DEPT. OF MATH. UNIVERSITY OF OSLO
STATISTICAL RESEARCH REPORT No. 2
ISBN 82–553–1293–5 June 2001
ISSN 0806–3842

Covariate adjustment of event histories estimated from Markov chains: The additive approach

Odd O. Aalen

Section of Medical Statistics, University of Oslo

P.O. Box 1122 Blindern, N-0317 Oslo, Norway

Ørnulf Borgan

Institute of Mathematics, University of Oslo

P.O. Box 1053 Blindern, N-0316 Oslo, Norway

Harald Fekjær

Section of Medical Statistics, University of Oslo

P.O. Box 1122 Blindern, N-0317 Oslo, Norway

and

The Cancer Registry of Norway, Montebello, N-0310 Oslo, Norway

Abstract

Markov chain models are much used for studying event histories that include transitions between several states. An empirical transition matrix for non-homogeneous Markov chains has previously been developed, including a detailed statistical theory based on counting processes and martingales. In this paper we show how to estimate transition probabilities dependent on covariates. This technique may, for instance, be used for making estimates of individual prognosis in epidemiological or clinical studies. The covariates are included through non-parametric additive models on the transition intensities of the Markov chain. The additive model allows for estimation of covariate dependent transition intensities, and again a detailed theory exists based on counting processes. The martingale setting now allows for a very natural combination of the empirical transition matrix and the additive model, resulting in estimates that can be expressed as stochastic integrals, and hence their properties are easily evaluated. Two medical examples will be given. In the first example we study how the lung cancer mortality of uranium miners depend on smoking and radon exposure. In the second example we study how the probability of being in response depends on patient group and prophylactic treatment for leukemia patients who have had a bone marrow transplantation. A program in R and S-PLUS that can carry out the analyses described here is being developed and will be freely available on the Internet.

1 Introduction

In clinical trials, epidemiology and other fields, one often observes a number of individuals passing through several states. Each individual may have a set of covariates measured, and one may want to estimate probabilities of interesting transitions within the state space and how these transitions depend on the covariates.

There are two steps in this problem. The first one is to estimate how transition rates between the various states depend on the covariates. The second one is to combine the estimated rates for a given set of covariates into appropriate transition probabilities. We shall point out how certain well-established methods for each of these steps fit very naturally together. For estimating the covariate effects we use the nonparametric additive regression model, and for calculating the transion probabilities we use an empirical transition matrix. Both have a crucial dependence on counting process and martingale theory, and their properties have been developed within this framework. It is precisely the martingale properties which make it easy to combine the two approaches, and the theory presented here is an example of how quite complex procedures may be handled elegantly in the counting process framework.

We shall assume that the transitions on the state space follow a Markov chain. A very simple example of a Markov chain is the competing risks model; here the Markov assumption is trivially fulfilled. A slightly more complex example is the illness/death model where an individual may move from a healthy state to an illness state before he eventually dies. Here the Markov property is less obvious, but still a common assumption. Both the models mentioned here are used as examples below.

It can be shown that estimators derived in a Markov context are sometimes less dependent on the Markov assumption than one might think. This seems first to have been recognized by Datta and Satten (2000). Although most of the paper will be kept within the Markov framework, the possibility of extending the results to the non-Markov case will be discussed in Section 7.

The issue of combining covariate information with estimation of transition probabilities has been studied before, using a Cox proportional hazards model (Andersen, Hansen and Keiding, 1991; Andersen, Borgan, Gill and Keiding, 1993, Section VI.2.3). Although useful, a difficulty with that approach is that no exact martingale properties are achieved. This considerably complicates the evaluation of the estimation procedure. Furthermore, the proportional hazards model is non-local, averaging possibly time-dependent covariate effects over a long time period. It is of interest to use a local model, which immediately picks up changes in the effects of covariates, and the additive model has this property.

We shall first review the empirical transition matrix for time-continuous Markov chains. Thereafter we show how to combine it with the additive regression model. Two real examples are then analyzed in detail. We shall presuppose some knowledge of the counting process approach to survival analysis (see Fleming and Harrington, 1991, or Andersen *et al.*, 1993).

A program in R and S-PLUS for analyzing survival data with the additive regression model is available on the web site `www.med.uio.no/imb/stat/addreg`. The program is being developed to fit the counting process framework, and will also be developed to analyze Markov chain models like those decribed in this paper.

2 The empirical transition matrix

The probability distribution of a Markov process is determined, apart from an initial distribution, by the transition probability matrix defined on the state space, which will be assumed to be finite. We will consider processes where time, denoted by t, runs continuously. Let us first look at a finite number of increasing time points $t_1 < t_2 < \ldots < t_n$. By the Markov assumption, the transition probability matrix from time $s = t_1$ to $t = t_n$ can be written as the following matrix product:

$$\mathbf{P}(s,t) = \mathbf{P}(t_1,t_2) \times \mathbf{P}(t_2,t_3) \times \cdots \times \mathbf{P}(t_{n-1},t_n) \tag{1}$$

If the number of time points increase, while the distance between them goes to zero in a uniform way, the product on the right hand side approaches a limit which is termed a product integral, see, for instance, Andersen *et al.* (1993, Section II.6). The product integral is most usefully written in terms of the intensity matrix, that is the matrix where the $(i,j)th$ off-diagonal element equals the transition intensity $\alpha_{ij}(t)$ between states i and j and the diagonal elements $\alpha_{ii}(t)$ are chosen so that all row sums are zero. Let $\mathbf{A}(t)$ denote the cumulative, or integrated, intensity matrix with elements $A_{ij}(t) = \int_0^t \alpha_{ij}(u)du$, and let $\mathbf{I}$ be an identity matrix. Then $\mathbf{P}(u, u+du) \approx \mathbf{I} + d\mathbf{A}(u)$. This explains why we may write the limit of equation (1) in product integral form as

$$\mathbf{P}(s,t) = \prod_{(s,t]}(\mathbf{I} + d\mathbf{A}(u)) \tag{2}$$

The product integral notation $\prod$ is used to suggest a limit of finite products $\prod$, just as the integral $\int$ is a limit of finite sums $\sum$.

We will now pass on to the question of estimating the transition probability matrix. Assume that one observes a number of individuals moving on the Markov chain. In event history analysis this could, for instance, be people moving between different states of health including death. One may then want to estimate the probability of moving from an initial state to some other state during a given time interval, that is, one wants to estimate the transition probability matrix between two points in time. A complication commonly arising is that of censoring, that is, individuals are lost to further observation at various times throughout the observation period.

The estimation may be tackled either within the time-discrete setting or time-continuously, corresponding to the two product formulas (1) and (2). The time-discrete version corresponds to the life-table (also called actuarial) estimator, while the time-continuous version generalizes the Kaplan-Meier estimator.

Looking first at equation (1), one would estimate the left hand side, through estimating each term in the product on the right. In the absence of censoring, the transition probability matrix over an interval would be estimated simply as the observed proportions of transition between various states. If censoring occurs, simple modifications can be made as is done in classical life-table analysis, for instance assuming that censored individuals are observed on the average over half the time interval.

In order to estimate the product integral in equation (2), a product limit argument is used just as in the derivation of the Kaplan Meier estimator. It is assumed that just one transition occurs at any given time, and that this does not coincide with a censoring. Censoring, however, may occur singly or in batches. Assume that a transition occurs at time t between states i and j, and that $Y_i(t)$ is the number of individuals at risk in state i just before time t. Then the following "transition matrix" corresponding to this event may be defined (as an example we give a 5×5 matrix):

$$\begin{array}{cccccc} & & & & & j: \\ & 1 & 0 & 0 & 0 & 0 \\ & 0 & 1 & 0 & 0 & 0 \\ i: & 0 & 0 & 1-\frac{1}{Y_i(t)} & 0 & \frac{1}{Y_i(t)} \\ & 0 & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 0 & 1 \end{array}$$

Hence, this is an identity matrix except for row i where a part of the probability mass is moved to column j. Such matrices are then multiplied together for all transitions over the relevant time interval. Note that this computation in a highly natural way generalizes the Kaplan-Meier estimator. The procedure was derived by Aalen and Johansen (1978) and is sometimes called the Aalen-Johansen estimator. We will here denote it the *empirical transition matrix*. A similar procedure, but not in the matrix form presented here, was studied by Fleming (1978a,b). A non-technical review is given by Borgan (1998).

The empirical transition matrix may be written as a product integral. In equation (2) one simply replaces $\mathbf{A}(t)$ with an estimate. The simplest way to do this is to substitute, for each pair of states (i, j), the cumulative intensity $A_{ij}(t)$ with the Nelson-Aalen estimate $\hat{A}_{ij}(t)$, and define $\hat{A}_{ii}(t) = -\sum_{j \neq i} \hat{A}_{ij}(t)$. The resulting estimated intensity matrix is denoted $\hat{\mathbf{A}}_0(t)$, which substituted into equation (2) yields the estimate:

$$\hat{\mathbf{P}}_0(s,t) = \prod_{(s,t]} (\mathbf{I} + d\hat{\mathbf{A}}_0(u)) \tag{3}$$

Notice that the product integral in (3) is simply a finite product over the matrices $\mathbf{I} + \Delta\hat{\mathbf{A}}_0(u)$ that arise when there is a transition taking place. Thus (3) is exactly the estimator discussed above. Based on martingale results, there is a nice theory for the empirical transition matrix including an estimate of its covariance matrix and asymptotic results. The reason why this works so well is the combination of the martingale properties of the Nelson-Aalen estimator with the product integral form, and the fact that product stochastic integrals, just like ordinary stochastic integrals, preserve the martingale structure; see Section IV.4 in Andersen *et al.* (1993) for a review.

3 Combining the empirical transition matrix with the additive regression model

We shall assume that for each transition intensity of the Markov model an additive regression model based on some covariates has been defined, and fitted from follow-up data for a number of individuals. More specifically, consider a pair of states (i,j) for which there is a positive transition intensity. Fixing for the moment (i,j), we introduce the notation $\alpha_{ij}(t,\mathbf{Z})$ for the intensity to indicate that it may depend on covariates, $\mathbf{Z}(t) = (Z_1(t),\ldots,Z_k(t))'$, that are specific to each individual. In the additive model one assumes that these covariates determine the intensity in the following way:

$$\alpha_{ij}(t,\mathbf{Z}) = \beta_0(t) + \beta_1(t)Z_1(t) + \cdots + \beta_k(t)Z_k(t) \tag{4}$$

Here, the regression functions $\beta_h(t)$, and possibly also the covariates, will depend on which pair of states (i,j) is being considered.

We now define a design matrix $\mathbf{Y}_{ij}(t)$ where there is one row for each individual. For a typical individual, if he is still at risk, the row is $(1, Z_1(t),\ldots,Z_k(t))$, and if the individual is not at risk, the row consists of zeroes. Then the multivariate intensity process for the multivariate counting process counting transitions between states i and j for all individuals can be written in the following matrix form:

$$\boldsymbol{\lambda}_{ij}(t) = \mathbf{Y}_{ij}(t)\boldsymbol{\beta}_{ij}(t) \tag{5}$$

where $\boldsymbol{\beta}_{ij}(t)$ is the vector $(\beta_0(t),\beta_1(t),\ldots,\beta_k(t))'$. We introduce the cumulative regression functions $B_h(t) = \int_0^t \beta_h(u)du$ and denote the vector of these functions for given (i,j) by $\mathbf{B}_{ij}(t)$.

Then, following Aalen (1980, 1989), see also Andersen *et al.* (1993, Section VII.4), the vector of cumulative regression functions are estimated by means of

$$\hat{\mathbf{B}}_{ij}(t) = \int_0^t \mathbf{X}_{ij}(s) d\mathbf{N}_{ij}(s) \tag{6}$$

Here $\mathbf{X}_{ij}(t)$ is a generalized inverse of $\mathbf{Y}_{ij}(t)$ and $\mathbf{N}_{ij}(t)$ is the multivariate counting process, considered as a column vector, counting for each individual the transitions from state i to state j. Note that this integral is simply a *finite sum* over the jumps in the counting process. Note also that the estimation procedure is only well defined at times for which $\mathbf{Y}_{ij}(t)$ has full rank, i.e. rank equal to the number of columns. The most common choice for generalized inverse is the least square inverse defined by $\mathbf{X}_{ij}(t) = (\mathbf{Y}'_{ij}(t)\ \mathbf{Y}_{ij}(t))^{-1}\mathbf{Y}'_{ij}(t)$, and we will use this choice in the examples in Section 6. With a slight modification to take care of time intervals when $\mathbf{Y}_{ij}(t)$ is not of full rank, $\hat{\mathbf{B}}_{ij}(t) - \mathbf{B}_{ij}(t)$ becomes a martingale, and this is the basis for the statistical theory of these estimators.

There exists a considerable amount of theory for the cumulative regression estimators defined above. There are tests for the null hypothesis $\beta_h(t) = 0$ over some time interval (Aalen, 1989; see also Andersen *et al.*, 1993, Section VII.4), and one may estimate the regression function $\beta_h(t)$ by smoothing techniques (Aalen, 1993). There also exists residual plots for checking the fit of the model (Aalen, 1993).

Our aim in this paper is to compute transition probabilities in the Markov chain dependent on covariate information. Hence, we wish to combine the empirical transition matrix with the estimators in the regression model. Consider a pair of states (i, j) and a vector of possibly time-dependent covariates $\mathbf{Z}^0_{ij}(t)$, where this denotes a specified (i.e. fixed) time-dependent covariate history. On the basis of data from a study we estimate the vectors $\hat{\mathbf{B}}_{ij}(t)$ for all relevant transitions in a Markov chain and then compute the estimates $\hat{A}_{ij}(t, \mathbf{Z}^0)$ for the integrated transition intensities by

$$\hat{A}_{ij}(t, \mathbf{Z}^0) = \int_0^t \mathbf{Z}^0_{ij}(s)'\, d\hat{\mathbf{B}}_{ij}(s) \tag{7}$$

Here the integral is again a finite sum.

The question discussed in this paper is the following one: If the matrix $\hat{\mathbf{A}}_0(t)$ in equation (3) is substituted by a matrix $\hat{\mathbf{A}}(t)$ consisting of the elements $\hat{A}_{ij}(t, \mathbf{Z}^0)$, what will be the properties of the corresponding estimator of the transition matrix? Notice that $\hat{\mathbf{A}}(t)$ depends on the covariate vector $\mathbf{Z}^0$, but this is suppressed in the notation in order not to overburden the formulas below. Because of the martingale property of $\hat{\mathbf{A}}(t)$ there is reason to assume that this should perform well and have nice theoretical properties. Hence, we would get an estimator of the transition matrix which is dependent on covariate information.

The proposed estimator is given as:

$$\hat{\mathbf{P}}(s,t) = \prod_{(s,t]} (\mathbf{I} + d\hat{\mathbf{A}}(u)) \tag{8}$$

Just as in equation (3) this product integral is a finite product of matrices, hence again it generalizes the Kaplan-Meier estimator.

4 Properties of the estimator

Let $\mathbf{A}^*(t)$ be the cumulative intensity matrix based on the following modified intensities:

$$\alpha^*_{ij}(t, \mathbf{Z}^0) = J_{ij}(t)\alpha_{ij}(t, \mathbf{Z}^0)$$

where $J_{ij}(t)$ is an indicator showing whether there is a sufficient set of individuals at risk in state i at time t to ensure that $\mathbf{Y}_{ij}(t)$ has full rank. Since the local estimation principle applied here can only be used at times when all $J_{ij}(t) = 1$, all we can sensibly estimate is $\mathbf{A}^*(t)$, and the exact martingale properties are also achieved by using this quantity. In fact $\hat{\mathbf{A}} - \mathbf{A}^*$ is an exact matrix valued martingale (see e.g. Andersen *et al.*, 1993, Section VII.4).

The martingale property carries over to the estimator $\hat{\mathbf{P}}(s,t)$ through Duhamel's equation (see Andersen *et al.*, 1993, p. 91):

$$\hat{\mathbf{P}}(s,t)\mathbf{P}^*(s,t)^{-1} - \mathbf{I} = \int_s^t \hat{\mathbf{P}}(s,u-)d(\hat{\mathbf{A}} - \mathbf{A}^*)(u)\mathbf{P}^*(s,u)^{-1}$$

where $\mathbf{P}^*$ is defined as in (2) with $\mathbf{A}$ substituted by $\mathbf{A}^*$. Since stochastic integration of a predictable process with respect to a martingale preserves the martingale property, the left hand side in the above equation is seen to be a martingale. This equation immediately yields important results. Firstly, under standard assumptions, the martingale central limit theorem guarantees the asymptotic normality of the estimator $\hat{\mathbf{P}}(s,t)$. Secondly, the estimated covariance matrix may be derived. In complete analogy with results of Andersen *et. al.* (1993, formula (4.4.14)) one can write up the following estimated covariance matrix of $\hat{\mathbf{P}}(s,t)$:

$$\int_s^t \hat{\mathbf{P}}(u,t)' \otimes \hat{\mathbf{P}}(s,u) d[\hat{\mathbf{A}} - \mathbf{A}^*](u) \hat{\mathbf{P}}(u,t) \otimes \hat{\mathbf{P}}(s,u)' \tag{9}$$

where $\otimes$ denotes the Kronecker product. In the above formula $[\hat{\mathbf{A}} - \mathbf{A}^*](t)$ denotes the quadratic variation process of the matrix valued martingale $\hat{\mathbf{A}} - \mathbf{A}^*$. By standard results for the additive regression model one has

$$[\hat{A}_{ij} - A^*_{ij}](t) = \int_0^t \mathbf{Z}^0_{ij}(u)' \mathbf{X}_{ij}(u) \mathrm{diag}\{d\mathbf{N}_{ij}(u)\} \, \mathbf{X}_{ij}(u)' \mathbf{Z}^0_{ij}(u) \tag{10}$$

where "diag" denotes the diagonal matrix with the relevant vector on the diagonal. From these elements the matrix $[\hat{\mathbf{A}} - \mathbf{A}^*](t)$ may be found as described in Andersen *et al.* (1993, pp. 290-293). In fact, the element-wise formula (4.4.16) in this reference is directly applicable if the factor in the third line of the formula is replaced by $d[\hat{A}_{gl} - A^*_{gl}](u)$; cf. (10).

5 Modifications for nested case-control data

The estimators described in the previous sections require information on the values of the covariates for all individuals under study. For large cohorts, this may be very expensive, or even logistically impossible. The nested case-control design (Thomas, 1977) offers an alternative. For this study design, if an $i \to j$ transition occurs at time t, one selects a simple random sample of size m_i from the $Y_i(t)$ individuals at risk in state i. Covariate information is collected for the individual who makes the transition (the case) and for the sampled controls, but are not needed for the remaining individuals.

Only minor modifications are needed to fit the additive model (4) for nested case-control data (Borgan and Langholz, 1997; Zhang and Borgan, 1999). For a pair of states (i,j), the estimator for the vector of cumulative regression functions has the same form as (6). But now only the case and its controls contribute to $\mathbf{N}_{ij}(t)$ and $\mathbf{Y}_{ij}(t)$, and the rows of the latter are multiplied by $Y_i(t)/m_i$ to compensate for this. The same modifications apply to (10), while the remaining estimators in the previous sections remain unchanged.

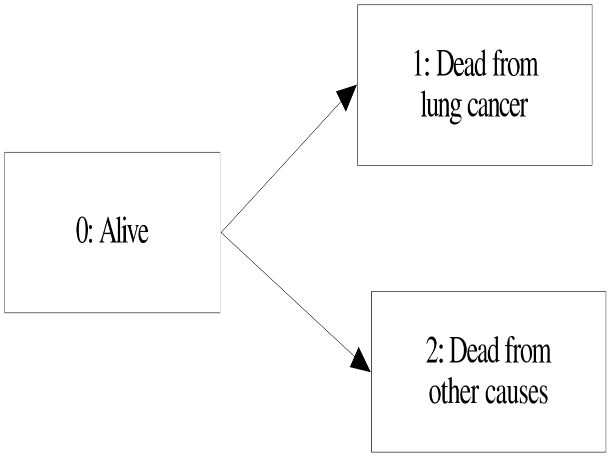

Figure 1: A model for competing risks

It is worth noting that even when covariate information is available for all individuals in a large cohort, it may be advantageous to use a nested case-control sample in order to reduce the burden of error checking and the numerical difficulties in dealing with high dimensional matrices.

6 Examples

6.1 Competing risks – lung cancer death for uranium miners

The Colorado Plateau uranium miners cohort data were collected to study the effects of radon and smoking on mortality. The data have been described in detail in earlier publications (e.g. Ludin, Wagoner, and Archer, 1971; Hornung and Meinhardt, 1987). We will focus on lung cancer mortality taking into account death from other causes.

The cohort consists of 3,347 Caucasian male miners recruited between 1950 and 1960. The cohort was traced for mortality outcomes through December 31, 1982, by which time 258 lung cancer deaths and 2087 deaths from other cause were observed. Exposure data include radon exposure, in working level months (WLM), and smoking histories, in number of packs of cigarettes (20 cigarettes per pack) smoked per day. In our analyses, we summarize these exposures into cumulative radon and cumulative packs of cigarettes lagged by two years, which we denote by $\mathbf{Z}(t) = (R(t), S(t))'$.

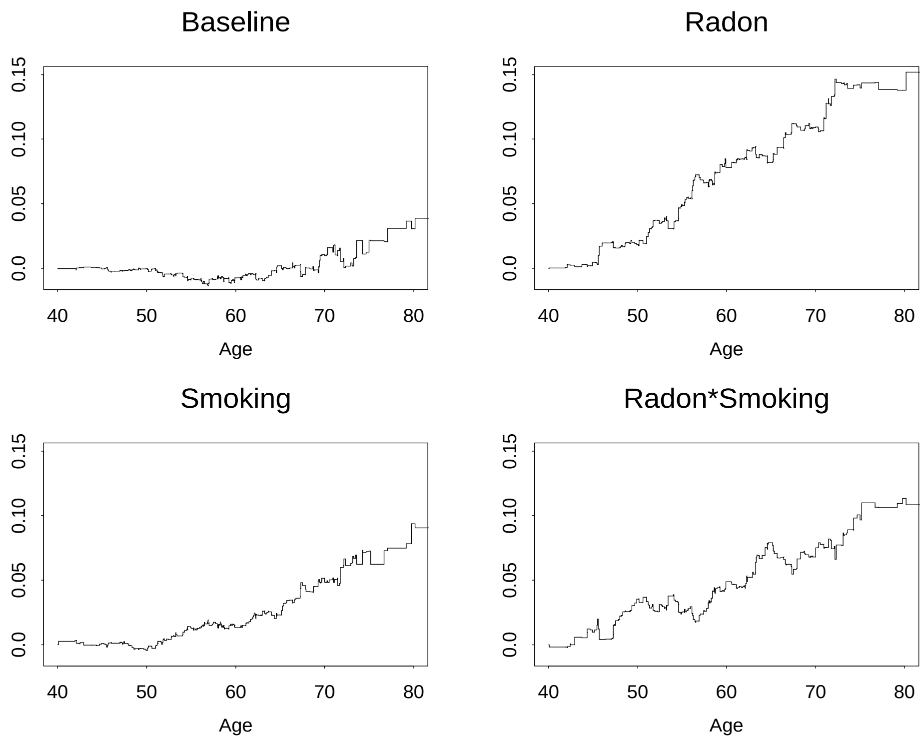

Figure 2: Estimated integrated regression functions for the uranum miners data

We adopt the competing risks model with the three states 0: alive, 1: dead from lung cancer, and 2: dead from other causes as illustrated in Figure 1. The lung cancer mortality at age t is assumed to take the form

$$\alpha_{01}(t;\mathbf{Z}) = \beta_0(t) + \beta_R(t)R(t) + \beta_S(t)S(t) + \beta_{RS}(t)R(t)S(t), \tag{11}$$

while the mortality $\alpha_{02}(t)$ from other causes, as an approximation, is assumed not to depend on the radon and smoking exposures. Note that our model for lung cancer mortality generalizes the semi-parametric relative excess risk model $\alpha_{01}(t;\mathbf{Z}) = \alpha_0(t)(1 + \beta_R R(t))(1+\beta_S S(t))$ which has been used a number of times in previous analyses of these data; see Langholz and Borgan (1997) and their references.

In order to simplify the calculations, a nested case-control sample with 40 controls per case was used to fit (11) as described in Section 5. Figure 2 shows the estimated integrated regression functions (starting at age 40 years) per 100 WLM radon exposure and per 1000 packs of cigarettes. The baseline is close to zero for ages below 70 years. The integrated regression functions for radon, smoking and the interaction are all fairly

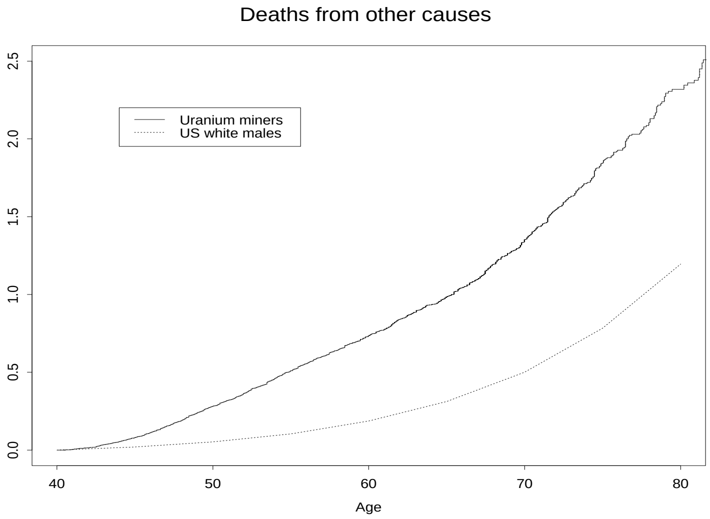

Figure 3: Estimated cumulative mortality from other causes for the uranum miners and the corresponding cumulative mortality for US white males

linear after age 50 years, corresponding to fairly constant excess risks: $\beta_R(t) \approx 0.0050$ per 100 WLM, $\beta_S(t) \approx 0.0025$ per 1000 packs and $\beta_{RS}(t) \approx 0.0030$. To estimate the cumulative mortality from other causes, a Nelson-Aalen estimator was used. Figure 3 shows the estimated cumulative mortality from other causes for the uranium miners together with the corresponding cumulative mortality for US white males (Breslow and Day, 1987, Appendix III). It is seen from the slopes of these curves that, for ages below 70 years, the mortality from other causes among the uranium miners is much higher than in the general population.

We then compute the integrated hazards and absolute risks of lung cancer death in the presence of death from other causes for four different radon and smoking histories. For all situations, we assume a constant exposure intensity for radon between ages 20 to 50. Thus, the two year lagged cumulative radon exposure $R(t)$ is zero for $t < 18$, then increases linearly up to the total dose of radon at $t = 48$, and is constant at the total dose thereafter. Smoking is described by the number of packs per day, and we assume that smoking begin at age 20 and continues throughout life at the same level. The following four exposure histories are considered:

- Total radon dose 960 WLM, smoking 1 pack per day.

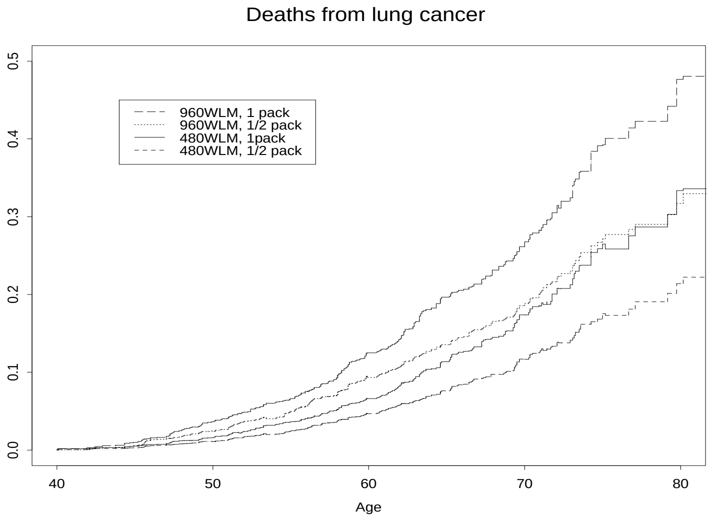

Figure 4: Estimated integrated hazard for lung cancer death for four exposure histories

- Total radon dose 960 WLM, smoking 1/2 pack per day.
- Total radon dose 480 WLM, smoking 1 pack per day.
- Total radon dose 480 WLM, smoking 1/2 pack per day.

Figure 4 shows the integrated hazards (7) for these exposure histories (with lower integration limit set to 40). It is seen that for ages below 60 years, a doubling of the radon exposure has a larger effect on the lung cancer death rates than a doubling of the amount of smoking. After about 60 years of age, the situation is reversed. This is a consequence of our assumption that radon exposure stops at age 50 years, while smoking continues throughout life.

At least for ages below 70 years, the integrated hazards in Figure 4 can roughly be interpreted as absolute risks of lung cancer deaths in the absence of deaths from other causes (e.g. Breslow and Day, 1987, Section 2.2.b). Absolute risks of lung cancer deaths in the presence of deaths from other causes are shown in Figure 5 with 95% point-wise confidence intervals. These are computed using (8) and (9) with $s = 40$. The effect of taking deaths from other causes into account substantially changes the death risks, especially for the higher ages where the risk of death from other causes becomes high.

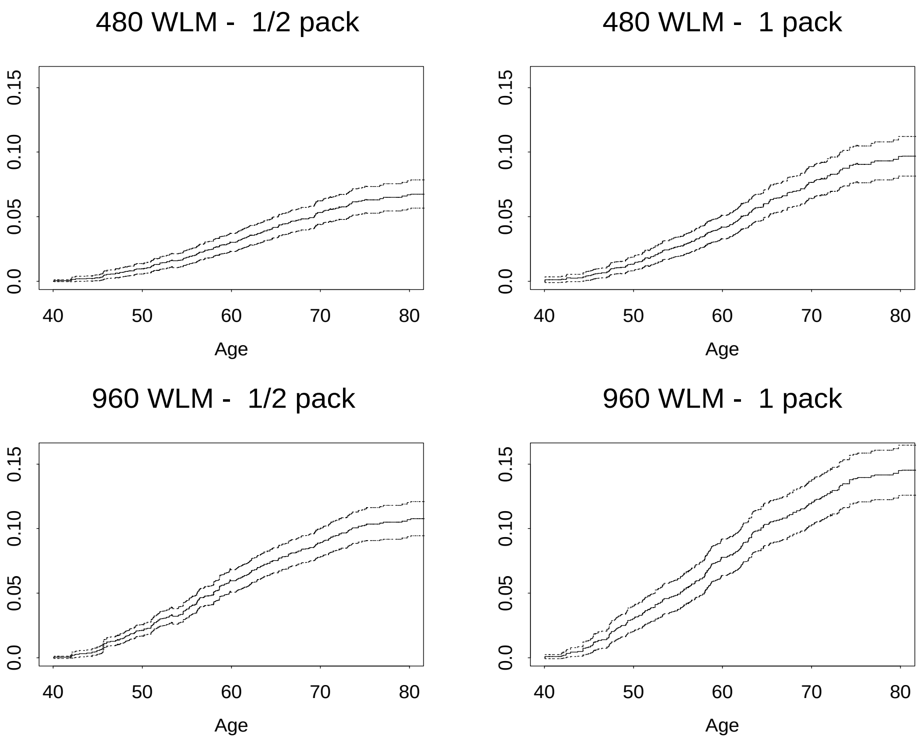

Figure 5: Estimated absolute risk of lung cancer death for four exposure histories

6.2 Probability of being in response – bone marrow transplantation for patients with acute leukemia

Klein and Moeschberger (1997, Section 1.3) describe a study where 137 patients with acute leukemia got a bone marrow transplantation during the period March 1, 1984 to June 30, 1989. The patients were followed for a maximum of seven years, and times to relapse and death were recorded. It was also recorded if and when the platelet count of a patient returned to a self-sustaining level. Such platelet recovery is a beneficial event: the rate at which the patients are relapsing or dying is smaller after their platelets recover (Klein and Moeschberger, 1997, p. 272).

The patients were grouped into three risk groups based on their status at the time of transplantation: acute lymphoblastic leukemia (ALL), low-risk acute myeloctic leukemia (AML low-risk), and high-risk acute myeloctic leukemia (AML high-risk). In our analysis the risk groups are represented by the indicators Z_1 for AML low-risk and Z_2 for AML high-risk patients. Other covariates we use are the indicator Z_3 of FAB (French-American-

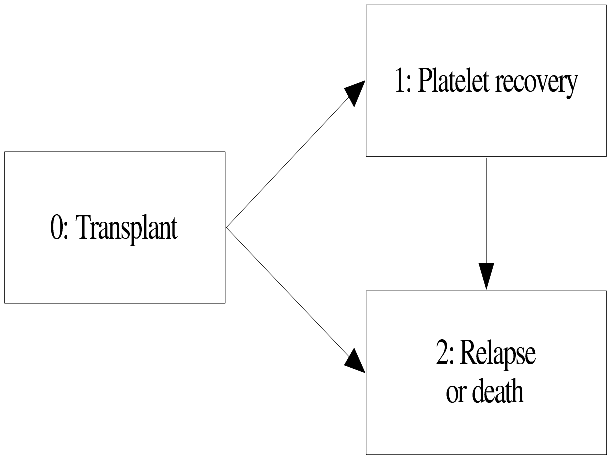

Figure 6: A Markov model for patients who have got a bone marrow transplatation

British) classification M4 or M5 for AML patients, the indicator Z_4 of whether a patient was given a graft-versus-host prophylactic treatment combining methotrexate (MTX) with cyclosporine and possibly methylprednisolone, and three covariates for the patient's and donor's age: Z_5 = patient's age $-$ 28, Z_6 = donor age $-$ 28, and $Z_7 = Z_5 \times Z_6$.

We will consider platelet recovery as a response to the bone marrow transplantation, and we will study how the "probability of being in response function" (Temkin, 1978) depends on covariates. To this end we adopt the Markov model with the three states 0: transplant, 1: platelet recovery, and 2: relapse or death as illustrated in Figure 6. A patient starts out in state 0 at time $t = 0$ when he gets the bone marrow transplant. If his platelets recover, the patient moves to state 1, and if he then relapses or dies, he moves on to state 2. If the patient relapses or dies without the platelets returning to a normal level, he moves directly from state 0 to state 2. We write $P_{01}(s, t)$ for the probability that a patient who is in state 0 at time s will be in state 1 at a later time t. This transition probability is the probability of being in response at time t for a patient who has not yet responded to the treatment at time s.

In order to estimate the probability of being in response for given values of the covariates, we first estimate the intensities for the three possible transitions $0 \to 1$, $1 \to 2$ and $0 \to 2$. (In this estimation, a number tied $0 \to 1$ transitions as well as a few other tied events were broken at random.) The first two of these transition intensities are assumed

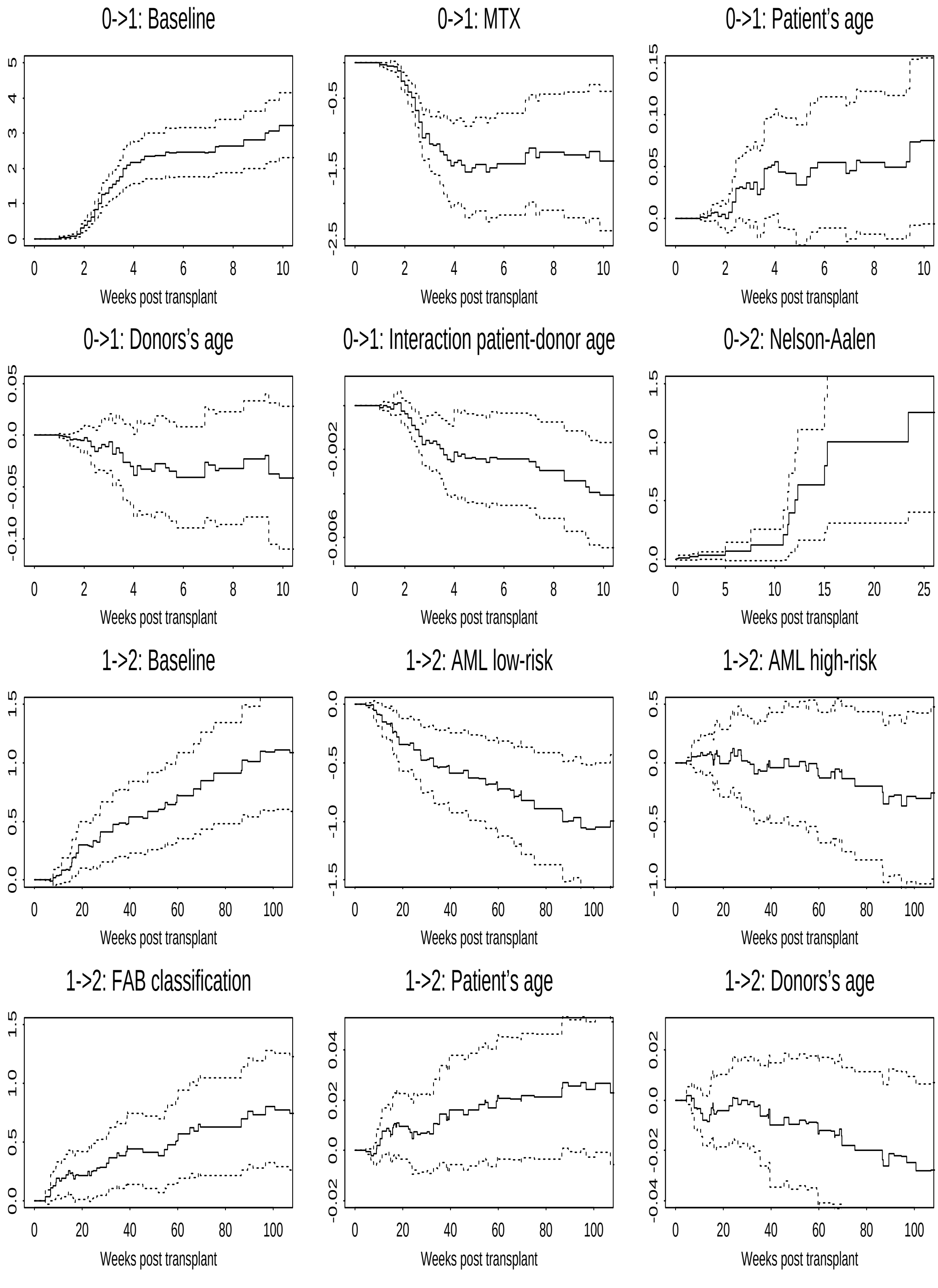

Figure 7: Integrated regression functions for the three possible transitions for the bone marrow transplantation data

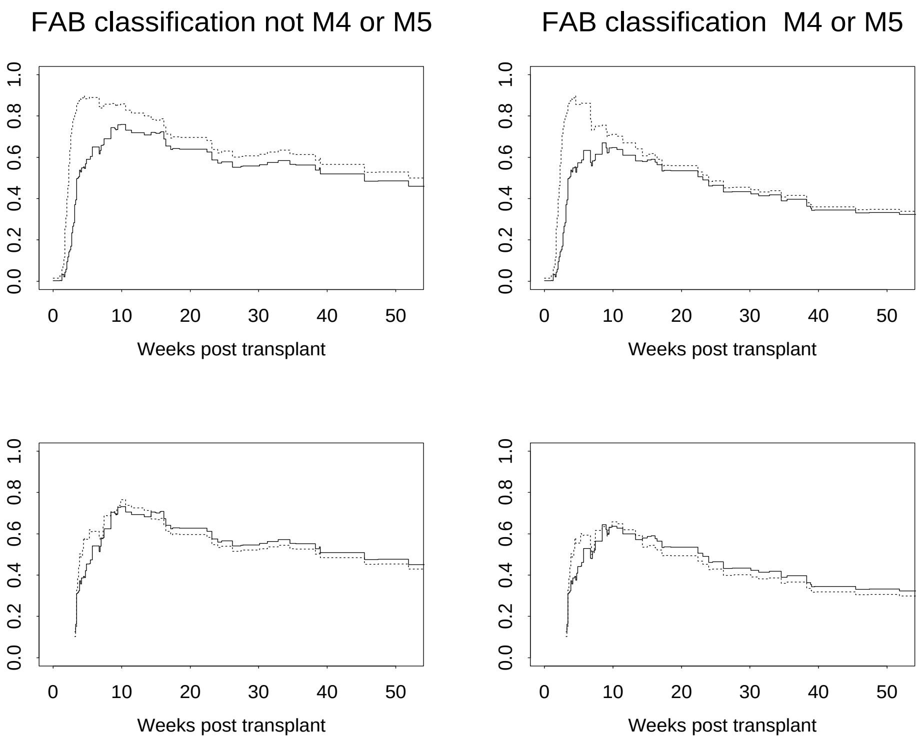

Figure 8: Estimated probability of being in response ($P_{01}(s,t)$) for AML high risk patients without use of MTX (drawn lines) and with use of MTX (dotted lines). Left-hand panel: FAB classification not M4 or M5; right-hand panel: FAB classification M4 or M5. Upper panel: $P_{01}(0,t)$; lower panel: $P_{01}(3,t)$

to be of the additive form (4) depending on the covariates:

- $0 \to 1$: MTX (Z_4), patient's age (Z_5), donor's age (Z_6), and the interaction patient-donor age (Z_7)
- $1 \to 2$: AML low-risk (Z_1), AML high-risk (Z_2), FAB classification (Z_3), patient's age (Z_5), and donor's age (Z_6)

These covariates are the same as the ones used by Klein and Moeschberger (1997, Section 9.5) in their analysis using proportional hazards models, with one exception. We do not include the interaction term Z_7 in our model for the $1 \to 2$ transition, since this interaction is not important on the additive scale. The estimated integrated regression functions for the $0 \to 1$ and $1 \to 2$ transitions are shown in Figure 7 with point-wise 95% confidence intervals.

For the $0 \to 1$ transition, one should in particular note the estimates of the integrated regression functions for MTX and the patient-donor age interaction. As seen from the slopes of these curves, both reduce the $0 \to 1$ transition intensity the first 3–4 weeks after transplantation. Use of MTX seems to have no effect later on, while the size of the age interaction is clearly reduced after about four weeks. For the $1 \to 2$ transition, all the estimated regression functions are fairly linear, corresponding to constant excess intensities of the covariates. Note in particular that AML low-risk patients have a reduced intensity for relapse or death, while this intensity is increased for patients with FAB classification M4 or M5.

As there are less than twenty patients who relapse or die without their platelets returning to a normal level, it is difficult in a meaningful way to estimate the effect of the covariates on the $0 \to 2$ transition using the additive model. We therefore adopt a model without covariates for this transition intensity, and estimate its integral using the Nelson-Aalen estimator. The estimate is shown in Figure 7. Note that the transition intensity is low and fairly constant the first 10 weeks after transplantation. Thereafter it increases to a higher level, but, as there are only 13 patients left in state 0 ten weeks after transplantation, the estimate now becomes quite uncertain.

To illustrate how (7) and (8) may be used to estimate the probability of being in response for a patient with specific covariates, we present in Figure 8 estimates for an AML high-risk patient assuming both patient and donor to be 28 years old ($Z_5 = Z_6 = Z_7 = 0$). The upper panel shows estimates of $P_{01}(0, t)$ for the first year after transplantation for the four combinations of MTX and FAB classification. In all cases, the estimated probability of being in response function increases steeply the first few weeks after transplantation as the platelets recover for most patients. Then the curves decrease due to relapse and death. Note that use of MTX reduces the probability of being in response, at least for the first 15-20 weeks, and that this probability is lower for patients with FAB classification M4 or M5 than for the other AML high-risk patients. The lower panel of Figure 8 shows the estimates of $P_{01}(3, t)$. For all four combinations of the covariates, the probability of being in response is reduced for a patient whose platelets has not yet recovered three weeks after transplantation. However, use of MTX is no longer of any importance. This clearly illustrate the value of the present approach which easily takes into account time-dependent

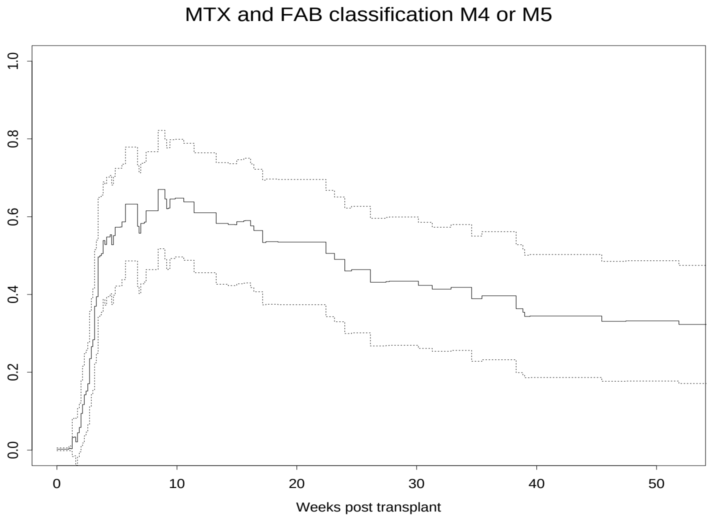

Figure 9: Estimated probability of being in response ($P_{01}(0,t)$) with pointwise 95% confidence intervals for an AML high risk patients with FAB classification M4 or M5 using MTX

effects of covariates.

Confidence intervals are not shown on Figure 8 in order not to overburden the figure. To give an indication of the estimation uncertainty, Figure 9 presents the estimate of $P_{01}(0,t)$ with pointwise 95% confidence intervals for a patient using MTX with FAB classification M4 or M5.

7 Discussion

We have shown that the empirical transition matrix and the additive model fit very nicely together. In the examples we have demonstrated that this also gives a useful statistical methodology. The computation of transition probabilities is often a useful way of summing up important clinical information, and it is important to be able to do this dependent on covariate values.

The present approach is an alternative to one based on the Cox model. This is needed because the proportional hazards idea, which so pervades survival analysis, does not allow any simple local estimation. It seems that the only way of achieving a local model, and at the same time preserving the martingale structure, is to assume that the intensities depend on the covariates in a linear, or additive, fashion. This has been criticized as being not in tune with the requirement that the intensities must be non-negative. However, there seems to be an unavoidable contradiction between guaranteed non-negativity on the one hand, and locality, martingale structure and simplicity on the other hand. It is clearly of interest to have a model which is simple, which picks up changes in covariate effects as time goes along and which preserves the martingale structure. Of course, the validity of the procedure will depend on whether an additive model can be made to reasonably approximate the data. In the present paper we had some negative contributions in the estimates, but they were mostly quite small, and did not seem to contradict the assumption that the model fitted reasonably well.

As mentioned in the introduction, the Markov assumption made throughout the paper may not be essential. This has been pointed out by Datta and Satten (2000) and Glidden (2001), and we shall discuss it briefly here. In order to justify writing up a product of matrices like the one in equation (1), one usually assumes that the process is Markovian. However, one would expect an expression of the same kind to be valid also for non-Markovian processes. We will consider here discrete time, but the same arguments hold for continuous time as well. Assume any kind of stochastic process on a finite state space. For each time $t_1 < \ t_2 < \ldots < t_n$ the process will have a certain probability distribution on the state space; denote these distributions by the row vectors $\mathbf{p}_1, \ \mathbf{p}_2, \ldots, \ \mathbf{p}_n$. From one time to the next the probability mass in a given state will be redistributed on the state space, and a transition probability matrix defines the redistribution that takes place between two times. Hence, as a *purely algebraic* result, without referring to any particular structure of the process, one may write the transition of probability mass from time 1 to time n in the following way for suitable transition probability matrices $\mathbf{P}(t_i, t_{i+1})$:

$$\mathbf{p}_n = \mathbf{p}_1' \times \mathbf{P}(t_1, t_2) \times \mathbf{P}(t_2, t_3) \times \ldots \times \mathbf{P}(t_{n-1}, t_n) \tag{12}$$

Consider now the observation of a number of individuals moving around on the state space. If individuals are only followed from one step to the next and then one forgets their history, then each element of the product above may be estimated from the observations and in particular the empirical transition matrix is still valid when estimating occupation probabilities.

There is one difficulty, however, and that is censoring. If censoring is entirely random, and in particular unrelated to the state the individual might be in, then everything is all right, and the Markov assumption is unnecessary. However, if censoring is related to how the process develops, for instance to state, then biases may arise when the process is not Markovian. If the process is semi-Markov, and censoring is related to state, then, since individuals who have stayed in a particular state for a long time may have greater likelihood to be censored, a biased picture of transition probabilities may result.

It is also not clear how the introduction of covariates may modify the arguments considered here. Since, however, one will often wonder how realistic the Markov assumption is, it is important to realize that this assumption is less essential than one may think.

References

Aalen, O. O. (1978). Non-parametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–726.

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes. *Lecture Notes in Statistics.* **2**, 1–25. Springer Verlag, New York.

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statist. Med.* **8**, pp. 907–925.

Aalen, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statist. Med.* **12**, 1569–1588.

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* **5**, 141–150.

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

Andersen, P. K., Hansen, L. S. and Keiding, N. (1991). Non- and semi-parametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process. *Scand. J. Statist.* **18**, 153–167.

Borgan Ø. (1998). Aalen-Johansen estimator. In: P. Armitage and T. Colton (eds.) *Encyclopedia of Biostatistics* **1**, 5–10. Wiley, New York.

Borgan, Ø and Langholz, B. (1997). Estimation of excess risk from case-control data using Aalen's linear regression model. *Biometrics* **53**, 690–697.

Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research. Volume II – The Design and Analysis of Cohort Studies*, *IARC Scientific Publications*, Vol. 82. International Agency for Research on Cancer, Lyon.

Datta, S. and Satten, G. A. (2000). Nonparametric estimation of stage occupation probabilities and transition hazards under dependent censoring. Technical Report 2000-4, Department of Statistics, University of Georgia.

Fleming, T. R. (1978a). Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks. *Ann. Statist.* **6**, 1057–1070.

Fleming, T. R. (1978b). Asymptotic distribution results in competing risks estimation. *Ann. Statist.* **6**, 1071–1079.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis.* Wiley, New York.

Glidden, D. V. (2001) Nonparametric estimation of event probabilities for non-Markov multistate event data. Technical Report # 60. Division of Biostatistics, University of California at San Fransisco.

Hornung, R. and Meinhardt, T. (1987). Quantitative risk assessment of lung cancer in U. S. uranium miners. *Health Physics*, **52**, 417–30.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival analysis. Techniques for censored and truncated data.* Springer-Verlag, New York

Langholz, B. and Borgan, Ø. (1997). Estimation of absolute risk from nested case-control data, *Biometrics* **53**, 768–775.

Lundin, F., Wagoner, J., and Archer, V. (1971). Radon daughter exposure and respiratory cancer, quantitative and temporal aspects. Joint Monograph 1, U.S. Public Health Service, Washington, D.C.

Temkin, N. R. (1978). An analysis for transient states with application to tumor shrinkage. *Biometrics* **34**, 571-580.

Thomas, D. C. (1977). Addendum to: Methods of cohort analysis: Appraisal by application to asbestos mining. By F. D. K. Liddell, J. C. McDonald and D. C. Thomas. *J. Roy. Statist. Soc. A* **140**, 469–491.

Zhang, J. and Borgan, Ø. (1999). Aalen's linear model for sampled risk set data: a large sample study. *Lifetime Data Anal.* **5**, 351–369.