TRANSACTIONS OF SOCIETY OF ACTUARIES
1994 VOL. 46

ACTUARIAL CALCULATIONS USING A MARKOV MODEL

BRUCE L. JONES

ABSTRACT

This paper develops a general approach to actuarial calculations in applications that can be modeled as multistate processes. Such situations arise when benefits are payable upon a change in the status of the insured or while the insured maintains a given status. Examples include life insurance, annuities, pensions, disability income insurance, and certain types of long-term-care insurance.

The method is based on convenient matrix results that are available when a continuous-time Markov model with constant forces of transition is assumed. In this case probabilities are easily obtained regardless of the number of states. This is of considerable benefit because it allows us to deal with very complicated actuarial problems.

Section 1 provides background on the kinds of problems for which the approach is suitable. The basic properties of the Markov process are presented in Section 2. Considered here are some useful results that hold under the assumption of constant or piecewise constant forces of transition. Section 3 addresses the situation in which the Markov assumption is inappropriate. Rather than using a more general semi-Markov model, one can reflect duration dependence by increasing the number of states in the model. This is justified by a limiting result and demonstrated by an example that applies the approach to select and ultimate mortality.

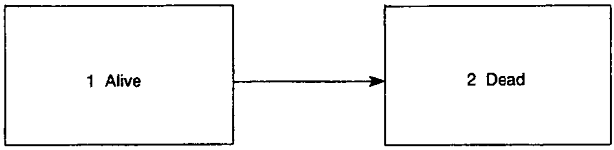
1. INTRODUCTION

1.1 Background

Some traditional problems in actuarial mathematics are conveniently viewed in terms of multistate processes. We assume that, at any time, an individual is in one of a number of states. The individual's presence in a given state or movement (transition) from one state to another may have some financial impact. Our task then is to quantify this impact, usually by estimating the expected value of future cash flows.

The simplest situation involves only two states: "alive" and "dead." As shown in Figure 1, an individual may make only one transition. For

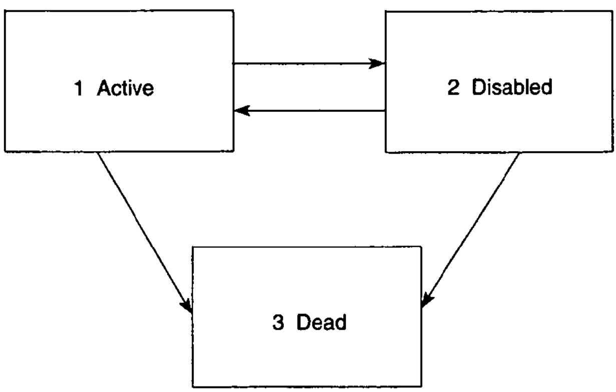
FIGURE 1
EXAMPLE OF TWO-STATE MODEL



a simple life annuity, benefits are payable while the annuitant is in state 1 and cease upon transition to state 2. In the case of a whole life insurance policy, premiums are payable while the insured is in state 1, and the death benefit is paid at the time of transition to state 2. Approaches to calculating actuarial values in these cases are simple and well-known (see Bowers et al. [1]).

A more complicated situation arises for processes with additional states. Figure 2 illustrates the three-state process commonly used to describe the state of an individual insured under a disability income policy. In this case, premiums are payable while the insured is in state 1, and benefits are payable while the insured is in state 2 (usually after a waiting period). Actuarial calculations for this example are more difficult because the individual can make repeat visits to each of states 1 and 2.

FIGURE 2
EXAMPLE OF THREE-STATE MODEL



For this reason it is often assumed that transitions from state 2 to state 1 are not possible.

A multistate model provides an intuitively pleasing description of the possible outcomes in numerous other areas. In examining a long-term-care system, we can represent the several levels of care available as states of a multistate model. Ongoing costs could then be associated with each state. We could use a multistate process in a life insurance context to describe the movement of individuals among various risk categories such as smoking status and blood pressure grouping; this is discussed by Tolley and Manton [23]. Pension plans can also be modeled within a multistate framework. In the simplest case, we would require states for working plan members, retirees, and those who have died. A more complicated model might require a disabled state and three retired states that reflect the status of a joint and last survivor annuity.

Many authors have used multistate models to analyze actuarial problems. Much of this work has drawn on the theory of stochastic processes to obtain new results of interest and to generalize results of more traditional methods. Such models are most tractable when it is assumed that the process satisfies the Markov property. Under this assumption, Hoem [5, 7] generalizes a number of standard results from life contingencies. Wolthuis and van Hoek [27] consider the expected value and variance of the loss function in a Markov model setting. The stochastic properties of the profit earned on an insurance policy are examined by Ramlau-Hansen [17, 18], who also analyzes the distribution of surplus [19]. Tolley and Manton [23] propose models for morbidity and mortality that include various risk factors in the model state space. In modeling the mortality of individuals infected with the HIV virus, Panjer [12] and Ramsay [16] use a Markov process with states that represent the stages of infection. Waters [24] discusses the development of formulas for probabilities and the estimation of parameters in a Markov model. The use of more general stochastic models is considered by Hoem [6], Hoem and Aalen [8], Ramsay [15], Seal [21], and Waters [25, 26].

1.2 Actuarial Calculations

Premiums and reserves for insurance and annuity contracts that are long term are usually based on the present value of payments to be made under the contract (both premiums and benefits). Typically, the occurrence, timing and/or amount of each payment is not known exactly in advance; it depends on some random outcome. Thus, it is customary to calculate the expectation of this present value. Such expected present values are easily obtained in the two-state case described earlier, since the random outcome can be represented as a single random variable, $T(x)$, measuring the time until death of an individual currently age x. For example, the expected present value of a continuous annuity paying 1 per annum for the remaining lifetime of an individual aged x is

$$E[\bar{a}_{\overline{T(x)}|}] = \int_0^\infty \bar{a}_{\overline{t}|}\, {}_tp_x\, \mu_{x+t}\, dt = \int_0^\infty v^t\, {}_tp_x\, dt.$$

The expected present value of 1 payable upon the death of an individual aged x is

$$E[v^{T(x)}] = \int_0^\infty v^t\, {}_tp_x\, \mu_{x+t}\, dt.$$

In the three-state case shown in Figure 2, the expected present value calculation is more difficult. Suppose we seek the expected present value of 1 payable continuously while in state 2 to an individual currently age x in state 1. This can be written

$$\int_0^\infty v^t\, p_{12}(x, x+t)\, dt,$$

where $p_{ij}(x, x+t)$ is the probability that an individual currently age x in state i will be in state j at age $x+t$. Unfortunately, this probability is not easy to obtain because it must allow for the possibility that the individual returns to state 1 one or more times between ages x and $x+t$. If we make the simplifying assumption that transitions from state 2 to state 1 cannot occur, then

$$p_{12}(x, x+t) = \int_0^t e^{-\int_0^s (\mu_{x+u}^{12} + \mu_{x+u}^{13}) du}\, \mu_{x+s}^{12}\, e^{-\int_s^t \mu_{x+u}^{23} du}\, ds,$$

where μ_y^{ij} is the force of transition from state i to state j at age y. The integrand in this expression can be interpreted as the probability that an individual in state 1 at age x moves to state 2 between age $x+s$ and $x+s+ds$ and remains in state 2 until age $x+t$. This "no recovery" assumption is often made in analyzing long-term disability insurance. However, for coverages in which transitions occur more frequently, such an assumption is inappropriate. We therefore need a general method of finding the probabilities $p_{ij}(t,x+t)$ in multistate models with three or more states.

Keyfitz and Rogers [9] provide a method for determining transition probabilities under a Markov process. The approach was developed by assuming that forces of transition are constant within age intervals of a fixed length. Transition probability matrices can then be calculated recursively for time periods that are multiples of this age interval. My method leads to expressions for the transition probabilities that are more convenient for certain types of calculations. I also provide a strategy for dealing with duration dependence, an issue not considered by Keyfitz and Rogers.

1.3 Outline of Paper

This paper presents a method for finding probabilities needed for actuarial calculations in applications that can be represented as multistate processes. Earlier research has focused on finding theoretical results under various multistate model assumptions. This paper provides techniques for calculating numerical results that are easily applied to a wide variety of problems. The approach is suitable for situations involving an arbitrary, but finite number of states. Thus, it may be a useful tool in analyzing complicated actuarial problems such as those presented by disability income insurance and long-term-care insurance.

Section 2 begins with a brief review of the properties of the Markov process. I then present a key result that exploits the mathematical tractability of Markov processes with constant forces of transition. I use a decomposition of the force of transition matrix that leads to a convenient representation of the transition probability matrix. The latter is expressed explicitly in terms of the time interval of interest. Furthermore, the probabilities are linear combinations of exponential functions. Therefore, the integration needed to compute expected sojourn times in the various states as well as actuarial values can be carried out analytically. This is also true when the forces of transition are piecewise constant. Thus, my approach avoids the numerical integration required by the method of Keyfitz and Rogers [9].

The idea of "duration dependence" is discussed in Section 3. Frequently actuaries encounter applications in which the forces of transition should depend on the time since entry to the current state. Such a process is called "semi-Markov." Unfortunately, the probabilities we seek are not easily obtained when using a semi-Markov model. However, by creating additional states, we can construct a Markov model that approximates the semi-Markov model. This approximation is justified by a limiting result that illustrates the convergence of the approximating Markov process to the semi-Markov process. I also provide a numerical example that allows for the duration dependence involved with select and ultimate mortality by including a third state. The two "alive" states are interpreted as "select" and "ultimate." This three-state Markov model with piecewise constant forces of transition yields probabilities that are very close to those obtained directly from the mortality table upon which parameter values were based.

Section 4 closes the paper with a brief summary.

2. THE MARKOV PROCESS

2.1 Basic Properties

As discussed in the previous section, we consider actuarial problems in which the cash flows depend on the outcome of a multistate process. To begin, let $X(t)$ represent the state of an individual at time (age) $t \geq 0$. We then denote the stochastic process by $\{X(t), t \geq 0\}$. We assume that there are a finite number of states labeled 1, 2, ..., k; that is, the process has state space $\{1, 2, \ldots, k\}$. Now, as defined by Ross [20, ch. 5], $\{X(t), t \geq 0\}$ is a Markov process if, for all $s, t \geq 0$ and $i, j, x(u) \in \{1, 2, \ldots, k\}$,

$$\begin{aligned}\Pr\{X(s+t) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s\} \\ = \Pr\{X(s+t) = j \mid X(s) = i\}.\end{aligned}$$

Thus, the future of the process (after time s) depends only on the state at time s and not on the history of the process up to time s.

The reasonableness of the Markov assumption depends somewhat on the level of detail in the state description. For example, consider the three-state process shown earlier in Figure 2. In this case, the Markov assumption may be inappropriate. The future health of a recently disabled individual is likely to differ from that of someone of the same age who has been disabled for a long time; this is discussed further in Section 3.

We define the transition probability function

$$p_{ij}(s, s+t) \equiv \Pr\{X(s+t) = j \mid X(s) = i\}, \; i, j \in \{1, 2, \ldots, k\},$$

and assume that

$$\sum_{j=1}^{k} p_{ij}(s, s+t) = 1 \text{ for all } t \geq 0.$$

We also assume the existence of the limits

$$\mu_{ij}(t) = \lim_{h \to 0^+} \frac{p_{ij}(t, t+h) - \delta_{ij}}{h}, \; i, j \in \{1, 2, \ldots, k\},$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

If $i \neq j$, $\mu_{ij}(t)$ represents the force of transition from state i to state j. It is easily seen that, for s, t, $u \geq 0$,

$$p_{ij}(s, s+t+u) = \sum_{l=1}^{k} p_{il}(s, s+t)\, p_{lj}(s+t, s+t+u),$$

$$i, j \in \{1, 2, \ldots, k\}. \tag{1}$$

These are known as the Chapman-Kolmogorov equations.

The transition probability functions are needed in the calculation of actuarial values. The forces of transition and the transition probability functions are related by the Kolmogorov forward and backward equations, which are

$$\frac{\partial}{\partial t} p_{ij}(s, s+t) = \sum_{l=1}^{k} p_{il}(s, s+t)\, \mu_{lj}(s+t), \tag{2}$$

and

$$\frac{\partial}{\partial s} p_{ij}(s, s+t) = -\sum_{l=1}^{k} \mu_{il}(s)\, p_{lj}(s, s+t), \tag{3}$$

respectively, with boundary conditions $p_{ij}(s, s) = \delta_{ij}$. In general, these systems of differential equations must be solved numerically to obtain the transition probability functions—a very tedious task.

2.2 Constant Forces of Transition

Explicit expressions for the transition probability functions are available when we assume that $\mu_{ij}(t) = \mu_{ij}$ for all t. Such a Markov process is referred to as time-homogeneous or stationary. The assumption of constant forces of transition implies that the time spent in each state is exponentially distributed. Also, the functions $p_{ij}(s, s+t)$ are the same for all $s \geq 0$ and therefore can be written $p_{ij}(t)$.

It is convenient to express the forces of transition and transition probability functions in matrix form. Let Q be the $k \times k$ matrix with (i, j) entry μ_{ij} and $P(t)$ be the $k \times k$ matrix with (i, j) entry $p_{ij}(t)$. Corresponding to Equation (1), the Chapman-Kolmogorov equations are given by

$$P(t + u) = P(t)\, P(u). \tag{4}$$

Also, corresponding to Equations (2) and (3), the Kolmogorov differential equations can be written

$$P'(t) = P(t)Q \tag{5}$$

and

$$P'(t) = QP(t), \tag{6}$$

with boundary condition $P(0) = I$. Equations (5) and (6) have the solution

$$\begin{aligned} P(t) &= e^{Qt} \\ &= I + Qt + \frac{Q^2 t^2}{2!} + \ldots \end{aligned}$$

This is of limited use because the series may converge rather slowly. However, as noted by Cox and Miller [3], if Q has distinct eigenvalues, $d_1, d_2, \ldots, d_k$, then $Q = ADC$ where $C = A^{-1}$, $D = \text{diag}(d_1, \ldots, d_k)$, and the i-th column of A is the right-eigenvector associated with d_i. Furthermore,

$$P(t) = A \,\text{diag}(e^{d_1 t}, \ldots, e^{d_k t})\, C. \tag{7}$$

Therefore, the problem of finding the transition probability functions is reduced to a problem of determining the eigenvalues and eigenvectors of the force of transition matrix Q. Software for performing this task is readily available. The requirement that Q have distinct eigenvalues imposes no practical restriction. In the situations we consider, this will be the case for almost all parameter values.

We can illustrate this in the case of the three-state model shown in Figure 2. We have

$$Q = \begin{bmatrix} -(\mu_{12} + \mu_{13}) & \mu_{12} & \mu_{13} \\ \mu_{21} & -(\mu_{21} + \mu_{23}) & \mu_{23} \\ 0 & 0 & 0 \end{bmatrix}.$$

The eigenvalues of Q are the solutions of

$$d[d^2 + (\mu_{12} + \mu_{13} + \mu_{21} + \mu_{23})d + \mu_{12}\mu_{23} + \mu_{13}\mu_{21} + \mu_{13}\mu_{23}] = 0.$$

Clearly, 0 is an eigenvalue. Neither of the other two eigenvalues can be 0, because this would require at least one force of mortality to be 0. Thus, if the three eigenvalues are not distinct, the quadratic in square brackets must have only one root. That is,

$$(\mu_{12} + \mu_{13} + \mu_{21} + \mu_{23})^2 - 4(\mu_{12}\mu_{23} + \mu_{13}\mu_{21} + \mu_{13}\mu_{23}) = 0.$$

This implies that, for any choice of three parameter values, the fourth must satisfy a quadratic equation. Therefore, at most two values of the fourth parameter will result in eigenvalues that are not distinct. It is then quite unlikely that the parameter estimates will result in nondistinct eigenvalues. In the event that this occurs, a slight change in one parameter will eliminate the problem.

If, in dealing with more complicated models, distinct eigenvalues cannot be achieved, an analogous decomposition to Jordan canonical form is possible (see Cox and Miller [13]).

From Equation (7), we can now write

$$p_{ij}(t) = \sum_{n=1}^{k} a_{in}\, c_{nj}\, e^{d_n t}, \tag{8}$$

where a_{ij} and c_{ij} are the (i, j) entries of A and C, respectively. Thus, we can express the transition probability functions explicitly as simple functions of t.

Note that in the two-state case shown in Figure 1, we have

$$Q = \begin{bmatrix} -\mu_{12} & \mu_{12} \\ 0 & 0 \end{bmatrix}.$$

Then $d_1 = -\mu_{12}$ and $d_2=0$. Corresponding eigenvectors are $(1, 0)'$ and $(1, 1)'$. Thus,

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

and

$$C = A^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

From Equation (7), we then have

$$\begin{aligned} P(t) &= A \operatorname{diag}(e^{d_1 t}, e^{d_2 t})C \\ &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^{-\mu_{12} t} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} e^{-\mu_{12} t} & 1 - e^{-\mu_{12} t} \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Hence, as expected, $p_{11}(t)= e^{-\mu_{12} t}$; $p_{12}(t)=1- e^{-\mu_{12} t}$; $p_{22}(t)=1$, and $p_{21}(t)=0$.

2.3 Piecewise Constant Forces

We assumed above that forces of transition were constant with respect to time. This permits convenient representation of the transition probability functions. Unfortunately, in many actuarial applications, this is impractical. We require forces that vary with age. In the two-state example shown in Figure 1, we clearly need a force of mortality that varies with the age of the individual. We can accomplish this, while preserving the tractability of constant forces, by using force of transition functions that are piecewise constant. We may wish to use forces of transition that vary with each year of age. In some instances, though, it will be reasonable to use broader age groups.

Let $\mu_{ij}(t)=\mu_{ij}^{(m)}$ if $t \in [t_{m-1}, t_m)$, for $m=1, 2, \ldots$, where $t_0=0$. Also, let $p_{ij}^{(m)}(t)$ be the transition probability function associated with time intervals $[u, u+t)$ contained in $[t_{m-1}, t_m)$. In matrix form, we have $Q^{(m)}$ and $P^{(m)}(t)$. Now define m_t to be the integer such that $t_{m_t-1} \le t < t_{m_t}$. Then from Equation (4), we have

$$P(s, t) = P^{(m_s)}(t_{m_s} - s)\, P^{(m_s+1)}(t_{m_s+1} - t_{m_s}) \ldots P^{(m_t)}(t - t_{m_t-1}). \quad (9)$$

Thus, given s and t, the transition probability matrix can be computed. We first determine $A^{(m)}$, $D^{(m)}$, and $C^{(m)}=(A^{(m)})^{-1}$ from $Q^{(m)}$ for each m, as described in Subsection 2.2; $P^{(m)}(t)$ is then obtained by using Equation (7). Finally, $P(s, t)$ can be found by using Equation (9).

As mentioned earlier, an advantage of the approach described in this paper is that transition probability functions are expressed in a very convenient form. To obtain quantities such as the expected time spent in a given state or the expected present value of payments made continuously while in a given state, the required integration can be performed analytically.

To illustrate this, consider a Markov process with forces of transition that are constant within each year of age. Let $\mu_{ij}^{(x)}$ be the force of transition from state i to state j for an individual between age x and age $x+1$, where x is a nonnegative integer and $i, j \in \{1, 2, \ldots, k\}$. These forces of transition can be used to construct the matrix $Q^{(x)}$, from which we can determine $A^{(x)}$, $C^{(x)}$, and $D^{(x)}$. Let $p_{ij}^{(x)}(t)$ be the i to j transition probability function associated with the age interval from x to $x+1$. Suppose we wish to determine the expected time spent in state j between ages 30 and 40 by an individual in state i at age 30. This quantity is given by

$$\int_{30}^{40} p_{ij}(30, t)\, dt = \sum_{x=30}^{39} \int_{x}^{x+1} p_{ij}(30, t)\, dt$$

$$= \sum_{x=30}^{39} \int_{x}^{x+1} \sum_{h=1}^{k} p_{ih}(30, x)\, p_{hj}^{(x)}(t - x)\, dt$$

from Equation (1)

$$= \sum_{x=30}^{39} \sum_{h=1}^{k} p_{ih}(30, x) \int_{x}^{x+1} \sum_{n=1}^{k} a_{hn}^{(x)}\, c_{nj}^{(x)}\, e^{d_n^{(x)}(t-x)}\, dt$$

from Equation (8)

$$= \sum_{x=30}^{39} \sum_{h=1}^{k} p_{ih}(30, x) \sum_{n=1}^{k} a_{hn}^{(x)}\, c_{nj}^{(x)} \int_{x}^{x+1} e^{d_n^{(x)}(t-x)}\, dt$$

$$= \sum_{x=30}^{39} \sum_{h=1}^{k} p_{ih}(30, x) \sum_{n=1}^{k} a_{hn}^{(x)}\, c_{nj}^{(x)}\, \frac{e^{d_n^{(x)}} - 1}{d_n^{(x)}}, \tag{10}$$

where $p_{ih}(30,x)$ is the (i, h) entry of

$$P(30, x) = \prod_{y=30}^{x-1} P^{(y)}(1)$$

$$= \prod_{y=30}^{x-1} A^{(y)} \operatorname{diag}(e^{d_1^{(y)}}, \ldots, e^{d_k^{(y)}}) C^{(y)};$$

$a_{hn}^{(x)}$ is the (h, n) entry of $A^{(x)}$; $c_{nj}^{(x)}$ is the (n, j) entry of $C^{(x)}$; and $d_n^{(x)}$ is the (n, n) entry of $D^{(x)}$. In matrix form, Equation (10) can be written as

$$\int_{30}^{40} P(30, t)dt = \sum_{x=30}^{39} P(30, x)\, A^{(x)} \operatorname{diag}\left(\frac{e^{d_1^{(x)}}-1}{d_1^{(x)}}, \ldots, \frac{e^{d_k^{(x)}}-1}{d_k^{(x)}}\right) C^{(x)}.$$

We also find that the expected present value of 1 payable continuously while in state j between ages 30 and 40 to an individual now in state i at age 30 is

$$\int_{30}^{40} e^{-\delta(t-30)} p_{ij}(30, t)dt = \sum_{x=30}^{39} e^{-\delta(x-30)} \sum_{h=1}^{k} p_{ih}(30, x) \sum_{n=1}^{k} a_{hn}^{(x)} c_{nj}^{(x)} \frac{e^{d_n^{(x)}-\delta} - 1}{d_n^{(x)} - \delta},$$

where δ is the force of interest. In matrix form,

$$\int_{30}^{40} e^{-\delta(t-30)} P(30, t)dt = \sum_{x=30}^{39} e^{-\delta(x-30)} P(30, x)\, A^{(x)}$$

$$\operatorname{diag}\left(\frac{e^{d_1^{(x)}-\delta}-1}{d_1^{(x)}-\delta}, \ldots, \frac{e^{d_k^{(x)}-\delta} - 1}{d_k^{(x)} - \delta}\right) C^{(x)}.$$

3. DURATION DEPENDENCE

3.1 Failure of the Markov Assumption

In the previous section I mentioned that, for some actuarial applications of multistate models, the Markov assumption is unsuitable. I cited the three-state disability case as a situation in which the probability of transition out of the disabled state may be influenced by the time since disablement as well as the age of the individual. Another example in which duration dependence arises is select mortality. In life insurance, it is generally assumed that mortality rates depend on the time since the individual became insured in addition to the individual's attained age.

Lapse rates are also heavily dependent on the duration of the insurance policy.

A stochastic model in which the future of the process depends on the time since transition to the current state is referred to as semi-Markov. Hoem [6] discusses a number of demographic and actuarial applications of semi-Markov models. Some aspects of the use of such models in sickness insurance are considered by Seal [21], Ramsay [15], and Waters [25, 26].

3.2 The General Semi-Markov Model

We can describe the general semi-Markov model in terms of the forces of transition. Let $\mu_{ij}(t, u)$ be the force of transition from state i to state j at time (age) t for an individual who has been in state i for a period of time u. We then require a more complicated definition of the transition probability functions, also involving the time since entry to state i. Hoem [6] defines these functions and points out a number of useful relationships.

Unfortunately, such a complicated stochastic model does not lead to convenient expressions for the probabilities needed to obtain actuarial values. Because this is the objective of this paper, we must seek some simplification. Seal [21] achieves such simplification in modeling the time spent in sickness of young and middle-aged individuals by restricting the model to two states. Because mortality rates are low at these ages, Seal assumes that mortality transitions can be ignored. He further assumes that the forces of transition to and from the sickness state are independent of attained age and, hence, depend only on the time since entry to the current state. The resulting stochastic process is called an alternating renewal process. The same model is used by Ramsay [15].

We wish to deal with more general multistate situations and therefore require some other form of simplification of the model. We suggest an approach that allows us to use the results for the Markov process discussed in the previous section. In particular, we propose the acceptance of a more complicated state space in exchange for the simpler Markov process.

3.3 Approximation by a Markov Model

An alternative approach to reflecting the duration dependence often present in actuarial applications is to treat each state as a collection of one or more substates. We then assume that future transitions are independent of the time of entry to the current substate. For example, we might assume that the "insured" state in a life insurance situation consists of two substates: "select" and "ultimate." This is suggested by Norberg [11] as a way of explaining select mortality. Norberg shows that, if

1. Only select lives may enter the insured state
2. Select lives may move to the ultimate state
3. Ultimate lives may not return to the select state
4. The force of mortality for select lives is less than that for ultimate lives,

then, for a fixed attained age, the force of mortality increases with duration since becoming insured. Møller [10] shows that the result also holds if assumption 3 is relaxed, and he explores the selection effect using more than one ultimate state.

The same approach could be used in the disability insurance model illustrated in Figure 2. Here we must allow the force of transition from the disabled state to depend on the time since disablement. To accomplish this, we can represent the disabled state by two substates, which might be interpreted as "unstable" and "stable." The unstable state would be the state entered upon disablement and would have fairly high forces of transition both to the active state and to the dead state. An individual could also move from unstable to stable, a state with lower forces of recovery and mortality. If necessary, the disabled state could comprise more than two substates.

The approximation of a semi-Markov process by a Markov process is discussed by Cox and Miller [3]. In particular, the method of stages allows the approximation of any failure time distribution by a combination of stages in series or parallel, where the time spent in each stage is exponentially distributed. Below we take such an approach in developing a limiting result that justifies the approximation in the context of select and ultimate mortality. The result can be generalized to more complicated processes.

Let $\mu(t)$, $t \geq 0$ be a bounded continuous function representing the force of mortality for a given age group, where t is the time since policy issue. Also, let

$$S(t) = e^{-\int_0^t \mu(s)ds}$$

be the corresponding survival function. We can construct a time-homogeneous Markov process with a survival function that converges to

$S(t)$ as the rate of transition from each state approaches infinity. To begin, suppose that an individual moves through a number of states labeled 1, 2, ... in sequence until death occurs. Let the total force of transition from each state (mortality and movement to next state) be given by λ at each point in time. Also, let the force of mortality in state i be $\mu(i/\lambda)$. Then we must have $\lambda > \mu(i/\lambda)$ for all i. Furthermore, $\lambda - \mu(i/\lambda)$ is the force of transition from state i to $i+1$. Note that $\mu(i/\lambda)$ equals the true force of mortality at a duration equal to the expected time of transition from state i.

Since the total force of transition from each state is λ, the Markov process can be thought of as a Poisson process in which, at each event time, the individual either dies or moves to the next state. Let $N(t)$ represent the number of events in $(0, t]$ and let $I(t)$ be 1 if the individual is alive at duration t and 0 otherwise. Then

$$\Pr\{N(t) = n\} = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$

Also,

$$\Pr\{I(t) = 1 | N(t) = n\} = \begin{cases} 1 & n = 0 \\ \displaystyle\prod_{i=1}^{n} \frac{\lambda - \mu(i/\lambda)}{\lambda} & n = 1, 2, \ldots . \end{cases}$$

Therefore,

$$\Pr\{I(t) = 1, N(t) = n\} = \begin{cases} e^{-\lambda t} & n = 0 \\ \displaystyle\frac{(\lambda t)^n e^{-\lambda t}}{n!} \prod_{i=1}^{n} \frac{\lambda - \mu(i/\lambda)}{\lambda} & n = 1, 2, \ldots \end{cases}$$

and

$$\Pr\{I(t) = 1\} = e^{-\lambda t} + \sum_{n=1}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} \prod_{i=1}^{n} \frac{\lambda - \mu(i/\lambda)}{\lambda}. \tag{11}$$

We now note that

$$\log\left\{\prod_{i=1}^{n}\frac{\lambda-\mu(i/\lambda)}{\lambda}\right\}=\sum_{i=1}^{n}\log\left\{\frac{\lambda-\mu(i/\lambda)}{\lambda}\right\}$$

$$=-\sum_{i=1}^{n}\frac{\mu(i/\lambda)}{\lambda}+o(1/\lambda)$$

$$=-\int_{0}^{n/\lambda}\mu(s)ds+o(1/\lambda)$$

by the definition of a Riemann integral

$$=\log\{S(n/\lambda)\}+o(1/\lambda).$$

It follows that

$$\prod_{i=1}^{n}\frac{\lambda-\mu(i/\lambda)}{\lambda}=e^{\log\{S(n/\lambda)\}+o(1/\lambda)}$$

$$=S(n/\lambda)\,e^{o(1/\lambda)}$$

$$=S(n/\lambda)\,\{1+o(1/\lambda)\}$$

$$=S(n/\lambda)+o(1/\lambda). \tag{12}$$

This is a special case of a more general result given by Pólya and Szegö [14, p. 47]. Now since

$$\sum_{n=0}^{\infty}\frac{(\lambda t)^{n}e^{-\lambda t}}{n!}$$

is bounded for all λ, from Equations (11) and (12) we have

$$\lim_{\lambda\to\infty}\Pr\{I(t)=1\}=\lim_{\lambda\to\infty}\sum_{n=0}^{\infty}\frac{(\lambda t)^{n}\,e^{-\lambda t}}{n!}\,S(n/\lambda)$$

$$=\lim_{\lambda\to\infty}E[S(N(t)/\lambda)].$$

Now consider the sequence $\{\lambda_r;\ r=1,2,\ldots\}$, where $\lambda_r=r/t$. For a given r, $N(t)/(\lambda_r t)$ can be viewed as the sample mean of r independent Poisson random variables with mean 1. By the weak law of large numbers it follows that $N(t)/(\lambda t)\to 1$ in probability as $\lambda\to\infty$. Hence,

$N(t)/\lambda \rightarrow t$ in probability as $\lambda \rightarrow \infty$. Furthermore, since $S(\cdot)$ is continuous, $S(N(t)/\lambda) \rightarrow S(t)$ in probability as $\lambda \rightarrow \infty$. Finally, since $S(\cdot)$ is bounded, this implies that $E[|S(N(t)/\lambda) - S(t)|] \rightarrow 0$ as $\lambda \rightarrow \infty$. Therefore,

$$\lim_{\lambda \rightarrow \infty} \Pr\{I(t) = 1\} = S(t).$$

By a similar construction, it is possible to approximate a more general process involving transitions other than death. In this case we have $\mu(t) = \Sigma_j \mu_j(t)$, where $\mu_j(t)$ is the force of transition to state j at duration t.

To examine the select and ultimate mortality model more closely, consider the setup shown in Figure 3. Suppose that the μ_{ij} represent forces of transition for some (attained) age group. According to this setup, the probability of surviving a period of time, t, is

$$\begin{aligned} p_{11}(t) + p_{12}(t) &= e^{-(\mu_{12}+\mu_{13})t} + \int_0^t e^{-(\mu_{12}+\mu_{13})x}\, \mu_{12} e^{-\mu_{23}(t-x)}\, dx \\ &= e^{-(\mu_{12}+\mu_{13})t} + \frac{\mu_{12}\, e^{-\mu_{23}t}\, [1 - e^{-(\mu_{12}+\mu_{13}-\mu_{23})t}]}{\mu_{12} + \mu_{13} - \mu_{23}} \\ &= \frac{(\mu_{13} - \mu_{23})\, e^{-(\mu_{12}+\mu_{13})t} + \mu_{12} e^{-\mu_{23}t}}{\mu_{12} + \mu_{13} - \mu_{23}}. \end{aligned}$$

This is clearly a weighted average of the survival functions associated with two exponential distributions. The corresponding force of mortality is

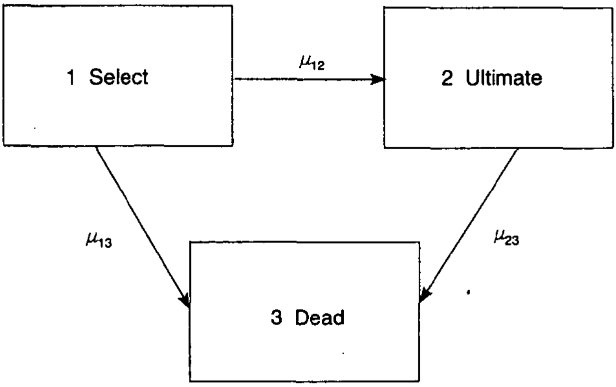
$$\begin{aligned} \mu(t) &= -\frac{\frac{d}{dt}[p_{11}(t) + p_{12}(t)]}{p_{11}(t) + p_{12}(t)} \\ &= \frac{(\mu_{23} - \mu_{13})(\mu_{12} + \mu_{13})\, e^{-(\mu_{12}+\mu_{13})t} - \mu_{12}\mu_{23} e^{-\mu_{23}t}}{(\mu_{23} - \mu_{13})\, e^{-(\mu_{12}+\mu_{13})t} - \mu_{12} e^{-\mu_{23}t}} \end{aligned} \tag{13}$$

Thus, μ_{12}, μ_{13} and μ_{23} should be chosen so that Equation (13) best represents the selection effect for this age group.

We find that, for any choice of the three parameter values, there is a second choice that produces exactly the same $\mu(t)$. That is, if $\mu_{12} = \hat{\mu}_{12}$, $\mu_{13} = \hat{\mu}_{13}$, and $\mu_{23} = \hat{\mu}_{23}$, then we can achieve the same $\mu(t)$ by letting

FIGURE 3

MODEL FOR SELECT MORTALITY



$\tilde{\mu}_{12}=\hat{\mu}_{23}-\hat{\mu}_{13}$, $\tilde{\mu}_{13}=\hat{\mu}_{13}$, and $\tilde{\mu}_{23}=\hat{\mu}_{12}+\hat{\mu}_{13}$. If we restrict our attention to the subset of the parameter space for which $\mu_{23}>\mu_{12}+\mu_{13}$ or $\mu_{23}<\mu_{12}+\mu_{13}$, then, for each $\mu(t)$, the parameterization is unique. In the absence of prior information about the parameter values, an arbitrary choice of subset may be made. Our objective is simply to find the best $\mu(t)$ based on this three-state setup. Ordinarily, we have no data on the three transitions shown in Figure 3, but only on transitions from states 1 and 2 combined to state 3.

Since the force of transition from state 1 to state 2 is likely to be quite large relative to the forces of mortality, it seems reasonable to assume that $\mu_{23}<\mu_{12}+\mu_{13}$. It follows from Equation (13) that $\lim_{t\to\infty} \mu(t)=\mu_{23}$. We also have $\mu(0)=\mu_{13}$. For $0<t<\infty$, $\mu(t)$ is weighted average of the select force, μ_{13}, and the ultimate force, μ_{23}. The weights are the conditional probabilities of being in the select and ultimate states at duration t given survival to duration t. Tenenbein and Vanderhoof [22] also modeled the force of mortality as an average of select and ultimate forces. However, in their models, the select and ultimate proportions at each duration are deterministic.

3.4 Numerical Example

To illustrate the procedure described above, we use the three-state model shown in Figure 3 to describe select mortality. Parameter values are obtained so that the model reflects the male aggregate mortality in the 1982–1988 Individual Ordinary Mortality Table published by the Canadian Institute of Actuaries [2]. The development of this table is described by Panjer and Russo [13].

We assumed that the forces of transition are constant within each year of age. Thus, for the age range $[x, x+1]$, we used the various tabular mortality rates for attained age x to determine estimates of $\mu_{12}^{(x)}$, $\mu_{13}^{(x)}$, and $\mu_{23}^{(x)}$. We first calculated tabular forces of mortality from the mortality rates by assuming that, for $k=0, 1, \ldots, 14$,

$$\mu_{[x-k]+k+1/2}^{T} = -\log(1 - q_{[x-k]+k}^{T})$$

and

$$\mu_{x+1/2}^{T} = -\log(1 - q_x^{T}).$$

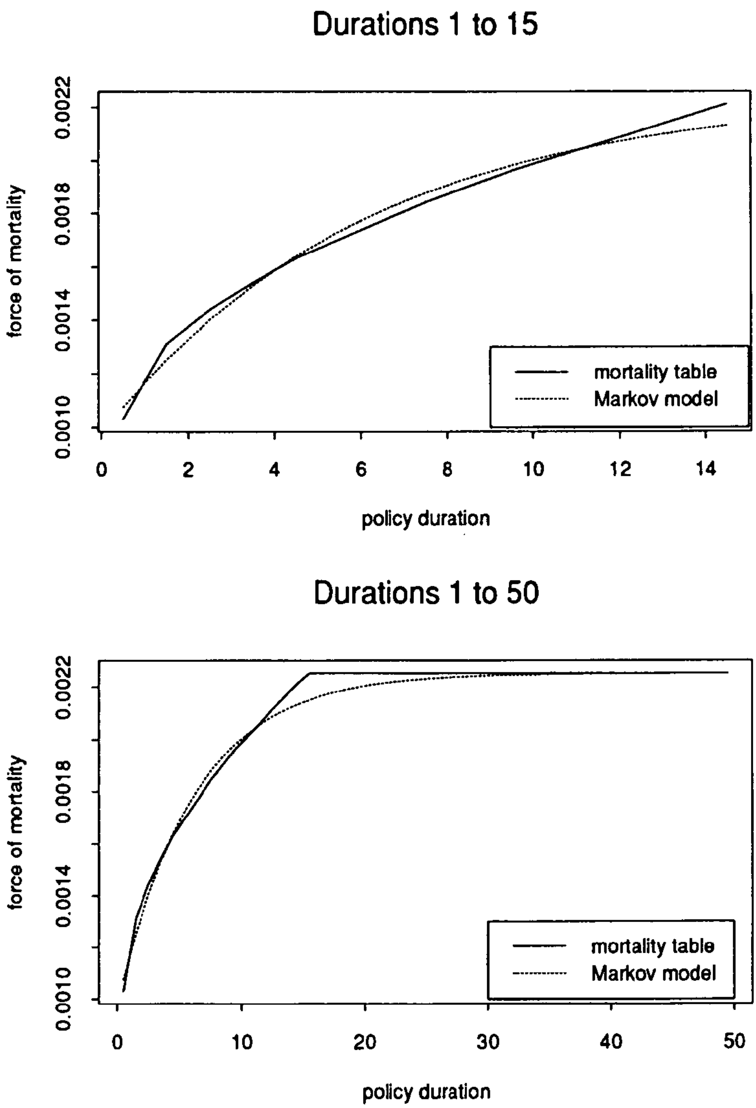
The T denotes a tabular value. Although the forces of mortality do not appear in the published table, we use the superscript to distinguish them from the forces of mortality that result from our Markov model.

As a function of policy duration (with attained age fixed), the tabular force of mortality increases during the 15-year select period and remains constant thereafter. Our corresponding "fitted" force of mortality, $\mu^{(x)}(t)$, determined by Equation (13), does not exhibit this behavior. It increases toward its limit as the policy duration approaches infinity. We let our estimate of $\mu_{23}^{(x)}$ equal the tabular ultimate force of mortality for age $x+1/2$. The estimates of $\mu_{12}^{(x)}$ and $\mu_{13}^{(x)}$ were obtained to minimize the squared deviations of $\mu^{(x)}(t)$ from the tabular forces of mortality at durations 0.5, 1.5, 2.5, ..., 14.5. That is, we minimized

$$\sum_{k=0}^{14} [\mu_{[x-k]+k+1/2}^{T} - \mu^{(x)}(k + 1/2)]^2.$$

The resulting $\mu^{(45)}(t)$ is plotted in Figure 4 along with $\mu_{[45-t]+t}$. The first graph shows only the select period. It demonstrates that, although we have used a model with just three parameters, our resulting force of mortality function is quite close to that based on the table at nearly all select durations. It is only near the end of the select period that our force of mortality appears to be significantly lower; this is quite noticeable in

FIGURE 4
COMPARISON OF FORCE OF MORATLITY (AGE 45)



the second graph. The reason is that our fitted force of mortality must have a much smoother path toward the ultimate level. The second graph also shows that the two curves are very close after 30 years and almost indistinguishable after 40. The results are similar for other attained ages.

The parameter estimates obtained for ages 45 through 70 are given in Table 1. As we might expect, there is very little variation in the estimates of $\mu_{12}^{(x)}$; all lie in the range from 0.163 to 0.174. The estimates of $\mu_{13}^{(x)}$ are slightly lower than the duration 1 tabular forces. As indicated above, the $\mu_{23}^{(x)}$ estimates are equal to the corresponding tabular ultimate forces.

TABLE 1

FORCES OF TRANSITION

x	$\mu_{12}^{(x)}$	$\mu_{13}^{(x)}$	$\mu_{23}^{(x)}$	x	$\mu_{12}^{(x)}$	$\mu_{13}^{(x)}$	$\mu_{23}^{(x)}$
45	0.164	0.00097	0.00225	58	0.165	0.00304	0.00933
46	0.164	0.00107	0.00251	59	0.165	0.00329	0.01036
47	0.163	0.00117	0.00280	60	0.167	0.00352	0.01150
48	0.163	0.00128	0.00313	61	0.167	0.00380	0.01274
49	0.163	0.00140	0.00350	62	0.167	0.00410	0.01411
50	0.163	0.00154	0.00391	63	0.167	0.00442	0.01560
51	0.163	0.00168	0.00437	64	0.168	0.00472	0.01725
52	0.164	0.00183	0.00488	65	0.169	0.00503	0.01905
53	0.164	0.00201	0.00544	66	0.169	0.00540	0.02102
54	0.164	0.00218	0.00608	67	0.170	0.00574	0.02318
55	0.164	0.00238	0.00677	68	0.171	0.00609	0.02553
56	0.164	0.00259	0.00755	69	0.173	0.00638	0.02811
57	0.164	0.00282	0.00840	70	0.174	0.00674	0.03093

The parameter estimates shown in Table 1 can be used to find various probabilities of interest. For the age range $(x, x+1)$, the force of transition matrix is

$$\begin{bmatrix} -(\mu_{12}^{(x)} + \mu_{13}^{(x)}) & \mu_{12}^{(x)} & \mu_{13}^{(x)} \\ 0 & -\mu_{23}^{(x)} & \mu_{23}^{(x)} \\ 0 & 0 & 0 \end{bmatrix}.$$

Upon finding the corresponding eigenvalues and right-eigenvectors, we can obtain the transition probability matrix for time intervals within this age range by using Equation (7). If this procedure is repeated for all x, then we can use Equation (9) to determine the transition probability matrix for any age range.

Table 2 shows survival probabilities obtained in this manner. The numbers represent the probability that an individual issued insurance at age 45 survives each of the next 50 years, that is, $1-p_{13}$ (45, 45+t). These are compared to the probabilities ${}_tp^T_{[45]}$, determined from the mortality table. Table 2 indicates that the survival probabilities obtained by using a simple three-state Markov model are very close to those obtained directly from the mortality table.

TABLE 2

COMPARISON OF SURVIVAL PROBABILITIES

t	${}_tp^T_{[45]}$	$1 - p_{13}(45, 45 + t)$	t	${}_tp^T_{[45]}$	$1 - p_{13}(45, 45 + t)$
1	0.9989700	0.9989312	14	0.9436032	0.9437097
2	0.9975215	0.9975511	15	0.9341106	0.9346207
3	0.9957659	0.9958410	16	0.9234337	0.9245411
4	0.9936747	0.9937680	17	0.9117430	0.9134101
5	0.9912204	0.9912964	18	0.8989695	0.9011527
6	0.9883657	0.9883858	19	0.8850534	0.8877119
7	0.9850546	0.9849978	20	0.8699190	0.8730093
8	0.9812228	0.9810886	21	0.8535037	0.8569861
9	0.9768171	0.9766096	22	0.8357508	0.8395813
10	0.9717767	0.9715020	23	0.8166037	0.8207379
11	0.9660141	0.9657210	24	0.7960171	0.8004175
12	0.9594548	0.9591973	25	0.7739516	0.7785747
13	0.9520191	0.9518781	26	0.7503770	0.7551810

Note that many other techniques are available for modeling select and ultimate mortality. Examples include those discussed by Currie and Waters [4], Panjer and Russo [13], and Tenenbein and Vanderhoof [22]. Select and ultimate mortality was discussed in this section because it is a simple case involving duration dependence. The techniques developed in this paper are most useful in dealing with applications requiring a greater number of states.

SUMMARY

This paper describes an approach whereby actuaries can determine probabilities required for calculations in applications that are represented as multistate processes. We have drawn on some very convenient mathematical results that are available when the process is assumed to be Markov with constant forces of transition (that is, a time-homogeneous Markov process). The extension to piecewise constant forces is straightforward. In cases involving duration dependence, rather than using the

less tractable semi-Markov process, it is possible to approximate the impact of duration by including additional states in the model.

REFERENCES

1. BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A., AND NESBITT, C.J. *Actuarial Mathematics*. Itasca, Ill.: Society of Actuaries, 1986.
2. CANADIAN INSTITUTE OF ACTUARIES. *1982–1988 Individual Ordinary Mortality Table*. Ottawa, 1992; Reprinted in *TSA 1991–92 Reports*, 701–70.
3. COX, D.R., AND MILLER, H.D. *The Theory of Stochastic Processes*. London: Chapman and Hall, 1965.
4. CURRIE, I.D., AND WATERS, H.R. ."On Modelling Select Mortality," *ARCH* 1990.1:85–101.
5. HOEM, J.M. "Markov Chain Models in Life Insurance," *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* IX (1969): 91–107.
6. HOEM, J.M. "Inhomogeneous Semi-Markov Processes, Select Actuarial Tables, and Duration-Dependence in Demography." In *Population Dynamics*, edited by T.N.E. Greville, 251–296. New York: Academic Press, 1972.
7. HOEM, J.M. "The Versatility of the Markov Chain as a Tool in the Mathematics of Life Insurance," *Record of Proceedings, International Congress of Actuaries*, Helsinki, Finland (1988): 171–202.
8. HOEM, J.M., AND AALEN, O.O. "Actuarial Values of Payment Streams," *Scandinavian Actuarial Journal* (1978): 38–47.
9. KEYFITZ, N., AND ROGERS, A. "Simplified Multiple Contingency Calculations," *Journal of Risk and Insurance* XLIX, no. 1 (1982): 59–72.
10. MØLLER, C. M. "Select Mortality and Other Durational Effects Modelled by Partially Observed Markov Chains," *Scandinavian Actuarial Journal* (1990): 177–99.
11. NORBERG, R. "Select Mortality: Possible Explanations," *Record of Proceedings, International Congress of Actuaries*, Helsinki, Finland (1988): 215–24.
12. PANJER, H.H. "AIDS: Survival Analysis of Persons Testing HIV+," *TSA* XL, Part I (1988): 517–30.
13. PANJER, H.H., AND RUSSO, G. "Parametric Graduation of Canadian Individual Insurance Mortality Experience: 1982–1988," *Proceedings of the Canadian Institute of Actuaries XXIII*, no. 2 (1992): 378–449.
14. PÓLYA, G., AND SZEGÖ, G. *Aufgaben und Lehrsätze aus der Analysis*. Berlin: Springer-Verlag, 1970.
15. RAMSAY, C.M. "The Asymptotic Ruin Problem When the Healthy and Sick Periods Form an Alternating Renewal Process," *Insurance: Mathematics and Economics* 3 (1984): 139–43.
16. RAMSAY, C.M. "AIDS and the Calculation of Life Insurance Functions," *TSA* XLI (1989): 393–422.

17. RAMLAU-HANSEN, H. "Hattendorf's Theorem: A Markov Chain and Counting Process Approach," *Scandinavian Actuarial Journal* (1988): 143–56.
18. RAMLAU-HANSEN, H. "The Emergence of Profit in Life Insurance," *Insurance: Mathematics and Economics* 7 (1988): 225–36.
19. RAMLAU-HANSEN, H. "Distribution of Surplus in Life Insurance," *ASTIN Bulletin* 21, no. 1 (1991): 57–71.
20. ROSS, S.M. *Stochastic Processes*. New York: Wiley, 1983.
21. SEAL, H.L. "Probability Distributions of Aggregate Sickness Durations," *Skandinavisk Aktuarietidskrift* 53 (1970): 193–204.
22. TENENBEIN, A., AND VANDERHOOF, I.T. "New Mathematical Laws of Select and Ultimate Mortality," *TSA* XXXII (1980): 119–58.
23. TOLLEY, H.D., AND MANTON, K.G. "Intervention Effects among a Collection of Risks," *TSA* XLIII (1991): 117–42.
24. WATERS, H.R. "An Approach to the Study of Multiple State Models," *Journal of the Institute of Actuaries* 111, part II (1984): 363–74.
25. WATERS, H.R. "Some Aspects of the Modelling of Permanent Health Insurance," *Journal of the Institute of Actuaries* 116, part III (1989): 611–24.
26. WATERS, H.R. "The Recursive Calculation of the Moments of the Profit on a Sickness Insurance Policy," *Insurance: Mathematics and Economics* 9 (1990): 101–13.
27. WOLTHUIS, H., AND VAN HOEK, I. "Stochastic Models for Life Contingencies," *Insurance: Mathematics and Economics* 5 (1986): 217–54.