scholarship. It was another piece of evidence in a detective story which had fascinated actuaries for almost a century. Since he (the President) had been a member of the C.M.I. Committee which had been responsible for the A 1949–52 tables, he had prepared a few remarks in case the discussion had flagged. But his points had already been made. He would confirm, however, that the C.M.I. Committee had had no difficulty in detecting a durational effect. The problem had been how to dispose of it. Roughly, mortality appeared to increase by 3 % each year after duration 2. How far that was due to true initial selection was debatable, and the answer was not known. The paper had provided more evidence and was more interesting than had perhaps been apparent, for the experience of other countries provided a helpful background.

He would like to think that the author's interest remained with activities in Britain, and that after teaching zoology students at Yale University he would still have time to continue studying the zoology of British policyholders. The meeting was a rather special occasion, for it was not often that one of the Institute's own offspring returned home to deliver a paper. To many members, perhaps, the paper had arrived with something like a supersonic bang; otherwise the audience would have been bigger. Those who were present were all delighted to see Dr Seal, and thanked him most warmly.

Mr W. Perks sent the following note in amplification of his remarks at page 200.

A simple arithmetical illustration may be useful. Taking the A 1949–52 ultimate table and starting at age 40, let us suppose that three groups of lives are abstracted: (i) ·001 l_{40} subject to 50 % per annum mortality (the 'death-bed' cases that do not propose for life assurance), (ii) ·005 l_{40} subject to 5 % per annum mortality (the declined lives) and (iii) ·05 l_{40} subject to double the A 1949–52 ultimate mortality (the lives accepted at substandard rates). The percentage ratios of the mortality rates of the residue (the lives accepted at ordinary rates) to the A 1949–52 ultimate rates at ages 40–49 (durations 0–9) work out as follows: 53, 70, 79, 84, 87, 88, 90, 91, 91, 92. The general course of these figures is not unlike the select figures in the 1949–52 data. The assumption about the mortality of the substandard lives would mean that it would be many years before the gap between 92% at duration 9 and the ultimate 100% would be bridged.

If it is further assumed that 10% of the adjusted l_{41}, 5% of the adjusted l_{42} and $2\frac{1}{2}$% of the adjusted $l_{43}, l_{44}, \ldots, l_{49}$ withdraw and that the withdrawing lives, from the age of withdrawal, are subject to 'select' mortality represented by the above percentages of the A 1949–52 ultimate mortality, then the remaining lives would show mortality percentages as follows: 53, 72, 82, 86, 89, 91, 92, 94, 95, 96. Clearly, this series will pass and remain beyond 100 at the later durations.

Mr H. A. R. Barnett submitted the following written contribution to the discussion.

My remarks will be directed against the middle of page 175, where the author states that he has obtained an overall picture by adding the values of X^2 over all the ages included in the computations.

To my mind he is too fond of overall pictures; evidence of this is given by his partiality for the χ^2 test which he described in his 1940 paper (*J.I.A.* **71**, 5) which I once criticized as 'too widely read'. Professor Jeffreys's views on the χ^2 test, that it tries to answer all the questions at once and in doing so may give a confused answer, apply equally to the author's approach in the present paper.

Several actuaries have examined human mortality with a feeling that the approach should be that the population at each age includes a proportion of lives already predisposed to death from certain causes. I believe that both Ogborn and R. D. Clarke followed this idea, and it was the basis of some recent research of mine; I should mention also a paper by L. Levinson (referred to by Mr Beard in the discussion). Now, what the author says in effect is that these predispositions can never be detectable at any age more than three years before death—for if they were so detectable, the lives would be rejected by temporary selection which would last for as long as those predisposed lives survive. Put in this way, I doubt whether he could obtain medical evidence to support that view.

It may well be that at the younger ages, where such detectable predispositions are probably proportionately rare, the proportion remaining alive but predisposed for more than three years is insignificant, but I doubt whether that would also be the case at the older ages.

In a recent paper (*J.I.A.* **81**, 105) I examined the select data of the experience upon which the A 1949–52 table was based. It may be recalled that I divided mortality into a geometric or 'Gompertz' curve plus a 'normal' curve. I concede that the extent to which the Gompertz curve differs between duration 2 and later durations may well be due to spurious selection; but spurious selection cannot explain away the behaviour of the normal curve, which is non-existent at duration 0, and has emerged as a little hump in the fifties of life at duration 1, and as a slightly larger hump a little later in life at durations 2–4, nowhere near the size of the normal curve at durations 5 and over. (I should mention here that durations 2–4 were combined only to save space, and that the data had not been collected in a form to permit further subdivision of the '5 and over' figures.)

These results were strongly suggestive that at the higher ages temporary selection rejects some predisposed cases where there are what the author calls 'symptoms of death' *more* than three years before death, and that therefore at these ages temporary selection remains beyond three and possibly even five years.

Unfortunately, he has begged the whole question by combining the ages, and obtaining one answer where there should possibly be two or more different answers. The size of the data at the younger ages, and the short duration of temporary selection at those ages, could well have swamped what evidence there might have been of more persistent temporary selection at the higher ages.

The author justifies himself on the grounds that published mortality tables have uniform select periods not varying with age. That argument fallaciously assumes that what has been done for purely practical reasons can point the way to theory. These practical reasons should show that the author could hardly expect anyone to construct a standard table with varying select periods—but this does not mean that the select period is, in fact, constant from age to age. Any research of this nature should take account of the possible age variation, even if it is not envisaged that a varying select period will ever be used in practice.

I would therefore urge the author to have his arithmetic reworked, dividing his assured lives' data between, perhaps, ages under 50 and ages 50 and over. This would only mean obtaining two totals instead of one for each experience, the basic arithmetic having already been performed, and should not involve much extra work. To my mind, without such subdivision the paper—for all the effort, algebra, arithmetic, and talk involved—is not of great value.

I would have thought it worth while, too, to include some up-to-date data.

Dr H. L. Seal, after acknowledging the vote of thanks, said that he would prefer to reply to the discussion in writing, and he subsequently sent the following comments.

It is important to remember that there is no such thing as a 'theory' of temporary selection (or of 'duration effects' in Mr Perks's terminology) in the sense that one can forecast, in any given experience and at any age x, the duration t at which $q_{[x-t]+t}$ will level-off or commence to decrease. All that one can do is to 'explain' why this levelling-off or decreasing tendency occurs at about the values of t observed from an analysis of the actual data. That this possibility of a decreasing series of values of $q_{[x-t]+t}$ must be taken seriously is shown by Table 2 of the paper which indicates that, in five of the first six sets of data thoroughly examined for long-term temporary selection, decreasing values of $q_{[x-t]+t}$ were found to occur after values of t extending from nine to thirty years—in spite of the use of quinary or other groupings of x-values which would inhibit such a result (see Table B).

The speakers in the discussion who argued in favour of a monotonically increasing mathematical form for $q_{[x-t]+t}$ (and thus for the estimation of its parameters) were *ignoring the possibility that* $q_x^{(3)}$ *(in the notation of the paper)* might well be smaller than $q_x^{(2)}$ because of a decrease in $q_{[x-t]+t}$ occurring at the later durations. It was, in fact, these considerations that led me to test the hypothesis of equality of $q_{[x-t]+t}$ for all values of $t \geqslant 4$ against the alternative that upward or downward trends were equally reasonable. It would, of course, have been possible to test the null hypothesis of equality against the alternative of a monotonic increase in $q_{[x-t]+t}$ throughout the range of t, the appropriate test being the extraction of a single degree of freedom from X_0^2 at each age x (see, for example, D. R. Cox (1958), *J. Roy. Statist. Soc.* B, **20**, 215–42). But such a procedure would presumably have been doomed to failure at least in those experiences (not limited to the five mentioned) where a downward trend developed at the later durations.

This may be expressed another way by saying that the actuary faced with a fresh set of mortality data is quite unable to decide whether the initial increase in $q_{[x-t]+t}$ continues through the higher t-values, levels off or changes to a decrease. Thus the most logical procedure is surely to test against the general alternative adopted in the paper. On the other hand, the use of the observations to guide the test (as suggested by Dr Vajda in his interesting contribution to the discussion) would require the development of a conditional, instead of an absolute, probability distribution for the statistical criterion adopted—a fact that I failed to notice in my Table G, so neatly demolished by Dr Moore.

A further fact to bear in mind concerns the statistical tests used in the paper. At any one age x the criteria used to test the acceptability of the null hypothesis of equality of the $q_{[x-t]+t}$ have been shown to be more powerful in the detection of non-equality than any others. However, it is admitted that the method adopted for summarizing (*a*) the individual ages of any mortality table, and (*b*) the results of the first 25, next 13, etc., tables, namely summation of the corresponding X^2 values, has not the same authority. Nevertheless, it is a reasonable (and convenient!) test and the 'weighting' proposal of Mr Perks would scarcely affect the results of Table 1, since the largest experiences are, for the most part, rather 'well-behaved'. We must thus conclude that the fact that a set of mortality data must comprise 100,000 deaths for there to be a 50 % chance of discovering whether the mean mortality at $t = 3$ and 4 is as much as 5 % greater or smaller than that at $t \geq 5$ (Table D) is inherent in the data themselves. However, if one has *a priori* reasons for thinking that the trend can only be upward, the number of deaths required falls to 4000 (Table C). Anyone who thinks he can improve on these results by suitable 'grouping' or otherwise is deluding himself, just as many of the authors cited in Tables 1 and 2 did.

A final word of caution to those who are convinced that, in a given set of data, secular improvements and class changes cannot have reversed the monotonic-upward trend in $q_{[x-t]+t}$. There is a small upward age-drift in x as t increases even when single ages are used. As Mr Ogborn hinted in the discussion, if lapses are heavy this upward drift will be accentuated. Furthermore, most of the X and X^2 values calculated in Table 1 are numerically overstated for the reasons given on pp. 171–73.

In conclusion, I would like to reply to some of the individual points raised in the discussion:

(1) It was not, perhaps, made clear that the 'previously published' values of q_x were used only in deciding which t-values to include, and not in the computations of any of the X or X^2 values.

(2) The heterogeneity mentioned by Mr Jones does not appear to disturb the tests of $\bar{\pi}_{[x-t]+t}$ against $\bar{\pi}_{[x-t-1]+t+1}$, etc.

(3) Jeffreys's criticisms of the χ^2 test of goodness-of-fit (pp. 91–2, *Theory of Probability*, Oxford, 1948) have no relevance to the X^2 criteria of Table 1 —which, by a coincidence, obey the same distribution law. In fact, at any one age x, Jeffreys himself would use tests very similar to those employed in this paper (pp. 232 *et seq.*, *loc. cit.*).

(4) The claims that recent British and American life-office data show strong selective effects up to high t-values need more evidence than that published in the form of q-values at quinquennially grouped x-values (see Table B and note the effects of heavy lapses and the implications of a gradually increasing skewness on account of a decreasing exposure of 'policies' or 'sums insured' at the larger t-values).

(5) I did not overlook the possibility of testing $q_{[x-5]+5}$ against $q_{[x-15]+15}$, for example, but felt that the age-drift and the 'down-turn' found by so many authors would prejudice the results.

(6) The good behaviour of 1743 values of X_1', in Table FF of Appendix II, indicates that the fears of Mr Perks about small y-values are groundless.

(7) In reply to Mr Beard, if the best available tests fail to identify non-chance variations in $q_{[x-t]+t}$, is it not presumptuous to say: *e pur si muove*?

(8) As to 'predisposition for death,' Dr Cuyler Hammond has told me that out of 2809 deaths registered at ages over 20 in New Haven, Conn., in 1951–2 only 22·3 % were stated to have died from causes foreseen by their doctors more than four years earlier; it would be interesting to speculate how many of these 22·3 % would have been accepted for assurance within four years of death.