ACTUARIAL RESEARCH CLEARING HOUSE
1996 VOL. 1

Analysis of CCRC Data

Bruce L. Jones
Department of Statistics and Actuarial Science
The University of Iowa
Iowa City, IA 52242

Abstract

This paper presents an approach to analyzing continuing care retirement community (CCRC) data, and demonstrates the methods using data from a large CCRC. It is assumed that residents make "transitions" among a number of "states" that represent the levels of care required by residents. There is randomness associated with both the transition times and the states entered at these times. The model is conveniently characterized in terms of "transition intensity functions", which represent the instantaneous rates of transition between pairs of states. Statistical methods for estimating these functions are discussed, and estimates are obtained from the data set. A simulation approach to determining probabilities and other interesting quantities based on the estimated intensity functions is also described and illustrated.

1 Introduction

1.1 Background on CCRCs

Continuing care retirement communities (CCRCs) offer housing and a wide range of services to elderly individuals. These services typically include daily meals, housekeeping, flat linen, maintenance of apartment and grounds, emergency nursing, security, scheduled transportation and activities (see Smith [22]). The most notable service from an actuarial perspective is long-term health care. CCRCs usually provide two or three levels of long-term care.

CCRCs generally charge a rather substantial entry fee as well as periodic fees paid throughout an individual's duration of residence. Additional fees may also be charged for some services. Many CCRC contracts provide for the refund of some portion of the entry fee in the event of death or withdrawal.

A key feature of most CCRC contracts is that some or all of the cost of long-term care is covered by the entry and periodic fees. Such contracts therefore provide a long-term care insurance benefit. For this reason, actuaries have a role to play in the financial management of CCRCs.

Further discussions of the characteristics of CCRCs and CCRC contracts are given by Brace [6], Hewitt and DeWeese [11], Rodermund [19, 20], Smith [22] and Winklevoss and Powell [23].

1.2 Actuarial Models for CCRCs

CCRCs offer a unique challenge for actuaries. Most communities provide two or more levels of care, and residents may transfer temporarily or permanently to the care units. Actuarial models must therefore permit "transitions" among a large number of "states", usually six or more. For pricing and valuation purposes, one should be able to estimate the probability that a resident is in any given state at any future time as well as the probability that a resident will move between any two states during any time interval. In order to perform cash flow and population projections, one should be able to estimate the expected number of residents in each state at any future time and the expected number of transitions between any two states during any future time interval. It is also important to be able to quantify the variation about these expected values.

Cumming and Bluhm [8] describe a CCRC population and financial model that uses a multiple decrement approach. Expected results can be calculated directly and random variation estimated by simulation. Jones [13, 14, 15] explores continuous-time multi-state stochastic models for analyzing CCRCs. Emphasis is on parsimonious models for which direct calculation is possible for many quantities of interest.

1.3 CCRC Data

Selection of parameter values for CCRC models is difficult. At present, there is no good source of CCRC industry data. Furthermore, the characteristics of CCRC residents may differ greatly between CCRCs. Some communities are like expensive resorts, affordable only to the wealthy. Others are much more modest with fees that reflect this. We expect to see differences in health care utilization between residents from different socio-economic classes. In addition, differing management philosophies with regard to resident transfers will affect CCRC experience. One should therefore choose carefully the data source upon which parameter values are to be based. Even if CCRC industry data were available, it would not necessarily be appropriate to use this data in choosing parameter values for a particular CCRC. Industry would combine the experience of a number of possibly quite different communities.

Ideally, parameter values for a given community should be based on its own experience. Many CCRCs have not maintained appropriate records for this purpose or are too small to have accumulated substantial recent experience. It is important, though, that as much information as possible be extracted from the data that is available. Modern statistical techniques can help in doing this.

Estimated rates of transition between model states should appropriately reflect the effect of various factors. These may include aspects of a resident's health history since entering the CCRC, as well as other information such as gender, marital status, fees paid, contract type, etc. One therefore requires data that provides this information.

1.4 Outline of Paper

The purpose of this paper is to present an approach to analyzing CCRC data and to demonstrate the methods using data collected from a large CCRC. Section 2 provides a description and some preliminary observations of the data set used. This will illustrate the nature of CCRC data and give an appreciation for the challenge presented by such data. Statistical methods for analyzing CCRC data are described in Section 3. The state occupied by a resident is modeled as a continuous-time stochastic process which is characterized by transition intensity functions. We discuss a non-parametric approach to estimating these functions and the Cox regression model for quantifying the effect of important variables on these functions. The methods are used in Section 4 to obtain estimates based on the data introduced in Section 2. Since these estimates are based on limited data from one CCRC, they are of illustrative value only and should not be used in actuarial analyses. In Section 5, we explain how probabilities and other quantities of interest can be obtained by simulation, and we use the results of Section 4 to illustrate the approach. Finally, some conclusions are discussed in Section 6.

2 The Pilot Study Data

2.1 Background

In 1991, the Society of Actuaries and the American Association of Homes for the Aging co-sponsored a pilot study that involved the collection of data from a large CCRC in Florida. The purpose of the pilot study was to gain insight into the considerations involved in collecting CCRC data with a view to larger scale data collection and analysis projects to be carried out in the future.

The CCRC under study houses (on average) 525 residents occupying three types of independent housing (single family, garden apartment and highrise units) and two levels of health care (assisted living and skilled care beds). Access to health care is guaranteed with an increased charge to the resident.

The data comprises information on all individuals who resided in the CCRC during the three year period from April 1, 1988 to March 31, 1991, which we shall call the "study period". Information was also coded for those who resided in the facility before this period with a spouse who remained in the CCRC through some or all of the study period. A total of 803 residents were included in the study. They spent a total of 1605 life years in the community during the study period.

Information recorded for each resident includes an identification number, name, birth date, sex, couple status, apartment type at entry, apartment type at beginning of study period (or entry for those who entered during the study period), entry fee, service fee, health status (at later of entry or beginning of study period), room mate identification number (if any), entry date, contract type and refund provision. In addition, for each change of health status that occurred during the study period, the new status, date of change and cause of change (if known) were recorded.

2.2 Preliminary Examination of Data

Of the 803 residents in the study, most were typical CCRC residents receiving residential services and possibly meals. Others, such as those admitted directly into assisted living, or those residing in assisted living units or skilled care beds on a per diem basis, were removed from the dataset. This left 722 residents who spent 1518 years in the CCRC during the study period.

Since the number of individuals in the study was fairly small, it was thought reasonable to combine the three types of independent living units for the purpose of estimating transition intensity functions. This reduces the number of functions to be estimated. In larger scale studies conducted in the future, it will be appropriate to test whether or not the type of independent living unit affects the intensity function.

While in the community, individuals transfer among the following "states":

1. Independent

   Residents in this state are capable of living alone or with a roommate without twenty-four hour supervision.

2. Assisted Living

   Residents in this state require some on-going, long-term supportive services in order to function. While some medical or nursing services may be provided, the emphasis is on personal care services.

3. Skilled Care (Temporary)

   Residents in this state require continuous or on-going nursing or medical care services provided by a licensed practical nurse, a registered nurse, or a physician. These residents are expected to recover and return to either the independent or the assisted living state.

4. Skilled Care (Permanent)

   This state is the same as state 3 except that residents in this state are not expected to recover. [1]

Departure from the CCRC during the study period occurs either by withdrawal or death.

[1] This is the traditional distinction between temporary and permanent. However, today it is common for residents to be classified as permanent only when the unit at the lower level of care is made available for another resident. If the individual was residing with a spouse, then this may not occur until the spouse vacates the unit. One must therefore recognize that the labeling of transfers to skilled care may differ between CCRCs.

Figure 1 illustrates the setup. The boxes represent states that may be occupied by an individual, and the arrows indicate the possible transitions. The total number of years spent in each state during the study period is shown in the appropriate box. Near the head of each arrow is the number of transitions of the indicated type during the study period. It is clear from Figure 1 that certain transitions occur with much greater frequency than others. For example, there were 371 transitions from state 1 to state 3, but only 2 transitions from state 1 to state 4. Thus, we should be able to say much more about the $1 \rightarrow 3$ transition intensity.

Certain transition should, in theory, not occur. There should be no recoveries from the assisted living state to the independent state. Although the word "permanent" has been omitted in describing the assisted living state, all visits to this state were coded as permanent. Also, there should be no recoveries from the skilled care (permanent) state to the independent, assisted living or skilled care (temporary) states. Therefore the numbers of $2 \rightarrow 1$, $4 \rightarrow 1$, $4 \rightarrow 2$ and $4 \rightarrow 3$ transitions should be zero. In practice, assessments of future health status cannot be performed with 100 percent accuracy. Figure 1 illustrates that some of these transitions did occur.

There were many more females in the dataset than males. Of the 722 individuals included in the analysis, only 198 were males. Table 1 shows the the total time spent in the CCRC during the study period by state and sex. Table 2 provides a breakdown of the number of transitions by transition type and sex.

One variable that is likely important in determining transition intensities is the age of the resident. Therefore, it is helpful to understand how the CCRC population is distributed by age. To this end, we prepared graphs showing the number of residents attaining each age during the study period (see Figure 2). We display separate graphs for all residents and residents in each state.

Another potentially relevant variable is duration since entry to the CCRC. This is due to the selection that occurs at the time of entry. Many CCRCs require residents to demonstrate that they are in good health before they are admitted. This is necessary because, as stated earlier, CCRCs provide a long-term care insurance benefit. Figure 3 shows the number of residents attaining each duration during the study period. The figure illustrates how the population was distributed by duration. The distribution is heavily skewed to the right. Roughly half of the total time that residents spent in the community during the study period was time spent during the first five years since entry. However, there were residents who had been in the community for as long as twenty-five years.

3 Statistical Methods for Analyzing CCRC Data

3.1 Introduction

As stated in Section 1, CCRCs present a challenge for actuaries because of the complexity of the possible outcomes for a given resident. A CCRC resident may transfer many times before leaving the community by death or withdrawal. Thus, it is easiest to think of the outcome as a realization of a stochastic process. We then attempt to find a model that reasonably describes the behavior of this process.

Suppose we have n residents in the study. For $j = 1, 2, \ldots, n$, let $X_j(t)$ represent the state occupied by resident j at time t. Then $\{X_j(t), t \geq 0\}$ is a continuous-time stochastic process with state space $\{1, 2, \ldots, 6\}$. Often t will represent age. However, it will sometimes be convenient to let t measure the time since some event such as entry to the CCRC or entry to a given state.

We can characterize the process in terms of transition intensity functions. These functions are also referred to as forces of transition since they are analogous to the force of mortality. Let

$$\alpha_{hij}(t; \mathbf{Z}_j(t)) = \lim_{u \to 0+} \Pr\left(X(t+u) = i \middle| X(t) = h, \mathbf{Z}_j(t)\right) \Big/ u, \qquad (1)$$

$$h, i = 1, 2, \ldots, 6, \ h \neq i, \ j = 1, 2, \ldots, n$$

be the transition intensity function for transitions from state h to state i by individual j. $\mathbf{Z}_j(t)$ is a vector of covariates containing relevant information about resident j that is available just prior to time t. Examples of possible components of $\mathbf{Z}_j(t)$ are the time since resident j entered state h and an indicator of the sex of resident j. The former depends on t, and is referred to as a time-dependent covariate. Initially, we consider the special case in which the transition intensity functions do not involve any covariates. We further assume that these functions are the same for all residents. That is,

$$\alpha_{hij}(t; \mathbf{Z}_j(t)) = \alpha_{hi}(t).$$

Our objective is to estimate the transition intensity functions. We attack this problem by first finding estimators for the corresponding cumulative intensity functions,

$$A_{hi}(t) = \int_0^t \alpha_{hi}(s)ds.$$

Let $Y_{hj}(t) = I(X_j(t-) = h)$ and $Y_h(t) = \sum_{j=1}^n Y_{hj}(t)$, where $I(A)$ is the indicator random variable of the event A. $Y_h(t)$ can be thought of as the number of residents "at risk" just prior to time t of a transition from state h. Also, let $N_{hij}(t)$ represent the number of observed $h \to i$ transitions made by resident j during [0,t], and let $N_{hi}(t) = \sum_{j=1}^n N_{hij}(t)$. Then $\{N_{hij}(t), t \geq 0\}$ and $\{N_{hi}(t), t \geq 0\}$ are counting processes. An elaborate theory has been developed for statistical models involving counting processes. This began with the work of Aalen [1], and is well described in books by Andersen et al. [4] and Fleming and Harrington [9]. The theory is based on the fact that the difference between a counting process and its integrated intensity process is a martingale. One can obtain variances of statistics that are stochastic integrals with respect to this martingale, and asymptotic distributions can be found using martingale central limit theory. The reader need not have an understanding of the theory of counting processes and martingales.

3.2 The Nelson-Aalen Estimator

A well known non-parametric estimator of $A_{hi}(t) = \int_0^t \alpha_{hi}(s)ds$ is

$$\hat{A}_{hi}(t) = \int_0^t Y_h(s)^{-1} dN_{hi}(s). \tag{2}$$

Intuitively, this estimator makes sense if we break up the interval $[0, t]$ into many sub-intervals of length ds. The probability that an individual in state h at time $s - ds$ moves to state i

by time s is $\alpha_{hi}(s)ds$. A reasonable estimator of this probability is the number of observed $h \to i$ transitions during $(s - ds, s]$, which is $dN_{hi}(s)$, divided by the number of individuals in state h at time $s - ds$, which is $Y_h(s)$. Summing the actual probabilities over all subintervals in $[0, t]$ gives $A_{hi}(t)$, and summing the estimators gives the right-hand side of (2). If $T_{hi1}, T_{hi2}, \ldots$ are the observed times of the $h \to i$ transitions, then $\hat{A}_{hi}(t)$ can be expressed as a simple sum,

$$\hat{A}_{hi}(t) = \sum_{k:T_{hik} \leq t} Y_h(T_{hik})^{-1}.$$

$\hat{A}_{hi}$ is the well known Nelson-Aalen estimator. It was introduced by Nelson [16] in the context of estimating the hazard function of failure time distributions using censored data. Nelson explored how to use plots of the estimates to gain information about the distribution. Aalen [2] discussed the estimator in a general counting process framework, and considered exact and asymptotic properties of the estimator.

The Nelson-Aalen estimator is not an unbiased estimator of A_{hi}, but is biased downward. Let

$$A^*_{hi}(t) = \int_0^t \alpha_{hi}(s) J_h(s) ds,$$

where $J_h(s) = I(Y_h(s) > 0)$. $A^*_{hi}(t)$ is almost the same as $A_{hi}(t)$ when $\Pr(Y_h(s) = 0)$ is small for all $s \leq t$. It turns out that

$$E[\hat{A}_{hi}(t)] = E[A^*_{hi}(t)] = \int_0^t \alpha_{hi}(s) \Pr(Y_h(s) > 0) ds.$$

Hence, the bias in using $\hat{A}_{hi}(t)$ to estimate $A_{hi}(t)$ is

$$E[\hat{A}_{hi}(t)] - A_{hi}(t) = -\int_0^t \alpha_{hi}(s) \Pr(Y_h(s) = 0) ds.$$

The implications of this in estimating transition intensity functions using the CCRC Pilot Study will be discussed shortly.

It is important to be able to quantify the variability of an estimator. The variance of the Nelson-Aalen estimator is

$$\begin{aligned} Var[\hat{A}_{hi}(t)] &= E[\{\hat{A}_{hi}(t) - A^{*}_{hi}(t)\}^2] \\ &= \int_0^t E\left[J_h(s)Y_h(s)^{-1}\right] dA_{hi}(s). \end{aligned}$$

An unbiased estimator of the variance is

$$\begin{aligned} \widehat{Var}[\hat{A}_{hi}(t)] &= \int_0^t Y_h(s)^{-1} d\hat{A}_{hi}(s) \\ &= \int_0^t Y_h(s)^{-2} dN_{hi}(s). \end{aligned} \tag{3}$$

As with equation (2), we can express the right-hand side of (3) as a sum,

$$\widehat{Var}[\hat{A}_{hi}(t)] = \sum_{k:T_{hik} \leq t} Y_h(T_{hik})^{-2}.$$

This variance estimator can be used to obtain approximate pointwise confidence limits for the cumulative intensity functions. In doing so we use the fact that the asymptotic distribution of $\hat{A}_{hi}(t)$ is normal. Since the distribution may depart significantly from the normal distribution when $Y_h(t)$ is small, confidence limits obtained using the normal distribution assumption are not reliable in this case.

To illustrate the ideas discussed in this subsection, we now examine the use of the Nelson-Aalen estimator in analyzing one transition type using the CCRC Pilot Study data. We consider transitions from state 1 to state 6. That is, deaths from the independent state.

Since only 31 such transitions occurred, we can clearly see how the Nelson-Aalen estimator works.

Table 3 shows, for both males and females, the ages at which each death from the independent state occurred, as well as the number of residents at risk of dying while in the independent state at each of those ages. The corresponding Nelson-Aalen estimates of the cumulative intensity functions for females and males are shown in Figure 4. The estimated cumulative intensity functions are step functions with jumps at each of the transition (death) ages. The size of each jump equals the number of transitions that occurred at that age divided by the number of residents at risk of making the transition at that age. Perhaps the most appealing aspect of using Nelson-Aalen estimates is the ability to plot the estimates and observe the general shape of the estimated cumulative intensity function. A cumulative intensity function that appears to increase linearly suggests a constant intensity function (because the cumulative intensity function is the integral of the intensity function). A cumulative intensity function that is convex (concave) suggests an increasing (decreasing) intensity function. Of course, one should keep in mind that estimates based on a small number of transitions, as in this example, are limited in how much information they can convey. Figure 4 seems to indicate that the female intensity function is increasing with age, and the male intensity function may be constant, though there are only ten male transitions.

As stated above pointwise confidence limits can be obtained assuming that estimators have a normal distribution. For example, an approximate 95 percent confidence interval for $A^*_{hi}(t)$ is given by $\hat{A}_{hi}(t) \pm 1.96\sqrt{\widehat{Var}[\hat{A}_{hi}(t)]}$. Figure 5 shows the estimated cumulative intensity functions along with these 95 percent confidence limits. Since, for males, the number at risk at each age is rather small, the confidence limits should not be trusted.

We mentioned earlier that $\hat{A}_{hi}(t)$ is an unbiased estimator of $E[A^*_{hi}(t)]$ and a biased estimator of $A_{hi}(t)$. A^*_{hi} and A_{hi} are quite different in the above example since there are no residents at the younger ages. In fact, $Y_h(s) = 0$ for all $s < 55$. Fortunately, we are less interested in estimating the function A_{hi} than we are in estimating $\alpha_{hi}(t)$ for values of t in the age range of the CCRC residents. Now $\alpha_{hi}(t)$ is the slope of A_{hi} at age t. For an age interval with $Y_h(t) > 0$, the slopes of A_{hi} and A^*_{hi} are the same. Thus, we can estimate $\alpha_{hi}(t)$ by estimating the rate of increase of A^*_{hi} at time t. This can be done by averaging the jumps in $\hat{A}_{hi}$ at ages near t. This is considered next.

3.3 Kernel Function Estimators

Smooth estimates of α_{hi} can be obtained using a kernel function estimator. This approach is discussed by Ramlau-Hansen [17, 18], Andersen et al. [4] and Gavin et al. [10]. The estimator is defined as follows.

$$\hat{\alpha}_{hi}(t) = \frac{1}{b}\int_{-\infty}^{\infty} K\left(\frac{t-s}{b}\right) d\hat{A}_{hi}(s), \tag{4}$$

where $\int_{-\infty}^{\infty} K(x)dx = 1$ and $K(x) = 0$ for $|x| > 1$. K is called the kernel function, and b is called the band width or window size. Viewing the real line as many small intervals of length ds, we see that $\hat{\alpha}_{hi}(t)$ is a weighted average of the jumps in the Nelson-Aalen estimator that occur in the interval $[t-b, t+b]$. The smoothness of the estimates increases as the value of

b increases. Again letting $T_{hi1}, T_{hi2}, \ldots$ be the observed $h \to i$ transition times, (4) can be written as a sum,

$$\hat{\alpha}_{hi}(t) = \frac{1}{b} \sum_k K\left(\frac{t - T_{hik}}{b}\right) Y_h(T_{hik})^{-1}.$$

A variance estimator of the kernel function estimator is given by

$$\widehat{Var}[\hat{\alpha}_{hi}(t)] = \frac{1}{b^2} \int_{-\infty}^{\infty} K^2\left(\frac{t-s}{b}\right) Y_h(s)^{-2} dN_{hi}(s),$$

which can also be written as

$$\widehat{Var}[\hat{\alpha}_{hi}(t)] = \frac{1}{b^2} \sum_k K^2\left(\frac{t - T_{hik}}{b}\right) Y_h(T_{hik})^{-2}.$$

A popular choice of kernel function is the Epanechnikov kernel function,

$$K(x) = 0.75(1 - x^2), \ |x| \leq 1.$$

We shall use this kernel function to obtain smooth transition intensity functions. Other kernel functions are discussed by Ramlau-Hansen [18].

Note that it is appropriate to use (4) only if $Y_h(s) > 0$ for all $s \in [t-b, t+b]$. Otherwise, a substantial downward bias could result since the absence of transitions in a given time range will produce a low transition intensity estimate. The actual transition intensity might be quite large, but no residents were at risk of making the transition. We discussed earlier that Nelson-Aalen estimates are informative only if calculated for a time interval $[t_1, t_2]$, where $Y_h(t) > 0$ for all $t \in [t_1, t_2]$. If $Y_h(t) = 0$ for t outside this interval, then we should restrict use of (4) to obtaining estimates of $\alpha_{hi}(t)$ for $t \in [t_1 + b, t_2 - b]$.

Smoothed transition intensities for deaths from the independent state were calculated using the CCRC Pilot Study data and appear in Figure 6 along with approximate 95 percent confidence limits. A window size of 6 was used for both females and males.

In this example, we have too few observations to make conclusions about the transition intensity functions. The confidence bands are quite wide (and for males they should not be trusted). We shall see later the usefulness of the kernel function estimator when more transitions are observed.

3.4 Regression Models

The transition intensity functions defined in (1) are resident-specific and depend on $\mathbf{Z}_j(t)$, a vector of covariates that provide relevant information about resident j that is available just prior to time t. One approach to reflecting the effect of covariates is by using a multiplicative hazards model. This approach was introduced by Cox [7] in the context of analyzing censored survival data, and is often referred to as the Cox regression model. Andersen and Gill [5] extended the ideas to general counting processes. Let

$$\alpha_{hij}(t; \mathbf{Z}_j(t)) = \alpha_{hi0}(t) \exp(\boldsymbol{\beta}_{hi}^{\mathsf{T}} \mathbf{Z}_j(t)), \tag{5}$$

where $\boldsymbol{\beta}_{hi}$ is a parameter vector, and α_{hi0} is the "baseline" intensity function. Under this model, the transition intensity functions for different values of a fixed covariate are proportional. For example, if $Z_j^1(t)$, the first component of the covariate vector for resident j, is 0 if the individual is a female and 1 if a male, then the model assumes that the male transition intensity function is $\exp(\beta_{hi}^1)$ times the female transition intensity function, where β_{hi}^1 is the first component of $\boldsymbol{\beta}_{hi}$.

In order to fit a model of this type we must estimate the components of $\boldsymbol{\beta}_{hi}$ and the baseline intensity function. One can argue (see Cox [7]) that little information about $\boldsymbol{\beta}_{hi}$ is provided by the transition times since the baseline transition intensity function could be very small except near the transition times where it could be very large. Hence, most of the information about $\boldsymbol{\beta}_{hi}$ is provided by the knowledge of which individuals made transitions given the transition times and the individuals at risk of transitions just prior to these times. This is the motivation for the partial likelihood,

$$L(\boldsymbol{\beta}_{hi}) = \prod_{k} \prod_{j=1}^{n} \frac{\exp(\boldsymbol{\beta}_{hi}^{\top} \mathbf{Z}_j(T_{hijk}))}{\sum_{l=1}^{n} \exp(\boldsymbol{\beta}_{hi}^{\top} \mathbf{Z}_l(T_{hilk})) Y_{hl}(T_{hilk})},$$

where $T_{hij1}, T_{hij2}, \ldots$ are the $h \to i$ transition times for individual j. The components of $\boldsymbol{\beta}_{hi}$ can be estimated by maximizing $L(\boldsymbol{\beta}_{hi})$. Inferences can then be made about $\boldsymbol{\beta}_{hi}$ as in the usual maximum likelihood setting (see Hogg and Craig [12]). Having determined estimates $\hat{\boldsymbol{\beta}}_{hi}$, the estimated baseline cumulative intensity function is

$$\hat{A}_{hi0}(t, \hat{\boldsymbol{\beta}}_{hi}) = \sum_{k: T_{hik} \le t} \left[\sum_{j=1}^{n} \exp(\hat{\boldsymbol{\beta}}_{hi}^{\top} \mathbf{Z}_j(T_{hik})) Y_{hj}(T_{hik}) \right]^{-1}. \tag{6}$$

This generalizes the Nelson-Aalen estimator and is often referred to as the Breslow estimator.

The statistical package S-PLUS is capable of fitting a Cox regression model. One need only specify, for each visit to state h (by all residents), the time at which the visit began, the time at which the visit ended, an indicator for whether or not the visit ended in a transition to state i, and the value of each covariate. If time-dependent covariates are being used (more specifically, covariates whose values may change during the visit to state h), each visit can be partitioned into several shorter visits based on the $h \to i$ transition times (by all residents). That is, the first such visit will end at the first transition time after the visit began. The covariate value used should be the actual covariate value just before this transition time.

The traditional actuarial approach to handling data with covariates is to group the data into homogeneous cells. All data in a given cell would have the same (or approximately the same) values of the important covariates. Separate estimates are then obtained for each cell using only data from that cell. Some smoothing across cells may then be done.

The Cox regression approach offers some advantages. Statistical tests can be performed to see which covariates are important. Accuracy is improved since transition intensity functions are estimated using all of the data, and not just the data from a given cell. The only cost is that one must be willing to assume that the transition intensity functions have the multiplicative form given in (5).

4 Estimation of Transition Intensity Functions using the Pilot Study Data

4.1 Introduction

In this section, we illustrate the techniques of Section 3 using the pilot study data described in Section 2. Our goal is both to demonstrate the methods of estimation and to show the forms we might obtain for the transition intensity functions. We use the results of this section to show how to determine probabilities and other important quantities in Section 5.

We stress at the outset that the pilot study data is not sufficient to estimate all of the intensity functions with reasonable accuracy. In practice, one should consult other sources of information when faced with this situation. Since our objective is to demonstrate the methods described earlier and to find estimates to be used later in the paper, we are content with estimates based solely on the pilot study data. We further remark that, in using the Cox regression model, we do not perform a thorough regression analysis. Our goal is simply to gain an understanding of what covariate information may influence the transition intensity functions. We use the regression coefficients that result to illustrate the nature of the intensity estimates one might obtain. A more thorough regression analysis, which is beyond the scope of this paper, should be undertaken if one wishes to use this approach in practice.

To begin, note that the intensity functions corresponding to certain transitions shown in Figure 1 should be zero. In particular, the $2 \to 1$, $4 \to 1$, $4 \to 2$ and $4 \to 3$ intensity functions should be zero in light of the permanent nature of states 2 and 4. In addition, the $1 \to 4$ transition intensity function will be set to zero. Only two such transitions occurred. Given the total time spent in state 1 by all residents, this suggests a very small intensity for this CCRC.

4.2 Withdrawals

Just twenty-eight withdrawals occurred during the study period. We do not expect to conclude very much based on this. We should, however, be able to get some idea of the behavior of the withdrawal intensity by age, and also test whether the differences in the withdrawal rates among the states from which withdrawal occurred are statistically significant. This can be accomplished by using a Cox regression model. Rather than model the four different withdrawal types separately as described in Subsection 3.4 and suggested by (5), we can model them together assuming that the withdrawal intensity functions are proportional. This leads to the following model.

$$\alpha_{\cdot 5j}(t; \mathbf{Z}_j(t)) = \alpha_{\cdot 50}(t) \exp(\boldsymbol{\beta}_{\cdot 5}^{\mathsf{T}} \mathbf{Z}_j(t)),$$

where $\alpha_{\cdot 50}$ is the baseline intensity function for withdrawals from any state. We assume that the last three components of the covariate vector, $\mathbf{Z}_j(t)$, are variables which indicate (1 if yes, 0 if no) whether resident j was in state 2, 3 and 4, respectively, just prior to time t. If the resident was in state 1, then all three variables are zero. Thus, if $Z_j^k(t)$, the kth component of $\mathbf{Z}_j(t)$, is the indicator for state 2, then the intensity for withdrawals from state 2 is $e^{\beta_{\cdot 5}^k}$ times the intensity for withdrawals from state 1, where $\beta_{\cdot 5}^k$ is the kth component of $\boldsymbol{\beta}_{\cdot 5}$.

Table 4 shows the results of a Cox regression run performed using S-PLUS. The input comprised an observation for each stay by a resident in one of the four states. For each observation the following information was provided: age at entry to the state (or beginning of study period if later), age at departure from the state (or end of study period if earlier), indicator of whether or not the stay ended with a withdrawal (1 if yes, 0 if no). Additional information about each stay was provided by five covariates that were included in the model. The first was an indicator of whether or not the resident was a male. The second was the age at which the individual entered the CCRC. The remaining three covariates were the indicators corresponding to states 2 (assisted living), 3 (skilled care temporary) and 4 (skilled care permanent), as described above. For each covariate, the table provides the estimate of the coefficient, $\beta^k_{\cdot 5}$, the corresponding proportionality factor, $e^{\beta^k_{\cdot 5}}$, the standard error of the coefficient, the p-value for a two-tailed test of the hypothesis that the coefficient equals zero, and upper and lower 95 percent confidence limits for the proportionality factor. At the bottom of the table are the likelihood ratio and efficient score statistics. These can be used for an overall test of whether or not the variables in the model are related to the withdrawal intensity.

Table 4 suggests that the indicator for skilled care (temporary) is significant with a p-value of 0.00126, and entry age is marginally significant with a p-value of 0.09137. The other three covariates are not significant. We should note that regression runs in which each of the five covariates were included separately produced similar results to those displayed in the table. The likelihood ratio statistic and the efficient score statistic both indicate that the regression is significant at the 5 percent level. This is largely due to the presence in the model of the skilled care (temporary) variable.

The results indicate that the withdrawal intensity for those in skilled care (temporary) is higher than that for residents in other states. However, this is based on just four withdrawals, and the 95 percent confidence interval for the proportionality factor is very large. Due to this uncertainty, we assume that the withdrawal intensity is the same for all four states.

The fact that the coefficient corresponding to entry age is positive suggests that the withdrawal intensity may be higher for those who entered the CCRC more recently. However, since this variable is only marginally significant, we shall ignore it and model the withdrawal intensity as a function of age only. It should be noted that this result is based on the limited data of one CCRC; the significance of age and duration to withdrawal may be different for other CCRCs.

Figure 7 shows the Breslow estimates of the baseline cumulative intensity function along with smooth kernel function estimates of the intensity function using a band width of 10. The concave behavior of the estimated cumulative intensity function suggests an intensity function that decreases with age, and we do observe a decrease in the intensity function estimates. To obtain a simple mathematical expression for the intensity function and to improve the smoothness, we assumed that the intensity function has the linear form $\alpha_{.50}(t) = at+b$. The corresponding cumulative intensity function is $A_{.50}(t) = at^2/2+bt$, and we assume that $A^*_{.50}(t) = at^2/2 + bt + c$, It is necessary to distinguish between $A_{.50}$ and $A^*_{.50}$ because $\hat{A}_{.50}$ is greatly influenced by the fact that there are no residents at the younger ages. The latter quadratic was fit to the cumulative intensity estimates using least squares (a routine for this is available in S-PLUS). The resulting parameter estimates are $a = -0.001346058$, $b = 0.1342243$ and $c = -6.09296$. The function is plotted in Figure 7 (dashed line). The relevant parameter values in determining the intensity function are a and b. The linear function based on these values is plotted (dashed line) along with the smoothed intensity estimates in Figure 7.

4.3 Mortality

We analyze mortality transitions in the same manner as withdrawals. However, 126 deaths occurred during the study period. This is considerably more than the 28 withdrawals. Also, we have more prior knowledge of mortality patterns. We expect the intensity (force of mortality) to be greater for males than females, and we expect the intensity to increase with the level of care provided.

A Cox regression run with the same covariates as in the analysis of withdrawals revealed that age of entry was not significant at all. We therefore dropped entry age from the model and obtained the results shown in Table 5. The table indicates that only the two skilled care variables are significant. The p-values for the male indicator and the assisted living indicator are 0.320 and 0.256, suggesting no evidence against the hypothesis that the corresponding two coefficients are zero. However, the proportionality factors for all four variables seem to make sense. It is reasonable for the male intensity to be 1.27 times the female intensity. We expect the intensity for those in assisted living to be somewhat larger than the intensity for those in the independent state. The intensity for those in skilled care should be considerably larger than the intensities for those in the assisted living or independent states. The numbers in the table are consistent with this. Note that the proportionality factor in the table for skilled care (temporary) is almost exactly twice that for skilled care (permanent). This is reasonable if we assume that the permanent designation implies some degree of stability that may not apply to the skilled care (temporary) residents.

Although the confidence intervals for the proportionality factors are quite large, the fact that the estimates of these factors conform to our expectations provides a level of comfort. We therefore estimate the intensity functions in terms of the estimated baseline intensity function, $\hat{\alpha}_{\cdot 60}(t)$, as follows:

$$
\begin{aligned}
\hat{\alpha}_{16j}(t;\mathbf{Z}_j(t)) &= \hat{\alpha}_{\cdot 60}(t),\\
\hat{\alpha}_{26j}(t;\mathbf{Z}_j(t)) &= 1.57\hat{\alpha}_{\cdot 60}(t),\\
\hat{\alpha}_{36j}(t;\mathbf{Z}_j(t)) &= 16.01\hat{\alpha}_{\cdot 60}(t),\\
\hat{\alpha}_{46j}(t;\mathbf{Z}_j(t)) &= 8.02\hat{\alpha}_{\cdot 60}(t),
\end{aligned}
$$

if resident j is a female. Each function should be multiplied by 1.27 if resident j is a male.

Figure 8 shows the Breslow estimates of the baseline cumulative intensity function as well as kernel function estimates of the baseline intensity function obtained using a band width of 6. The estimated cumulative intensity function has a convex shape through the eighties and early nineties, suggesting an increasing intensity function. The smooth intensity function estimates exhibit this increasing behavior. As with the withdrawal intensity, we can obtain a simple mathematical expression for the intensity by assuming an intensity function of the form $\alpha_{\cdot 60}(t) = a + bc^t$ (Makeham's law). The corresponding cumulative intensity function is $A_{\cdot 60}(t) = at + b(c^t - 1)/\log c$, and we assume that $A^*_{\cdot 60}(t) = d + at + b(c^t - 1)/\log c$. The latter function was fit to the Breslow estimates by least squares. The resulting parameter values are $a = 0.01200684$, $b = 7.075078 \times 10^{-7}$, $c = 1.122718$ and $d = -0.8505983$, and the function is represented by the dashed line that appears along with the baseline cumulative intensity estimates in Figure 8. The corresponding intensity function is also plotted along with the smoothed intensity estimates.

4.4 Other Transitions from Independent

We have two types of transition from the independent state for which we have not yet estimated the intensity function. They are transitions to assisted living and transitions to skilled care (temporary). The latter is the transition type for which we observed the most transitions (371) during the study period. We discuss this transition type first. Since we observed a relatively large number of transitions, we should be able to say more about the impact of covariates on the transition intensity function, and perhaps reflect one or more covariates in our intensity function estimates. Elements of a resident's health status history that might affect the intensity function are the time since the current stay in independent began, and the number of previous visits to skilled care (temporary). Unfortunately, for many residents, we do not observe one or both of these quantities. For a stay in independent that began before the beginning of the study period, we only know that this stay has lasted at least since the beginning of the study period and no longer than since the date of entry to the community. Also, for residents who entered the CCRC before the beginning of the study period, we do not know how many visits have been made to skilled care. We only know how many visits have been made since the beginning of the study period.

Recognizing these limitations, it is still informative to examine the results of regression runs that include as time-dependent covariates the "observed" duration of the current stay and the number of previous "observed" visits to skilled care. This provides information as to whether duration of the current stay and number of previous visits to skilled care influence the intensity. However, the coefficients obtained from such runs cannot be used to obtain transition intensity estimates. This is because the covariate values do not correctly represent the quantities we wish our intensity function to reflect.

Other covariates considered were sex and duration since entry to the CCRC. The latter has the same effect as entry age. Duration in the community was found not to be significant. Sex, number of observed visits to skilled care, and observed duration of the current stay in independent were found to be significant, the latter being most important in terms of its effect on the likelihood ratio and efficient score statistics.

While the number of prior visits to skilled care may be relevant in estimating the transition intensity, we cannot appropriately reflect its impact using the pilot study data. As stated above, we do not know the number of prior visits for residents who entered the community before the beginning of the study period. Thus, we would be required to base our estimates on new entrants only. Since these residents gave rise to only 49 of the 371 transitions, we would have to discard a substantial portion of the data on these transitions. For this reason, we have chosen not to incorporate the number of prior visits to skilled care into our intensity estimates.

A similar problem exists in producing intensity estimates that are functions of the duration of the current stay in the independent state. This duration is known only for stays that began during the study period. However, these stays resulted in 197 of the 371 transitions. Given the importance of this variable, we chose to accept the loss of information in order to obtain estimates that reflect the duration of the current stay as well as age and sex.

We found that the intensity decreases with duration of the current stay. We would expect this to be the case for those who have returned from a stay in skilled care. However, we would not expect this for those who have just entered the CCRC. Since we are considering only stays that began during the study period, we can determine whether or not the stays follow visits to skilled care. Larger likelihood ratio and efficient score statistics were achieved when the duration of stay covariate was multiplied by an indicator of whether or not the stay followed a visit to skilled care. These statistics also increased when we replaced the duration of the current stay by its natural logarithm, reflecting the fact that the effect of this duration wears off quickly. We considered including an additional covariate indicating whether or not the current stay was preceded by a visit to skilled care. This allows for effects of this not related to duration. However, these effects were not statistically significant. In order to ensure that the intensity for those who previously visited skilled care was not smaller than the intensity for new entrants, we capped the duration of current stays at one year. Table 6 shows the results of a Cox regression run. The covariate labeled "dur" is actually log(min(duration of current stay, 1)) times an indicator that is 1 if the stay was preceded by a visit to skilled care and 0 otherwise.

It is interesting that the transition intensity for males is smaller than that for females.

This may be because a higher proportion of males live with a spouse.

The baseline cumulative intensity estimates and kernel function intensity estimates using a band width of 4 are shown in Figure 9. The smooth intensity estimates appear to be somewhat linear. We therefore fitted a quadratic function to the cumulative intensity estimates using least squares. The function appears on the graph (dashed line) in Figure 9. The derivative of this function is linear with intercept and slope given by -1.059995 and 0.01789773, respectively. This function is plotted in Figure 9 along with the smoothed intensity estimates.

We now consider transitions from independent to assisted living ($1 \to 2$). There were 32 such transitions, 29 were females and 3 were males. Cox regression runs indicated that sex and duration of the current stay in independent are important in estimating the transition intensity. Unfortunately, we have the same problem reflecting the duration of the current stay as we did in analyzing the transitions from independent to skilled care. If we consider only stays that began during the study period, we are left with only 14 transitions, making estimation of the baseline intensity function difficult. We therefore chose to ignore duration of the current stay in estimating the transition intensity.

The results of a Cox regression run in which the only covariate is an indicator of whether or not the individual is a male is shown in Table 7. We see that the estimated proportionality factor for males, 0.284, is quite small, and that the 95 percent confidence interval, $(0.0863, 0.936)$ is rather large. This wide confidence interval along with the fact that only three male transitions occurred cause us to be somewhat skeptical about the estimated proportionality factor. We decided to instead use the corresponding proportionality factor obtained for the $1 \to 3$ transition intensity. That is, 0.590.

The baseline cumulative intensity estimates and kernel function intensity estimates using a band width of 4 appear in Figure 10. We again fitted a Makeham curve to the Breslow estimates (dashed line). The Makeham parameter values are $a = -0.001871268$, $b = 4.312047 \times 10^{-8}$ and $c = 1.173223$.

4.5 Other Transitions from Assisted Living

There are two types of transition from assisted living that we have not yet considered, those to skilled care (both temporary and permanent). We first examine transitions to skilled care (temporary). The number of observed transitions of this type was 174. Again, we found that the observed duration of the current stay in assisted living and the number of observed prior visits to skilled care both had a statistically significant impact on the transition intensity. For the reason stated in the previous subsection, we reflect only the duration of the current stay in assisted living (as well as the sex of the resident) in our intensity estimates. Also, we once again base our estimates only on those stays that began during the study period. Such stays gave rise to 148 of the 174 transitions. Table 8 shows the results of a Cox regression run in which the covariates are an indicator of whether or not the resident is male and the natural logarithm of the duration of the current stay in assisted living. Our estimates will be based on the coefficients shown in the table.

The Breslow estimates of the baseline cumulative intensity function and kernel function estimates of the intensity function using a band width of 4 are shown in Figure 11. We note that the estimated cumulative intensity function is approximately linear, suggesting a constant intensity function. The dashed line added to the plot of the cumulative intensity estimates was fitted by least squares. The intercept and slope of the line are -53.4908 and 0.7498699, respectively. The slope is the relevant quantity as it represents the value of the resulting transition intensity. A (dashed) horizontal line at this level appears on the graph of the smoothed intensity estimates. We shall use this constant intensity as our baseline intensity estimate.

Only twelve transitions occurred between assisted living and skilled care (permanent). This was too few to discern the importance of any covariates or even a pattern in the intensity by age. We therefore simply assume that the intensity is constant and is the same for all residents. Our intensity estimate is the number of transitions divided by the total time spent in assisted living, that is, $12/163 = 0.0736$.

4.6 Other Transitions from Skilled Care (Temporary)

Other than mortality and withdrawal, three types of transition can occur from the skilled care (temporary) state. They are, transitions to independent, transitions to assisted living and transitions to skilled care (permanent). It is important to note that the transition intensities for these transition types depend on the previous state visited. For example, if the previous state was assisted living, then the intensity of transition from skilled care (temporary) to independent is zero. However, if the previous was independent, then the transition intensity is not zero. In fact, for all 288 residents who moved from skilled care (temporary) to independent during the study period, the previous state was independent.

Figure 2 indicates that, for a substantial portion of the relevant age range, the number of residents in skilled care (temporary) was zero. This suggests that the Nelson-Aalen (and Breslow) estimator would have a significant downward bias if age is used as the time variable. To avoid this problem, we can use the time since entry to skilled care as the time variable. The effect of age can be reflected by including the age at which the stay began as a covariate. For stays in skilled care that began before the beginning of the study period, we do not know the time since entry to this state. We treat these stays as "censored" observations, since we know that the stays have lasted at least since the beginning of the study period. Fortunately, there were only six residents in skilled care (temporary) at the beginning of the study period.

We first examine transitions to the independent state. As discussed above, the intensity is zero for for those whose previous state was assisted living. We therefore consider only those stays in skilled care (temporary) for which the previous state was independent. We found that the age at which the stay began and the age at which the individual entered the CCRC were significant, while the sex of the resident was marginally significant. The number of observed prior visits to skilled care was also significant, but, for reasons discussed earlier, we do not reflect this in our intensity estimates. Table 9 shows the results of a Cox regression run. The coefficients shown in the table will be used in determining intensity estimates.

The Breslow estimates of the cumulative intensity function are shown in Figure 12. Since most of the transitions occurred quite close to time zero, the tail problem associated with using a kernel function estimator is relevant in this case. However, we recognize from the concave nature of the estimates, that the intensity function will decrease over time. We could therefore use Makeham's law with $0 < c < 1$ to describe the intensity function. Hence, the intensity function would be given by $a + bc^t$, and the cumulative intensity function by $at + b(c^t - 1)/\log c$. Fitting the latter function to the cumulative intensity estimates by least squares, we obtained the parameter values $a = 3.79241$, $b = 63.4064$ and $c = 8.317622 \times 10^{-15}$. The function appears in Figure 12 (dashed line).

Of the 173 transitions from skilled care (temporary) to assisted living, 37 were by residents who previously resided in the independent state, 133 were by residents who previously resided in assisted living, and 3 were by residents who previously resided in skilled care (permanent) We ignore the latter, and focus on the former two. For both we found that the age at which the current stay in skilled care began and the age at which the resident entered the CCRC were significant. However, almost all of the effect of these covariates is contained in their difference. Tables 10 and 11 show the Cox regression results with a single covariate representing this difference.

Breslow estimates of the cumulative intensity functions appear in Figure 13. We see that for those whose previous state was independent, the cumulative intensity estimates are quite linear, while for those whose previous state was assisted living, the estimated cumulative intensity function is concave. In the former case, we fitted a straight line with intercept zero to the cumulative intensity estimates (dashed line in Figure 13). The resulting slope, which gives our constant intensity estimate is 0.73112. In the latter case we fitted a Makeham curve (dashed line in Figure 13), and obtained the parameters $a = 2.89537$, $b = 45.5908$ and $c = 9.39390 \times 10^{-16}$.

The number of observed transitions from skilled care (temporary) to skilled care (permanent) was 52. Of these, 23 resided in independent before entering skilled care (temporary), 28 resided is assisted living and 1 resided in skilled care (permanent). Due to the small numbers of transitions, rather than performing separate analyses for the different previous states, we used a covariate to reflect the difference. Examination of the Nelson-Aalen estimates for each confirmed that it was reasonable to do this. We considered only those stays in skilled care (temporary) for which the previous state was independent or assisted living, and used an indicator of whether or not the previous state was assisted living as a covariate. Cox regression runs showed that this variable is in fact significant. The only other covariate found to be significant was the age at which the resident entered the CCRC. Neither the duration since entry to the CCRC nor the age at which the stay in skilled care (temporary) began were significant. Since it did not seem to make sense to allow the transition intensity to depend on entry age without regard for current age, we chose not to include entry age in the model. Table 12 shows the results of a Cox regression run in which the only covariate is the indicator of whether or not the previous state was assisted living.

The Breslow estimates of the baseline cumulative intensity function appear in Figure 14. The shape of this function is such that it is difficult to fit a simple curve. This is because the function is convex until about time 0.2 and then becomes somewhat linear and less steep. We therefore fitted a Makeham curve to the estimates for $t < 0.2$ and a straight line to the estimates for $0.2 \le t < 0.5$. We assume that this line also applies to values of t larger than 0.5. The parameter values for the Makeham curve are $a = 0.1195655$, $b = 0.114535$ and $c = 2.825737 \times 10^8$. The parameter values for the straight line are $a = 1.347154$ and $b = -0.01725844$. The dashed line in Figure 14 illustrates the fit of these functions.

5 Determining Probabilities and Other Quantities

5.1 Introduction

As stated in Subsection 1.2, for pricing and valuation purposes, one should be able to estimate the probability that a resident is in any given state at any future time, as well as the probability that a resident will move between any two states during any time interval. Depending on the complexity of the transition intensity functions, these probabilities may be difficult to calculate directly. In this section, we present an approach to determining probabilities and other quantities using simulation. The method can be used for very general forms of the transition intensity functions. The approach is described in Subsection 5.2, and numerical results obtained using the intensity functions found in Section 4 are provided in Subsection 5.3.

5.2 Simulation Approach

Consider the general setup in which the transition intensity functions are given by $\alpha_{hi}(t; \mathbf{Z}_j(t))$. We assume that the components of $\mathbf{Z}_j(s)$ are either constant over time, or depend only on the history of $\{X_j(t)\}$ up to time $s-$. Thus, if we know $X_j(t)$ for $0 \leq t \leq s$, then we also know $\alpha_{hi}(t; \mathbf{Z}_j(t))$ for values of t up to time s. Furthermore, if we know $X_j(t)$ up to time s, and we assume that no transitions occur during the next w years, then we know $\alpha_{hi}(t; \mathbf{Z}_j(t))$ up to time $s + w$.

For example, suppose that $\alpha_{hi}(t; \mathbf{Z}_j(t)) = \alpha_{hi}(t, u)$ where t is the age of the individual, and u is the time since the individual entered state h. Now if, at age s, the individual has been in state h for v years, then assuming no transitions occur during the next w years the $h \to i$ transition intensity at age $s + w$ is $\alpha_{hi}(s + w, v + w)$.

Assume that resident j is in state h at time s, and we wish to determine the probability that this individual will be in state k at time $r > s$. We can do this by simulating the time and state entered at each transition time up to time r. Then if we repeat this a large number of times, the proportion of times that the individual is in state k at time r approximates the desired probability.

We use a method known as *thinning* (See Ross [21, p. 73]). As stated above, if no transitions occur by time $s + w$, then the values of $\alpha_{hi}(t; \mathbf{Z}_j(t))$ are known for $s \leq t \leq s + w$ and $i = 1, 2, \ldots, 6$, $i \neq h$. Let

$$\alpha = \max_{s \leq t \leq r} \left(\sum_{i: i \neq h} \alpha_{hi}(t; \mathbf{Z}_j(t)) \right),$$

assuming that no transitions occur before time r. Now $\sum_{i:i \neq h} \alpha_{hi}(t; \mathbf{Z}_j(t))$ is the intensity of transition out of state h at time t, and α is no less than this intensity for $t \in [s, r]$. To determine the first transition time after time s, we successively generate the event times,

$T_1, T_2, \ldots$, of a Poisson process with intensity α, and accept T_l with a probability equal to

$$\sum_{i:i\neq h} \alpha_{hi}(s+T_l; \mathbf{Z}_j(s+T_l))/\alpha.$$

The event times of a Poisson process with intensity α are easily generated since the times between successive event are exponentially distributed with mean $1/\alpha$. An event time can be accepted with a given probability by generating a random number that is uniform on $(0,1)$, and accepting the event time if the number is no greater than this probability. Let T_1^*, the first accepted time, be the time until the next transition. The state entered at time $s+T_1^*$ can then be generated based on its conditional distribution. For $i^* \neq h$, the probability that i^* is the state entered at time $s+T_1^*$ equals

$$\alpha_{hi^*}(s+T_1^*; \mathbf{Z}_j(s+T_1^*))/\sum_{i:i\neq h} \alpha_{hi}(s+T_1^*; \mathbf{Z}_j(s+T_1^*)).$$

The following algorithm provides the steps outlined above. The statement "generate U" means " generate a random number U whose distribution is uniform on $(0,1)$."

Algorithm:

Step 1: $T_0^* = T_0 = s$

Step 2: $l = 0$

Step 3: $l = l+1$

Step 4: generate U

Step 5: $T_l = T_{l-1} - \frac{1}{\alpha}\log U$

Step 6: generate U

Step 7: if $U > \sum_{i:i\neq h} \alpha_{hi}(s+T_l; \mathbf{Z}_j(s+T_l))/\alpha$, go to Step 3

Step 8: $T_1^* = T_l$

Step 9: generate U

Step 10: $i^* = \max\left\{m : \sum_{i<m:i\neq h} \alpha_{hi}(s+T_1^*; \mathbf{Z}_j(s+T_1^*))/\sum_{i:i\neq h} \alpha_{hi}(s+T_1^*; \mathbf{Z}_j(s+T_1^*)) < U\right\}$

Once the time of and state entered upon the next transition are found, we determine the intensity functions that will apply after this transition, recalculate α, and repeat the procedure. We continue until we have a transition that occurs after time r, or until death or withdrawal. We then know the value of $X_j(t)$ for $s \leq t \leq r$.

By repeating the procedure a large number of times, not only can we estimate various probabilities, but we can estimate other interesting quantities. For example, for each simulation outcome we could compute the present value of the fee income that would result. Then the average of these present values provides an estimate of the actuarial (expected) present value of the fee income.

5.3 Numerical Illustration

Using the method described in Subsection 5.2 and the intensity function estimates found in Section 4, 10,000 simulations were performed for a 75 year-old female and for a 75 year-old male, both having just entered the CCRC. Table 13 summarizes the results of these simulations. The columns of the table provide the probability that the resident is in each of the six states at the end of each of the next twenty years. Since 10,000 simulations were performed, for each probability estimate, the standard deviation is at most $\sqrt{(0.5)^2/10000} = 0.005$. Hence, the estimates should be within 0.01 of the true value (that is, true according to our estimated intensity functions) with probability at least 0.95.

One interesting observation we make from Table 13 is that the probability that the resident is in state 6 (dead) is higher for females than for males. This can be explained by the fact that our estimated transition intensities from state 1 to states 2 and 3 are smaller (by a factor of 0.590) for males than for females. Therefore, females tend to move more quickly to states in which the mortality rate is higher. As stated in Subsection 4.4, this may be because a higher proportion of males live with a spouse. We have not allowed for this in estimating intensity functions. One should not conclude from this that female CCRC residents experience higher overall mortality than male residents. Rather, some feature of this data set or this CCRC has caused us to obtain intensity estimates that produce this anomalous result.

6 Conclusions

This paper has described and demonstrated an approach to analyzing CCRC data. We conclude the paper with some observations that are relevant to those wishing to conduct such an analysis.

Given the number of transitions that can be made by a given resident and the frequency with which these transitions occur, it is natural to use a continuous-time multi-state stochastic model in order to fully capture the randomness in resident transitions. It is convenient to characterize such a model in terms of the transition intensity functions. The methods described in Section 3 can then be used to obtain estimates for these functions.

The analysis discussed in Section 4 suggests that a model that incorporates the important sources of variation in resident outcomes will be rather complicated. For example, we have seen that duration in the current state, duration since entering the CCRC, and number of previous visits to skilled care as well as age and sex may affect transition intensities. Furthermore, other variables that we did not explore may have some impact. Fitting such a complicated model requires good data and careful use of statistical methods. We have introduced some useful methods in this paper.

We pointed out in Section 4, that the pilot study data is not sufficient to accurately estimate all of the transition intensity functions. We should therefore consider how much data is required. The accuracy with which we can estimate a transition intensity function depends greatly on the number of observed transitions. As we have seen, this varies considerably across transition types. We observed 371 transitions from independent to skilled care (temporary), and are able to estimate the transition intensity fairly accurately. The number of deaths from each state was rather small. However, if we are willing to assume that the intensity functions for mortality are proportional, we can combine the transitions. This allows us to obtain reasonable intensity estimates. Unfortunately, for some of the transition types, such as those from assisted living to skilled care (permanent), there were very few observed transitions, and we have little confidence in our estimates. Based on our analysis with this data set, we believe that, with five to ten times as much data, one could reasonably estimate all of the transition intensity functions.

Another important issue in deciding how much data is required relates to the completeness of the data. We determined that the intensity function for transitions from independent to skilled care (temporary) is influenced by the number of previous visits to skilled care. Unfortunately, for most residents, we do not know this number because information is only available for transitions that occur during a three-year study period. A similar problem exists in quantifying the effect of duration in the current state on the transition intensity. It would be ideal to have complete health status histories for all residents involved in the study. In summary, more data on a given group of residents is better than data on more residents.

Finally, we remind the reader that the numerical results presented in this paper are intended to be illustrative. They have been obtained using a small amount of data from one CCRC. One should not assume that these results are representative of the experience of other CCRCs.

Acknowledgment

This research was funded by the Society of Actuaries through a grant from the Committee on Knowledge Extension Research.

References

[1] Aalen, O.O. "Statistical inference for a family of counting processes." PhD Thesis, University of California, Berkeley, (1975).

[2] Aalen, O.O. "Nonparametric inference for a family of counting process," *Annals of Statistics* 6, (1978), 701–726.

[3] Actuarial Standards Board. Actuarial Standard of Practice No. 3, Practices Relating to Continuing Care Retirement Communities (Revised Edition), July 1994.

[4] Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. *Statistical Models Based on Counting Processes.* New York: Springer-Verlag, 1993.

[5] Andersen, P.K. and Gill, R.D. "Cox's regression model for counting processes: a large sample study," *Annals of Statistics* 10, (1982), 1100–1120.

[6] Brace, G.L. "Continuing Care Retirement Communities: A Primer," *Health Section News*, Society of Actuaries, No. 28, 4–6.

[7] Cox, D.R. "Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. B.* 34, (1972), 187–220.

[8] Cumming, R.B., and Bluhm, W.F. "CCRC Population and Financial Model," *The Actuary* Vol. 26, No. 3, 3–5.

[9] Fleming, T.R. and Harrington, D.P. *Counting Processes and Survival Analysis.* New York: Wiley, 1991.

[10] Gavin, J., Haberman, S. and Verrall, R. "Graduation by Kernel and Adaptive Kernel Methods with Boundary Correction," *TSA* XLVII, (1995), to appear.

[11] Hewitt, D.L. and DeWeese, C.C. "Long Life! Financially Sound CCRCs Can Help You Manage It and Enjoy It," *Contingencies*, Vol. 2, No. 5, September/October 1990, 28–30.

[12] Hogg, R.V. and Craig, A.T. *Introduction to Mathematical Statistics.* Englewood Cliffs, NJ: Prentice Hall, 1995.

[13] Jones, B.L. "A Stochastic Population Model for High Demand CCRCs," *Insurance: Mathematics and Economics* 16, (1995), 69–77.

[14] Jones, B.L. "Transient Results for a High Demand CCRC Model," *Scandinavian Actuarial Journal*, (1995), to appear.

[15] Jones, B.L. "Stochastic Models for Continuing Care Retirement Communities," submitted, (1995).

[16] Nelson, W. "Hazard plotting for incomplete failure data," *Journal of Quality Technology* 1, (1969), 27–52.

[17] Ramlau-Hansen, H. "Smoothing counting process intensities by means of kernel functions," *Annals of Statistics* 11, (1983), 453–466.

[18] Ramlau-Hansen, H. "The Choice of a Kernel Function in the Graduation of Counting Process Intensities," *Scandinavian Actuarial Journal* (1983), 165–182.

[19] Rodermund, M. "Letter from Medford: Security in a CCRC," *Contingencies*, Vol. 2, No. 5, September/October 1990, 41–45.

[20] Rodermund, M. "Letter from Medford," *Contingencies*, Vol. 5, No.5, September/October 1993, 8–9.

[21] Ross, S.M. *A Course in Simulation.* New York: Macmillan, 1990.

[22] Smith, K. "Continuing Care Retirement Communities: Should Consumers be Better Protected?" *Emphasis*, Summer 1988, 14–15.

[23] Winklevoss, H.E. and A.V. Powell. *Continuing Care Retirement Communities: An Empirical, Financial, and Legal Analysis.* Homewood, IL: Richard D. Irwin, 1984.

A Tables

Table 1: Time spent in each state by sex

State	Females	Males	Total
Independent	807.7	288.8	1096.5
Assisted Living	141.1	22.1	163.2
Skilled Care (Temporary)	37.4	6.9	44.3
Skilled Care (Permanent)	185.3	28.4	213.7
Total	1171.4	346.2	1517.6

Table 2: Number of transitions by type and sex

Type	Females	Males	Total
$1 \to 2$	29	3	32
$1 \to 3$	308	63	371
$1 \to 4$	2	0	2
$1 \to 5$	11	6	17
$1 \to 6$	21	10	31
$2 \to 1$	3	0	3
$2 \to 3$	156	18	174
$2 \to 4$	10	2	12
$2 \to 5$	3	1	4
$2 \to 6$	8	1	9
$3 \to 1$	244	44	288
$3 \to 2$	155	18	173
$3 \to 4$	42	10	52
$3 \to 5$	4	0	4
$3 \to 6$	16	7	23
$4 \to 1$	0	0	0
$4 \to 2$	3	2	5
$4 \to 3$	4	0	4
$4 \to 5$	2	1	3
$4 \to 6$	56	7	63

Table 3: Deaths from the independent state

Females		Males	
Age	At Risk	Age	At Risk
74.00958	17	75.69884	11
76.21903	24	77.27584	10
77.67830	30	80.18891	17
77.67830	30	80.38877	18
79.29363	45	80.60780	17
79.29911	44	83.38946	24
82.02053	61	85.13621	22
83.14305	62	87.34565	13
83.36482	62	89.84531	12
84.27926	53	90.14648	8
85.17180	46		
85.60986	46		
86.25873	46		
87.09651	53		
87.72621	53		
89.06776	35		
90.62834	23		
92.06571	19		
92.90075	14		
94.06160	9		
101.54689	1		

Table 4: Cox Regression Results for Withdrawals

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
male	0.2361	1.27	0.4270	0.58040	0.548	2.92
entry age	0.0968	1.10	0.0573	0.09137	0.985	1.23
assisted	0.6103	1.84	0.5985	0.30785	0.570	5.95
skilled (t)	1.8725	6.50	0.5808	0.00126	2.084	20.31
skilled (p)	0.3135	1.37	0.6547	0.63204	0.379	4.94
Likelihood ratio statistic = 11.4 on 5 df, p=0.0439						
Efficient score statistic = 16.5 on 5 df, p=0.00551						

Table 5: Cox Regression Results for Mortality

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
male	0.239	1.27	0.240	0.320	0.793	2.03
assisted	0.448	1.57	0.394	0.256	0.723	3.39
skilled (t)	2.773	16.01	0.292	0	9.042	28.35
skilled (p)	2.082	8.02	0.245	0	4.964	12.95
Likelihood ratio statistic = 123 on 4 df, p=0						
Efficient score statistic = 166 on 4 df, p=0						

Table 6: Cox Regression Results for 1 $\rightarrow$ 3 Transitions

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
male	-0.527	0.590	0.2091	0.0117	0.392	0.890
dur	-0.504	0.604	0.0532	0	0.545	0.671
Likelihood ratio statistic = 91.4 on 2 df, p=0						
Efficient score statistic = 125 on 2 df, p=0						

Table 7: Cox Regression Results for 1 $\rightarrow$ 2 Transitions

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
male	-1.26	0.284	0.608	0.0386	0.0863	0.936
Likelihood ratio statistic = 5.96 on 1 df, p=0.0147						
Efficient score statistic = 4.87 on 1 df, p=00274						

Table 8: Cox Regression Results for 2 $\rightarrow$ 3 Transitions

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
male	-0.849	0.428	0.3256	0.00915	0.226	0.81
log(dur of stay)	-0.349	0.706	0.0641	0	0.622	0.80
Likelihood ratio statistic = 44.8 on 2 df, p=0						
Efficient score statistic = 46.4 on 2 df, p=0						

Table 9: Cox Regression Results for (1 →)3 → 1 Transitions

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
male	-0.2871	0.750	0.1661	0.0839	0.542	1.039
age ent skilled	-0.0657	0.936	0.0168	0.00009	0.906	0.968
age ent CCRC	0.0339	1.035	0.0165	0.0399	1.002	1.069
Likelihood ratio statistic = 18.6 on 3 df, p=0.0003						
Efficient score statistic = 17.8 on 3 df, p=0.0005						

Table 10: Cox Regression Results for (1 →)3 → 2 Transitions

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
dur in CCRC at start of stay	0.0893	1.09	0.029	0.00209	1.03	1.16
Likelihood ratio statistic = 8.13 on 1 df, p=0.00435						
Efficient score statistic = 9.86 on 1 df, p=0.00169						

Table 11: Cox Regression Results for (2 →)3 → 2 Transitions

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
dur in CCRC at start of stay	-0.0341	0.966	0.0169	0.0436	0.935	0.999
Likelihood ratio statistic = 4.32 on 1 df, p=0.0376						
Efficient score statistic = 4.1 on 1 df, p=0.0428						

Table 12: Cox Regression Results for 3 → 4 Transitions

Covariate	Coefficient	exp(Coef)	std error	p-value	lower 95%	upper .95%
prev state = 2	0.955	2.6	0.286	0.00082	1.49	4.55
Likelihood ratio statistic = 11.2 on 1 df, p=0.0008						
Efficient score statistic = 12.1 on 1 df, p=0.0005						

Table 13: Simulation Results for 75 Year-Old New Residents

Females						
	Probability of being in					
Age	state 1	state 2	state 3	state 4	state 5	state 6
75	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
76	0.7492	0.0407	0.0759	0.0704	0.0284	0.0354
77	0.5474	0.0802	0.0728	0.1606	0.0566	0.0824
78	0.3906	0.1050	0.0561	0.2305	0.0804	0.1374
79	0.2734	0.1095	0.0442	0.2732	0.1002	0.1995
80	0.1912	0.1087	0.0311	0.2913	0.1167	0.2610
81	0.1339	0.0954	0.0219	0.2942	0.1315	0.3231
82	0.0885	0.0833	0.0178	0.2839	0.1434	0.3831
83	0.0554	0.0691	0.0151	0.2632	0.1528	0.4444
84	0.0363	0.0577	0.0088	0.2393	0.1609	0.4970
85	0.0236	0.0443	0.0049	0.2120	0.1668	0.5484
86	0.0154	0.0349	0.0035	0.1821	0.1718	0.5923
87	0.0097	0.0261	0.0037	0.1515	0.1754	0.6336
88	0.0051	0.0203	0.0020	0.1246	0.1771	0.6709
89	0.0024	0.0154	0.0015	0.1018	0.1790	0.6999
90	0.0012	0.0107	0.0011	0.0803	0.1802	0.7265
91	0.0006	0.0069	0.0007	0.0616	0.1811	0.7491
92	0.0003	0.0052	0.0005	0.0441	0.1817	0.7682
93	0.0002	0.0032	0.0008	0.0328	0.1819	0.7811
94	0.0001	0.0020	0.0000	0.0232	0.1819	0.7928
95	0.0000	0.0013	0.0001	0.0160	0.1819	0.8007
Males						
	Probability of being in					
Age	state 1	state 2	state 3	state 4	state 5	state 6
75	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
76	0.8198	0.0256	0.0461	0.0417	0.0337	0.0331
77	0.6650	0.0567	0.0424	0.0949	0.0641	0.0769
78	0.5311	0.0764	0.0416	0.1327	0.0887	0.1295
79	0.4195	0.0931	0.0345	0.1551	0.1135	0.1843
80	0.3283	0.1039	0.0281	0.1639	0.1313	0.2445
81	0.2500	0.1092	0.0223	0.1746	0.1476	0.2963
82	0.1908	0.1079	0.0175	0.1691	0.1601	0.3546
83	0.1443	0.1054	0.0137	0.1585	0.1729	0.4052
84	0.1022	0.1021	0.0112	0.1465	0.1809	0.4571
85	0.0762	0.0936	0.0074	0.1328	0.1876	0.5024
86	0.0572	0.0822	0.0054	0.1162	0.1948	0.5442
87	0.0403	0.0742	0.0061	0.0985	0.1985	0.5824
88	0.0289	0.0650	0.0036	0.0841	0.2022	0.6162
89	0.0189	0.0549	0.0032	0.0691	0.2042	0.6497
90	0.0121	0.0454	0.0024	0.0582	0.2057	0.6762
91	0.0084	0.0391	0.0013	0.0438	0.2074	0.7000
92	0.0058	0.0308	0.0013	0.0319	0.2080	0.7222
93	0.0029	0.0264	0.0013	0.0236	0.2082	0.7376
94	0.0019	0.0202	0.0009	0.0176	0.2089	0.7505
95	0.0013	0.0165	0.0002	0.0119	0.2090	0.7611

B Figures

Figure 1: State transition diagram for CCRC residents

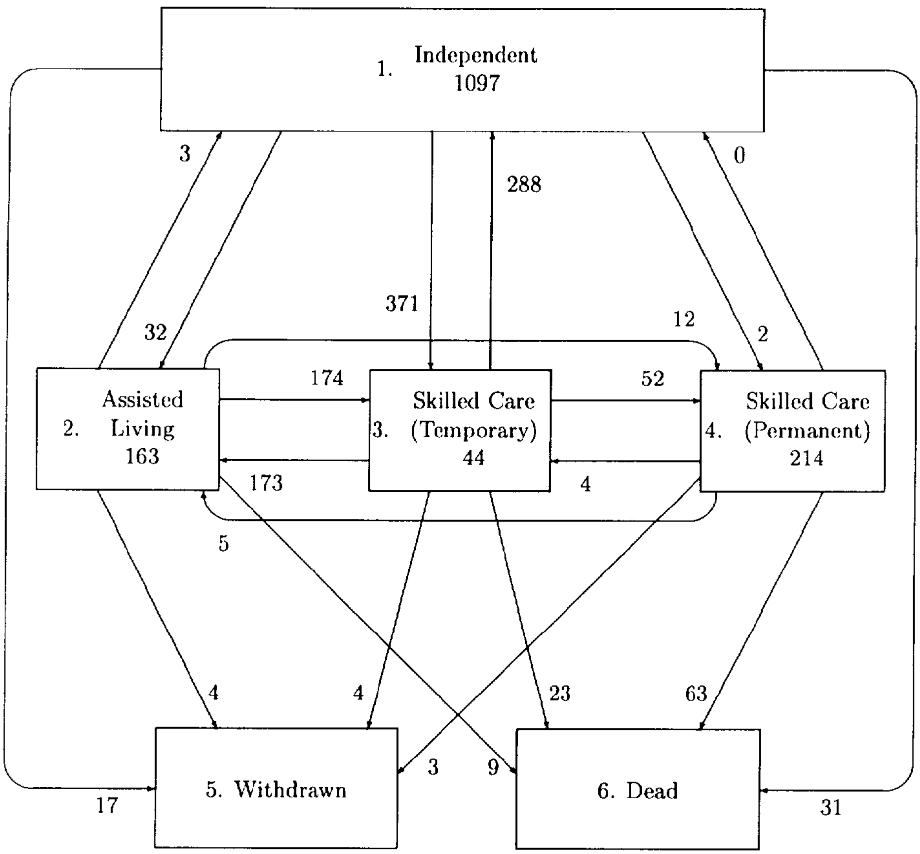


Figure 2: Number of residents attaining each age during the study period

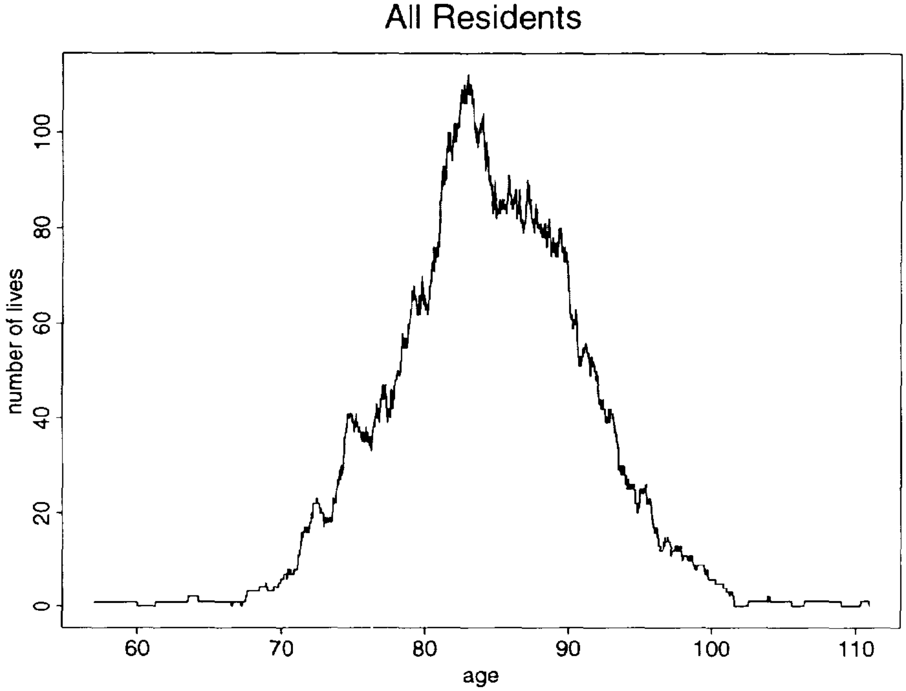


Figure 2: Number of residents attaining each age during the study period

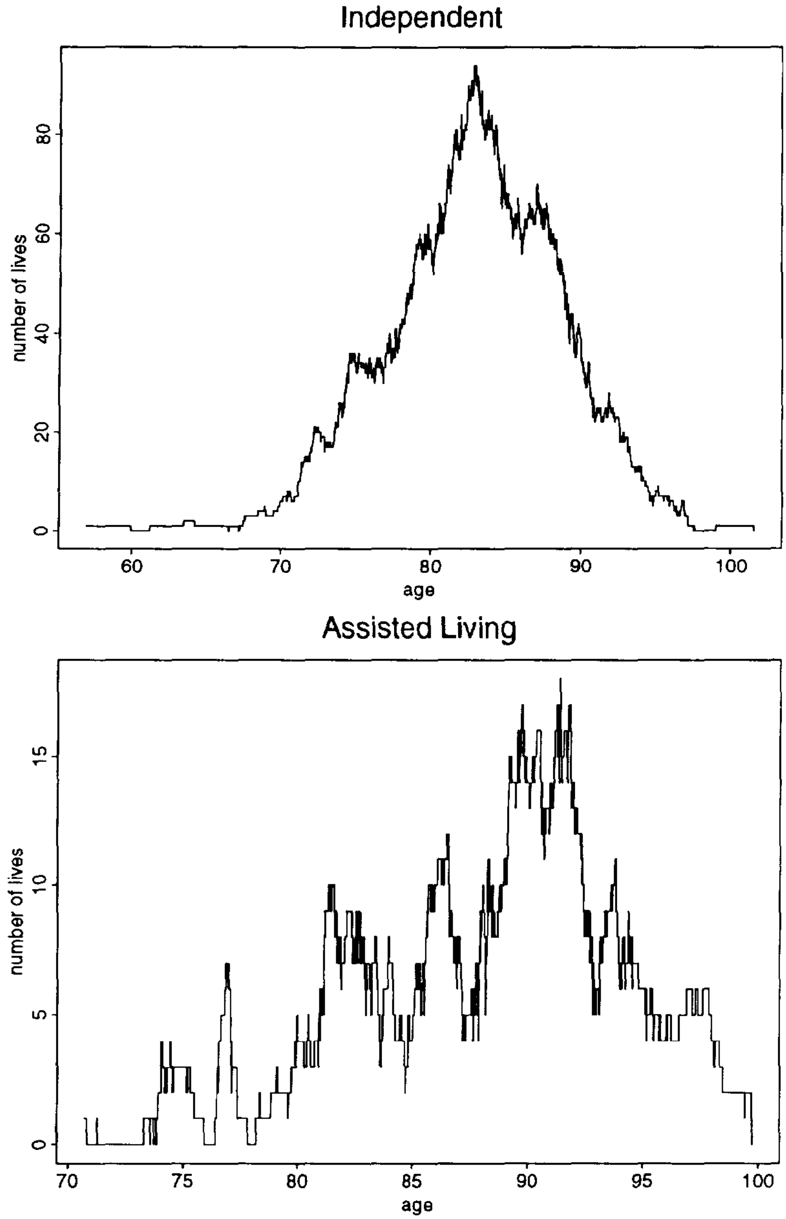


Figure 2: Number of residents attaining each age during the study period

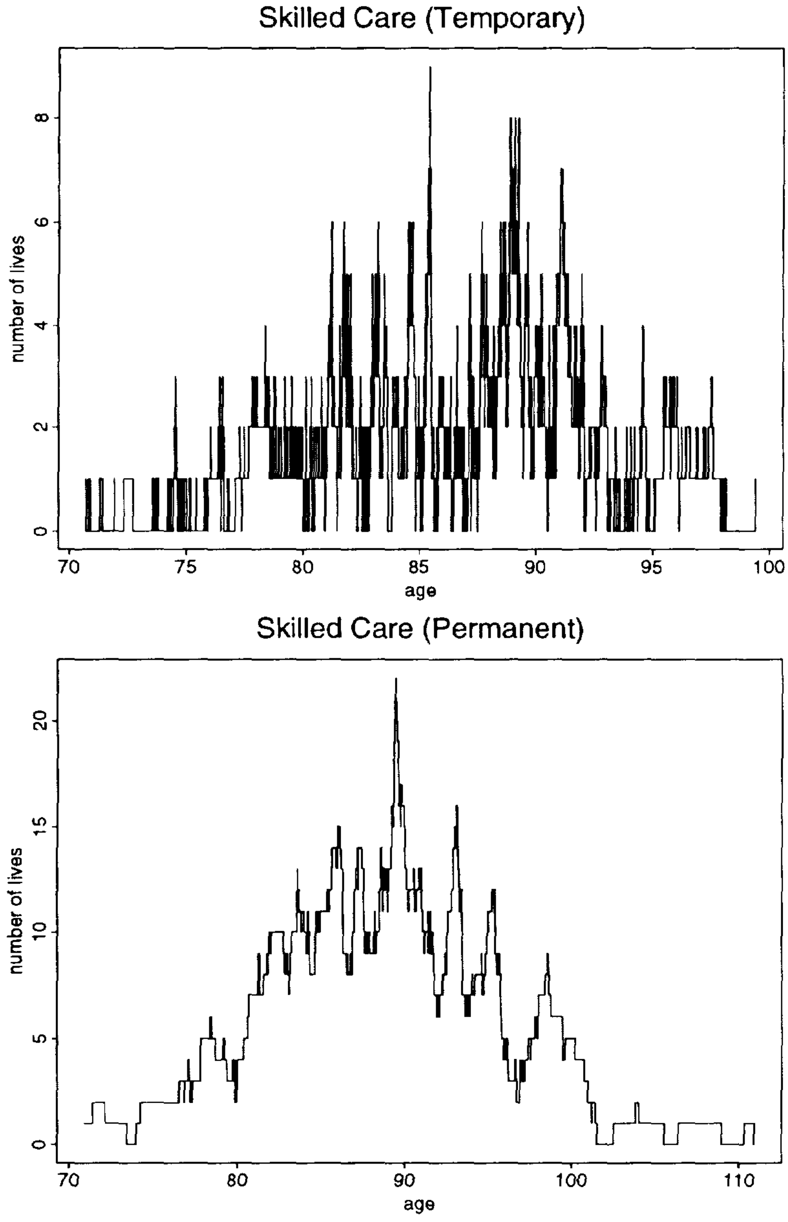


Figure 3: Number of residents attaining each duration during the study period

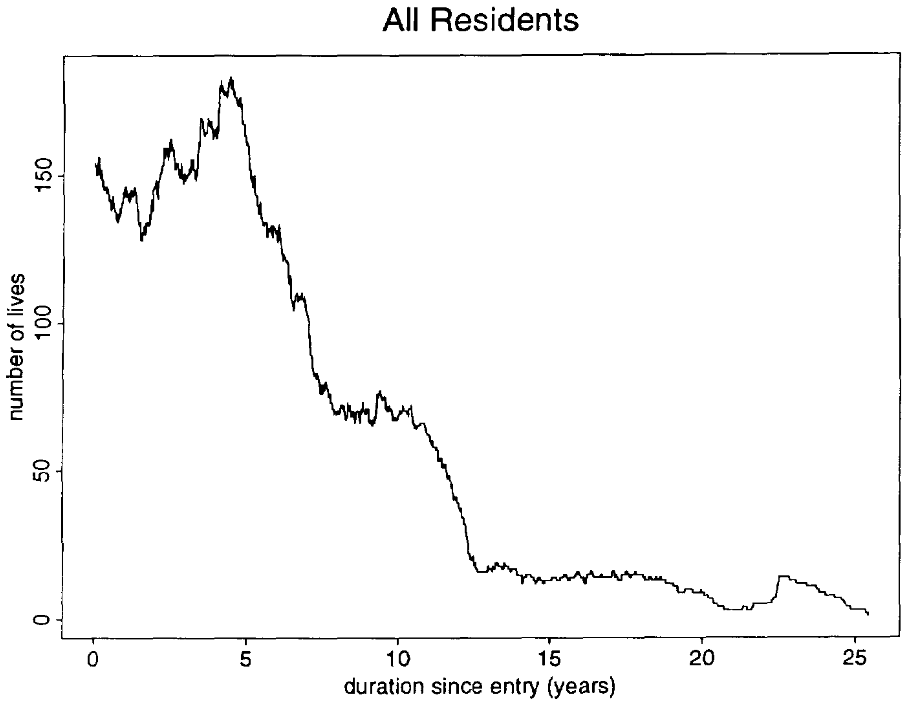


Figure 4: Estimated cumulative intensity functions for female and males (dotted)

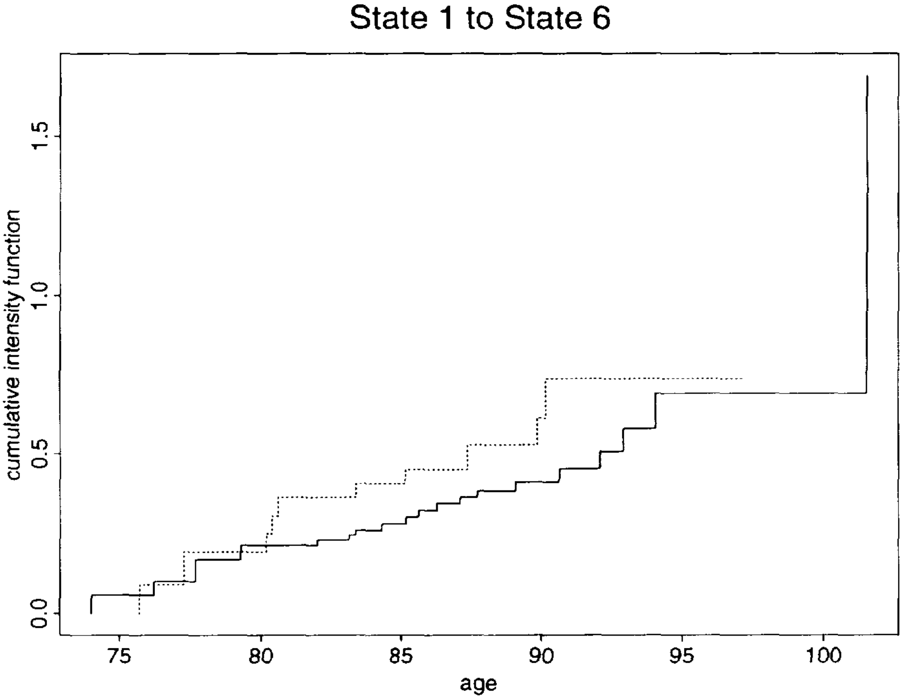
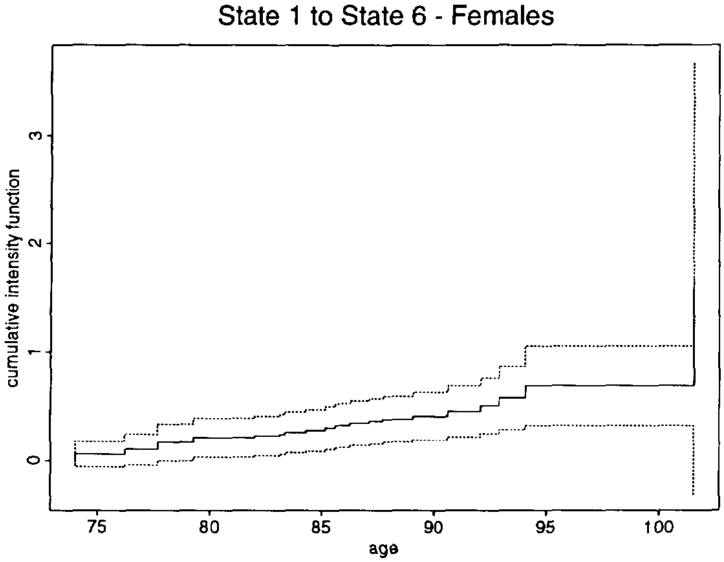


Figure 5: Estimated cumulative intensity functions with approximate 95 percent pointwise confidence limits (dotted)



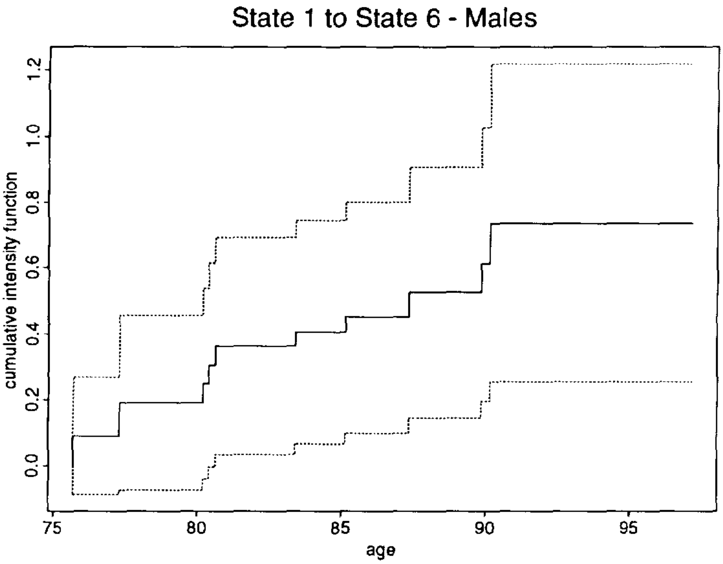


Figure 6: Smoothed intensity functions with approximate 95 percent pointwise confidence limits (dotted)

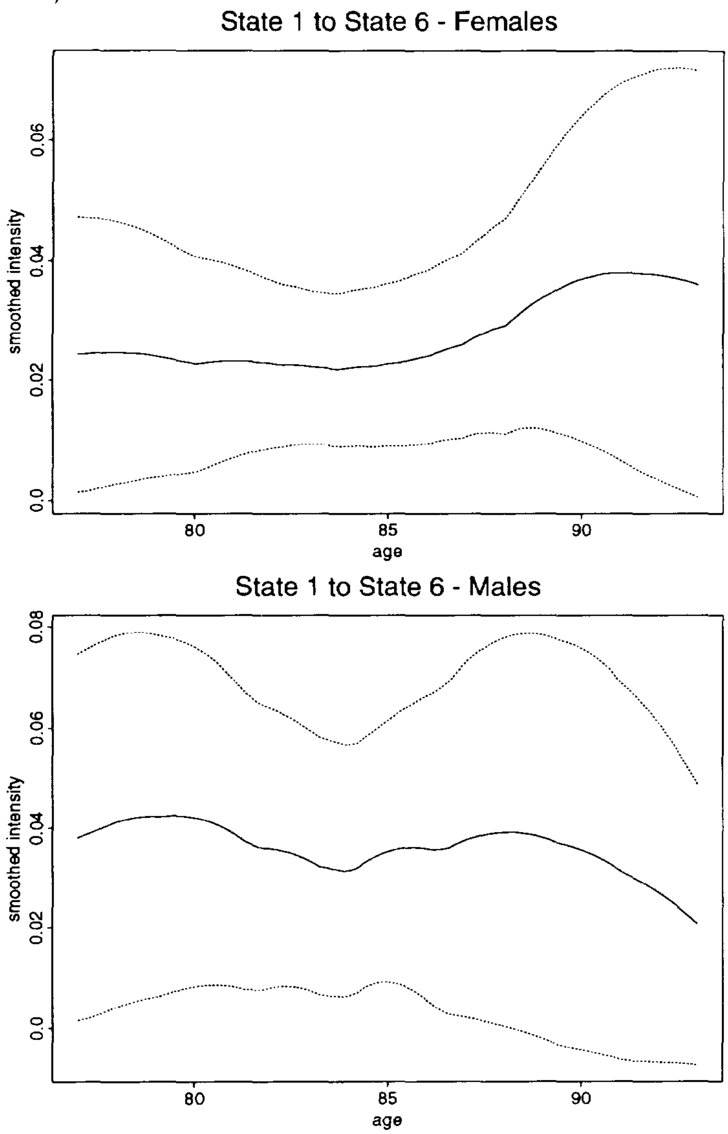


Figure 7: Cumulative and Smoothed Baseline Withdrawal Intensity Estimates

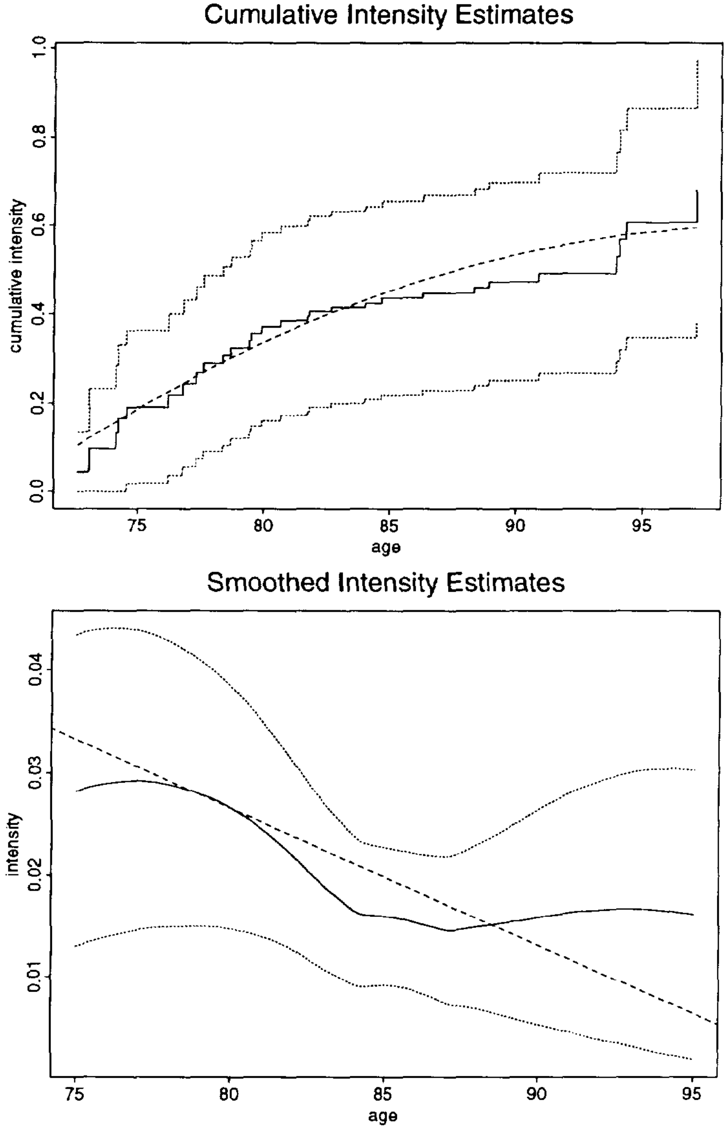


Figure 8: Cumulative and Smoothed Baseline Mortality Intensity Estimates

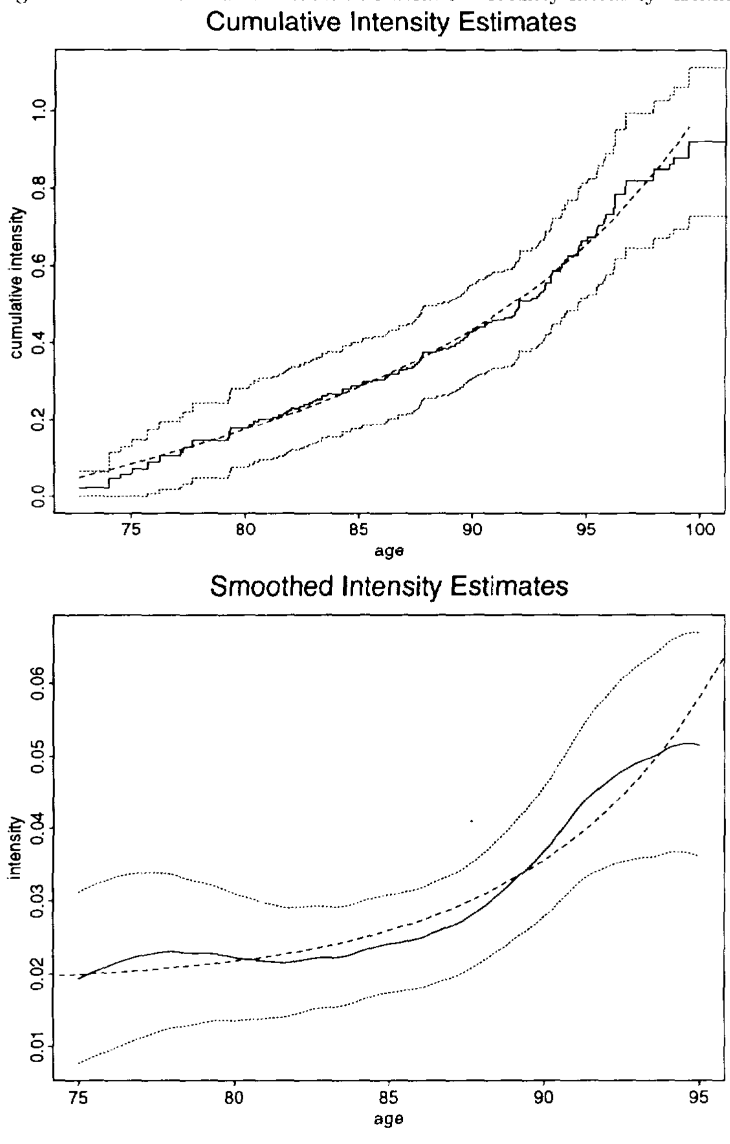


Figure 9: Cumulative and Smoothed Baseline 1 → 3 Intensity Estimates

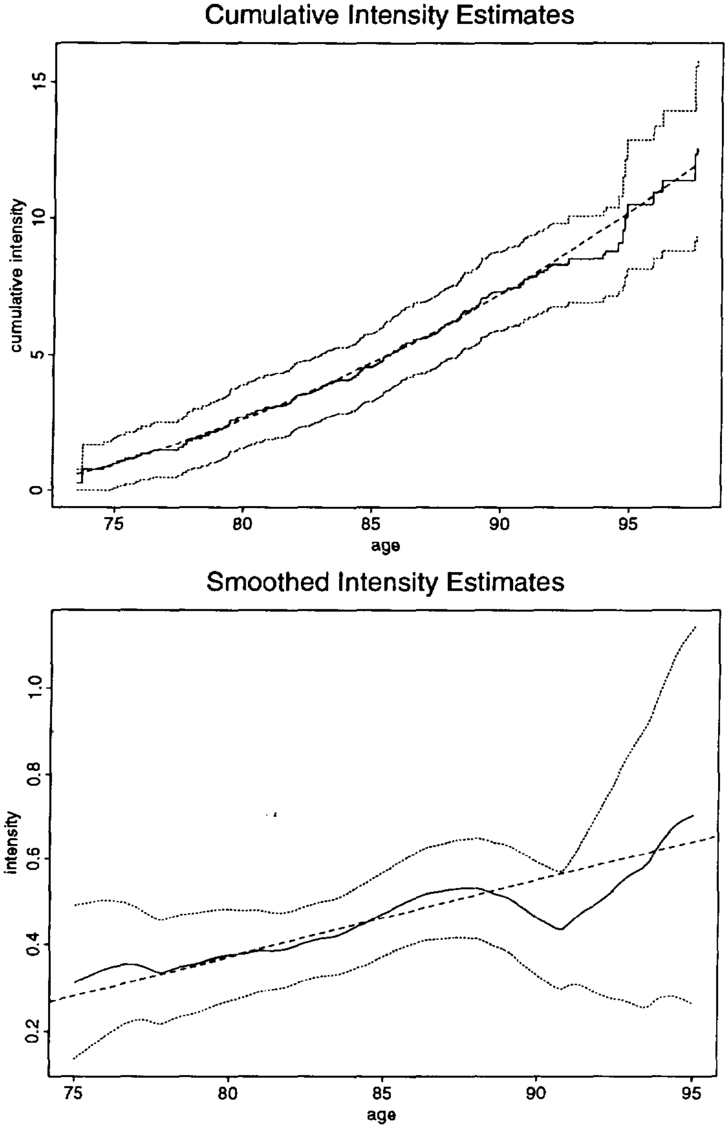


Figure 10: Cumulative and Smoothed Baseline 1 → 2 Intensity Estimates

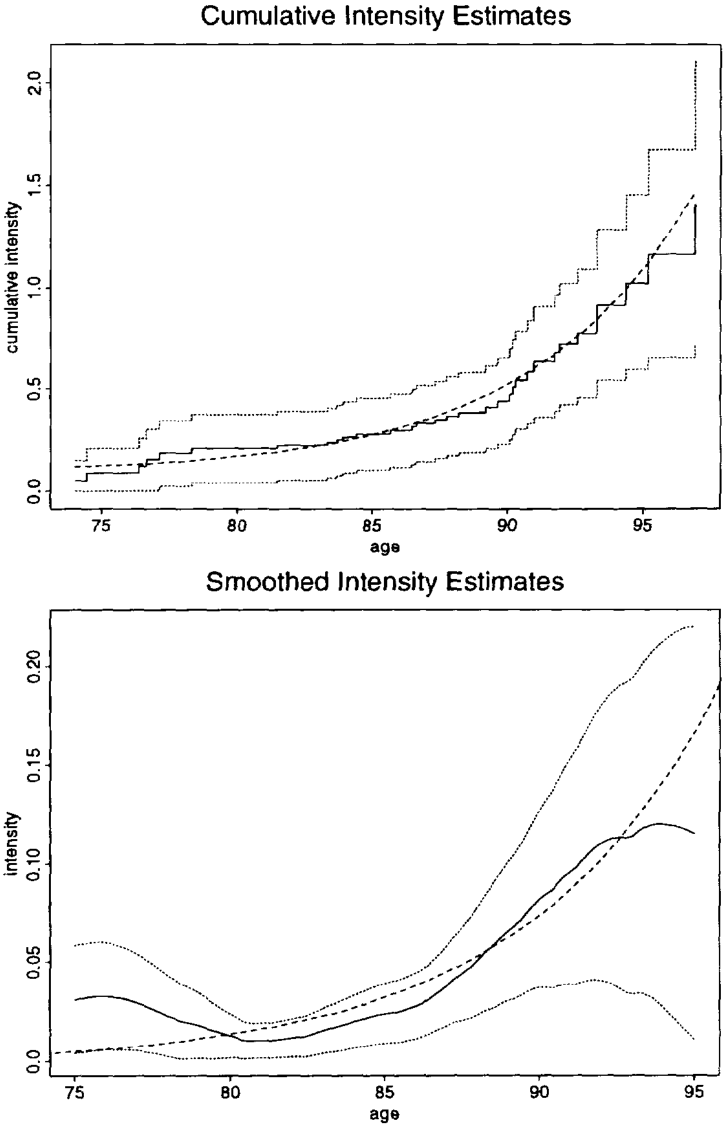


Figure 11: Cumulative and Smoothed Baseline 2 → 3 Intensity Estimates

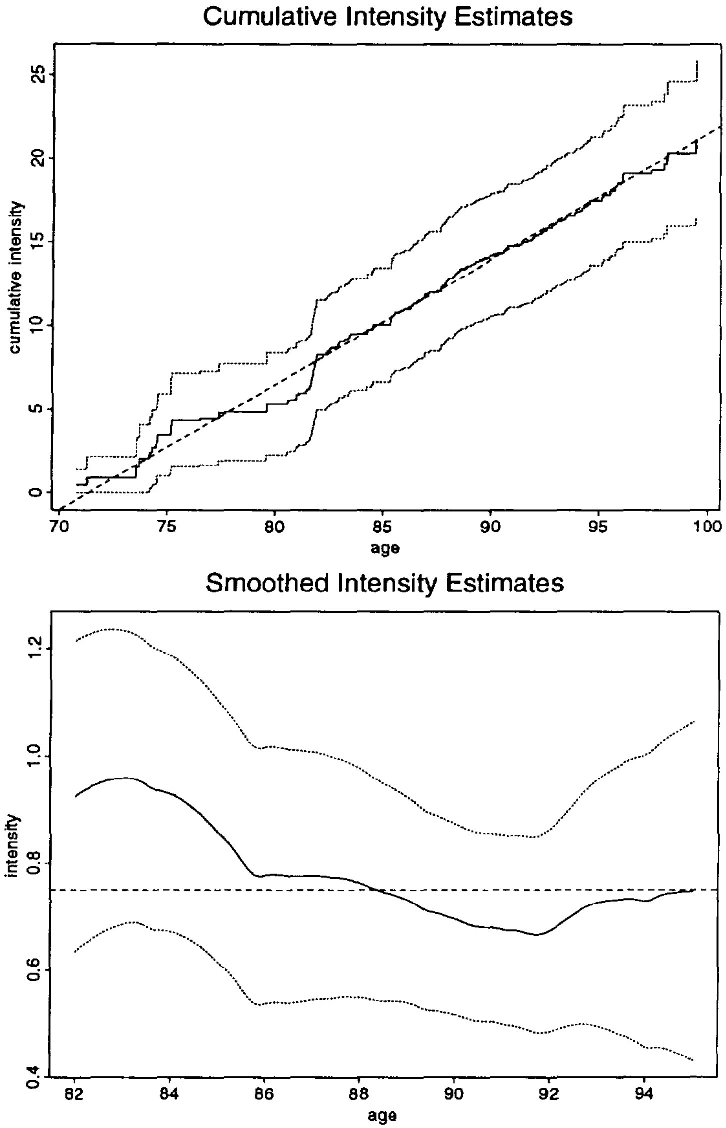


Figure 12: Cumulative Baseline 3 → 1 Intensity Estimates

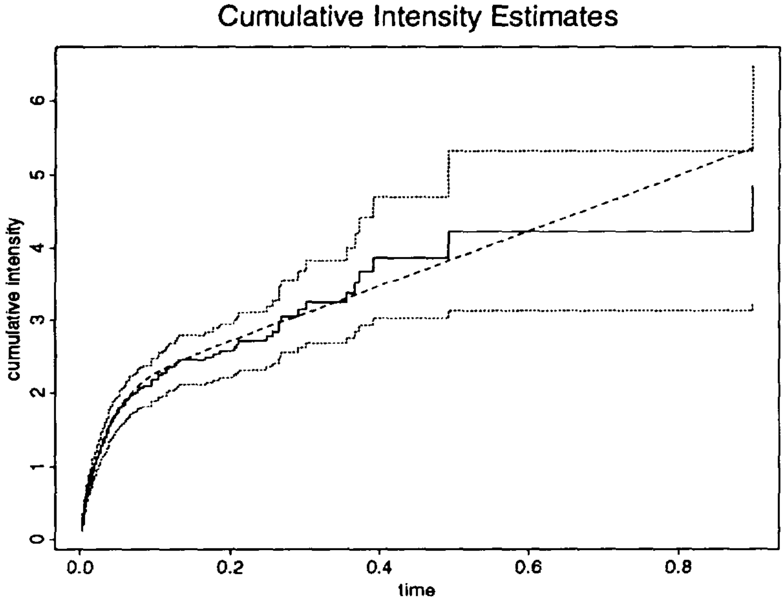


Figure 13: Cumulative Baseline 3 → 2 Intensity Estimates

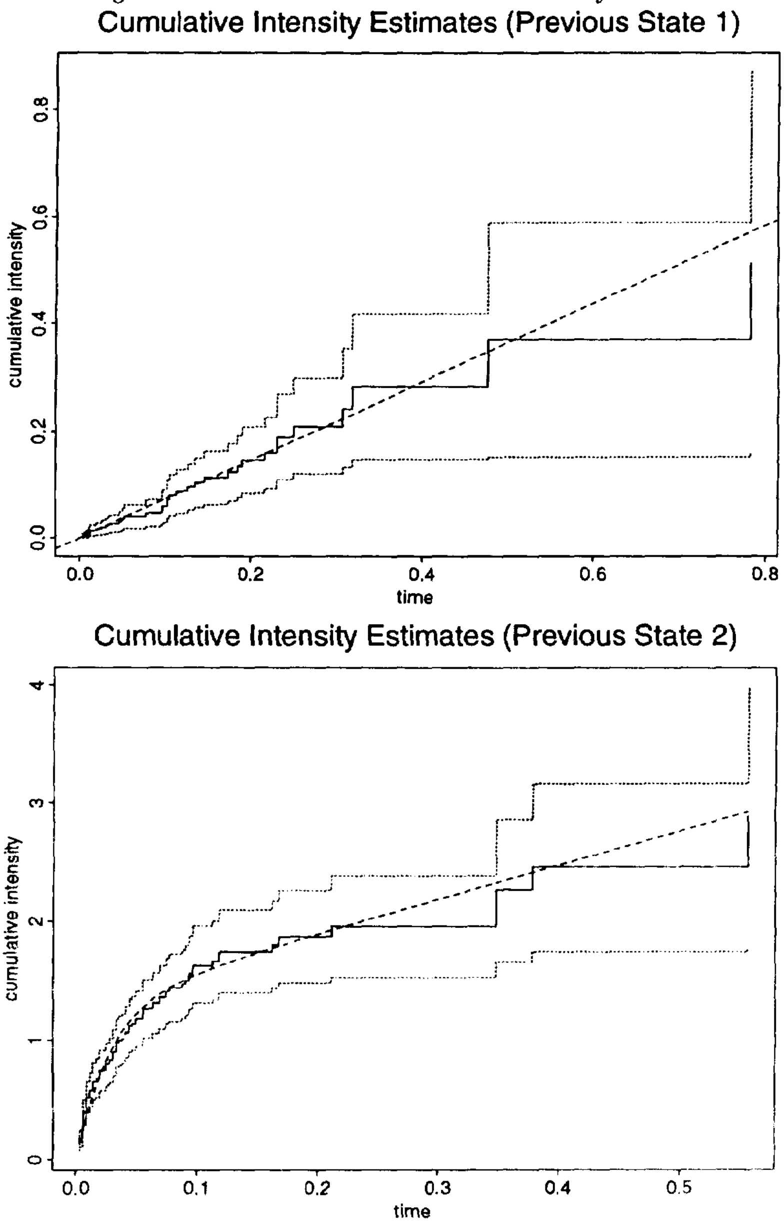


Figure 14: Cumulative Baseline 3 → 4 Intensity Estimates