ACTUARIAL RESEARCH CLEARING HOUSE
1980 VOL. 1

LABORATORY OF ACTUARIAL MATHEMATICS
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN Ø
DENMARK
PHONE: (01) 35 31 33

Mr. Arnold F. Shapiro
Joint Editor, ARCH
509 Business Administration Building
Pennsylvania State University
University Park, Penn 16802
U S A

JMH/UE, 12/12-79

Dear Sir,

Exposed-to-risk considerations based on the Balducci assumption and other assumptions in the analysis of mortality.

In a paper called "Estimation of the rate of mortality in the presence of in-and-out movement" which appeared on pp. 41-56 in the 1978.2 ARCH issue, T. N. E. Greville described concepts of "exposure" which have grown up around the assumption of a uniform distribution of deaths, the Balducci assumption, and the assumption of a constant force of mortality, called assumptions A, B, and C, respectively. The purpose is to estimate q_x, and Greville's starting point is a set of three moment relations, one for each mortality assumption, based on the consideration that "expected deaths" should be equal to aggregate actual deaths. The three relations, called (A), (B), and (C), had been picked up from two Society of Actuaries text-books on the measurement of mortality.

I should like to point out that

1^o. the argument for the three relations seems to be less than completely accurate, and that

2^o. procedures based on correct reasoning can be devised on the basis of straightforward, non-subtle statistical theory.

The upshot of the following substantiation is that there can be no reason to prefer the Balducci assumption (or the assumption of a uniform distribution of deaths for that matter) to the assumption of a constant force of mortality.

To argue these points, let n independent lives contribute to our information about the age interval $(x,x+1)$, and let the observation of individual no. i actually start at age $x + s_i$ and end at $x+t_i$. Assume that each individual can only contribute a single interval of this kind. Of course, $s_i = 0$ for those under observation right from the outset (at age x), and $t_i = 1$ for those present at age x+1. For later entrants, $0 < s_i < 1$, while for exitants and for those who die, $0 \leqq s_i < t_i < 1$. Let $D_i = 1$ if individual no. i dies under observation, $D_i = 0$ otherwise, and $D = \Sigma D_i$. Then all the three formulas mentioned can be written in the form

$$\sum_i \{ {}_{1-s_i}q_{x+s_i} - {}_{1-t_i}q_{x+t_i}(1-D_i) \} = D \ , \tag{1}$$

which essentially is based on the argument that the general term on the left here is the "expected" number of deaths for individual no. i. But how can that be true? Assume that there is some external mechanism which determines some maximal age $x+\tau_i$ under exposure for individual no. i. Then τ_i is unknown to the actuary when individual no. i enters the cohort, and it is observed to be

equal to t_i if $t_i < 1$ and $D_i = 0$,

larger than $t_i < 1$ if $D_i = 1$, and

equal to 1 if $t_i = 1$ (in which case $D_i = 0$).

Thus, the precise value of τ_i gets to be known (and to equal t_i) if $D_i = 0$, otherwise not. For simplicity, let us argue <u>as if</u> $\tau_i = 1$ when $D_i = 1$. Then

$$\tau_i = t_i + (1-t_i)D_i \ . \tag{2}$$

Whether we use this formula or not, the expected number of deaths to individual no. i (for the Laplacian demon who knows the value of τ_i) is

$${}_{\tau_i-s_i}q_{x+s_i} = {}_{1-s_i}q_{x+s_i} - {}_{\tau_i-s_i}p_{x+s_i}\ {}_{1-\tau_i}q_{x+\tau_i} \tag{3}$$

at entry. This resembles the general term in (1) but does not coincide with it. We easily see that (2) makes ${}_{1-\tau_i}q_{x+\tau_i} = {}_{1-t_i}q_{x+t_i}(1-D_i)$, so if one is willing to take ${}_{\tau_i-s_i}p_{x+s_i}$ to be approximately equal to 1, then perhaps one might say that the general term in (1) is approximately OK under (2). But its acceptance rests on a stretch of imagination.

This seems to bring out a similar fault in Greville's argument for (U) on page 44. In both cases, the error consists in treating exitants on a par with entrants (except for a change of sign in m_t) and thus disregarding the fact that one must be an entrant (possibly at age x) before one can ever become an exitant. By writing $m_t\ {}_{1-t}q_{x+t} = D_t$, he argues conditionally on the fact that the individua is alive at age x+t, but for a potential exitant, the fact that he will survive to some time of exit is not known at his time of entry. To get a moment relation of the kind sought, one must treat all lives equivalently, for instance by only conditioning on survival until age at entry. Instead of his $m_t \, {}_{1-t}q_{x+t} = D_t$, what one gets is then my (3) above. By reorganizing that relation, we have

$$ {}_{\tau_i - s_i}q_{x+s_i} = \frac{{}_{1-s_i}q_{x+s_i} - {}_{1-\tau_i}q_{x+\tau_i}}{1 - {}_{1-\tau_i}q_{x+\tau_i}} \, , \tag{4} $$

from which wholly correct moment relations may be derived. Under Assumption B, for instance,

$$ {}_{\tau_i - s_i}q_{x+s_i} = \frac{\tau_i - s_i}{1 - (1-\tau_i)q_x} \, q_x \, , $$

which, by aggregation, leads to

$$ \sum_i \frac{\tau_i - s_i}{1 - (1-\tau_i)q_x^*} \, q_x^* = D \, . $$

If we use (2), this becomes

$$ \sum_i \frac{(t_i - s_i) + (1-t_i)D_i}{1 - (1-t_i)(1-D_i)q_x^*} \, q_x^* = D \, , \tag{5} $$

which, in principle, is a replacement for (B). Perhaps the correspondence with (B) is seen more clearly if we note that $(t-s_i) + (1-t_i)D_i = 1-s_i$ if $t_i = 1$ or $D_i = 1$, which enables us to write the left hand side of (5) as

$$ \sum_{\{i:t_i=1 \text{ or } D_i=1\}} (1-s_i)q_x^* + \sum_{\{i:t_i<1 \text{ and } D_i=0\}} \frac{(1-s_i) - (1-t_i)}{1 - (1-t_i)q_x^*} \, q_x^* \, . \tag{6} $$

Thus, the inaccuracy in (B) stems from replacing $(1-t_i)q_x^*$ by 0 in the second member here for each individual i who exits alive before age x+1. I guess the problem of solving (5) for q_x^* has the same not overwhelming level of difficulty as that of solving Greville's (A).

Under Assumption A, $\mu(x+t) = q_x/(1-tq_x)$, which makes

$$ {}_{1-t}q_{x+t} = \frac{1-t}{1-tq_x} \, q_x \, . \tag{7} $$

By (4), this makes

$$_{t_i-s_i}q_{x+s_i} = \frac{t_i-s_i}{1-s_iq_x}\, q_x \ .$$

Aggregation results in the following moment relation when we use (2):

$$\sum_i \frac{(t_i-s_i) + (1-t_i)D_i}{1-s_iq_x^*}\, q_x^* = D \ . \tag{8}$$

A decomposition of the left hand side of (8) corresponding to that of (6) looks like this:

$$\sum_{\{i:t_i=1 \text{ or } D_i=1\}} \frac{1-s_i}{1-s_iq_x^*}\, q_x^* + \sum_{\{i:t_i<1 \text{ and } D_i=0\}} \frac{(1-s_i)-(1-t_i)}{1-s_iq_x^*}\, q_x^* \ .$$

This may of course be written as

$$\sum_i \frac{1-s_i}{1-s_iq_x^*}\, q_x^* - \sum_{\left\{\substack{i:\text{individual } i \text{ exits}\\ \text{alive before age } x+1}\right\}} \frac{1-t_i}{1-s_iq_x^*}\, q_x^* \ . \tag{9}$$

Again, the inaccuracy in (A) stems from the treatment of those who exit alive before age x+1, in that in (A), $s_iq_x^*$ is replaced by $t_iq_x^*$ in the denominator of the second sum of (9).

Although the moment relations (5) and (8) have been derived by an accurate argument, there is still the remaining difficulty that we have used (2) as a kind of practical approximation. By contrast, maximum likelihood estimation bypasses even this minor hurdle. The likelihood is

$$\Lambda = \exp\left[-\sum_i \int_{s_i}^{t_i} \mu(x+u)du\right] \prod_{i=1}^{n} \{\mu(x+t_i)\}^{D_i} \ .$$

[By the way, note that $\Sigma_i \int_{s_i}^{t_i}\mu(x+u)du = \int_0^1 h(x+u)\mu(x+u)du$, and compare this with formula (5) on Greville's page 51.] Under Assumption A, the log-likelihood becomes

$$\ln \Lambda = D \ln q_x + \Sigma_i(1-D_i)\ln(1-t_iq_x) - \Sigma_i \ln(1-s_iq_x) \ ,$$

so the likelihood equation may be written

$$\Sigma_i \frac{(1-D_i)t_i\hat{q}_x}{1-t_i\hat{q}_x} - \Sigma_i \frac{s_i\hat{q}_x}{1-s_i\hat{q}_x} = D \ , \tag{10}$$

or

$$\Sigma_i \left\{ \frac{t_i - s_i}{1 - s_i \hat{q}_x} - D_i t_i \right\} \frac{\hat{q}_x}{1 - t_i \hat{q}_x} = D \ , \tag{10}$$

which is reminiscent of (8) but does not coincide with it.

Under Assumption B, the log-likelihood becomes

$$\ln \Lambda = D \ln q_x - \Sigma_i \ (1+D_i) \ \ln\{1-(1-t_i)q_x\} + \Sigma_i \ \ln\{1-(1-s_i)q_x\} \ ,$$

and the likelihood equation is

$$\Sigma_i \frac{(1-s_i)\hat{q}_x}{1-(1-s_i)\hat{q}_x} - \Sigma_i \frac{(1+D_i)(1-t_i)\hat{q}_x}{1-(1-t_i)\hat{q}_x} = D \ ,$$

or

$$\Sigma_i \left\{ \frac{t_i - s_i}{1-(1-s_i)\hat{q}_x} - D_i(1-t_i) \right\} \frac{\hat{q}_x}{1-(1-t_i)\hat{q}_x} = D \ , \tag{11}$$

which looks much like (10), and which is not all that much different from (5).

Either of these relations must be solved by numerical iteration. Assumption C notably does not lead to any of this trouble. For all practical purposes, it seems superior to Assumptions A and B. I do not know whether there is an a priori theoretical reason to prefer any particular one of these assumptions; they all look like they have been introduced to simplify matters, and only Assumption C seems to really do so. In addition, Assumption B implies a force of mortality like $\mu(x+t)=q_x/[1-(1-t)q_x]$ which decreases when t increases, and that must be a counterargument against the Balducci assumption in its own right. Thus, I am reinforced in my conviction that the piecewise constant force assumption is the sensible one to make in the circumstances discussed in Greville's paper.

Yours respectfully,

Jan M. Hoem