JIA 96 (1970) 251-264

RANDOM MORTALITY FLUCTUATIONS AND THE BINOMIAL HYPOTHESIS

By A. H. POLLARD, M.Sc., M.Sc. (Econ.), Ph.D., F.I.A.

1. INTRODUCTION

A great deal of research has been carried out into expected values of q_x or into expected number of deaths. Little attention has however been paid to the random variations in mortality rates or to the random variations in the number of deaths. Research in this direction might very well further our knowledge of the mortality process.

The mortality of a country in a given year may be thought of as a sample from some hypothetical large population. If the population mortality rate q_x at age x were constant, the *observed* rate of mortality of such a sample will vary randomly about q_x. Similarly the *observed* number of deaths, with constant exposed to risk, will also show these random variations.

There are, however, sound practical reasons to suspect that, even if there are no changes in the level of mortality, the population value of q_x may not be constant but may itself be subject to random fluctuations. The population value of q_x may be considered as the weighted sum of the rates of mortality of groups of persons suffering from particular disabilities. If there is any variation in the proportion suffering, for example, from particular heart conditions or if there is any variation in the degree of such impairments then variations in the *population* value of q_x must be expected in addition to the random variations which occur in the *observed* rate of mortality when q_x is constant. Also, total mortality at any age is the sum of mortality due to various causes. For certain causes of death, for example diseases of the respiratory system, bronchitis, etc. the population value of q_x might be expected to vary with climate. Random variations might be expected in the population rate of mortality from infectious diseases. With assured lives data variations may occur from year to year in the proportion of female lives. These are but examples of the many reasons which could be put forward to justify the expectation that random variation in the population value of q_x rather than constant q_x is the usual situation.

We shall first derive formulae for the variance of the actual number of deaths under various mortality hypotheses and then attempt to measure the variation which has actually occurred by analysing certain recent Australian data.

2. VARIATION UNDER DIFFERENT MORTALITY HYPOTHESES

Hypothesis 1. That all persons of a given age have the same *constant* chance of dying Q. This is the simple binomial hypothesis and the variance of the number of deaths D is given by

$$\text{Var.}(D) = NPQ \tag{1}$$

where

$$P = 1 - Q$$

It is, however, an accepted fact of life that all persons of a given age do not have the same chance of dying within a year. Some, because of their particular disabilities, have a higher chance of dying than others. Hence we are led to:

Hypothesis 2. That the number N of persons aged x is made up of r sub-groups, the ith sub-group consisting of n_i persons each having a chance q_i of dying within a year where $N = \Sigma n_i$ and the numbers n_i and chances q are constant.

The total deaths D are given by

$$D = d_1 + d_2 + \ldots + d_r$$

where $d_1, d_2 \ldots d_r$ are assumed to be independent.

$$\mathrm{E}(D) = \Sigma \mathrm{E}(d_i) = \Sigma n_i q_i = NQ$$

where

$$Q = \frac{1}{N} \Sigma n_i \, q_i$$

is the average rate of mortality.

$$\begin{aligned}
\mathrm{Var.}(D) &= \Sigma \mathrm{Var.}(d_i) \\
&= \Sigma n_i \, q_i \, (1 - q_i) \\
&= NQ - \Sigma n_i \, q_i^2 \\
&= NQ - \Sigma n_i \, (q_i - Q)^2 + Q^2 \Sigma n_i - 2Q \Sigma n_i \, q_i \\
&= NQ - \Sigma n_i \, (q_i - Q)^2 - NQ^2 \\
&= NPQ - \Sigma n_i \, (q_i - Q)^2 \qquad (2)
\end{aligned}$$

Thus we obtain the perhaps unexpected result that heterogeneity in mortality *lowers* the variance in the number of deaths. This result, however, appears less unreasonable when we consider the limiting case in which some of the population are certain to die ($q = 1$) and the balance are certain to live ($q = 0$). In this limiting case the outcome is certain and the variance of the total number of deaths is zero. In view of the known heterogeneity of mortality we might well expect the variance of the number of deaths to be less than the value NPQ resulting from the binomial hypothesis. We shall consider this again later in this section.

Hypothesis 3. That all persons of a given age have the same chance q of dying but that q itself is a random variable which varies stochastically with mean Q and variance equal to Var.(q).

Let the number exposed be N, the number of deaths be D and the frequency function of q be $f(q)$.

$$\mathrm{E}(D|q) = Nq$$

$$\text{Hence } \mathrm{E}(D) = \int_0^1 N \, q \, f(q) \, dq = NQ$$

Since, with the usual notation, $\mu_2 = \mu_2' - \mu_1'^2$

$$\begin{aligned} \mathrm{E}(D^2|q) &= Npq + N^2 q^2 \\ &= N(N-1)q^2 + Nq \end{aligned}$$

$$\text{Hence } \mathrm{E}(D^2) = \int_0^1 [N(N-1)q^2 + Nq] f(q)\, dq$$

$$= N(N-1)\,[\mathrm{Var.}(q) + Q^2] + NQ$$

$$\begin{aligned} \text{Therefore } \mathrm{Var.}(D) &= \mathrm{E}(D^2) - \mathrm{E}^2(D) \\ &= NPQ + N(N-1)\,\mathrm{Var.}(q) \end{aligned} \tag{3}$$

Thus, if there is random variation in q, considerable random variations in D can occur. A numerical illustration of this has been given by J. H. Pollard.[1]

Hypothesis 4. (This is a combination of Hypothesis 2 and Hypothesis 3.) That the number N of persons aged x is made up of r sub-groups, the ith sub-group consisting of n_i persons each having a chance q_i of dying, where $N = \Sigma n_i$ and the numbers n_i are constant but each q_i is a random variable varying stochastically with mean $\bar{q}_i$ and variance $\mathrm{Var.}(q_i)$.

As before

$$D = d_1 + d_2 + \ldots + d_i$$

$$\mathrm{E}(D) = \Sigma \mathrm{E}(d_i) = \Sigma n_i\, \bar{q}_i = NQ$$

where

$$Q = \frac{1}{N} \Sigma n_i\, \bar{q}_i$$

$$\begin{aligned} \mathrm{Var.}(D) &= \Sigma \mathrm{Var.}(d_i) \\ &= \Sigma n_i\, \bar{p}_i\, \bar{q}_i + \Sigma n_i\, (n_i - 1)\,\mathrm{Var.}\,(q_i) \text{ from (3)} \\ &= NPQ - \Sigma n_i\, (\bar{q}_i - Q)^2 + \Sigma n_i\, (n_i - 1)\,\mathrm{Var.}(q_i) \end{aligned} \tag{4}$$

following the same algebraic steps which led to (2).

Thus heterogeneity in mortality reduces variation in the number of deaths but random variation in q increases this variation.

Hypothesis 5. That the population aged x consists of r sub-classes, the ith sub-class consisting of n_i persons all with a chance q_i of dying within a year, where the q_i are constant but the n_i are random variables varying stochastically with mean $\bar{n}_i$ and variance $\mathrm{Var.}(n_i)$.

Let d = the actual number of deaths in the ith sub-class = $n\,q$
n and q being the *experienced* numbers and mortality rate.

$$\begin{aligned} \text{Then, } \mathrm{E}(d) &= \mathrm{E}(n\,q) \\ &= \bar{n}_i\, q_i \text{ since } n \text{ and } q \text{ are independent} \\ \mathrm{E}(d^2) &= \mathrm{E}(n^2\, q^2) \\ &= (\bar{n}_i + \mathrm{Var.}(n_i))\,(q_i^2 + \mathrm{Var.}(q)) \\ &= (\bar{n}_i^2 + \mathrm{Var.}(n_i)) \left(q_i^2 + \frac{p_i\, q_i}{\bar{n}_i} \right) \end{aligned}$$

Therefore,

$$\text{Var.}(d) = \bar{n}_i\, p_i\, q_i + \left(q_i^2 + \frac{p_i\, q_i}{\bar{n}_i}\right) \text{Var.}(n_i)$$

Assuming the deaths are independent (an assumption about which there is some doubt if $\Sigma n_i = N$ where N is fixed)

$$\text{Var.}(D) = \Sigma \bar{n}_i\, p_i\, q_i + \Sigma\left(q_i^2 + \frac{p_i\, q_i}{\bar{n}_i}\right)\text{Var.}(n_i)$$

which from (2)

$$= NPQ - \Sigma \bar{n}_i\,(q_i - Q)^2 + \Sigma\left(q_i^2 + \frac{p_i\, q_i}{\bar{n}_i}\right)\text{Var.}(n_i) \tag{5}$$

Hypothesis 6. That the population consists of r sub-classes, the ith sub-class consisting of n_i persons all of whom have a chance q_i of dying within a year, both n_i and q_i being random variables varying independently with means $\bar{n}_i$ and $\bar{q}_i$ and variances $\text{Var.}(n)_i$ and $\text{Var.}(q_i)$ respectively.
Here, for the ith sub-class

$$\text{E}(d|q_i) = \text{E}(n\, q|q_i) = \bar{n}_i\, q_i$$

Therefore, $\text{E}(d) = \int_0^1 \bar{n}_i\, q_i\, f(q_i)\, dq_i$

where $f(q_i)$ is the frequency function of q_i

Hence

$$\text{E}(d) = \bar{n}_i\, \bar{q}_i$$

Now

$$\begin{aligned}\text{E}(d^2|q_i) &= \text{E}(n^2\, q^2|q_i)\\ &= (\bar{n}_i^2 + \text{Var.}(n_i))(q_i^2 + \text{Var.}(q))\\ &= (\bar{n}_i^2 + \text{Var.}(n_i))\left(q_i^2 + \frac{p_i\, q_i}{\bar{n}_i}\right)\end{aligned}$$

Therefore,

$$\begin{aligned}\text{E}(d^2) &= \int_0^1 (\bar{n}_i^2 + \text{Var.}(n_i))\left(q_i^2 + \frac{q_i - q_i^2}{\bar{n}_i}\right) f(q_i)\, dq_i\\ &= (\bar{n}_i^2 + \text{Var.}(n_i)) \int_0^1 \left(\frac{q_i}{\bar{n}_i} + \frac{\bar{n}_i - 1}{\bar{n}_i}\, q_i^2\right) f(q_i)\, dq_i\\ &= (\bar{n}_i^2 + \text{Var.}(\text{n}_i))\left(\frac{\bar{q}_i}{\bar{n}_i} + \frac{\bar{n}_i - 1}{\bar{n}_i}(\bar{q}_i^2 + \text{Var.}(q_i))\right)\end{aligned}$$

$$= (\bar{n}_i^2 + \text{Var.}(n_i))\left(\frac{\bar{p}_i\bar{q}_i}{\bar{n}_i} + \bar{q}_i^2 + \frac{\bar{n}_i - 1}{\bar{n}_i}\text{Var.}(q_i)\right)$$

Therefore,

$$\text{Var.}d = \bar{n}_i\bar{p}_i\bar{q}_i + \bar{n}_i(\bar{n}_i - 1)\,\text{Var.}(q_i) + \left(\bar{q}_i^2 + \frac{\bar{p}_i\bar{q}_i}{\bar{n}_i}\right)\text{Var.}(n_i) + \frac{\bar{n}_i - 1}{\bar{n}_i}\text{Var.}(n_i)\,\text{Var.}(q_i)$$

and hence from (2), and assuming the d's to be independent:

$$\text{Var.}(D) = NPQ - \Sigma\bar{n}_i(\bar{q}_i - Q)^2 + \Sigma\left(\bar{q}_i^2 + \frac{\bar{p}_i\bar{q}_i}{\bar{n}_i}\right)\text{Var.}(n_i) + \Sigma\bar{n}_i(\bar{n}_i - 1)\,\text{Var.}(q_i) + \Sigma\frac{\bar{n}_i - 1}{\bar{n}_i}\text{Var.}(n_i)\,\text{Var.}(q_i) \qquad (6)$$

This formula is a generalization of formulae (1) to (5) inclusive and they can each be obtained from (6).

Thus, if $i = 1$ and n_i and q_i are fixed, Var.(n_i) and Var.(q_i) are zero and $\bar{q}_i = Q$. Hence we have the binomial where

$$\text{Var.}(D) = NPQ \qquad (1)$$

If we have r classes in which n_i and q_i are constant, Var.(n_i) and Var.(q_i) are zero and we have

$$\text{Var.}(D) = NPQ - \Sigma n_i\,(q_i - Q)^2 \qquad (2)$$

If $i = 1$, and $n_i = N$ is constant, $q_i = q$ and $\bar{q}_i = Q$, Var.$(n_i) = 0$ and we have

$$\text{Var.}(D) = NPQ + N(N-1)\,\text{Var.}(q) \qquad (3)$$

If the n_i are constant Var. $(n_i) = 0$ and we have

$$\text{Var.}(\text{D}) = NPQ - \Sigma n_i\,(\bar{q}_i - Q)^2 + \Sigma n_i\,(n_i - 1)\,\text{Var.}\,(q_i) \qquad (4)$$

If the q_i are constant Var. $(q_i) = 0$ and we have

$$\text{Var.}(D) = NPQ - \Sigma\bar{n}_i(q_i - Q)^2 + \Sigma\left(q_i^2 + \frac{p_iq_i}{\bar{n}_i}\right)\text{Var.}(n_i) \qquad (5)$$

Simplification of the formulae

For the particular values of q which normally occur in practice, some of the terms are of the second order of smallness and may be neglected.

Thus, the second term in formulae (2), (4), (5) and (6), which is due to heterogeneity in the mortality experience, is of the order NQ^2 whereas the first term is of the order NQ. Hence for normal values of the rate of mortality this term is small and may be neglected.

Also since n_i is in all cases large, $n_i - 1$ may be replaced by n_i in formulae (4) and (6).

Again, since

$$q_i^2 + \frac{p_i q_i}{\bar{n}_i} = q_i\left(\frac{\bar{n}_i q_i + p_i}{\bar{n}_i}\right),$$

the second term would normally be much less than the first, as $p_i < 1$ and $\bar{n}_i q_i$ is equal to the expected deaths.

Hence

$$\left(q_i^2 + \frac{p_i q_i}{\bar{n}_i}\right)\text{Var.}(n_i)$$

may be replaced by q_i^2 Var.(n_i).

The variance of the number of deaths may thus be taken as NPQ with the addition of

(1) N^2 Var.(q) or Σn_i^2 Var.(q_i) if q only varies stochastically;

(2) Σq_i^2 Var.(n_i) if n_i only varies stochastically;

and

(3) $\Sigma \bar{n}^2$ Var.$(q_i) + \Sigma \bar{q}_i^2$ Var.$(n_i) + \Sigma$ Var.(n_i) Var.(q_i) if both n_i and q_i vary independently and stochastically.

Thus

$$\text{Var.}(D) \doteqdot NPQ + \Sigma \bar{n}_i^2 \text{ Var.}(q_i) + \Sigma \bar{q}_i^2 \text{ Var.}(n_i) + \Sigma \text{Var.}(n_i) \text{ Var.}(q_i) \quad (7)$$

Now $$\text{E}(n_i q_i) = \bar{n}_i \bar{q}_i$$

and $$\text{E}(n^i q_i^2) = (\bar{n}^2 + \text{Var.}(n_i))(\bar{q}_i^2 + \text{Var.}(q_i))$$

Therefore, $\text{Var.}(n_i q_i) = \bar{q}_i^2$ Var.$(n_i) + \bar{n}_i$ Var.$(q_i) +$ Var.(n_i) Var.(q_i)

Hence

$$\text{Var.}(\Sigma n_i q_i) = \Sigma \bar{q}_i^2 \text{ Var.}(n_i) + \Sigma \bar{n}_i \text{ Var.}(q_i) + \Sigma \text{ Var. } (n_i) \text{ Var.}(q_i).$$

Therefore, substituting in (7) we obtain a simple general formula

$$\text{Var.}(D) = NPQ + \text{Var. } (\Sigma n_i q_i) \quad (8)$$

From formula (8) the variance of the total deaths will exceed NPQ if there is any variation in $\Sigma n_i q_i$, i.e. in NQ. The variation could be in n_i or q_i or both.

3. ESTIMATION OF RANDOM VARIATION FROM ACTUAL DATA

The estimation of the amount of random variation in actual data is a problem calling for some ingenuity. Each situation calls for its own

particular approach. Ideally, we require a large number of identical samples each consisting of the same (large) number of persons exposed to risk of death, at the same period of time and all of the same age. We are not usually presented with such an ideal situation and therefore have to estimate the size of the random component of variation from the total variation which exists in a real life situation. Given a sufficiently large population, estimates, adequate for this purpose, of the random variation in the number of deaths can be obtained. Three methods of estimation have been used in the numerical work of this paper.

METHOD 1. Australian population data for the calendar years 1961–65 inclusive[2] was the first subject of study. The number of persons exposed to risk of death at each age within each quinquennial age group did not vary greatly nor did the number exposed at any age vary greatly from one calendar year to another within the 5 calendar years being considered. The exposed to risk was taken as the average for the 5 calendar years and for the five ages within the quinquennial age group. The actual number of deaths each year and at each age was rated up or down in simple proportion to this assumed exposed to risk to produce what we might call 'adjusted' deaths. A multiple linear regression function was then fitted to these 'adjusted' deaths to eliminate both any time variation and the known age variation. For this short 5-year period and for the five ages, variation was in each case assumed to be linear. The variation from this linear regression function was taken as an estimate of the random variation with 22 degrees of freedom.

Tests of significance

If the simple binomial hypothesis were true these estimates would be estimates of NPQ and the extent of any departure from NPQ would be due solely to random variation. We should firstly like some indication whether the actual departures from the binomial hypothesis are significant or not. A study of the signs of actual variance minus NPQ was made as these should be randomly positive and negative if the binomial hypothesis were true. Two other tests were used.

Test 1. The estimate of variance k_2 obtained from a sample of size r from a binomial population has a variance[3]

$$\frac{NPQ(1-6PQ)}{r}+\frac{2N^2P^2Q^2}{r-1}$$

which for large N and small Q becomes

$$\frac{2N^2P^2Q^2}{r-1}$$

Hence its standard deviation $= \dfrac{NPQ}{\sqrt{(r-1)/2}}$

which in our case is approximately equal to

$$\frac{NPQ}{3{\cdot}3}$$

Therefore, if our estimate of variance approaches twice NPQ it would be more than three standard deviations above the expected value NPQ and must be considered significantly large.

Test 2. If k_2 is the estimate of variance obtained from a sample from a normal population then

$$\frac{vk_2}{\sigma_2} \frown \chi_v^2$$

where v is the number of degrees of freedom.

We might therefore use as another guide to significance the values of $22k_2/NPQ$ and treat as significant values exceeding 40 (the 1% level for χ_{22}^2).

METHOD 2. Deaths in Australia from specific causes were only available in quinquennial age groups.[4] The 6 years 1961–66 inclusive were chosen as there was no change in classification during this period and there appeared to be no great secular change in the rates. 'Adjusted' deaths were calculated in the manner mentioned earlier and these 'adjusted' deaths were used to obtain estimates of variance with 5 degrees of freedom. As an indication of significance, tests similar to those mentioned earlier were used.

METHOD 3. The results of a mortality investigation of assured lives in Australia for the 5 years ended 1963 were published by the Institute of Actuaries of Australia and New Zealand in May, 1968.[5] For policies with durations 2 years and over the actual deaths and exposed to risk were published for individual ages. The expected deaths by the A1949–52 ultimate table were also published and for the principal age groups the actual deaths were about 87% of the expected on this basis. Each quinquennial age group was therefore taken and the A1949–52 expected deaths at each age were rated up so that the total for the quinquennial age group was equal to the total actual deaths for that age group. The sum of (actual − expected)^2 on this basis, divided by 4 because of the one constraint, was used as an estimate of random variation. Similar indications of significance were again used.

4. AN ANALYSIS OF AUSTRALIAN DATA

Population data

The estimate of variance of the number of deaths in a population of given age and size based on Australian population data for the years 1961–65, together with the variance according to the binomial hypothesis, is given for males and females separately in Tables 1 and 2. The estimates have been made by Method 1.

Figures are only given for ages above 25 because values of q could not be considered linear within quinquennial groups at younger ages. In the case of males, for all ages above 25 the estimated variance exceeds NPQ and for all ages over 50 it significantly exceeds NPQ. In the case of females, for ages over 40 the actual variance appears to exceed NPQ significantly.

Table 1. *Australia* 1961–65. *Males*

(k_2 *stands for estimate of variance*)

Age group	NPQ	k_2	$\frac{k_2}{NPQ}$	$\frac{22k_2}{NPQ}$	k_2 greater (+) or less (−) than NPQ	Significant or not
25–29	102	112	1·1	24	+	No
30–34	126	160	1·3	28	+	No
35–39	186	227	1·2	27	+	No
40–44	275	341	1·2	27	+	No
45–49	407	509	1·3	28	+	No
50–54	626	1148	1·8	40	+	Yes
55–59	834	1579	1·9	42	+	Yes
60–64	1062	2460	2·3	51	+	Yes
65–69	1228	4285	3·5	77	+	Yes
70–74	1422	8689	6·1	134	+	Yes
75–79	1295	4041	3·1	69	+	Yes
80–84	869	6004	6·9	152	+	Yes

Table 2. *Australia* 1961–65. *Females*

(k_2 *stands for estimate of variance*)

Age group	NPQ	k_2	$\frac{k_2}{NPQ}$	$\frac{22k_2}{NPQ}$	k_2 greater (+) or less (−) than NPQ	Significant or not
25–29	57	30	·6	14	−	No
30–34	66	57	·9	19	−	No
35–39	113	110	1·0	22	−	No
40–44	162	322	2·0	44	+	Yes
45–49	241	363	1·5	33	+	Barely
50–54	327	640	2·0	43	+	Yes
55–59	403	639	1·6	35	+	Barely
60–64	568	1962	3·5	76	+	Yes
65–69	795	2050	2·6	57	+	Yes
70–74	1117	10446	9·4	206	+	Yes
75–79	1241	6683	5·4	119	+	Yes
80–84	1098	12144	11·0	243	+	Yes

The amount by which the actual variance exceeds NPQ is described as 'excess variance'; the corresponding standard deviation is described as 'excess S.D.' This excess standard deviation of the number of deaths is given, for males and females, in Table 3. When expressed as a percentage of the average deaths, NQ, this 'excess S.D.' does not vary much with age. It appears to be about ·04 NQ for males and ·06 NQ for females, at the significant ages.

Hence we conclude that

$$\left.\begin{aligned} \text{Var.}(D) &= NPQ+(\cdot 04NQ)^2 \text{ for males.} \\ \text{Var.}(D) &= NPQ+(\cdot 06NQ)^2 \text{ for females.} \end{aligned}\right\} \tag{9}$$

These formulae hold for ages over 40. They *may* hold at all ages, as the observed variance at young ages does not differ *significantly* from that given by *these* formulae. However, the simple binomial is an adequate hypothesis at most of the younger ages.

Table 3. *Australia* 1961–65 '*Excess*' *standard deviation in number of deaths*

Age group	Males		Females	
	excess S.D.	excess S.D. / NQ	excess S.D.	excess S.D. / NQ
25–29	3·2	·031		
30–34	5·8	·046		
35–39	6·4	·034		
40–44	8·2	·030	12·7	·078
45–49	10·1	·025	11·0	·046
50–54	22·9	·036	17·7	·054
55–59	27·3	·032	15·4	·038
60–64	37·4	·034	37·3	·065
65–69	55·3	·043	35·4	·044
70–74	85·2	·056	96·6	·083
75–79	52·4	·037	73·8	·056
80–84	71·7	·070	105·1	·085

Source of excess variation—cause of death study

It is not possible to be specific about the causes of this excess variation but some light may be thrown on the problem by studying the variation in deaths from specific causes.

Again ages below 25 were ignored, in this case because of paucity of deaths. Values of $5k_2/Npq$ for certain principal causes of death for each age group for males and females using Method 2 applied to Australian data for the years 1961–66 inclusive are given in Table 4. These figures may be compared with χ^2_5 which equals 15 for 1% level of significance.

Table 4 confirms that for 'all causes' the actual variance is significantly greater than the binomial at ages over 40 for both sexes.

For some causes of death (diseases of the nervous system and sense organs, neoplasms, and in the case of males only, accidents) the binomial hypothesis would appear to be adequate. For others (diseases of the circulatory system at ages over 55, diseases of the respiratory system at most ages and accidents, in the case of females only at the middle ages) the variance is significantly greater than the binomial.

No very clear-cut conclusions emerge from this analysis. However, it suggests that for some diseases (e.g. neoplasms) the rate of mortality and the incidence in the community do not vary to any extent from year to year. For others (e.g. diseases of the circulatory system at higher ages) either the rate of mortality or the incidence of the relative impairments or both vary significantly from year to year with resultant effect on the deaths from 'all causes'. The different result obtained for the two sexes in the case of accidental deaths is of interest.

Assured lives data

The result of applying Method 3 to the Australian assured lives data[5] for durations 2 and over for the 5 years ended 1963 is given in Table 5.

Table 4. *Values of* $5k_2/Npq$ *for various causes of death for Australia* 1961–66

(k_2 *stands for estimate of variance*)

Age group	Cause of death 1	2	3	4	5	6	All causes
Males							
25–29	5	4	2	19	7	4	5
30–34	5	14	4	27	21	17	9
35–39	1	7	6	4	4	1	10
40–44	16	9	2	5	10	10	16
45–49	3	6	19	33	7	15	13
50–54	2	4	6	12	12	12	18
55–59	9	2	9	24	15	29	17
60–64	3	9	11	11	37	25	47
65–69	7	2	8	35	34	43	24
70–74	2	4	5	11	45	84	65
75–79	16	2	17	16	26	78	51
80–84	9	15	9	3	14	87	37
Total	78	78	98	200	232	405	312
Females							
25–29	3	13	3	7	6	14	4
30–34	1	14	5	4	4	12	3
35–39	4	11	8	6	3	12	10
40–44	4	21	5	12	8	7	14
45–49	8	37	11	8	12	13	21
50–54	4	23	11	17	9	20	11
55–59	4	8	7	3	23	11	25

Age group	Cause of death 1	2	3	4	5	6	All causes
60–64	3	12	8	13	21	6	16
65–69	7	5	13	5	26	17	14
70–74	3	2	8	10	26	18	16
75–79	5	10	9	10	9	49	24
80–84	13	9	7	9	29	80	57
Total	59	165	95	104	176	259	215

Cause 1. Diseases of nervous system and sense organs.
Cause 2. Accidents, poisonings and violence.
Cause 3. Neoplasms.
Cause 4. Infective and parasitic diseases.
Cause 5. Diseases of circulatory system.
Cause 6. Diseases of respiratory system.

Table 5. *Australian assured lives* 1958–63. *Estimate of random variation in number of deaths* (k_2) *and ratio of 'excess' S.D. to NQ*

Age group	k_2	NPQ	$\frac{k_2}{NPQ}$	Excess S.D. NQ
$37\frac{1}{2}$–$41\frac{1}{2}$	133	97	1·37	·062
$42\frac{1}{2}$–$46\frac{1}{2}$	339	194	1·75	·062
$47\frac{1}{2}$–$51\frac{1}{2}$	374	369	1·02	·006
$52\frac{1}{2}$–$56\frac{1}{2}$	655	559	1·17	·017
$57\frac{1}{2}$–$61\frac{1}{2}$	1934	768	2·52	·044
$62\frac{1}{2}$–$66\frac{1}{2}$	1683	893	1·89	·031
$67\frac{1}{2}$–$71\frac{1}{2}$	2733	1064	2·57	·037
$72\frac{1}{2}$–$76\frac{1}{2}$	4076	1103	3·69	·047
$77\frac{1}{2}$–$81\frac{1}{2}$	2355	854	2·76	·041
$82\frac{1}{2}$–$86\frac{1}{2}$	775	479	1·62	·031

Again the estimate of variance exceeds NPQ in all cases. As the estimate is based only on 4 degrees of freedom it is not easy to establish significant variation. For some of the higher age groups the variation is significantly greater than NPQ. If a combined test were applied to several age groups significant results would be obtained.

It is interesting to note that these results (from an experience consisting mainly of male lives) support the male formula (9) that

$$\text{Var.}(D) = NPQ + (\cdot 04NQ)^2$$

The process of selection adopted by life offices, resulting in the elimina-

tion of impaired lives, might also eliminate much of the excess variation in the number of deaths if this excess variation is due to variations from year to year in the incidence of various impairments. This might lead one to expect the binomial hypothesis to be appropriate for the first couple of years after selection but to become less and less appropriate as the period since selection increases.

The Australian assured lives experience[5] excluded impaired lives as far as possible. It was carried out on a select basis and although data for individual ages by duration was not published it was kindly made available to the author. The number of deaths in some cells was fewer than one would wish but nevertheless a fairly clear pattern did emerge from the data.

The analysis was limited, because of paucity of data, to the six quinquennial age groups $37\frac{1}{2}$–$41\frac{1}{2}$ to $62\frac{1}{2}$–$66\frac{1}{2}$ inclusive. The method adopted was that described earlier as Method 3. Although values of the ratio k_2/NPQ for the different age groups fluctuated widely, for duration 0 for all age groups except one the values were greater than unity, for durations 1 and 2 all values except for one age group were less than unity, and the ratios then increased with duration. Some indication of the pattern is given by the values of $\Sigma k_2/\Sigma NPQ$ in Table 6 (the Σ extending over the six age groups). Values of χ^2_{24} measuring the departure of actual deaths from expected on the basis of Method 3 are also given in Table 6. As the expected value of χ^2_{24} is 24 it can be seen that for durations 0, 1, 2 and 3 the values obtained fluctuate reasonably about the expected. For duration 4 and durations 5 and over (where the data is large) the values of χ^2 are significant at the 5% level.

The process of selection does appear to eliminate the 'excess' variation but as selection wears off significant 'excess' variation appears. For durations 5 and over only, the data in the higher age groups $67\frac{1}{2}$–$71\frac{1}{2}$, $72\frac{1}{2}$–$76\frac{1}{2}$, and $77\frac{1}{2}$–$81\frac{1}{2}$ was substantial. Here values of k_2/NPQ of 2·83, 4·95 and 2·74 were obtained and a highly significant value of χ^2_{12} of 39·7 for the three age groups. The *average* period since selection in these three older age groups would be much larger than the average period since selection in the younger age groups although both are in the duration category '5 and over'. An analysis for individual durations 5, 6, 7 etc. would have been of interest, had the data been available, to confirm the obvious trend of increased 'excess' variation with duration.

The assured lives experience, based on number of policies, would include duplicates which should increase the value of the estimate of variance. The magnitude of this effect is not known. However, the general impression left from the analysis is that at higher ages with population data the variance of the number of deaths exceeds NPQ, and with assured lives data the binomial hypothesis is adequate during the period of selection, but after selection has worn off the binomial hypothesis again underestimates the variance.

Table 6. *Australian assured lives* 1958–63 *Age groups* $37\frac{1}{2}$–$41\frac{1}{2}$ *to* $62\frac{1}{2}$–$66\frac{1}{2}$ *inclusive*

Period since selection (years)	$\frac{\Sigma k_2}{\Sigma NPQ}$	Measure of discrepancy χ^2_{24}	Significant at 5% level
0	1·29	29·5	No
1	·75	19·1	No
2	·61	12·4	No
3	1·01	23·9	No
4	1·65	34·8	Yes
5 and over	1·63	35·4	Yes

5. PRACTICAL IMPLICATIONS

Tests of a mortality graduation

Most tests of a mortality graduation are based on the binomial hypothesis rather than on formulae (8) or (9) and hence there would be a tendency to over graduate.

Stop-loss reinsurance

If reinsurance treaties were arranged so that the reinsurer pays any amount by which the year's claims exceed an agreed figure then the premium for such reinsurance would depend on the variance of the amount of death claims. If this is calculated on the binomial hypothesis then such a reinsurance premium might be understated.

6. GENERAL COMMENTS

Criticisms could fairly be made concerning points of detail in this paper, e.g. the effect of errors of age so well known in population data, paucity of deaths in some cells, the independence assumptions, the consistency of diagnosis of cause of death, the tests used, the effect of duplicates. However, none of these is likely to invalidate the general conclusions reached.

The author is of the opinion that more attention should be paid to the variation in the number of deaths in mortality studies, particularly in the case of assured lives, as this could contribute to our knowledge of the mortality process.

REFERENCES

(1) POLLARD, J. H. A note on multi-type Galton–Watson processes with random branching probabilities. *Biometrika* 1968, **55**, 589.

(2) *Demography Bulletins*, Nos. 79–83. Commonwealth Bureau of Census and Statistics, Australia.

(3) FISHER, R. A. *Statistical methods for research workers*, p. 75.

(4) *Causes of Death*. Bulletins Nos 1–4. Commonwealth Bureau of Census and Statistics, Australia.

(5) Mortality Investigation. *Transactions of the Institute of Actuaries of Australia and New Zealand*, 1968, p. 1.