JIA 118 (1991) 483-488

J.I.A. **118**, III, 483–488

SOME STATISTICAL ASPECTS OF THE CONTINUOUS MORTALITY INVESTIGATION BUREAU'S MORTALITY INVESTIGATIONS

BY W. F. SCOTT, M.A., Ph.D., F.F.A.

ABSTRACT

This article discusses the formulae for the select and ultimate exposed-to-risk in the CMIB's mortality studies, and certain statistical aspects of these investigations. It is shown that there are difficulties in the traditional binomial approach to the distribution of deaths, particularly for select rates, and the use of the Poisson distribution, rather than the binomial, is advocated.

KEYWORDS

Exposed-to-risk; Graduation; Continuous Mortality Investigation

1. INTRODUCTION AND NOTATION

The system of classification of policies used by the Continuous Mortality Investigation Bureau (CMIB) since its establishment in 1923[1] is according to:

$$y = \text{age nearest birthday}$$

and

$$t = \text{curtate duration.}$$

Thus, for example, $\theta(y,t)$ denotes the number of deaths in a given period among policyholders aged y nearest birthday with curtate duration t years. Similarly, the in force at time r years from a given 1 January, denoted by ${}_rP(y,t)$, is the number of policies in force at this time on lives aged y nearest birthday with curtate duration t years. The CMIB calculates the 'in force' on each 1 January, and hence obtains the 'mean in force' for any given calendar year by the formula:

$$\tfrac{1}{2}\left[{}_0P(y,t) + {}_1P(y,t)\right] \tag{1.1}$$

where time is measured in years from the start of the year in question. Formula (1.1) is, of course, an approximation to the central exposed-to-risk for the calendar year, viz.

$$E^c(y,t) = \int_0^1 {}_rP(y,t)dr. \tag{1.2}$$

The central exposed-to-risk, and the corresponding deaths, may be aggregated over several calendar years; in practice, the CMIB uses quadrennia (1975–78, 1979–82 and so on).

According to the 'traditional' approach (that used before 1988), the 'initial' (or

'q-type') exposed-to-risk in respect of lives aged y nearest birthday with curtate duration t years is calculated by the formula:

$$E(y,t) = E^c(y,t) + \tfrac{1}{2}\,\theta(y,t). \qquad (1.3)$$

For a given select period, s years, the ultimate deaths, $\theta(y)$, and central exposed-to-risk, $E^c(y)$, at age y nearest birthday are calculated by the formulae:

$$\theta(y) = \sum_{t \geqslant s} \theta(y,t)$$

and

$$E^c(y) = \sum_{t \geqslant s} E^c(y,t)$$

$$\simeq \tfrac{1}{2}\,[{}_0P(y) + {}_1P(y)]$$

where ${}_rP(y)$ denotes the number of ultimate policies in force at age y nearest birthday at time r, i.e.:

$${}_rP(y) = \sum_{t \geqslant s} {}_rP(y,t).$$

$E(y)$ is defined as:

$$\sum_{t \geqslant s} E(y,t)$$

which equals

$$E^c(y) + \tfrac{1}{2}\,\theta(y).$$

We note that $\theta(y) = \theta_{y-\frac{1}{2}}$, $E^c(y) = E^c_{y-\frac{1}{2}}$, and so on, because the ultimate lives 'aged y' are aged y nearest birthday, and are therefore followed from age $(y - \frac{1}{2})$ to age $(y + \frac{1}{2})$. For select lives, we note that lives aged $(y - \frac{1}{2})$ to $(y + \frac{1}{2})$ on death must have entered into assurance, etc., between ages $((y - \frac{1}{2}) - (t - 1))$ and $((y + \frac{1}{2}) - t)$, i.e. at age $(y - t - \frac{1}{2})$ on average. It is therefore assumed that:

$$E^c(y,t) = E^c_{[y-t-\frac{1}{2}]+t},\ \theta(y,t) = \theta_{[y-t-\frac{1}{2}]+t}, \text{ etc.}$$

The maximum select period available in respect of CMI data is 5 years. By suitable aggregation, however, a shorter select period (or none) may be used. Mortality data may be collected on the basis of 'lives' or 'amounts' (i.e. weighted by size of benefit); except where otherwise stated, we shall refer to 'lives' data.

2. STATISTICAL PROBLEMS OF THE TRADITIONAL APPROACH

We now discuss the statistical basis employed in the CMI Reports, and by those constructing mortality tables from CMI data, before 1988. For the purpose of graduation and other statistical calculations it was assumed that (ignoring any duplicate policies):

$$\theta(y,t) \sim \text{binomial } (E(y,t),q) \tag{2.1}$$

where:

$$q = q_{[y-t-\frac{1}{2}]+t} \tag{2.2}$$

If these formulae hold, the crude q-type death rate:

$$\hat{q} = \frac{\theta(y,t)}{E(y,t)} \tag{2.3}$$

is an unbiased estimator of the true rate q, and the variance of $\hat{q}$ is $q(1-q)/E(y,t)$.

We shall consider the accuracy of formula (2.1). Let us first assume, for the sake of simplicity, that there are no surrenders or other terminations except by death, and that there are no fractional contributions to the exposure from 'beginners' and 'enders'. Each policy contributes a period of between 0 and 1 year to $E^c(y,t)$, as shown in Figure 2.1.

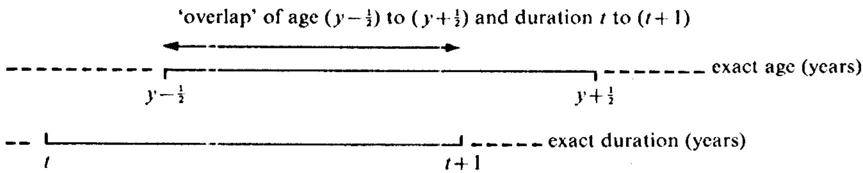

Figure 2.1.

In practice, there may be a tendency for assurance policies to be effected just before a birthday, due to premium increases at birthdays; likewise, there may be a tendency for pensions and annuity policies to be effected just after a birthday. If this were always the case, all the contributors to $E^c(y,t)$ would each give $\frac{1}{2}$ year less losses due to deaths. If there are no such tendencies, it is likely that the average contribution to $E^c(y,t)$ will be approximately $\frac{1}{2}$ year, less losses due to deaths, as policies will probably be taken out more or less uniformly over the year of age.

We therefore consider the true position to be similar to that in which all lives contribute exactly $\frac{1}{2}$ year, ignoring 'losses' of exposure caused by deaths. The exact entry age is either $(y-t)$ (with exact duration t since entry), as shown in

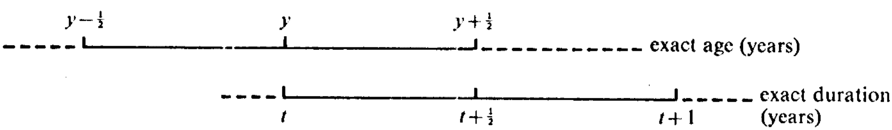

Figure 2.2.

Figure 2.2, or $(y - t - 1)$ (with exact duration $(t + \frac{1}{2})$ since entry). We assume that, in each case, the probability of death within $\frac{1}{2}$ year is $\frac{1}{2}q$, where q is defined by formula (2.2).

Since the average extra time that would have been spent as a (y,t) life by each such policyholder who dies may be taken as $\frac{1}{4}$ year, we have the following alternative model. Let us define the q-type exposed-to-risk as:

$$\bar{E}(y,t) = E^c(y,t) + \theta(y,t)/4. \tag{2.4}$$

Instead of formula (2.1), we have:

$$\theta(y,t) \sim \text{binomial } [2\bar{E}(y,t), \tfrac{1}{2}q]. \tag{2.5}$$

In particular, $\theta(y,t)$ has mean $\bar{E}(y,t)q$ and variance $\bar{E}(y,t)q(1 - \frac{1}{2}q)$. Thus:

$$\hat{q} = \frac{\theta(y,t)}{\bar{E}(y,t)}$$

is, approximately, an unbiased estimator of $q = q_{[y-t-\frac{1}{2}]+t}$, and its variance is equal to $q(1 - \frac{1}{2}q)/\bar{E}(y,t)$. The actual position is further complicated by the existence of surrenders and other terminations besides deaths, and by the fractional contributions due to 'beginners' and 'enders'. These lead to even greater departures from formula (2.1) (see Hoem[2] for a discussion of the problems associated with fractional contributions to the q-type exposed-to-risk in non-select tables).

The foregoing remarks do not apply with the same force to the 'ultimate' exposed-to-risk because each policyholder is generally an ultimate life aged $(y - \frac{1}{2})$ to $(y + \frac{1}{2})$ over parts of two policy years, although the problems mentioned by Hoem apply.

3. THE POISSON THEORY

We now consider the use of the central exposed-to-risk (called the 'person-years of exposure' by many statisticians) rather than the q-type exposed-to-risk. This approach has been considered independently by several writers (e.g. Weck[3], Scott[4,5] and is largely adopted by Forfar *et al.*[6] and in a recent CMI paper[7]. In the theory developed by Scott[4], it is not assumed that the force of mortality is constant over any age-range: the force of mortality is assumed to be continuous with respect to age, as one would expect from biological considerations (except in special circumstances, e.g. at the exact ages at which young people are first permitted to drive motor vehicles).

According to this theory, the number of deaths, $\theta(y,t)$, is assumed to be approximately the same as would occur if $E^c(y,t)$ 'replaceable' lives aged $(y - t - \frac{1}{2})$ on entry and with exact duration t were followed until duration $(t + 1)$. That is:

$$\theta(y,t) \sim \text{Poisson } [E^c(y,t)m(y,t)] \tag{3.1}$$

where:

$$m(y,t) = m_{[y-t-\frac{1}{2}]+t} = \int_0^1 \mu_{[y-t-\frac{1}{2}]+t+r}\,dr. \tag{3.2}$$

Using properties of the Poisson distribution:

$$\hat{m}(y,t) = \frac{\theta(y,t)}{E^c(y,t)} \tag{3.3}$$

is an unbiased and efficient estimator of $m(y,t)$. Formula (3.2) is not the usual definition of the central death rate, but is that which arises naturally from the Poisson theory. An alternative symbol is $\bar{\mu}(y,t)$. Similar remarks apply to formula (3.6). Having graduated $\{\hat{m}(y,t)\}$, the values of $q_{[y-t-\frac{1}{2}]+t}$ may be found by the formula:

$$q_{[y-t-\frac{1}{2}]+t} = 1 - \exp[-m(y,t)]. \tag{3.4}$$

These results may be extended to multiple-decrement tables simply by adding the appropriate affix: α, β, etc.

This approach also applies to the ultimate rates. We have:

$$\theta(y) \sim \text{Poisson } [E^c(y)m(y)] \tag{3.5}$$

where:

$$m(y) = m_{y-\frac{1}{2}} = \int_0^1 \mu_{y-\frac{1}{2}+t}\,dt. \tag{3.6}$$

It follows that:

$$\hat{m}(y) = \frac{\theta(y)}{E^c(y)} \tag{3.7}$$

is an unbiased and efficient estimator of $m_{y-\frac{1}{2}}$. After one has obtained graduated values of $m_{y-\frac{1}{2}}$, $q_{y-\frac{1}{2}}$ may be found by the formula:

$$q_{y-\frac{1}{2}} = 1 - \exp(-m_{y-\frac{1}{2}}). \tag{3.8}$$

The Poisson theory has a number of statistical advantages, particularly in connection with the allowance for duplicate policies and for mortality by 'amounts' (see Scott[(5)]). It is also much simpler than the binomial theory when there is more than one mode of decrement, or when the population under investigation is inhomogeneous.

REFERENCES

(1) *CMIR* **1** (1973). Consolidated Rules of the Continuous Mortality Investigation Bureau. 7–14. Institute of Actuaries and Faculty of Actuaries.

(2) HOEM, J. M. (1984). A flaw in actuarial exposed-to-risk theory. *Stockholm Research Reports in Demography*, **15**, University of Stockholm.

(3) WECK, F. A. (1947). The mortality rate and its derivation from actual experience. *Rec. Am. Inst. Actuaries*, **36**, 23–54.

(4) SCOTT, W. F. (1982). Some applications of the Poisson distribution in mortality studies. *T.F.A.* **38**, 255–63 and *T.F.A.* **39**, 419–20.

(5) SCOTT, W. F. (1988). On the theory of graduation of mortality tables. *Proc. XXIIIrd International Congress of Actuaries, Finland*, 341–72.

(6) FORFAR, D. O., MCCUTCHEON, J. J. & WILKIE, A. D. (1988 & 1990). On graduation by a mathematical formula (with discussion.) *J.I.A.* **115**, 1–149 and *T.F.A.* **41**, 97–269.

(7) *CMIR* **9** (1988). The graduation of the 1979–82 mortality experience. 1–102. Institute of Actuaries and Faculty of Actuaries.