ESTIMATION OF THE RATE OF MORTALITY IN THE PRESENCE OF IN-AND-OUT MOVEMENT*

T. N. E. Greville
Mathematics Research Center
University of Wisconsin-Madison
Madison, Wisconsin 53706

1. Introduction.

In estimating mortality rates from the experience of a life insurance company or employee-benefit plan, it is expedient, because of movement into and out of the covered group on dates other than birthdays or policy anniversaries, to make some assumption about mortality over a fractional part of a year of age. R. W. Batten in Chapter One of his text Mortality Table Construction (Prentice-Hall, 1978) discusses in detail three well known "simple" assumptions of this kind. These are:

Assumption A. The assumption of uniform distribution of deaths,

$${}_tq_x = tq_x \qquad (0 \le t \le 1).$$

Assumption B. The Balducci assumption,

$${}_{1-t}q_{x+t} = (1 - t)q_x \qquad (0 \le t \le 1).$$

Assumption C. The assumption of a constant force of mortality,

$$\mu_{x+t} = \bar{\mu}_x \qquad (0 < t < 1).$$

For reasons that will emerge presently, Assumption B is very generally used in the construction of mortality tables by actuaries, and in connection with it the concept of "exposure" or "exposed to risk" has been developed. The other two assumptions, though less often resorted to in mortality table construction, crop up frequently in the theory of life contingencies in other connections. A possibly more meaningful way of characterizing the three assumptions is in terms of the resulting shape of the ℓ_x curve between ages x and x + 1. Under Assumption A this is linear, under Assumption B it is hyperbolic, and under Assumption C it is exponential.

In this paper we shall briefly describe the concept of "exposure" that has grown up around Assumption B, and shall then develop and elaborate analogous concepts of "exposure" under Assumptions A and C. Finally we shall propose a fourth method of estimating the rate of mortality, not involving in any way the notion of "exposure," which I call the product method.

Little that is in the paper is original with me. The notion of "exposure"

*Sponsored by the U. S. Army under Contract No. DAAG29-75-C-0024.

under the assumption of uniform distribution of deaths seems to have been first proposed by Ralph Edwards in his discussion of a paper of H. H. Wolfenden (TASA, XLIV, 33-35). The underlying formulas were given by Harry Gershenson (in Section 6.4, pages 183-186, of his book Measurement of Mortality, especially Exercise 6.5). I had overlooked this material in Gershenson's book and I am indebted to Robert Batten for pointing it out. A different approach to the concept of "exposure" under the uniform-distribution assumption was proposed by Donald Schuette after reading an early draft of this paper, and this too is described.

The basic idea underlying the concept of exposure under the assumption of a constant force of mortality probably was first suggested by Frank Weck (RAIA, XXXVI, 23-54), and has been further developed and utilized in a practical way by Cecil Nesbitt and Hans Gerber.

The "product method" appears to be a slight variation of a technique known among biostaticians as the Kaplan-Meier product-limit method. (For references see Section 11.) However, in the form in which it is presented here no limit is involved.

As originally presented to the Actuarial Research Conference at Ball State University in the summer of 1978, this was exclusively a paper on the concept of "exposure" under the uniform-distribution assumption. After hearing the comments of other participants in the Conference I decided that it would serve a useful purpose to expand it into a general survey of various approaches to estimation of the rate of mortality in the presence of in-and-out movement. I am especially grateful to Robert Batten, Hans Gerber, and Donald Schuette for permission to incorporate some of their valuable comments and suggestions (but they should not be held responsible for any errors that I may have made in paraphrasing and interpreting their thoughts). I have made no attempt to discuss the statistical aspects of the estimation of the mortality rate, so ably discussed at the Conference by Jim Hickman. I sincerely hope that he will find it possible to submit to ARCH a separate paper on the subject.

2. Exposure under the Balducci Assumption.

As background for what follows, it seems advisable to review briefly the usual development of the concept of exposure. For this purpose, we adapt Batten's notation as follows. Let A denote the number of lives under observation at age x (an integer), and let m_t denote the net "migration" (i. e., entrants minus exitants) at exact age $x + t$. Let Σ_t denote summation over all ages $x + t$ between x and $x + 1$ at which there is movement into or out of the group, and let D denote the number of deaths between ages x and x + 1 of persons under observation. Then it is shown by Batten (and also by Gershenson in Measurement of Mortality) that, if in the aggregate actual deaths are equal to expected deaths, the three assumptions lead to the following three equations:

$$\text{(A)} \qquad Aq_x + \Sigma_t\, m_t \frac{(1 - t)q_x}{1 - tq_x} = D ,$$

$$\text{(B)} \qquad Aq_x + \Sigma_t\, m_t(1 - t)q_x = D ,$$

$$\text{(C)} \qquad Aq_x + \Sigma_t\, m_t\left[1 - (1 - q_x)^{1-t}\right] = D .$$

Assumption A leads to an algebraic equation in q_x of degree one greater than the number of different ages between x and x + 1 at which "migration" occurs. Assumption C yields a transcendental equation in q_x. Assumption B produces a simple linear equation in q_x. Solution of this equation gives a formula for q_x in which the numerator is D and the denominator is an expression that has come to be known as the "exposure," because it lends itself to a simple interpretation in terms of person-years of exposure to the risk of death. It is, in fact, the aggregate number of years lived by members of the group while under observation between exact ages x and x + 1, except that deaths of persons under observation require special treatment. Persons who die between these ages while under observation are credited with exposure from age x or age at entry, whichever is later, up to age x + 1, irrespective of the exact age at which death occurs. This is the case even if the (x + 1)th birthday is beyond the end of the observation period.

In the preceding paragraph we have summarized the usual argument leading to the concept of "exposure" under the Balducci assumption and to the exclusion of Assumptions A and C. As we shall see presently, this argument is too simplistic, and artifices can be emplIyed to recast equations (A) and (C) into a form that lends itself to easy solution. In the case of Assumption A, serious practical disadvantages will still remain (but can be avoided by a different treatment). Under Assumption C these will be minimal.

3. Exposure under the Uniform-Distribution Assumption.

In Chapter Two of Batten's text the reader is left with the impression that the concept of "exposure" has relevance only to Assumption B, the Balducci assumption, as explained in the preceding section. However, toward the end of Chapter Five (on valuation-schedule exposure formulas) there is a section called "Formulas Based on Uniform Deaths," in which reference is made here and there to "exposures" under the uniform-distribution assumption. Moreover, in Tables 5-6, 5-7, and 5-8 formulas are given for E_x (the notation for "exposure") under that assumption, and in Exercises 16 and 17 at the end of the chapter, the student is asked to calculate such "exposures." The explanation is given on page 154 that the possibility arises of introducing this assumption because in valuation-schedule formulas for the rate of mortality it it is assumed that there is no migration or that migration occurs only at year-end or on birthdays.

Notwithstanding these references to the subject, I do not find anywhere in Batten's text an explicit definition of "exposure" under the uniform-distribution assumption. The one that appears to be implicit, and which I shall adopt, is that "exposure" is the denominator of a formula for the rate of mortality in which the numerator is the number of deaths occurring between ages x and $x + 1$ to persons under observation ("observed deaths" as Gershenson calls them).

We shall show that it is possible to develop a usable (though not very practical) theory of "exposure" under the uniform-distribution assumption. To this end, we extend slightly the previous notation. For uniformity we shall use m_o, rather than A, for the number under observation at age x. Let D_t denote the expected number of deaths (positive or negative) between ages $x + t$ and $x + 1$ among the migrants at exact age $x + t$, and let $S_t = m_t - D_t$ denote the expected survivors (positive or negative) at age $x + 1$ among such migrants. Then

$$m_t \; {}_{1-t}q_{x+t} = D_t \; ,$$

and multiplication by ${}_tp_x$ gives

$$m_t(q_x - {}_tq_x) = D_t(1 - {}_tq_x) \; .$$

Introduction of the uniform-distribution assumption now gives

$$m_t(1 - t)q_x = D_t(1 - tq_x) \; ,$$

or

$$[(1 - t)m_t + tD_t]q_x = D \; .$$

Thus, the exposure is

$$\text{(U)} \qquad E_x^{(A)} = \Sigma_t[(1 - t)m_t + tD_t] = D + \Sigma_t(1 - t)S_t \; .$$

If both the aggregate number of deaths and the average age at death are assumed to be the same for the actual and the expected deaths, D_t and S_t may now be reinterpreted as the actual deaths and survivors in the various migration cohorts.

Clearly this is not a practical approach (except possibly in very special circumstances), since it requires knowledge of deaths among persons who withdraw from the experience, occurring between the date of withdrawal and the next following birthday. (Gershenson calls these "unobserved deaths.") Nevertheless, this approach has some interesting properties, as we shall see in the next section. A different approach to "exposure" under the uniform-distribution assumption is described in Section 8.

4. Characteristics of the "Uniform" Exposure Concept.

The amount of exposure provided by formula (U) for different groups of lives may be summarized as follows:

(1) Persons who do not die between exact ages x and x + 1 have the same exposure as under the conventional method. An individual who enters the group at age x + s (s may be zero) and leaves at age x + t and is alive at age x + 1 has exposure of t - s. (If he does not leave, we may take t = 1.)

(2) An individual who enters at age x + s and dies before age x + 1 as a member of the group has exposure of unity.

(3) An individual who enters at age x + s, leaves at age x + t, and then dies before age x + 1 has exposure of zero.

We recall that under the Balducci assumption, when an individual dies while under observation, his exposure does not terminate on death, but is continued to the next birthday (even if that birthday falls after the closing date of the investigation). The uniform assumption implies a principle that is, in a mathematical sense, the dual of the one just stated. When an individual dies after having entered the experience (even if he has subsequently left it, provided a birthday has not intervened), his exposure (whether positive or negative) is extended backward to the preceding birthday (even if the latter antedates his entry into the group or the starting date of the investigation).

In one respect the "uniform" concept of exposure differs sharply from the conventional one. Under the latter a new entrant who dies in the same year of age may have a fractional exposure in the year of death. Under the "uniform" concept, no one has a fractional exposure in the year of death. The net exposure of a decedent is either unity or zero.

5. Application to Valuation-Schedule Formulas.

The principles enunciated in Sections 3 and 4 can be used to derive valuation-schedule formulas based on the "uniform" assumption. However, there are pitfalls, and caution must be exercised, as Robert Batten has cogently demonstrated in his remarks at the Ball State Conference.

In this connection it must be pointed out that there is one situation in which it is not necessary to know about deaths occurring after exit from the group and before the succeeding birthday. This is the case in which there are A starters at at age x and D deaths between ages x and $x + t$, and all the $A - D$ survivors at age $x + t$ exit from the experience at that age. This would be the case, for example, if the A starters all had the same birthday and their age was $x + t$ on the closing date of the investigation. We would then have

$$A\ {}_tq_x = Atq_x = D\ ,$$

and consequently,

(U') $$E_x^{(A)} = At\ .$$

This result is different from the exposure that would be calculated under the conventional (Balducci) approach, which would be

$$A - (A - D)(1 - t) = At + (1 - t)D\ .$$

As illustrations of the derivation of valuation-schedule formulas based on the "uniform" assumption, let us consider the derivation of formulas (5-20) and (5-23) of <u>Mortality Table Construction</u>. These two examples were cited by Batten in his comments at the Ball State Conference. The appropriate diagram for formula (5-20) is shown in Figure 1. The observation period is the calendar year z, deaths are grouped by age last birthday, births are assumed to fall on October 1, α-migration is assumed to occur just after birthdays, and δ-migration just before birthdays. The exposure along the upper diagonal is

$$\tfrac{1}{4}(E_x^z + {}_\alpha m_x^z) = \tfrac{1}{4}(P_x^{z+1} + {}_\alpha D_x^z)$$

by formula (U'); along the lower diagonal we have by formula (U)

$$\tfrac{3}{4}(P_x^z - {}_\delta D_x^z) + {}_\delta D_x^z = \tfrac{3}{4}P_x^z + \tfrac{1}{4}\ {}_\delta D_x^z\ .$$

Adding the contributions gives

$$E_x\Big]_x^{x+1} = \tfrac{3}{4}P_x^z + \tfrac{1}{4}P_x^{z+1} + \tfrac{1}{4}D_x^z\ ,$$

which is the denominator of formula (5-20).

Figure 2 is the single-diagonal diagram for formula (5-23). Here we are seeking an expression for $q_{x-1/3}$ under the assumptions of May 1 births, δ-migration on January 1, and α-migration on December 31. The observation period is calendar year z and deaths are grouped by calendar age.

Because the age range for deaths is from $x - 1/3$ to $x + 2/3$, and the quantities ${}_\delta D_{x-1}^z$ and ${}_\alpha D_x^z$ are not separately available, we must think of the E_x^z lives attaining exact age x as leaving and immediately re-entering the group at that age. By formula (U'), the upper one-third of the diagonal gives the exposure

$$\tfrac{1}{3}(P_{x-1}^z + {}_\delta m_{x-1}^z) = \tfrac{1}{3}(E_x^z + {}_\delta D_{x-1}^z)\ .$$

The lower two-thirds gives, by formula (U),

$$\tfrac{2}{3}(E_x^z - {}_\alpha D_x^z) + {}_\alpha D_x^z = \tfrac{2}{3}E_x^z + \tfrac{1}{3}\ {}_\alpha D_x^z\ ,$$

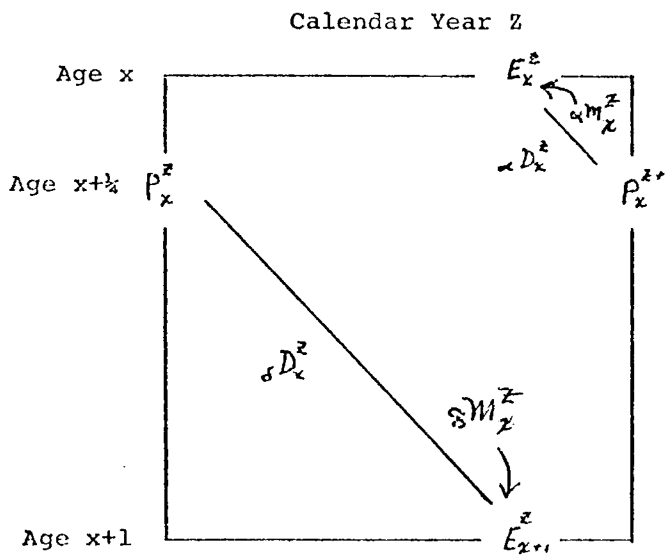
Calendar Year Z
Age x
Age x+½
Age x+1
E_x^z
$_\alpha m_x^z$
$_\alpha D_x^z$
P_x^z
P_x^{z+1}
$_\delta D_x^z$
$_\delta m_x^z$
E_{x+1}^z

Figure 1

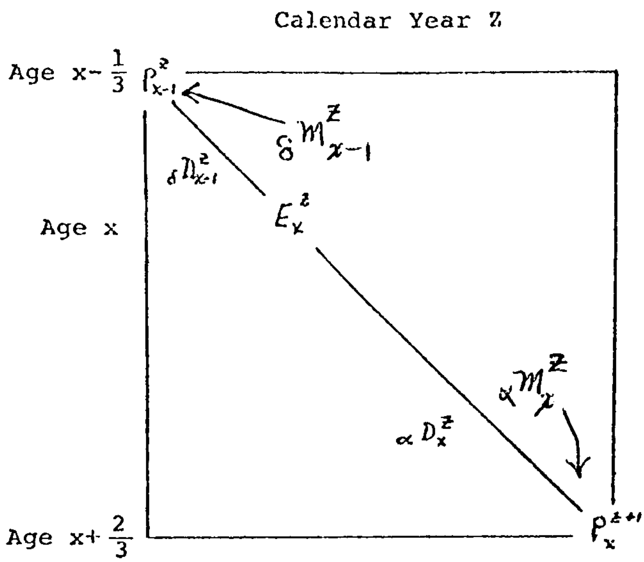
Calendar Year Z
Age x- 1/3
Age x
Age x+ 2/3
P x-1 z
δ M x-1 Z
δ N x-1 z
E x z
α M x Z
α D x Z
P x z+1

Figure 2

and adding the two contributions gives

$$E_x\Big]_{x-1/3}^{x+2/3} = E_x^z + \tfrac{1}{3}D_x^z$$

as in Batten's formula (5-23).

In connection with these examples one technical point should be mentioned. The fact that the death terms in the final formulas have fractional coefficients does not negate the statement in Section 4 that no decedent has a fractional exposure in the year of death. The reader will recall that in the derivation of valuation-schedule formulas, migration terms are eliminated by making use of demographic relationships between "populations," "survivors," migration, and deaths. Thus the coefficient of the death term in the final formula is not necessarily the exposure given to the decedents in question.

6. Exposure under the Constant-Force Assumption.

Suppose the force of mortality μ_{x+t} has the constant value $\bar{\mu}_x$ for $0 < t < 1$, and, instead of attempting to solve the transcendental equation (C) of Section 2 for q_x, we seek an estimate $\hat{\mu}_x$ of the constant force $\bar{\mu}_x$. We can then estimate q_x by

(T) $$\hat{q}_x^{(C)} = 1 - e^{-\hat{\mu}_x}.$$

Using the notation of Section 3, we consider the m_t "migrants" at exact age $x + t$. The expected number of deaths among these individuals between ages $x + t$ and $x + 1$ is

$$D_t = m_t \int_0^{1-t} {}_s p_{x+t}\, \mu_{x+t+s}\, ds = \bar{\mu}_x\, m_t \int_0^{1-t} {}_s p_{x+t}\, ds .$$

The aggregate number of years lived between ages $x + t$ and $x + 1$ by the m_t migrants is expected to be

$$m_t \int_0^{1-t} {}_s p_{x+t}\, ds = D_t/\bar{\mu}_x .$$

Summing over all values of t, we find that the expected number of "observed" deaths is

$$\bar{D} = \Sigma_t\, D_t ,$$

while the expected aggregate number of years lived between ages x and $x + 1$ while under observation is $\bar{D}/\bar{\mu}_x$. Therefore, an estimate of $\bar{\mu}_x$ is obtained by

(CF) $$\hat{\mu}_x = D/E_x^{(C)} ,$$

where D is, as before, the actual number of "observed" deaths, and $E_x^{(C)}$ is defined as the actual aggregate number of years lived between ages x and $x + 1$ by lives under observation, with deaths being treated like any other decrement. The final step is to estimate q_x by formula (T).

Nesbitt and Gerber have used this method with clinical data involving cystic fibrosis patients. It would seem entirely possible to use it in mortality investigations of insured lives or pension funds, and tabulations of deaths by exact age at death could be avoided by simplifying assumptions concerning average age at death, just as is commonly done with migration.

7. The Product Method.

Suppose that "migration" occurs at times $t_1, t_2, \ldots, t_{N-1}$ after exact age x, where $0 < t_i < 1$, and let $t_o = 0$ and $t_N = 1$. For $i = 0, 1, \ldots, N-1$, let $s_i = t_{i+1} - t_i$. Then,

$$q_x = 1 - \prod_{i=0}^{N-1} {}_{s_i}p_{x+t_i} \; .$$

Let m_i denote the number of "migrants" at exact age $x + t_i$, and let h_{i-o} and h_{i+o} denote the number of persons under observation, respectively, just before and just after the migration at age $x + t_i$ occurs. Then $h_{i+o} - h_{i-o} = m_i$, and an estimate of ${}_{s_i}p_{x+t_i}$ is h_{i+1-o}/h_{i+o}. An estimate of q_x is therefore

(P) $$\hat{q}_x^{(P)} = 1 - \prod_{i=0}^{N-1} h_{i-o}/h_{i-1+o} \; .$$

This procedure is merely an adaptation of a very basic actuarial technique. The same principle is used, for example, in calculating ${}_nq_x$ from the q_x values for the n single ages involved. It does not involve in any way the notion of "exposure," and makes no assumption about the shape of the ℓ_x curve. It probably would not be considered practical for the usual mortality investigation of insured lives or pension-fund participants. However, it is easy to imagine a situation in which it would not be too laborious to apply, and might be expected to give reasonable results. This would be a case in which N is small and the volume of data is moderately large. Consider, for example, a birthday-to-birthday study of insured lives in which deaths are tabulated by insuring age last birthday, and suppose that entry and exit are permitted only on premium-due-dates, with premiums payable annually, semiannually, or quarterly (but not monthly).

The same principle could be used in valuation-schedule formulas to avoid making restrictive assumptions about birthdays. (Assumptions about the dates of occurrence of migration would still be needed.) Consider, for example, the situation depicted in Figure 1, but without any assumption as to the day of the year on which birthdays occur. We could estimate q_x by

$$\hat{q}_x^{(P)} = 1 - \frac{P_x^{z+1}}{P_x^{z+1} + {}_{\alpha}D_x^z} \; \frac{P_x^z - {}_{\delta}D_x^z}{P_x^z} \; .$$

It would be necessary to cross-tabulate deaths by valuation year of birth and in[illegible]ing age at death, so that ${}_{\alpha}D_x^z$ and ${}_{\delta}D_x^z$ would be known. This idea, of course, has been frequently used in demographic applications. (See, for example, page 118 of Spiegelman's Introduction to Demography, Revised Edition.)

8. An Approach to the Exposure Concept Using Integration.

Donald Schuette has proposed an ingenious approach to the exposure concept based on integration. Modifying slightly the notation of Section 7, let $h(x + t)$ denote the number of persons under observation at exact age $x + t$. (This function has jump discontinuities at those exact ages where migration occurs.) If, in the aggregate, actual deaths are equal to expected deaths,

(S) $$D = \int_0^1 h(x + t)\mu_{x+t}\, dt \ .$$

By reference to Figure 1-4 on page 12 of Batten's Mortality Table Construction or otherwise, one notes that under Assumption A,

$$\mu_{x+t} = \frac{q_x}{1 - tq_x} \qquad (0 < t < 1),$$

from which it follows that

(A') $$\mu_{x+t} = q_x(1 + t\mu_{x+t}) \qquad (0 < t < 1).$$

Similarly, under Assumption B we obtain

(B') $$\mu_{x+t} = q_x[1 + (1 - t)\mu_{x+t}] \qquad (0 < t < 1),$$

and under Assumption C

(C') $$\mu_{x+t} = \bar{\mu}_x \qquad (0 < t < 1)$$

as on page 1 of this paper.

Substitution of the respective right members for μ_{x+t} in (S) gives in the first two cases $D = \hat{q}_x\, E_x$, and in the third case $D = \bar{\mu}_x\, E_x$, where the respective "exposures" are given by

(A") $$E_x^{(A)} = \int_0^1 h(x + t)\, dt + \int_0^1 th(x + t)\mu_{x+t}\, dt \ ,$$

(B") $$E_x^{(B)} = \int_0^1 h(x + t)\, dt + \int_0^1 (1 - t)\, h(x + t)\mu_{x+t}\, dt \ ,$$

(C") $$E_x^{(C)} = \int_0^1 h(x + t)\, dt \ .$$

It is curious to note that while Assumptions B and C lead to essentially the same result obtained previously, Assumption A leads to a different result. Formula (A") does not seem to require information about "unobserved" deaths. The exposure for those who do not die is the same as under the conventional approach, while for "observed" deaths the exposure is the period from entry to death plus the period from the preceding birthday to death (even if that birthday precedes entry). Thus there is some "double exposure" for "observed" deaths.

With the hint provided by formula (A") essentially the same result can be obtained by discrete methods. Let D_i denote the expected deaths between exact ages $x + t_i$ and $x + t_{i+1}$. Then, reverting to the notation of Section 7, we have

$$D_i = h_{i+o}\ {}_{s_i}q_{x+t_i}\ ,$$

and multiplication by ${}_{t_i}p_x$ gives

$$h_{i+o}({}_{t_{i+1}}q_x - {}_{t_i}q_x) = (1 - {}_{t_i}q_x)D_i\ .$$

Introduction of Assumption A reduces this to

$$D_i = q_x(s_i\ h_{i+o} + t_i\ D_i)\ .$$

Thus, if the aggregate number of deaths and the average age at death are the same for actual and expected deaths, $q_x = D/E_x^{(S)}$, where

(A"') $$E_x^{(S)} = \sum_{i=0}^{N-1} s_i\ h_{i+o} + \sum_{i=0}^{N-1} t_i\ D_i\ ,$$

with D_i now reinterpreted as actual deaths between ages $x + t_i$ and $x + t_{i+1}$.

The very slight difference between formulas (A") and (A"') as regards exposure in the interval between ages $x + t_i$ and $x + t_{i+1}$ (where the latter provides exposure of s_i and the former twice the period from age $x + t_i$ until death--the same if death occurs exactly in the middle of the interval) is due to the fact that equation (S) treats death as a continuous process, even though it allows jump discontinuities for other decrements and for increments. The Schuette formula (A"') differs from the Edwards formula (U) because deaths have been classified by the age interval in which they occur rather than by the migration cohort to which they belong.

9. Numerical Example.

Suppose there are 1,000 starters at age x, 300 new entrants at age $x + 1/3$, and 180 withdrawals at age $x + 2/3$. At age $x + 1$ there are 990 survivors of the 1,000 starters (including some who subsequently withdrew), 295 survivors of the new entrants, and 179 survivors of the withdrawals. The number of "observed" deaths is $10 + 5 - 1 = 14$. Of these, 4 occurred between ages x and $x + 1/3$ at the average age of $x + 1/4$, 5 occurred between ages $x + 1/3$ and $x + 2/3$ at the average age of $x + 1/2$, and 5 occurred between ages $x + 2/3$ and $x + 1$ at the average age of $x + 3/4$.

The exposure by the conventional method is

$$E_x^{(B)} = 1(1000) + \tfrac{2}{3}(300) - \tfrac{1}{3}(180) = 1140\ ,$$

and the estimated rate of mortality is

$$\hat{q}_x^{(B)} = \frac{14}{1140} = .012281 \ .$$

The exposure under the Edwards "uniform" assumption is

$$E_x^{(A)} = 14 + 1(990) + \tfrac{2}{3}(295) - \tfrac{1}{3}(179) = 1141 \ ,$$

and the estimated rate of mortality is

$$\hat{q}_x^{(A)} = \frac{14}{1141} = .012270 \ .$$

The difference in the exposure is negligible, and is accounted for by the fact that each of the 5 deaths among new entrants receives additional exposure of 1/3, while the one death among the withdrawals has exposure less by 2/3.

This example shows that the proposed "uniform" exposure concept would be troublesome to apply on an individual-record basis. Not only would it require information about "unobserved" deaths, but also detailed tabulations of deaths (or survivors) by mode and date of entry and/or exit.

The exposure under the constant-force assumption can be obtained by subtracting from that under the Balducci assumption the canceled exposure for the deaths. This gives

$$E_x^{(C)} = 1140 - \tfrac{3}{4}(4) - \tfrac{1}{2}(5) - \tfrac{1}{4}(5) = 1133.25 \ ,$$

and the estimated rate of mortality is

$$\hat{q}_x^{(C)} = 1 - \exp(-14/1133.25) = .012278 \ .$$

The data needed to apply the product method are shown in the following table:

i	h_{i-o}	h_{i+o}
0	—	1000
1	996	1296
2	1291	1111
3	1106	—

and the estimate of q_x is

$$\hat{q}_x^{(P)} = 1 - \frac{996}{1000}\,\frac{1291}{1296}\,\frac{1106}{1111} = .012308 \ .$$

Finally, Schuette's approach gives

$$E_x^{(S)} = \tfrac{1}{3}(1000 + 1296 + 1111) + \tfrac{1}{3}(5) + \tfrac{2}{3}(5) = 1140\tfrac{2}{3}$$

and

$$\hat{q}_x^{(S)} = 14/1140\tfrac{2}{3} = .012274 \ .$$

This is quite close to $\hat{q}_x^{(A)}$.

In this example the results are close together and the differences are not of practical significance. However, it is interesting to note that $\hat{q}_x^{(C)}$ is intermediate between $\hat{q}_x^{(A)}$ and $\hat{q}_x^{(B)}$ (as might be expected in view of the fact that the geometric mean of a number of quantities is always between the arithmetic and harmonic means), while $\hat{q}_x^{(P)}$ exceeds all the others by an amount that is several times greater than the differences between the others. (Substitution of $\hat{q}_x^{(S)}$ for $\hat{q}_x^{(A)}$ would not change this statement.) I believe this finding is typical. In this connection it will be noted that the product term in formula (P) is a product of quantities each of which is less than or equal to 1. Therefore the product is less than (or at most equal to) the smallest of the factors.

10. Anomalous Values.

It has sometimes been pointed out that under the conventional method it is possible in extreme cases to have the deaths exceed the exposures, so that an estimated mortality rate greater than unity is obtained. To take a very extreme illustration, suppose there are no starters at age x, but only new entrants at age x + 1/2, and these all die before attaining age x + 1. Then the estimated rate of mortality by the conventional method is 2.

This kind of anomaly cannot occur under the Edwards "uniform" assumption when the exposure is calculated from individual records without simplifying assumptions, because under formula (U) every "observed" death is credited with a full year of exposure in the year of death. However, as Batten has so clearly shown, no such statement applies to formula (U'), in which it is entirely possible for D to exceed At, and the anomaly in question can arise in valuation-schedule formulas derived by means of formula (U'). Taking an illustration used by Batten, we suppose in Figure 1 that there is no migration and that $P_x^z = {}_{\delta}D_x^z = 0$, while ${}_{\alpha}D_x^z = 4$ and $P_x^{z+1} = 0$. Then, in Batten's formula (5-20), derived in Section 5 of this paper, D = 4 and the exposure is 1.

Schuette's approach to the uniform-distribution assumption is not free from the anomaly, though its occurrence would be rare.

The constant-force assumption is virtually free from the possibility of anomalous results. If μ is any positive quantity whatever, $1 - e^{-\mu}$ is between 0 and 1.

The product method, by its very nature, cannot produce a value of $\hat{q}_x^{(P)}$ greater than 1, but it has its own brand of anomalous results, and cannot be safely applied to scanty data. To take an extreme example, suppose the group under observation should dwindle at some point to one person, and that person should happen to die before the next "migration point." Then $\hat{q}_x^{(P)} = 1$. If, in one of the subintervals into which the year of age is divided, there is no exposure (and no deaths), the method fails to give any result.

11. References to Literature.

I have tried to avoid discussing subtle questions of statistical theory, because I am not competent to deal with them, and in any case this paper is long enough as it is. However, a few references may be helpful to the reader who is inclined to pursue such questions. A good beginning is a paper by Jim Hickman "Some Actuarial Views of Life Table Construction," which was presented at the 1970 annual meeting of the American Statistical Association and appeared in the first issue of ARCH. Another very useful reference, especially with regard to the estimates based on Assumption C, which most resemble demographic rates, is the paper of Jan Hoem "The Statistical Theory of Demographic Rates," Scandinavian Journal of Statistics, 3 (1976), 169-185, especially Section 2C, "Occurrence/ exposure rates and their statistical properties."

The Kaplan-Meier product method is well known among biostatisticians (though I did not know this until I submitted an earlier draft of this paper to others for comment) and it has some properties dear to the hearts of statisticians. The basic reference is E. L. Kaplan and P. Meier, "Nonparametric Estimation from Incomplete Observations." Kaplan and Meier acknowledge that the method was proposed in 1912 by P. E. Böhmer ("Theorie der unabhängigen Wahrscheinlichkeiten," Rapports, Mémoires et Procès-verbaux du Septième Congrès International d'Actuaires, Amsterdam, 2 (1912), 327-343. An excellent example of the application of this technique is A. W. Kimball, "Estimation of Mortality Intensities in Animal Experiments," Biometrics, 16 (1960), 505-521.

With regard to the fact that the commonly used actuarial estimates do not have the properties of unbiasedness and maximum likelihood, the interested reader may consult H. L. Seal, "Deaths among Prospective Existings," Proceedings of the Conference of Actuaries in Public Practice, 11 (1961), and L. Elveback "Estimating Survivorship in Chronic Disease: the 'Actuarial" Method," Journal of the American Statistical Association, 53 (1958), 420-440.

12. Summary and Conclusions.

This survey tends to reinforce the conventional wisdom that the Balducci approach leading to the usual exposure concept is to be preferred in mortality investigations of life-insurance or pension-fund experience. In most practical situations the proportion of lives affected by in-and-out movement is small, and the differences between the results obtained by the different approaches will not be significant, so that the great convenience in data handling under the conventional procedures far outweighs any slight theoretical objections.

The constant-force method is a close competitor. It is theoretically superior and could probably be applied without real difficulty even in large-scale mortality investigations. It is the method of choice for clinical data involving limited experience and high rates of mortality.

Both the Edwards uniform-distribution approach and the product method, especially the latter, are theoretically interesting, but would seem to have little to recommend them from a practical standpoint. Both impose excessive data-handling requirements without compensating advantages.

Schuette's approach to the uniform-distribution assumption does not involve serious data-handling problems, especially if an average age at death is assumed, but there might be a mental obstacle in seeking to explain and justify the "double exposure" for deaths that it appears to involve.

Acknowledgments.

I wish to express my appreciation to the members of my class in Business 663 at the University of Wisconsin-Madison in the spring semester of 1978, whose interest in the problem of mortality-rate estimation stimulated me to pursue this investigation. For the references in Section 11 I am indebted to John Beekman and Jim Hickman. My gratitude to Robert Batten, Hans Gerber, and Donald Schuette for permission to include in the paper material developed by them was previously mentioned.