B.A.J. 2, I, 157–183 (1996)

BIAS IN SELECT MORTALITY INVESTIGATIONS

BY R. G. CHADBURN, Ph.D., F.I.A.

ABSTRACT

The bias inherent in select mortality investigations where data are grouped by coincident or non-coincident rate intervals is analysed and compared. The key to the bias is shown to be the particular patterns of non-uniform 'exposure frequency' inherent in each method. It is shown that the bias using non-coincident rate intervals is sensitive to a non-uniform distribution of new entrants by age, while the use of coincident rate intervals is sensitive to making appropriate assumptions regarding the distribution of policy anniversaries over life-time, $p(s)$. Under both methods a significant proportion of the bias is retained after graduation. It is shown that non-coincident rate intervals may be preferred where $p(s)$ is unknown, as in the CMI Bureau's investigations, while coincident rate intervals would be preferred where $p(s)$ is known and taken into account in deriving the age for the estimated rates.

KEYWORDS

Estimation; Exposure Frequency; Mortality; Rate Interval

1. INTRODUCTION

1.1 Roberts (1986) has observed that bias in non-select mortality (or other) investigations depends upon the underlying shape of the survival curve and of the relative frequency of exposure over the rate interval. In particular, he observes that in all cases bias is minimised when exposure over the rate interval is uniform.

1.2 All mortality investigations, in practice, are generally subject to data censoring: that is where some individuals are only observed for part of the year of age (or duration) at a particular age (or duration) label, due to commencing or ceasing observation during the rate interval involved. Individuals present at the start or end of the investigation will contribute such partial exposures, as will those who are observed as increments or decrements during the investigation period. This will, therefore, cause unequal exposure frequencies of the kind studied by Roberts. Such bias may well arise in select mortality investigations, particularly in the first policy year, during which voluntary discontinuance rates may be especially high. As all policyholders in their first policy year are observed from exact duration 0 (apart from those existing at the start of the investigation period), the relative exposure towards the start of the year of duration [0,1] may be considerably higher than that towards the end of that policy year, causing bias.

1.3 In the present paper I am going to neglect the effect on bias caused by censoring. The purpose of this paper is to identify and to analyse the effect of the unequal exposure frequencies inherent in select mortality investigations, even when the effect of censoring is excluded. The bias caused by censoring in select mortality investigations is, therefore, a subject for further research.

1.4 It is usual, in select mortality investigations, to assume either that the force of mortality is constant over the rate interval(s) involved (see Forfar, McCutcheon & Wilkie, 1988 (referred to as FMW)) or, less restrictively, to assume that exposure is uniform over the rate interval(s) involved (see Scott, 1982, 1991). This latter assumption can also be criticised, however, particularly for select investigations. The main argument against its use is that it is unrealistic to assume that a life, who exits from observation at exact duration n, will be replaced by another life entering observation elsewhere in the investigation at the exact same duration n, even in the limit, as more lives will (in general) enter observation at duration 0 than at any later duration. Nevertheless, it probably remains preferable to the assumption of a constant force of mortality *for select rate estimation* because:

(1) it is accepted that mortality is a smoothly progressing continuous function of age and duration; and
(2) mortality is known to vary relatively rapidly and non-linearly at early durations in particular.

1.5 The assumption of replacement would intuitively suggest that all parts of the rate interval(s) are equally represented in the exposed to risk, and that, as a consequence, no bias could arise; that this is not true for select investigations is clearly demonstrated in this paper.

1.6 This paper, therefore, analyses the bias inherent in select mortality investigations where the replacement assumption is made.

2. Exposure Frequency

2.1 *Method 1: Current Age Method with Non-Coincident Rate Intervals*

2.1.1 *Definition*

Consider a select mortality investigation where data are grouped by current age (and by current duration), which implies grouping by 'non-coincident' rate intervals (see Chadburn, Cooper & Haberman, 1995 (referred to as CCH)). The simplest example would be where data are grouped by age last birthday y, and by curtate duration t, where y and t are integers.

We can obtain a central exposed to risk (or 'waiting time') for the joint label y,t, by summing the individual periods of observation for which individuals are both within the rate interval y (the year of age from y to $y+1$) and the rate interval t (the year of duration from t to $t+1$) during the investigation period. It is easy to see that these individual contributions will vary between almost zero and almost unity, depending upon where exactly the yth birthday occurs relative to the tth policy aniversary. However, following CCH, if it is assumed that the yth birthday is equally likely to fall on any particular day within the period of duration $t-1$ to $t+1$ for any life labelled y and t, then, on average, the rate intervals can be *considered* to be coincident; that is, the central exposed to risk can be *assumed*

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

to represent the sum of individual contributions to exposure over a year of age and duration, at the beginning of which lives were aged exactly y and had exact duration t.

If we now make the following replacement assumption (after Scott, 1982): *all exits during the investigation with the label* y,t *are immediately replaced by lives who join the investigation at exactly the same age and duration as the lives who left*; then this, *and* the assumption that the rate intervals are coincident, would imply that all parts of the rate intervals y and t are equally represented in the exposed to risk. Hence, following Scott (1991), we can say that the random number of deaths labelled y,t in the investigation ($D(y,t,1)$, say) is Poisson distributed with parameter:

$$E(y,t,1)\int_0^1 \mu(y+r,t+r)\,dr$$

where:

$\mu(y+x,t+r)$ = the force of mortality appropriate to a life of exact age $y+x$ and exact duration $t+r$
(= $\mu_{[y+x-t-r]+t+r}$ in conventional notation)

and

$E(y,t,1)$ = the central exposed to risk for method 1, as described above.

Hence, assuming the rate intervals are coincident, the above Poisson parameter can be written:

$$E(y,t,1)\int_0^1 \mu_{[y-t]+t+r}\,dr.$$

2.1.2 *The effect of non-coincident rate intervals*

We will now relax the assumption that the rate intervals are coincident. That not all parts of the rate intervals y and t are equally represented in the exposed to risk will be clear by examining the exposure time contributed to each rate interval by individual lives. This is best illustrated by the use of a Lexis diagram, such as described in Hill, Laplanche & Rezvani (1985), and used by Renshaw (1988). In Figure 1, policy duration is shown on the horizontal axis and age on the vertical axis. An individual commences exposure at exact duration 0 at a particular exact age, and that individual's age and duration development through time is tracked by following the appropriate diagonal from south-west to north-east across the diagram.

For example, the life shown in Figure 1 enters exposure at exact duration 0 and exact age $y-\frac{1}{4}$. This life is therefore labelled $y-1,0$ (i.e. $y-1$ last birthday and curtate duration 0) until he or she reaches his or her yth birthday. By tracking along the diagonal it can be seen that the life is exposed:

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

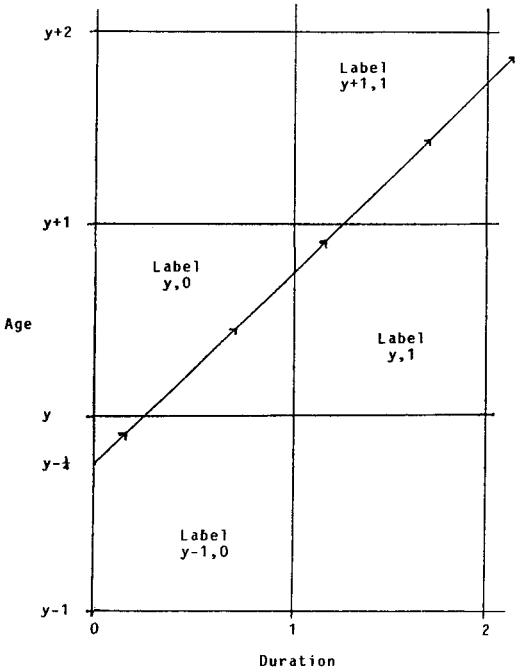

Figure 1. Lexis diagram showing the progress of a life effecting a policy at exact age y–¼, using method 1

— for a quarter of a year as label *y*–1,0;
— for three-quarters of a year as label *y*,0; and
— for a quarter of a year as label *y*,1;

and so on. In general, it can be seen that lives contribute to exposure with the label *y,t* if they pass through any part of the square (*t*,*y*), (*t*,*y*+1), (*t*+1, *y*+1), (*t*+1, *y*), and this is shown in Figure 2.

Lives with the label *y,t* can be grouped into two types, depending on whether they enter the *y,t* square through its *base* (for example, life 1 in Figure 2), in which case they are referred to as type *A*, or through its *side* (life 2 in Figure 2), in which case they are referred to as type *B*. It will be apparent that the two types have the following features.

For type *A*, labelled *y,t*:

(1) all lives enter label *y,t* on passing the *y*th birthday;
(2) all lives leave label *y,t* on passing the *t*+1th policy anniversary;
(3) all lives are exposed at exact age *y*, none at age *y*+1, with a decreasing gradient of exposure at intermediate ages; and
(4) no life is exposed at exact duration *t*, while all are exposed at duration *t*+1, with an increasing gradient of exposure at intermediate durations.

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

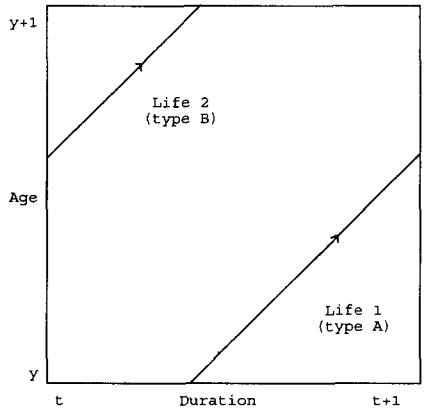

Figure 2. Lexis diagram showing the exposure square for label y, t using method 1

For type B, labelled y,t:

(1) all lives enter label y,t on passing the tth policy anniversary;
(2) all lives leave label y,t on passing the y+1th birthday;
(3) no life is exposed at exact age y, while all are exposed at age y+1, with an increasing gradient of exposure at intermediate ages; and
(4) all lives are exposed at exact duration t, none at duration t+1, with a decreasing gradient of exposure at intermediate durations.

2.1.3 *Patterns of exposure frequency*

Let us define:

$f(y+x, A)$ = proportion of type A lives exposed to risk at exact age $y+x$ $(0 \leq x \leq 1)$;

$g(t+r, A)$ = proportion of type A lives exposed to risk at exact duration $t+r$ $(0 \leq r \leq 1)$

= $f(y+1-r, A)$;

$f(y+x, B)$ = proportion of type B lives exposed to risk at exact age $y+x$ $(0 \leq x \leq 1)$; and

$g(t+r, B)$ = proportion of type B lives exposed to risk at exact duration $t+r$ $(0 \leq r \leq 1)$

= $f(y+1-r,B)$.

These functions can be described as 'exposure frequency' functions. Typical patterns of exposure frequency for a very small investigation, involving 10 lives of each type, are illustrated in Figure 3.

The pattern of exposure frequency in any investigation depends on the location of the yth birthday in relation to the tth policy anniversary for each life. For the ith life labelled y,t in the investigation, we could therefore define:

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

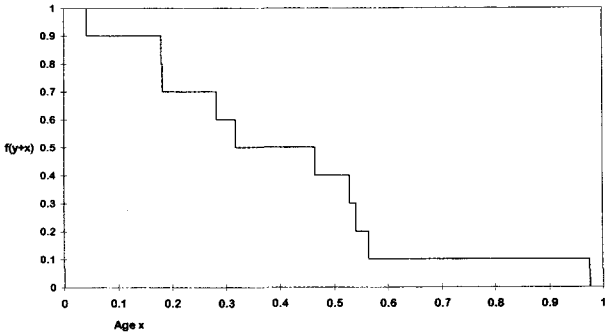

Figure 3a. Age, type A

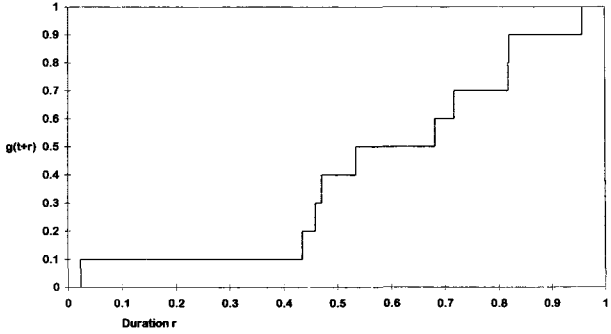

Figure 3b. Duration, type A

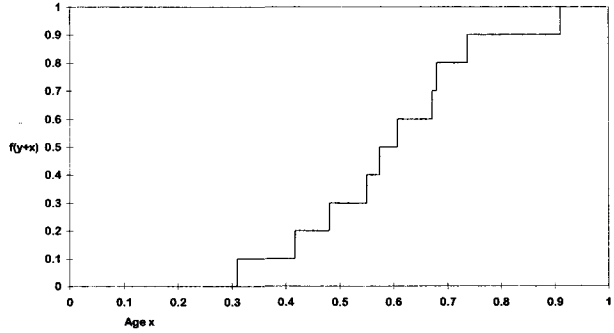

Figure 3c. Age, type B

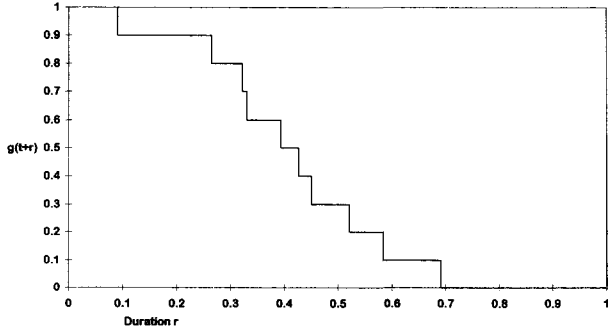

Figure 3d. Duration, type B

Figure 3. Simulated actual exposure frequencies, $N = 10$, $p(s)=1$, method 1

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

s_i = time (in years) between any birthday and the following policy anniversary

which is a random variable with probability density $p(s)$, say. The exposure frequency functions are, therefore, random proportions, which can be shown to have the following expected values (see Chadburn, 1993):

$$\mathrm{E}\{\, f(y+x,A)\,\} = \int_{x}^{1} p(s).ds \tag{2.1}$$

$$\mathrm{E}\{\, g(t+r,A)\,\} = \int_{1-r}^{1} p(s).ds \tag{2.2}$$

$$\mathrm{E}\{\, f(y+x,B)\,\} = \int_{0}^{x} p(s).ds \tag{2.3}$$

$$\mathrm{E}\{\, g(t+r,B)\,\} = \int_{0}^{1-r} p(s).ds. \tag{2.4}$$

2.1.4 *Assumptions for* p(s)

In practice the form of $p(s)$ might be assessed by reference to the data. In this paper two alternative models for $p(s)$ will be assumed, in order to illustrate the possible effect that varying $p(s)$ might have on bias. These will be:

(1) $p(s) = 1$; and

(2) $p(s) = 2s$.

Model (1) is clearly the assumption that policy anniversaries are uniformly distributed over the year of age. In model (2) a policy anniversary has zero probability of occurring on a birthday, and has a linearly increasing probability of occurring through the year of age. The expected (or average) age at which the policy anniversary occurs is:

$$\mathrm{E}\{\, y+s_i\,\} = y + \int_{0}^{1} s.\, p(s).ds$$

$$= y + \tfrac{1}{2} \qquad \text{for } p(s)= 1$$

$$= y + \tfrac{2}{3} \qquad \text{for } p(s) = 2s.$$

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

$p(s) = 2s$ is then not a particularly extreme model to assume for the distribution of policy anniversaries in relation to birthdays.

The resulting expected exposure frequencies for each function and type are given in Table 1 and shown in Figures 4 and 5.

Table 1. Expected exposure frequencies for method 1

Exposure frequency function	Expected exposure frequencies	
	$p(s) = 1$	$p(s) = 2s$
$f(y+x,A)$	$1 - x$	$1 - x^2$
$f(y+x,B)$	x	x^2
$g(t+r,A)$	r	$1 - (1-r)^2$
$g(t+r,B)$	$1 - r$	$(1-r)^2$

Chadburn (1991) demonstrated that simulated random exposures to risk of as few as 100 lives produce exposure frequency patterns which are barely distinguishable from these expected patterns. In the rest of this paper the exposure frequency will therefore be taken as its appropriate expected value.

2.2 *Method 2: the Entry Age Method, or Current Age Method with Coincident Rate Intervals*

2.2.1 Consider a select mortality investigation where:

$D(y,t,2)$ = number of deaths during the investigation aged $y-t$ last birthday at entry and curtate duration t at death.

2.2.2 This definition, along with the corresponding definition for the exposed to risk $E(y,t,2)$, implies a policy year rate interval running from exact duration t, where lives are aged y last birthday, i.e:

t	Duration →	$t+1$
y last birthday	Age →	$y+1$ last birthday

2.2.3 If we now assume that the yth birthday is equally likely to be located anywhere between exact durations $t-1$ and t, then $E(y,t,2)$ can be assumed to represent the sum of individual contributions to exposure over a year of age, at the beginning of which lives were aged exactly $y+\frac{1}{2}$ and had exact duration t. (Note: the assumption of a uniform distribution of birthdays over the policy year is not always made, as described in ¶3.3.3.)

2.2.4 On this basis, and assuming replacement (see Section 2.1.1, ¶3), we would conclude that the random number of deaths $D(y,t,2)$ is Poisson distributed with parameter:

$$E(y,t,2)\int_0^1 \mu(y+\tfrac{1}{2}+r,\ t+r)\ dr.$$

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

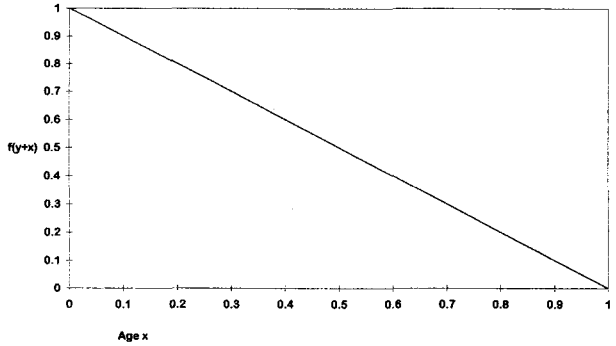

Figure 4a. Age, type A

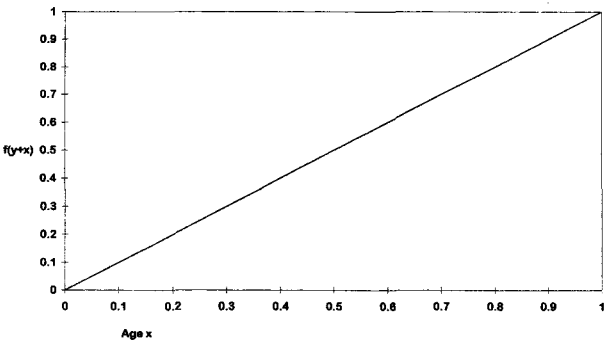

Figure 4b. Age, type B

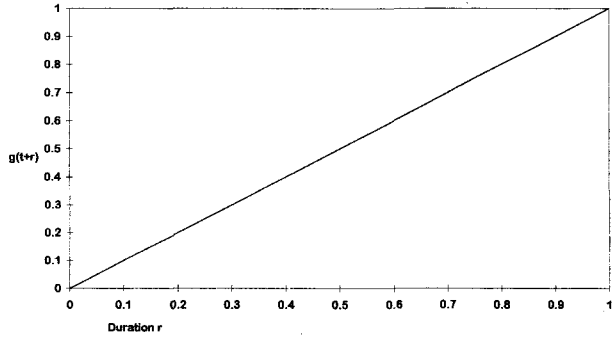

Figure 4c. Duration, type A

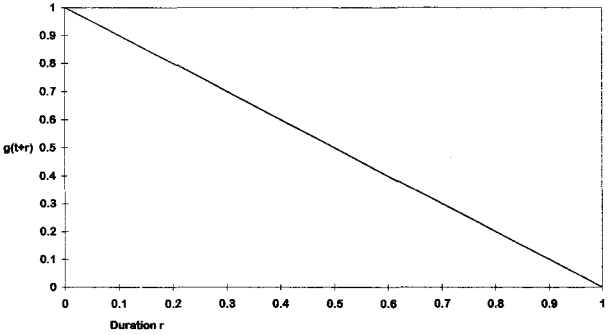

Figure 4d. Duration, type B

Figure 4. Expected exposure frequencies assuming $p(s) = 1$, method 1

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

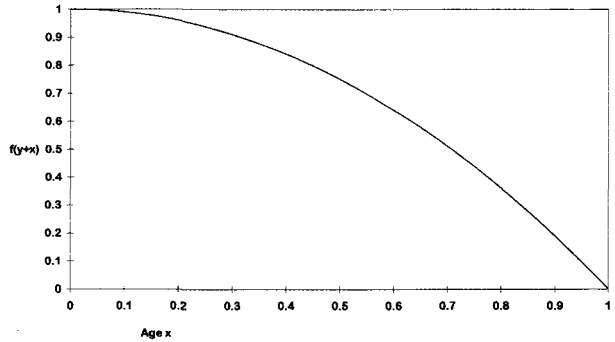

Figure 5a. Age, type A

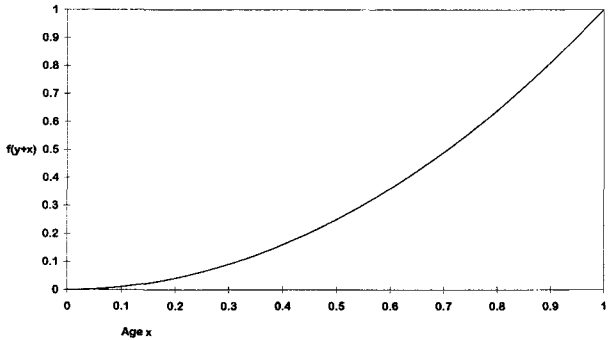

Figure 5b. Age, type B

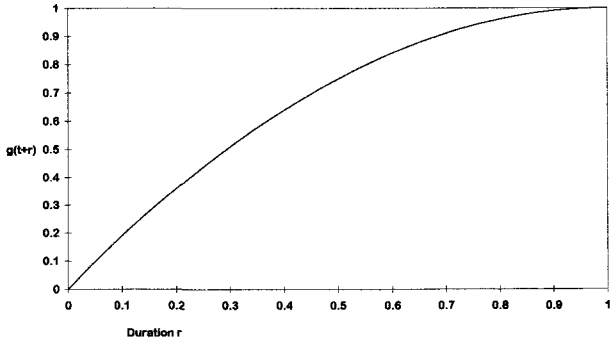

Figure 5c. Duration, type A

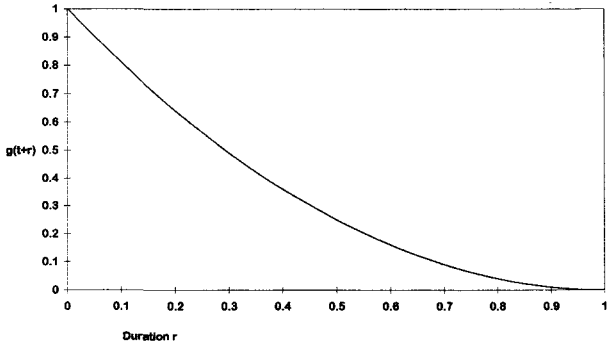

Figure 5d. Duration, type B

Figure 5. Expected exposure frequencies assuming $p(s) = 2s$, method 1

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

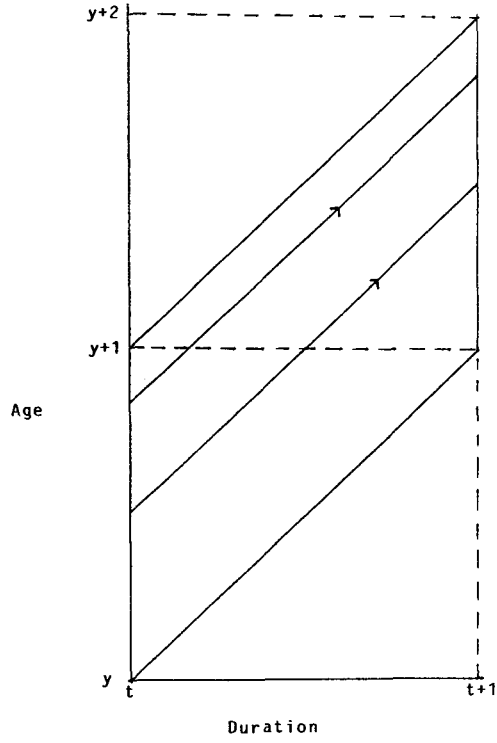

Figure 6. Lexis diagram showing the exposure rhombus for label y,t using method 2

2.2.5 The reality of the situation is, however, shown by the Lexis diagram for this method (Figure 6), which takes the form of a rhombus rather than a square as under method 1.

2.2.6 It is clear from Figure 6 that all lives are exposed at all durations between t and $t+1$, which is, of course, easily apparent from general reasoning; i.e:

$$g(t+r) = 1 \text{ for } 0 \leq r \leq 1.$$

2.2.7 Now consider $f(y+x)$:

(1) no life is exposed at exact age y and at exact age $y+2$, while all lives are exposed at exact age $y+1$; and

(2) there is an increasing gradient of exposure between ages y and $y+1$, and a decreasing gradient of exposure between ages $y+1$ and $y+2$.

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

Hence for $(0 \le x \le 1)$:

$$\mathrm{E}\{\, f(y+x)\,\} = \int_0^x p(s)\,ds. \tag{2.5}$$

For $(1 \le x \le 2)$:

$$\mathrm{E}\{\, f(y+x)\,\} = \int_{x-1}^1 p(s)\,ds. \tag{2.6}$$

2.2.8 Using the same assumptions for $p(s)$ as in Section 2.1.4, ¶1, the expected exposure frequencies are given in Table 2 and shown in Figure 7.

Table 2. Expected exposure frequencies for method 2

Exposure frequency function = $f(y+x)$

Range	$p(s) = 1$	$p(s) = 2s$
$0 \le x \le 1$	x	x^2
$1 \le x \le 2$	$2-x$	$1 - (x-1)^2$

3. Calculating the Bias Caused by Unequal Exposure Frequencies

3.1 *The Mortality Model*

3.1.1 A model for mortality must be assumed in which the age and duration components can be separately identified. Currie & Waters (1991) describe a multiplicative model for select mortality, involving separate functions for age and duration, plus an interaction term. This model was used successfully by them to graduate CMI data over the whole range of age and duration.

3.1.2 For the present investigation, however, it is only necessary to model mortality over short ranges of age and duration. It was considered that a simple additive model would be adequate for this purpose, to be fitted to existing standard tabular rates, although it would be interesting to observe the extent to which a multiplicative model might produce different results. It is possible that the differences would not be great.

3.1.3 Hence we define:

$$\mu(y+x,t+r) = \mu(y,t) + \mu(y+x,t) + \mu(y,t+r) \tag{3.1}$$

where $\mu(y,t)$ is a constant (with respect to x and r), and $\mu(y+x,t)$ and $\mu(y,t+r)$ are continuous functions of age and duration respectively.

3.2 *Calculating the Expected Deaths*

3.2.1 *Method 1*

Assume that the investigation involves $N(y,t,A)$ type A lives labelled y (last birthday) and (curtate) duration t. Consider the number of these type A lives

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

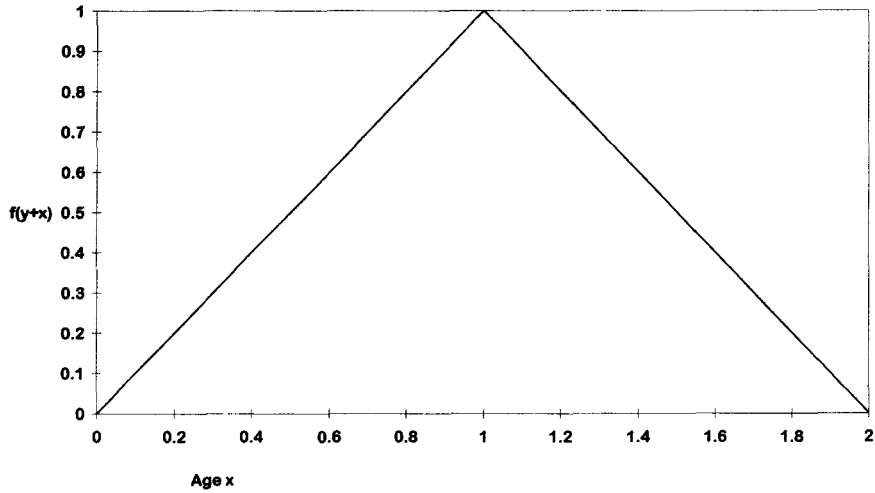

Figure 7a. $p(s) = 1$

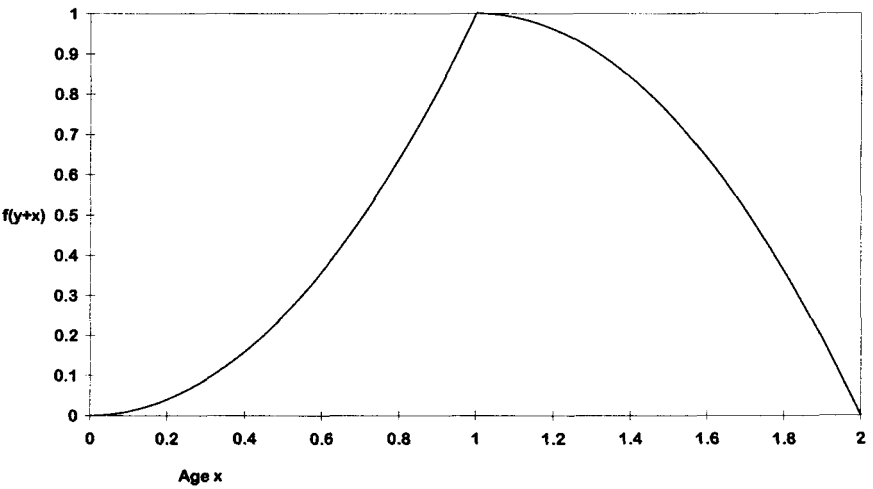

Figure 7b. $p(s) = 2s$

Figure 7. Expected exposure frequencies for method 2

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

attaining age y between exact durations $[t+r,t+r+dr]$, where dr is small (see Figure 2). Given the distribution of policy anniversaries over lifetime as defined by $p(s)$, then the *expected* number of such lives is:

$$N(y,t,A).p(1-r).dr. \tag{3.2}$$

Now the assumption of replacement implies that the number of lives at *every* point along the diagonal from duration $t+r$ and age $y+1-r$ (i.e. between durations $[t+r,t+1]$ and ages $[y,y+1-r]$) is equal to the amount in equation (3.2). Hence, in the limit, as $N(y,t,A) \to \infty$, the number of deaths arising from this 'line of exposure' has a Poisson distribution with parameter:

$$\lambda(y,t+r,A) = N(y,t,A).p(1-r).dr \int_0^{1-r} \mu(y+s,t+r+s)\,ds$$

$$= N(y,t,A).p(1-r).dr \left(\int_0^{1-r} \mu(y,t)\,ds + \int_0^{1-r} \mu(y,t+r+s)\,ds + \int_0^{1-r} \mu(y+s,t)\,ds \right) \tag{3.3}$$

which is, therefore, the expected number of deaths from this exposure. (Note: I will refer to parameters $\lambda(\ )$ as *asymptotic* Poisson parameters, reflecting the requirement for $N(\ ) \to \infty$.)

Summing equation (3.3) over all possible durations at entry into the y,t exposure square, we obtain the total expected deaths from individuals of type A:

$$\lambda(y,t,A) = N(y,t,A) \left(\begin{array}{l} \displaystyle\int_0^1 p(r) \int_0^r \mu(y,t)\,ds\,dr + \int_0^1 p(1-r) \int_r^1 \mu(y,t+s)\,ds\,dr \\ \qquad\qquad + \displaystyle\int_0^1 p(r) \int_0^r \mu(y+s,t)\,ds\,dr \end{array} \right)$$

which can be rearranged as:

$$\lambda(y,t,A) = N(y,t,A) \left(\begin{array}{l} \mu(y,t) \displaystyle\int_r^1 p(s)\,ds + \int_0^1 \mu(y,t+r) \int_{1-r}^1 p(s)\,ds\,dr \\ \qquad\qquad + \displaystyle\int_0^1 \mu(y+r,t) \int_r^1 p(s)\,ds\,dr \end{array} \right).$$

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

Using equations (2.1) and (2.2), and omitting the expected value notation for exposure frequencies for brevity, we obtain:

$$\lambda(y,t,A) = N(y,t,A)\left(\begin{array}{l}\mu(y,t)\int_0^1 f(y+r,A)\,dr + \int_0^1 \mu(y,t+r).g(t+r,A)\,dr \\ \qquad\qquad + \int_0^1 \mu(y+r,t).f(y+r,A)\,dr\end{array}\right). \quad (3.4)$$

Note that, as $f(y+r,A) = g(t+1-r,A)$, then we can equally write:

$$\mu(y,t)\int_0^1 g(t+r,A)\,dr$$

for the first element of $\lambda(y,t,A)$.

Since:

$$\lambda(y,t,A) = \int_0^1 \lambda(y,t+r,A)\,dr$$

it follows that equation (3.4) is the asymptotic Poisson parameter for the total number of type A deaths.

Repeating the above with respect to type B deaths, we obtain the following for the total expected number of deaths under method 1:

$$\lambda(y,t,1) = N(y,t,A)\left(\begin{array}{l}\mu(y,t)\int_0^1 f(y+r,A)\,dr + \int_0^1 \mu(y,t+r).g(t+r,A)\,dr \\ \qquad\qquad + \int_0^1 \mu(y+r,t).f(y+r,A)\,dr\end{array}\right)$$

$$+N(y,t,B)\left(\begin{array}{l}\mu(y,t)\int_0^1 f(y+r,B)\,dr + \int_0^1 \mu(y,t+r).g(t+r,B)\,dr \\ \qquad\qquad + \int_0^1 \mu(y+r,t).f(y+r,B)\,dr\end{array}\right) \quad (3.5)$$

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

which is the asymptotic Poisson parameter for the total number of deaths under method 1, where $N(y,t,B)$ is the total number of type B lives observed in the investigation with the label y,t.

3.2.2 *Method 2*

Assume that there are $E(y,t,2)$ lives labelled y,t in the investigation. Consider the number of these lives attaining exact duration t between ages $[y+r,y+r+dr]$, where dr is small (see Figure 6.) The expected number of such lives, given $p(s)$, is:

$$E(y,t,2).p(r).dr.$$

Assuming replacement as before, then the expected number of deaths from the diagonal of exposure between ages $[y+r,y+1+r]$ and durations $[t,t+1]$ is:

$$\lambda(y+r,t,2) = E(y,t,2).p(r).dr \int_0^1 \mu(y+r+s,t+s)\,ds$$

which is the asymptotic Poisson parameter for the number of deaths arising from this exposure. Hence the expected total number of deaths arising under method 2 will be:

$$\lambda(y,t,2) = \int_0^1 \lambda(y+r,t,2)\,dr = E(y,t,2)\int_0^1 p(r) \int_0^1 \mu(y+r+s,t+s)\,ds\,dr$$

$$= E(y,t,2)\left(\mu(y,t) + \int_0^1 \mu(y,t+s)\,ds + \int_0^1 p(r) \int_r^{r+1} \mu(y+s,t)\,ds\,dr \right)$$

$$= E(y,t,2)\left(\begin{array}{l} \mu(y,t) + \int_0^1 \mu(y,t+s)\,ds + \int_0^1 \mu(y+r,t) \int_0^r p(s)\,ds\,dr \\ \qquad + \int_1^2 \mu(y+r,t) \int_{r-1}^1 p(s)\,ds\,dr \end{array} \right)$$

$$= E(y,t,2)\left(\mu(y,t) + \int_0^1 \mu(y,t+s)\,ds + \int_0^2 \mu(y+r,t).\, f(y+r)\,dr \right) \quad (3.6)$$

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

using equations (2.5) and (2.6). The asymptotic Poisson parameter for the total number of deaths labelled y,t under method 2 is given by equation (3.6).

3.3 *Calculating Bias*

3.3.1 Bias can now be calculated as the proportionate difference between the actual expected deaths and the expected deaths according to the usual Poisson models in each case (see Section 2.1.1, ¶3 and Section 2.2.4.)

3.3.2 Hence, if $B(y,t,1)$ and $B(y,t,2)$ represent bias under methods 1 and 2 respectively, we have:

$$B(y,t,1) = \frac{\lambda(y,t,1)}{E(y,t,1)\int_0^1 \mu(y+r,t+r)\,dr} - 1 \tag{3.7}$$

and

$$B(y,t,2) = \frac{\lambda(y,t,2)}{E(y,t,2)\int_0^1 \mu(y+\frac{1}{2}+r,t+r)\,dr} - 1. \tag{3.8}$$

3.3.3 A refinement of method 2 is possible if the form of $p(s)$ is taken into account when determining the average entry age to which the estimated rates relate. For example, in the case where $p(s) = 2s$, investigation of the data will lead to the conclusion that the average age at entry is $y\text{-}t+\frac{2}{3}$ rather than $y\text{-}t+\frac{1}{2}$ (see Section 2.1.4, ¶2.) If the investigator takes this factor into account when interpreting the data, then in the bias formula the denominator becomes:

$$E(y,t,2)\int_0^1 \mu(y+\tfrac{2}{3}+r,t+r)\,dr. \tag{3.9}$$

4.. Some Observations

4.1 *Method 1*

(1) If $N(y,t,A) = N(y,t,B)$ and f and g are both equal to their expected values, then $B(y,t,1) = 0$. Hence, if there are equal numbers of individuals of types A and B in the exposed to risk, then the expected bias is zero (i.e. $E\{\ B(y,t,1)\ \}=0$).

(2) There will be zero bias if the force of mortality is constant over both rate years.

It is usually unrealistic to make the assumption of constant mortality, particularly at the shorter durations, as discussed in ¶1.4. It is, therefore, necessary to investigate the expected bias when $N(y,t,A) \neq N(y,t,B)$.

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

4.2 *Method 2*

(1) There will be zero bias if the force of mortality is constant over the age range y to $y+2$.

(2) There will be zero expected bias if the force of mortality varies linearly over the age range y to $y+2$, *and* policy anniversaries are uniformly distributed over the life year (i.e. $p(s) = 1$).

Situation 1 is likely to be unrealistic. Situation 2 is more plausible, however, and this suggests that the method is likely to lead to less potential bias than method 1. The point mentioned in ¶3.3.3 should also be borne in mind.

4.3 An important difference between method 1 and method 2 is that in the latter (the age at entry method) no assumption is necessary regarding the variation in mortality by duration in order to avoid bias.

5. NUMERICAL EXAMPLES

5.1 *Calculating Expected Bias*

5.1.1 Mortality was assumed to be according to the A1967-70[5] table. The model was fitted to the tabular values using simple quadratic functions (see Chadburn, 1993, for details.)

5.1.2 The full range of results is shown in Chadburn (1993.) The main findings are presented here.

5.2 *Method 1*

5.2.1 Table 3 shows how the expected bias per cent using method 1 varies with the proportion of type A lives in the exposed to risk (denoted by $N(A)/N$), for age 25 and duration 0, for $p(s) = 1$. Bias at duration 1 showed an identical pattern to duration 0, but at a consistently lower level:

$$B(25,1,1) \approx (0.6).B(25,0,1) \text{ for all } N(A)/N.$$

Table 3. $B = 100 \times \mathrm{E}\{\ B(25,0,1)\ \}$ for $p(s) = 1$

$N(A)/N$	0	0.2	0.4	0.6	0.8	1.0
B	−5.23	−3.14	−1.05	1.05	3.14	5.23

5.2.2 Clearly the greater the deviation of $N(A)/N$ from 50%, the greater the absolute values of the expected bias.

5.2.3 Similar patterns are observed at other ages, but at different overall levels of bias. Examples are shown in Table 4 for $N(A)/N = 0$ (i.e. assuming only type B lives are present in the exposed to risk).

5.2.4 Hence, according to this mortality assumption, at nearly all ages bias is negative where $N(y,t,A) < N(y,t,B)$ and positive where $N(y,t,A) > N(y,t,B)$. Expected absolute bias decreases between ages 25 and 40, increasing with increasing age.

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

Table 4. $B = 100 \times \mathrm{E}\{ B(y,t,1) \}$ for $N(A)/N = 0$, $p(s) = 1$

Age	Duration 0	Duration 1
20	-4.46	-3.47
25	-5.23	-3.11
30	-4.24	-1.58
35	-2.30	-0.14
40	-1.47	0.28
45	-1.65	0.09
50	-2.32	-0.30
55	-3.21	-0.73
60	-4.25	-1.17
65	-5.40	-1.59
70	-6.66	-2.01
75	-8.03	-2.40
80	-9.50	-2.79

5.2.5 It was also found that the bias for $p(s) = 2s$ bears a fixed relationship to the bias for $p(s) = 1$ *for a particular value of* $N(A)/N$, regardless of age or duration. Examples of this relationship are shown in Table 5.

Table 5. Comparison of bias between $p(s) = 2s$ and $p(s) = 1$

$$R = \frac{B(y,t,1) \quad [p(s) = 2s]}{B(y,t,1) \quad [p(s) = 1]}$$

$N(A)/N$	0	0.2	0.4	0.6	0.8	1.0
R	1.51	1.26	1.08	0.94	0.84	0.76

5.2.6 The asymmetrical nature of the bias when $p(s) = 2s$ should be noted.

5.2.7 Explanation for these observations lies in the patterns of exposure frequencies (as shown in Figures 4 and 5) and the way in which mortality in the A1967-70[5] mortality table varies by age and duration. Where mortality increases with both age and duration, as it does over most of the age range above about age 30, then the effect of exposure frequencies on bias will cancel out, to an extent, for each type. Thus mortality for type *A* will be weighted towards the start of the year of age (where mortality is lower), and towards the end of the year of duration (where mortality is higher), and vice versa for type *B*. Where mortality decreases with increasing age, and increases with duration, then there will be a compounding effect on bias. This explains the relatively high levels of bias observed below age 30 in Table 4, where mortality decreases with decreasing age following the 'accident hump' at ages 17-20. The increasingly negative bias among the type *B* lives with increasing age thereafter (Table 4), reflects the increasing rate of change of mortality with duration at the older ages, compared with the rate at which it varies by age. (Increasingly positive bias would be observed for type *A*).

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

5.2.8 At later durations the rate of increase in mortality with duration becomes much slower. This explains the positive shift in bias observed for the type *B* lives in Table 4 between durations 0 and 1, (with an equivalent negative shift for type *A*). This positive shift will continue for still later durations, as the influence of the change in mortality with age becomes dominant over duration.

5.2.9 It should be noted that of paramount importance in the overall extent of bias is the ratio of types *A* and *B* in the exposed to risk, and that the effect of the other factors ($p(s)$, mortality basis, age and duration) is conditional upon the extent to which the numbers of *A*-type and *B*-type individuals differ: no expected bias is obtained under any circumstance where $N(y,t,A) = N(y,t,B)$.

5.3 *Method 2*

5.3.1 Expected values of $B(y,0,2)$ for $p(s) = 1$ are shown in Table 6, alongside a range of example values of $B(y,0,1)$ for comparison.

5.3.2 Table 7 shows equivalent figures where $p(s) = 2s$. Column *T* shows the total bias involved when the investigator simply assumes a uniform distribution of policy anniversaries over lifetime, i.e. that the average age at entry is $y+\frac{1}{2}$ according to equation (3.8). Column *R* shows the residual bias when the investigator correctly allows for the average entry age of $y+\frac{2}{3}$ using equation (3.9).

5.3.3 The following observations can be made:

(1) The bias for $p(s) = 1$ is extremely small, less than one quarter of one per cent at all ages, broadly similar in absolute terms to method 1 bias, where only a slight deviation from equal numbers of types *A* and *B* is assumed (e.g. 48% type *A*).
(2) The total bias for $p(s) = 2s$ (column *T* in Table 7) is an order of magnitude higher than for $p(s) = 1$. At nearly all ages this equates with method 1, assuming 0 - 40% type *A*, depending on age.
(3) The residual bias for $p(s) = 2s$ is negligible, at all ages smaller in absolute terms than for $p(s) = 1$.
(4) The bias is positive at nearly all ages.

5.3.4 Provided the distribution of policy anniversaries over lifetime is taken into account when fixing the age to which the rates apply, the bias from method 2 is clearly negligible and can be safely ignored. Where, however, $p(s)$ is assumed incorrectly to be unity when determining the age to which the rates apply, more significant bias can be introduced.

5.3.5 The bias is mainly positive, due to the fact that the first and second derivatives of the force of mortality are positive at most ages. The implication is that none of the bias will be removed by the graduation process.

Table 6. $100 \times E\{ B(y,0) \}$ under methods 1 and 2, assuming $p(s) = 1$

Age	Method 1 for $N(A)/N$ = 0	0.2	0.4	0.48	Method 2
20	-4.46	-2.68	-0.89	-0.18	-0.02
25	-5.23	-3.14	-1.05	-0.21	0.03
30	-4.24	-2.54	-0.85	-0.17	0.12
35	-2.30	-1.38	-0.46	-0.09	0.19
40	-1.47	-0.88	-0.29	-0.06	0.21
45	-1.65	-0.99	-0.33	-0.07	0.21
50	-2.32	-1.39	-0.46	-0.09	0.21
55	-3.21	-1.93	-0.64	-0.13	0.20
60	-4.25	-2.55	-0.85	-0.17	0.21
65	-5.40	-3.24	-1.08	-0.22	0.21
70	-6.66	-4.00	-1.33	-0.27	0.22
75	-8.03	-4.82	-1.61	-0.32	0.23
80	-9.50	-5.70	-1.90	-0.38	0.24

Table 7. $100 \times E\{ B(y,0) \}$ under methods 1 and 2, assuming $p(s) = 2s$

	Method 1 for $N(A)/N$ =				Method 2	
Age	0	0.2	0.4	0.48	T	R
20	-6.73	-3.37	-0.96	-0.18	-1.41	-0.01
25	-7.91	-3.95	-1.13	-0.21	-0.79	0.02
30	-6.42	-3.21	-0.92	-0.17	0.81	0.08
35	-3.47	-1.74	-0.50	-0.09	2.14	0.13
40	-2.20	-1.10	-0.31	-0.06	2.51	0.14
45	-2.48	-1.24	-0.35	-0.07	2.47	0.14
50	-3.49	-1.74	-0.50	-0.09	2.34	0.13
55	-4.84	-2.42	-0.69	-0.13	2.23	0.13
60	-6.42	-3.21	-0.92	-0.17	2.15	0.14
65	-8.17	-4.08	-1.17	-0.22	2.11	0.14
70	-10.08	-5.04	-1.44	-0.27	2.09	0.15
75	-12.16	-6.08	-1.74	-0.33	2.09	0.15
80	-14.40	-7.20	-2.06	-0.39	2.11	0.16

where R = Residual bias
and T = Total bias
(see ¶5.3.2 for explanation.)

6. Implications for Mortality Investigations

6.1 *Use of Method 1*

6.1.1 *General*

The overall bias inherent in method 1 is somewhat hard to determine, as it is so dependent on the proportions of types A and B in the exposed to risk. Clearly,

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

in extreme situations (e.g. 100% of either type) the bias can be considerable. However, if the ratio were to fluctuate randomly, in the region of 50:50 from age to age, bias at individual ages would be low and would also be likely to be largely removed by graduation.

6.1.2 *The proportions of types A and B in method 1 investigations*

In order to assess the bias that can be expected using method 1 in practice, some idea of the proportions of types A and B in the actual experience needs to be assessed.

Chadburn (1991, 1993) shows that, assuming replacement:

$$N(y,0,B) = U(y)$$

and $$N(y,0,A) = U(y-1)$$

where: $U(y)$ = the number of new entrants aged y last birthday at entry.

The distribution of new entrants during the investigation by age last birthday at entry can, therefore, be used to assess fairly accurately the numbers of types A and B for duration 0 at each age.

A similar approach can be devised for other durations, but, as bias is less at these durations, consideration will here be given to duration 0 only.

6.1.3 *Empirical example: CMI data*

The example chosen is the duration 0 data from the CMI investigation into male Permanent Assured lives mortality for 1979-82. The data are presented by FMW, which forms the basis of this analysis.

FMW give the central exposed to risk and the observed deaths at each age. It will be assumed that the central exposed to risk at duration 0 can be taken to be approximately proportionate to the number of new entrants at each age. This will then be used as a realistic example of the distribution of new entrants by age for this type of investigation: no further implication for this *particular* CMI investigation is being made (indeed, the mortality assumptions used to calculate the bias would be inappropriate for this purpose).

A summary of the results is given in Table 8 (full details can be found in Chadburn, 1993).

Table 8. Empirical results for expected bias using CMI 'new entrant' data

Age range	100 × expected bias						Percentage of ages where bias is positive
	$p(s) = 1$			$p(s) = 2s$			
	Min	Median	Max	Min	Median	Max	
18 - 35	-0.44	-0.06	0.07	-0.46	-0.06	0.07	22%
36 - 55	0	0.06	0.24	0	0.06	0.23	100%
56 - 75	-1.32	0.59	1.62	-1.45	0.58	1.50	90%

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

The mostly negative bias up to the mid-thirties reflects generally increasing new entrant numbers by age, and the mainly positive bias thereafter reflects generally declining numbers. The most important features of these results are:

(a) actual levels of expected bias at *most* ages are generally very low;
(b) the distribution of bias according to its sign is distinctly non-random over the age range; and
(c) compared with method 2, the difference in bias caused by a change in the form of $p(s)$ is negligible.

Observation (b) means that the bias, however small, will only be partly removed by the graduation process. The result will be that, according to the assured lives mortality basis, the graduated rates will understate the true mortality at the younger ages and will overstate the true mortality at the older ages.

6.1.4 *Bias after graduation*

For a description of the graduation process, see FMW.

Example deaths were generated by random simulation assuming $D(y,0,1)$ is Poisson distributed with parameter $\lambda(y,0,1)$ with mortality according to A1967-70[5]. Hence the expected bias at each age was fully consistent with the underlying mortality of the experience.

The same set of random numbers was used to calculate actual deaths under the two assumptions for $p(s)$. The following crude rates were graduated in each case, where $\theta(y,0,1)$ is the simulated realised value of $D(y,0,1)$:

$$\mathring{\mu}_{(y)} = \frac{\theta(y,0,1)}{E(y,0,1)}$$

$$\mathring{\mu}^{*}_{(y)} = \frac{\mathring{\mu}_{(y)}}{1 + B(y,0,1)}$$

producing graduated functions $\breve{\mu}_{(y)}$ and $\breve{\mu}^{*}_{(y)}$ respectively. Following FMW, $\breve{\mu}_{(y)}$ and $\breve{\mu}^{*}_{(y)}$ were determined by fitting $GM(2,2)$ functions to the respective crude rates by the method of maximum likelihood, where:

$$GM(2,2) = \alpha(0) + \alpha(1).y + \exp(\beta(0) + \beta(1).y)$$

where $\alpha(0)$, $\alpha(1)$, $\beta(0)$, and $\beta(1)$ are the fitted parameters.

The 'graduated' bias at each age could then be obtained from the ratio:

$$GB(y) = \frac{\breve{\mu}_{(y)}}{\breve{\mu}^{*}_{(y)}}$$

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

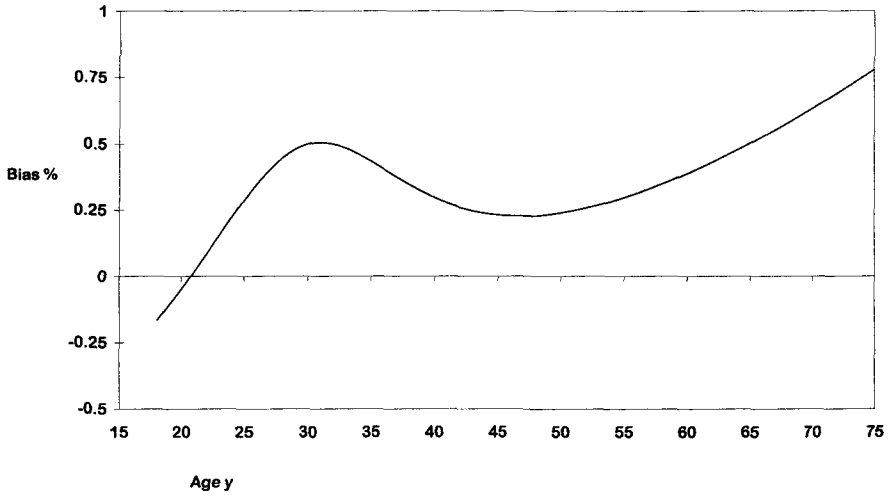

Figure 8a. $p(s) = 1$

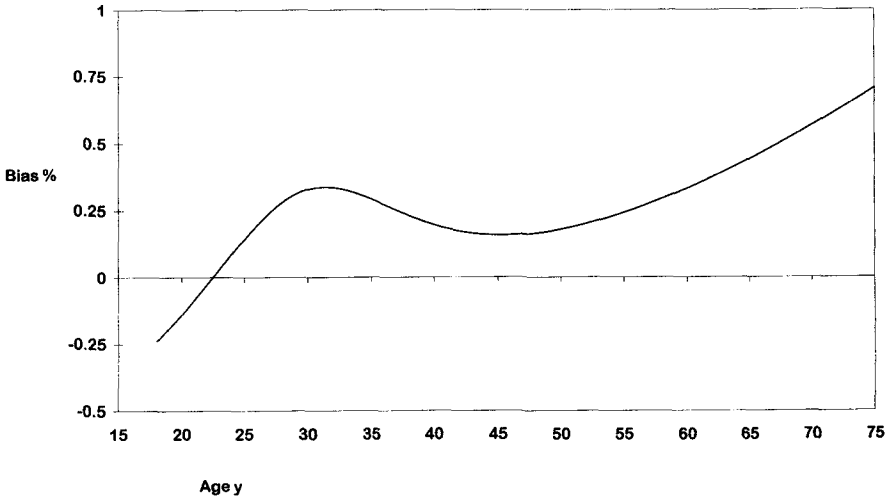

Figure 8b. $p(s) = 2s$

Figure 8. Graduated bias using method 1, assuming new entrant distribution according to CMI data

and the results are shown in Figure 8.

It can be seen from Figure 8 that graduated bias has a curious shape, with an initial peak at around age 30, falling then rising again at increasingly older ages. Graduated bias is positive for this experience at nearly all ages.

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

6.2 *Method 2*

6.2.1 The figures for method 2, shown in Tables 6 and 7, effectively give the graduated bias for this method. These figures are compared with the equivalent figures for method 1 in Table 9.

Table 9. Comparison of 'graduated bias' between methods 1 and 2

	$p(s) = 1$		$p(s) = 2s$		
Method	1	2	1	2(*R*)	2(*T*)
Age					
20	-0.05	-0.02	-0.14	-0.01	-1.41
25	0.29	0.03	0.14	0.02	-0.79
30	0.50	0.12	0.33	0.08	0.81
35	0.43	0.19	0.29	0.13	2.14
40	0.30	0.21	0.20	0.14	2.51
45	0.23	0.21	0.16	0.14	2.47
50	0.24	0.21	0.18	0.13	2.34
55	0.30	0.20	0.24	0.13	2.23
60	0.39	0.21	0.33	0.14	2.15
65	0.50	0.21	0.44	0.14	2.11
70	0.63	0.22	0.57	0.15	2.09
75	0.78	0.23	0.70	0.15	2.09

where R = Residual bias
T = Total bias

6.2.2 Where $p(s) = 1$ method 2 gives lower absolute values of bias at every age than method 1. Over about age 35 the bias for method 2 is also much more stable with increasing age than method 1, with a total range of 0.19% to 0.23% compared with 0.23% to 0.78% for method 1.

6.2.3 For $p(s) = 2s$ the comparison is similar to that described in ¶6.2.2, except that all the values are slightly lower, *provided* that in method 2 the expected average age for the rates is recognised as being $y+\frac{2}{3}$. Where the age is taken as $y+\frac{1}{2}$, however, the bias under method 2 is significantly higher than for method 1.

7. CONCLUSIONS

7.1 The significant factors affecting bias under the two methods are summarised in ¶¶7.2 and 7.3.

7.2 *Method 1*:

(1) bias is effectively dependent upon the distribution of new entrants from age to age;

(2) part of the bias will be removed by graduation; however, to the extent that

https://doi.org/10.1017/S1357321700003378 Published online by Cambridge University Press

new entrants are non-randomly distributed by age, bias will be retained in the graduated rates; and

(3) *apart from where the distribution of new entrants shows severe discontinuities with age*, overall bias after graduation appears likely to be very low, probably well less than 1% over most of the age range.

7.3 *Method 2*:

(1) bias is effectively dependent upon the extent to which the assumed age at entry differs from the actual average age at entry;

(2) where the assumed and actual average entry ages are different, bias can be relatively and consistently high, for example of the order of 2% where the assumed entry age is $y+\frac{1}{2}$ compared with an actual average entry age of $y+\frac{2}{3}$; and

(3) where actual and assumed entry ages are equal, bias is negligible and quite stable over the age range.

7.4 *Where $p(s)$ is unknown or unobtainable*, method 1 appears the more robust of the two methods in that it shows least sensitivity to variations in $p(s)$. Under these circumstances it would seem that method 1 should be preferred. Given that the CMI Bureau is likely to be in this position regarding its data, it would be concluded that the CMI Bureau is happily correct in having adopted method 1 for its select investigations. (Note that the age definition used in the CMI implies exposure between ages $[y-\frac{1}{2},y+\frac{1}{2}]$ rather than between $[y,y+1]$: bias observed for this age definition would, of course, be almost indistinguishable from that reported here. However, the CMI committee's choice of age definition also helps to reduce the effect of the possible non-uniform distribution of policy anniversaries over the life year, caused by many policyholders effecting their policies just before (or just after) a birthday. By grouping by age nearest birthday, this increased density of policy anniversaries near the birthday occurs around the middle of the rate interval, rather than at one end. This will more closely approximate to an assumption of $p(s) = 1$ than if an age last (or next) birthday definition had been used.)

7.5 For investigations on a smaller scale, (for example those carried out by individual life offices or for pension schemes), method 2 might be considered preferable, provided the average ages at entry can be assessed. In smaller experiences, the distribution of new entrants from age to age is likely to fluctuate more widely than is the case for the CMI, for example, while the distribution of policy anniversaries over lifetime may be relatively more stable.

7.6 It is, perhaps, hard to draw firmer conclusions without some knowledge of the form of $p(s)$ in empirical examples. This is a possible area for further research.

7.7 A few additional points regarding method 1 are worth noting.

(1) Where there are significant discontinuities in the progression of new entrants by age, the exposed to risk will be almost entirely either of type A or of type

B, which can lead to bias at these individual ages of the order of 5% or more in some cases. For example the exposed to risk at the minimum entry age would be 100% type *B*, while that for the age one more than the maximum entry age would be 100% type *A*. It is suggested that these data should either be excluded from the graduation, or the inherent bias at these ages allowed for individually in some way. It should be noted that the CMI data used in this paper *exclude* the sharp discontinuity in new entrants that occurs between ages 16 and 17.

(2) The situation where mortality varies by age and duration in opposite directions gives rise to somewhat higher levels of bias. Possible examples where this might occur include the mortality experience of ill-health retired pensioners, with duration measured from retirement, and the mortality experience of 'sick' lives in sickness investigations, with duration measured from the date of sickness inception. This effect does not occur when using coincident rate intervals. This situation is also investigated in Chadburn (1993).

ACKNOWLEDGEMENTS

I am grateful for the assistance of a number of my colleagues at The City University in the production of this paper. In particular I would like to thank Dr A. S. Puzey for useful discussion, Dr A. E. Renshaw for the loan of his GLIM programs in performing the various graduations, Professor S. Haberman for commenting on the paper and to Miss M. Brown for typing the original manuscript. The scrutineers' comments were also found to be extremely helpful.

REFERENCES

CHADBURN, R.G. (1991). Bias in select mortality investigations where data are subdivided by age at death. *City University Actuarial Research Paper* No. 31. (Institute of Actuaries Library, Oxford.)

CHADBURN, R.G. (1993). Bias in select mortality investigations. *City University Actuarial Research Paper* No. 56. (Institute of Actuaries Library, Oxford.)

CHADBURN, R.G., COOPER, D.R. & HABERMAN, S. (1995). *Actuarial mathematics: core reading for subject D*. Institute and Faculty of Actuaries, Oxford.

CURRIE, I.D. & WATERS, H.R. (1991). On modelling select mortality. *J.I.A.* **118**, 453-81.

FORFAR, D.O., MCCUTCHEON, J.J. & WILKIE, A.D. (1988). On graduation by mathematical formula. *J.I.A.* **115**, 1-149 and 693-698 and *T.F.A.* **41**, 97-269.

HILL, C., LAPLANCHE, A. & REZVANI, A. (1985). Comparison of a cohort with the mortality of a reference population in a prognostic study. *Statistics in Medicine*, **4**, 295.

RENSHAW, A.E. (1988). Modelling excess mortality using GLIM. *J.I.A.* **115**, 299-315.

ROBERTS, L.A. (1986). Bias in decremental rate estimates. *O.A.R.D.* No. 38, Institute of Actuaries.

SCOTT, W.F. (1982). Some applications of the Poisson distribution in mortality studies. *T.F.A.* **38**, 255-63.

SCOTT, W.F. (1991). Some statistical aspects of the Continuous Mortality Investigation Bureau's mortality investigations. *J.I.A.* **118**, 483-8.