The Korean Communications
in Statistics Vol. 5, No. 1, 1998
pp. 239-263

Semiparametric Inference for a Multistate Stochastic Survival Model[1)]

Sung Chil Yeo[2)]

Abstract

In this paper, we consider a multistate survival model which incorporates covariates and contains two illness states and two death states. The underlying stochastic process is assumed to follow nonhomogeneous Markov process. The estimates of survival, transition and competing risks probabilities are given via the methods of partial likelihood and nonparametric maximum likelihood. Our discussion is based on the statistical theory of counting process. An illustration is given to the data of patients in a heart transplant program. The goodness of fit procedures are also discussed to check the adequacy of the model.

1. Introduction

The usual survival model can be viewed as a simple two-state stochastic process where the force of transition from a transient 'alive' state to an absorbing 'death' state is the hazard rate function for the survival time distribution.

In clinical trials, however, the one transient 'alive' state and the one absorbing 'death' state can often be split into two or more transient states and two or more absorbing states which typically corresponding to occurrence of various nonfatal complications and to different causes of death, respectively. The concept of competing risks, with each risk a potential cause of death, can be involved in this situation. The statistical analysis of such resulting data is often called multi-state survival analysis in the literature. Such multi-state models enable us to study and analyse the life history data in more detail which occur frequently in practice. For example, in medical follow-up studies, we may begin with healthy persons who are followed to determine which of them develop a specific disease, such as coronary heart disease, how it affects their survival, and which covariates(e.g., age, blood pressure, smoking, physical

1) This research was supported by Kon-Kuk University Foundation, 1996.
2) Professor, Department of Applied Statistics, Kon-Kuk University, Seoul, 143-701, Korea.

condition, and etc.) are important for predicting who gets the disease and who may die from it.

In this paper, we consider a multistate survival model which contains two 'alive' states and two 'death' states. Sometimes 'death' may not be literal but may mean progression to some other defined state.

During the last few decades, there has been an extensive work on the problems of a nature similar to the present work. For example, Fix and Neyman(1951) first studied the problem from a statistical point of view, and they introduced the concept of the net and crude probabilities. Such probabilities and their relation compose the concept of competing risks and are described more completely in Chiang(1961a, 1961b, 1968). Other work, more closely related to the present work is due to Sacks and Chiang(1977) and Beck(1979). Sacks and Chiang studied the stochastic model and applied to coronary heart disease, but in the absence of covariates. Beck investigated the model with the addition of covariates and applied to the analysis of survival data from patients accepted into a heart transplant program. But they all assume that the underlying transition rates are constant. Other related articles include Kodell and Nelson(1980) who consider the model with Weibull intensity functions, Mcknight and Crowly(1984), Dinse and Lagakos(1982), Kodell et al.(1982), Turnbull and Mitchell(1984), and Dewanji and Kalbfleisch(1986) who all consider the model without covariates and analyse the carcinogenesis data from survival/sacrifice experiments by the classical likelihood methods.

In this paper, we develope a multistate stochastic survival model which incorporates covariates. We regard that the baseline intensity function is time dependent and left completely unspecified. Thus, the present work extends the previous studies mentioned above.

In a theoretical point of view, a description of the multistate survival data as a stochastic process has proven useful for the purpose of studying large sample properties of many parametric, non-parametric and semi-parametric statistical procedures frequently used for censored data. For example, Aalen and Johansen(1978) developed the method for non-parametric estimation of transition probabilities in a non-homogeneous finite state Markov processes, and used product-integration combined with results on counting process, martingales, and stochastic integrals to study their small and large sample properties. For the case of a simple two state stochastic model, Aalen(1980) suggested the additive regression model formulated for counting processes which complements the proportional hazard model proposed by Cox(1972). Andersen and Gill(1982) elaborated the counting process approach to the Cox regression model. Later, Andersen and Borgan(1985) and Andersen et al.(1991) extended the work of Andersen and Gill to Cox type regression models for the transition intensities of the nonhomogeneous Markov process.

The plan of this paper is as follows: In Section2, we describe the illness-death model based on the work of Sacks and Chiang(1977) and Beck(1979). we also give the expressions of the transition probabilities via the Kolmogorov forward equations. In Section 3, we estimate the vectors of regression parameters and the baseline integrated intensity function based on the methods of multistate Cox's partial likelihood and multistate nonparametric likelihood,

respectively. We then estimate the transition, survival and competing risks probabilities. In Section 4, we explore the asymptotic properties of the estimators given in Section 3 based on the work of Aalen and Johansen(1978) and Andersen et al.(1991). In Section 5, as an application of the methods, we analyse the data of patients from a heart transplant program. Finally, Section 6 contains discussion and some concluding remarks.

2. The model

In this paper, we consider a model which describes a population of healthy individuals who may develop a disease and later then die. In this model there are two transient states 1 and 2 and two absorbing states 3 and 4, where 1 is a healthy state, 2 is a specific disease state, 3 is the state of death from all causes other than the disease, and 4 represents death due to the disease under study. An individual is said to be in state 1 if he or she is free from the disease, or in state 2 if he or she is affected with the disease. A person enters absorbing state 3 if he or she dies from other causes, and enters the absorbing state 4 if he or she dies from the disease under study. We shall assume that the disease is irreversible, so that a person entering state 2 can not return to state 1. Also, since a person can not die from the disease without first having developed the disease under study, the transition from state 1 to state 4 is not allowed. The transitions an individual makes from one state to another are governed by the intensity functions. In order to incorporate covariates, we use the proportional intensities regression model. Moreover, in order to still have a finite state Markov process for given values of the covariates entering into the regression models for the transition intensities, we restrict attention to the case where these covariates are time-independent.

The model, with the intensity function for each transition, is exhibited in Figure 1.

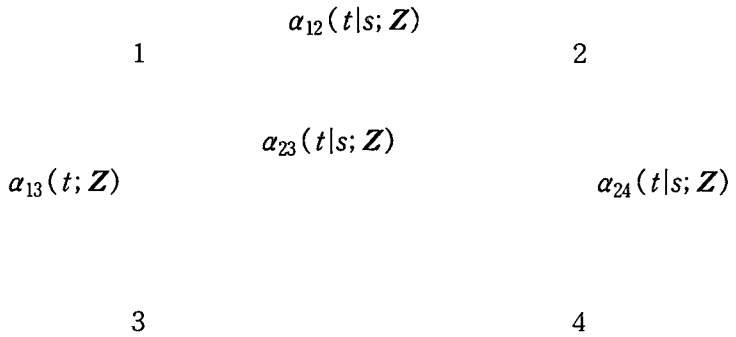


Figure 1. An illness-death model with two transient and two absorbing states

Now we suppose that n individuals initially stay in state 1 and the individuals behave independently of each other between states according to a non-homogeneous Markov process.

Let $X(t)$ be the state occupied at time $t \in \mathcal{T} = [0, \tau]$ for some $\tau > 0$, by a randomly chosen individual with a given covariates vector $\boldsymbol{Z}$. Then the process $\{X(t), t \in T\}$ is governed by the two conditions below:

(*i*) $X(0) = 1$, that is any individual is in state 1 at the time of diagnosis, and

(*ii*) the 4×4 matrix of transition probabilities is given by $\boldsymbol{P}(s, t; \boldsymbol{Z})$, $0 \leq s \leq t$, with entries

$$P_{jk}(s, t; \boldsymbol{Z}) = P(X(t) = k \mid X(s) = j; \boldsymbol{Z}), \tag{2.1}$$

where $(j, k) \in E = \{(j, k) \in S^2 \mid j \leq k\}$, $S = \{1, 2, 3, 4\}$, $P_{jk}(s, t; \boldsymbol{Z}) = 0$ otherwise. That is, given that a randomly chosen person with covariates $\boldsymbol{Z}$ is in state j at time s (present), the conditional probability that this person will be in state k at time t (future) is independent of $X(u)$, $0 \leq u \leq s$ (past), (Markovian property). In (2.1) we note that

$$P_{33}(s, t; \boldsymbol{Z}) = P_{44}(s, t; \boldsymbol{Z}) = 1 \text{ and } \sum_{k=1}^{4} P_{1k}(s, t; \boldsymbol{Z}) = \sum_{k=2}^{4} P_{2k}(s, t; \boldsymbol{Z}) = 1.$$

Alternatively, this process can be specified in terms of the intensity functions of a transition from state j to state k at time t:

$$\alpha_{jk}(t; \boldsymbol{Z}) = \lim_{h \downarrow 0} \frac{1}{h} P_{jk}(t, t+h; \boldsymbol{Z}), \tag{2.2}$$

where $(j, k) \in \Gamma = \{(1, 2), (1, 3), (2, 3), (2, 4)\}$; $\alpha_{jk}(t; \boldsymbol{Z}) = 0$ otherwise.

For the intensity function (2.2) with the basic covariates $\boldsymbol{Z}$, following Cox(1972), we use the log-linear form for the covariates adjustment for its mathematical convenience as well as to avoid a negative intensity function. That is, we let

$$\alpha_{jk}(t; \boldsymbol{Z}) = \alpha_{jk0}(t) \exp(\boldsymbol{\beta}_{jk}^{T} \boldsymbol{Z}), \tag{2.3}$$

where $\alpha_{jk0}(t)$ is the so-called baseline intensity function of $j \rightarrow k$ transition and $\boldsymbol{\beta}_{jk}$ is a vector of unknown regression parameters of $j \rightarrow k$ transition corresponding to the basic covariates vector $\boldsymbol{Z}$.

Alternatively, it is convenient to write $\alpha_{jk}(t; \boldsymbol{Z})$ not as the form of (2.13) but as

$$\alpha_{jk}(t; \boldsymbol{Z}) = \alpha_{jk0}(t) \exp(\boldsymbol{\beta}^{T} \boldsymbol{Z}_{jk}), \tag{2.4}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^{T}$ is one vector of regression parameters containing all the different parameters in the $\boldsymbol{\beta}_{jk}$ vectors and hence, not depending on the type of transitions, and $\boldsymbol{Z}_{jk}$

is a vector of type specific covariates given by the basic covariates vector Z. The models on the form of (2.3) can always be written as (2.4) if in the entire model (i.e. for all types together) there are p regression parameters to be estimated, then each of the type specific covariate vectors Z_{jk} should be made p-variate possibly by including extra components equal to zero. We have been chosen to do so because the formulations and arguments of the large sample properties are simplified in this case.

Let $A(t;Z)$ be the 4×4 matrix of the integrated intensity functions $A_{jk}(t;Z)=\int_0^t \alpha_{jk}(u;Z)du$, where $(j,k)\in E$. And let the base-line integrated intensity function be $A_{jk0}(t)=\int_0^t \alpha_{jk0}(u)du$. Then the integrated intensity functions becomes

$$A_{jk}(t;Z)=A_{jk0}(t)\exp(\beta^T Z_{jk}),\ (j,k)\in E \tag{2.5}$$

We also define

$$\alpha_{jj}(t;Z)=-\sum_{k\neq j}\alpha_{jk}(t;Z),\ j=1,2\ ;\ (j,k)\in E \tag{2.6}$$

and

$$A_{jj}(t;Z)=-\sum_{k\neq j}A_{jk}(t;Z),\ j=1,2;\ (j,k)\in E \tag{2.7}$$

Then the relationship between the intensity functions and the transition probabilities can be found by the Kolmogorov forward equation:

$$\begin{aligned} &P(s,s;Z)=\mathbf{I},\\ &P(s,dt;Z)=P(s,t^-;Z)A(dt;Z) \end{aligned} \tag{2.8}$$

The solution to the equation (2.8) can be expressed in terms of product integrals as

$$P(s,t;Z)=\mathbf{P}_{(s,t]}\{\mathbf{I}+dA(u;Z)\}, \tag{2.9}$$

where $\mathbf{P}$ denotes the product integral(e.g. Aalen and Johansen(1978), Gill and Johansen(1990)).

Although the expression (2.9) provides general formula for the transition probability matrix, it may also be of some interest to derive explicit expressions for each transition probability $P_{jk}(s,t;Z)$, where $(j,k)\in E$. From the equation (2.8), we get the following equations:

$$dP_{11}(s,t;Z)=-P_{11}(s,t^-;Z)\{dA_{12}(t;Z)+dA_{13}(t;Z)\}, \tag{2.10}$$

$$\begin{aligned} dP_{12}(s,t;Z)=&P_{11}(s,t^-;Z)dA_{12}(t;Z)\\ &-P_{12}(s,t^-;Z)\{dA_{23}(t;Z)+dA_{24}(t;Z)\}, \end{aligned} \tag{2.11}$$

$$dP_{13}(s,t;Z)=P_{11}(s,t^-;Z)dA_{13}(t;Z)+P_{12}(s,t^-;Z)dA_{23}(t;Z), \tag{2.12}$$

$$dP_{14}(s,t;Z) = P_{12}(s,t^-;Z)dA_{24}(t;Z), \tag{2.13}$$

$$dP_{22}(s,t;Z) = -P_{22}(s,t^-;Z)\{dA_{23}(t;Z)+dA_{24}(t;Z)\}, \tag{2.14}$$

$$dP_{23}(s,t;Z) = P_{22}(s,t^-;Z)dA_{23}(t;Z), \tag{2.15}$$

$$dP_{24}(s,t;Z) = P_{22}(s,t^-;Z)dA_{24}(t;Z). \tag{2.16}$$

Using the initial condition $P_{jk}(s,s;Z)=\delta_{jk}$, where δ_{jk} is the Kronecker delta, and solving the equations (2.10) - (2.16), we obtain the following results:

$$P_{11}(s,t;Z) = \mathbf{P}_{(s,t]}\{1-dA_{12}(u;Z)-dA_{13}(u;Z)\}, \tag{2.17}$$

$$P_{12}(s,t;Z) = \int_{(s,t]} P_{11}(s,u^-;Z)dA_{12}(u;Z)P_{22}(u,t;Z), \tag{2.18}$$

$$P_{13}(s,t;Z) = P_{13}^{(1)}(s,t;Z) + P_{13}^{(2)}(s,t;Z), \tag{2.19}$$

where

$$P_{13}^{(1)}(s,t;Z) = \int_{(s,t]} P_{11}(s,u^-;Z)dA_{13}(u;Z)\ , \tag{2.19a}$$

$$P_{13}^{(2)}(s,t;Z) = \int_{(s,t]} P_{12}(s,u^-;Z)dA_{23}(u;Z), \tag{2.19b}$$

$$P_{14}(s,t;Z) = \int_{(s,t]} P_{12}(s,u^-;Z)dA_{24}(u;Z)\ , \tag{2.20}$$

$$P_{22}(s,t;Z) = \mathbf{P}_{(s,t]}\{1-dA_{23}(u;Z)-dA_{24}(u;Z)\}, \tag{2.21}$$

$$P_{23}(s,t;Z) = \int_{(s,t]} P_{22}(s,u^-;Z)dA_{23}(u;Z), \tag{2.22}$$

$$P_{24}(s,t;Z) = \int_{(s,t]} P_{22}(s,u^-;Z)dA_{24}(u;Z). \tag{2.23}$$

From the above equations (2.17) - (2.23), For a person with the covariaes vector Z who starts in state 1 at time 0, the survival probability is

$$S(t;Z) = P_{11}(0,t;Z) + P_{12}(0,t;Z), \tag{2.24}$$

where $P_{11}(0,t;Z)$ is the probability that this person is still in state 1 at time t and $P_{12}(0,t;Z)$ is the probability that this person is now in state 2 at time t; And the death probabilities are $P_{13}^{(1)}(0,t;Z)$, $P_{13}^{(2)}(0,t;Z)$ and $P_{14}(0,t;Z)$, respectively, where $P_{13}^{(1)}(0,t;Z)$ is the probability that this person has been dead at time t from other causes without having developed the specific disease under study, and $P_{13}^{(2)}(0,t;Z)$ is the probability

that this person has been dead at time t from other causes after having developed the disease, and $P_{14}(0,t;\boldsymbol{Z})$ is the probability that this person has bee dead at time t from the specific disease.

On the other hand, in the competing risks theory, there are several probabilities of dying such as crude, net, and partial crude probabilities due to a specific causes of death (e.g. Chiang, 1968, *pp.* 242-247). Since the present model allow the covariates, these probabilities are defined for a particular set of covariates value. The cause-specific death transition probabilities $P_{13}^{(1)}(0,t;\boldsymbol{Z})$, $P_{13}^{(2)}(0,t;\boldsymbol{Z})$ and $P_{14}(0,t;\boldsymbol{Z})$ are identical to what are called the crude probabilities.

In order to obtain the net probabilities that are uniquely identified, the competing risks of death are assumed to act independently (e.g. Tsiatis(1975)). The net and partial crude probabilities may be useful in estimating certain death probabilities which cannot normally be observed but are calculated through their relationships with the crude probabilities. For example, it may be meaningful to ask in the present model how the probabilities of dying is changed if the risk of dying from the disease(state 2) is eliminated. Since the present model includes only two death states, this probability is a net probability as well as a partial crude probability given by

$$Q_{13}^{(2)}(0,t;\boldsymbol{Z}) = \int_0^t Q_{12}(0,u-;\boldsymbol{Z})\,dA_{23}(u;\boldsymbol{Z}), \tag{2.25}$$

where

$$Q_{12}(0,t;\boldsymbol{Z}) = \int_0^t P_{11}(0,u-;\boldsymbol{Z})\,dA_{12}(u;\boldsymbol{Z})\,Q_{22}(u,t;\boldsymbol{Z}), \tag{2.26}$$

$$Q_{22}(s,t;\boldsymbol{Z}) = \mathrm{P}_{(s,t]}\{1 - dA_{23}(u;\boldsymbol{Z})\}. \tag{2.27}$$

3. Estimation

The transition intensities of the Markov process also appears in the compensator of the counting processes recording each type of transitions. Let $\boldsymbol{N} = \{N_{jki}(t); t\in T, (j,k)\in\Gamma, i=1,\cdots,n\}$ be the multivariate counting process for the model considered in this paper. That is, each component $N_{jki}(t)$ represents the number of direct transitions from j to k ($(j,k)\in\Gamma$) observed for individual i with the covariates vector $\boldsymbol{Z}_i$ in the time interval $[0,t]$.

Define $Y_{ji}(t) = I_{(X_i(t-)=j)}$, where $I_{(A)}$ denotes the indicator function of a set A. That

is, $Y_{ji}(t)=1$ if the individual number i is observed to be in state i just before time t; $Y_{ji}(t)=0$ otherwise. Thus $\{Y_{ji}(t); t\in T, j=1,2, i=1,\cdots,n\}$ is a predictable indicator process. Then under the assumption of independent censoring mechanisms (e.g. Andersen et al. (1988), Kalbfleisch and Prentice (1980, *pp* 119-122), we may decompose $N_{jki}(t)$ into its compensator $\Lambda_{jki}(t)$ and a local square integrable martingale $M_{jki}(t)$ with respect to $\mathcal{F}_t^i$, where $\mathcal{F}_t^i = \sigma\{X_i(s); 0\le s\le t, t\in T\}$ is the smallest σ-field making all of random variables $X_i(s)$, $0\le s\le t$, measurable. Or equivalently we may denote

$$\mathcal{F}_t^i = \sigma\{N_{jki}(u), Y_{ji}(u+), Z_i, (j,k)\in\Gamma\} \tag{3.1}$$

since the initial distribution of $X_i(t)$ is degenerate at time $t=0$. Thus, we may write

$$N_{jki}(t) = \Lambda_{jki}(t) + M_{jki}(t) \tag{3.2}$$

(This is the Doob-Meyer decomposition of the local submartingale N_{jki}).

Under the regularity conditions (e.g. Aalen(1978, Section3.2) $\Lambda_{jki}(t)$ is absolutely continuous, so that there exists predictable process $\boldsymbol{\lambda}=\{\lambda_{jki}(t); t\in T, ((j,k)\in\Gamma, i=1,\cdots,n\}$ such that

$$\Lambda_{jki}(t) = \int_0^t \lambda_{jki}(u)\,du. \tag{3.3}$$

Thus, (3.2) may be expressed as

$$dN_{jki}(t) = \lambda_{jki}(t)\,dt + dM_{jki}(t). \tag{3.4}$$

Furthermore, we have

$$\lambda_{jki}(t+) = \lim_{h\downarrow 0}\frac{1}{h} P\{N_{jki}(t+h) - N_{jki}(t)\,|\,\mathcal{F}_t\}, \tag{3.5}$$

where $\mathcal{F}_t = \bigvee_{i=1}^{n}\mathcal{F}_t^i$ which is the smallest σ-field containing $\mathcal{F}_t^i$.

The expression (3.5) may also informally be written as

$$\lambda_{jki}(t)\,dt = P\{dN_{jki}(t)=1\,|\,\mathcal{F}_{t-}\}, \tag{3.6}$$

where $dN_{jki}(t)$ denotes the increments of $N_{jki}(t)$ over a small time interval of length dt around time t. Thus, provided that the underlying stochastic process is Markovian, we have

$$\lambda_{jki}(t) = Y_{ji}(t)\alpha_{jki}(t; Z_i). \tag{3.7}$$

For the proportional intensities regression model as given in (2.3), we write

$$\lambda_{jki}(t) = Y_{ji}(t)\alpha_{jk0}(t)\exp(\beta_{jk}^T Z_i). \tag{3.8}$$

Alternatively, as given in (2.4), we may write

$$\lambda_{jki}(t) = Y_{ji}(t)\alpha_{jk0}(t)\exp(\beta^T Z_{jki}) \ . \tag{3.9}$$

Thus, the expression (3.2) can be written as

$$N_{jki}(t) = \int_0^t \alpha_{jk0}(u)\exp(\beta^T Z_{jki})\, Y_{ji}(u)\,du + M_{jki}(t). \tag{3.10}$$

Summing (3.10) over all $i = 1, 2, \cdots, n$, we have

$$N_{jk}(t) = \int_0^t \alpha_{jk0}(u) S_{jk}^{(0)}(\beta, u)du + M_{jk}(t), \tag{3.11}$$

where
$$N_{jk}(t) = \sum_{i=1}^{n} N_{jki}(t), \tag{3.12}$$

$$M_{jk}(t) = \sum_{i=1}^{n} M_{jki}(t), \tag{3.13}$$

and
$$S_{jk}^{(0)}(\beta, t) = \sum_{i=1}^{n} \exp(\beta^T Z_{jki})\, Y_{ji}(t). \tag{3.14}$$

We note that $N_{jk}(t)$ represents the total number of direct transitions from j to k observed in $[0, t]$, and $M_{jk}(t)$ is a martingale with mean zero, and $S_{jk}^{(0)}(\beta, t)$ represents the sum of the quantities $\exp(\beta^T Z_{jki})$ over the risk set in state j at time t.

On the other hand, (3.11) may be written as

$$dN_{jk}(t) = S_{jk}^{(0)}(\beta, t)\, dA_{jk0}(t) + dM_{jk}(t), \tag{3.15}$$

where

$$dA_{jk0}(t) = \alpha_{jk0}(t)dt. \tag{3.16}$$

Since $dM_{jk}(t)$ represents a zero mean 'random noise' component, a natural estimator of $A_{jk0}(t)$, the baseline integrated intensity function, is

$$\hat{A}_{jk0}(t;\beta) = \int_0^t \frac{1}{S_{jk}^{(0)}(\beta, u)}\, dN_{jk}(u). \tag{3.17}$$

However, one may have $Y_{ji}(t) = 0$ for some t, and in order to take care of this possibility, we introduce the indicator $J_j(t) = I_{(Y_j(t)=0)}$, where $Y_j = Y_{j1} + Y_{j2} + \cdots + Y_{jn}$, and define the estimator of $A_{jk0}(t)$ formally by

$$\hat{A}_{jk0}(t;\beta) = \int_0^t \frac{J_j(u)}{S_{jk}^{(0)}(\beta, u)}\, dN_{jk}(u). \tag{3.18}$$

On the other hand, the estimator of $A_{jk0}(t)$ can also be obtained from nonparametric

maximum likelihood for $\boldsymbol{N}$, we may interpret $dN_{jki}(t)$ as conditionally independent multinormial or Poisson distribution depending on whether $A_{jki}(t;\boldsymbol{Z}_i)$ is discrete or continuous. That is, given $\mathcal{F}_{t-}$, $\{dN_{jki}(t);(j,k)\in\Gamma, i=1,\cdots,n\}$ are independent and multinomially $\{Y_{ji}(t), dA_{jki}(t;\boldsymbol{Z}_i)\}$ distributed or independent Poisson $\{Y_{ji}(t)dA_{jki}(t;\boldsymbol{Z}_i)\}$ distributed. Thus, under noninformative censoring the 'multinomial' likelihood for $\boldsymbol{N}$ can be written, in terms of product integral, as

$$L_\tau = \mathbf{P}_{t\in\mathcal{T}}\left\{\prod_{(j,k)\in\Gamma}\prod_{i=1}^{n}(Y_{ji}(t)dA_{jki}(t))^{dN_{jki}(t)}\left(1-\sum_{(j,k)\in\Gamma}\sum_{i=1}^{n}dA_{jki}(t)\right)^{1-dN_{\cdots}(t)}\right\}, \quad (3.19)$$

where $N_{\cdots}(t) = \sum_{(j,k)\in\Gamma}\sum_{i=1}^{n}N_{jki}(t)$ and $dN_{jki}(t) = N_{jki}(t) - N_{jki}(t-)$.

Since $dA_{jki}(t;\boldsymbol{Z}_i) = dA_{jk0}(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_{jki})$, (3.19) becomes

$$\begin{aligned}L_\tau = \mathbf{P}_{t\in\mathcal{T}}\Bigg\{&\prod_{(j,k)\in\Gamma}\prod_{i=1}^{n}\left(Y_{ji}(t)dA_{jk0}(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_{jki})\right)^{dN_{jki}(t)}\\ &\times\left(1-\sum_{(j,k)\in\Gamma}dA_{jk0}(t)S_{jk}^{(0)}(\boldsymbol{\beta},t)\right)^{1-dN_{\cdots}(t)}\Bigg\}. \quad (3.20)\end{aligned}$$

For continuous case of $A_{jk0}(t)$, the 'multinomial' likelihood (3.20) can be expressed as the 'Poisson' likelihood

$$\begin{aligned}L_\tau = &\prod_{t\in\mathcal{T}}\left\{\prod_{(j,k)\in\Gamma}\prod_{i=1}^{n}\left(dA_{jk0}(t)\exp(\boldsymbol{\beta}^T\boldsymbol{Z}_{jki})\right)^{dN_{jki}(t)}\right\}\\ &\times\exp\left\{-\sum_{(j,k)\in\Gamma}\int_0^\tau S_{jk}^{(0)}(\boldsymbol{\beta},u)dA_{jk0}(u)\right\}. \quad (3.21)\end{aligned}$$

We note that the product over $t\in\mathcal{T}$ is a product over disjoint intervals. So (3.20) and (3.21) reduce to the finite products over t for which $N_{jki}(t)$ jumps at time t, i.e. $dN_{jki}(t)=1$; elsewhere $dN_{jki}(t)=0$.

Given the value of $\boldsymbol{\beta}$, maximization of (3.20) or (3.21) with respect to $dA_{jk0}(t)$ leads to

$$d\hat{A}_{jk0}(t;\boldsymbol{\beta}) = \frac{dN_{jk}(t)}{S_{jk}^{(0)}(\boldsymbol{\beta},t)}. \quad (3.22)$$

Thus, we again estimate $A_{jk0}(t)$ by (3.18).

Now we want to find the estimator of the regression parameters vector $\boldsymbol{\beta}$. If we substitute (3.22) into (3.20) or (3.21), we obtain the following partially maximized (partial) likelihoods (likelihood profile) only depending on $\boldsymbol{\beta}$ (assuming no ties) :

$$L_\tau^* = L(\beta, \tau) \times \mathbf{P}_{t \in \mathcal{T}} \{ \prod_{(j,k) \in \Gamma} (dN_{jk}(t))^{dN_{jk}(t)} (1 - dN_{\cdot\cdot}(t))^{1 - dN_{\cdot\cdot}(t)} \} \tag{3.23}$$

or

$$L_\tau^* = L(\beta, \tau) \times \prod_{t \in \mathcal{T}} \prod_{(j,k) \in \Gamma} dN_{jk}(t)^{dN_{jk}(t)} \exp(-N_{\cdot\cdot}(\tau)), \tag{3.24}$$

where

$$L(\beta, \tau) = \prod_{t \in \mathcal{T}} \prod_{(j,k) \in \Gamma} \prod_{i=1}^{n} \left(\frac{Y_{ji}(t) \exp(\beta^T Z_{jki})}{S_{jk}^{(0)}(\beta, t)} \right)^{dN_{jki}(t)} \tag{3.25}$$

which is the multistate Cox partial likelihood. The estimation of β can be based on (3.25), or equivalently on the logarithm of (3.25) such that

$$\begin{aligned} \log L(\beta, \tau) &= C(\beta, \tau) \\ &= \sum_{(j,k) \in \Gamma} \left\{ \sum_{i=1}^{n} \int_0^\tau \beta^T Z_{jki} \, dN_{jki}(t) - \int_0^\tau \log S_{jk}^{(0)}(\beta, t) \, dN_{jk}(t) \right\}. \end{aligned} \tag{3.26}$$

We denote $\hat{\beta}$ for the value of β which maximizes (3.26) (if such a value exists). Usually the estimate $\hat{\beta}$ may be obtained by the Newton-Raphson method as the solution of $U(\beta, \tau) = \frac{\partial}{\partial \beta} C(\beta, \tau) = 0$, where $U(\beta, \tau)$ is given by (4.8). We then estimate $A_{jk0}(t)$ by $\hat{A}_{jk0}(t, \hat{\beta})$ which is obtained from substituting $\hat{\beta}$ into (3.18).

Based on the preceding results, we now estimate the transition probabilities of the non-homogeneous Markov process given in Section 2. To do this we let $\hat{A}(t; Z)$ be the 4×4 matrix of the estimated integrated intensity functions $\hat{A}_{jk}(t; Z)$, $(j, k) \in E$, where

$$\hat{A}_{jk}(t; Z) = \hat{A}_{jk0}(t) \exp(\hat{\beta}^T Z_{jk}), \quad (j, k) \in I, \tag{3.27}$$

and

$$\hat{A}_{jj}(t; Z) = - \sum_{k \neq j} \hat{A}_{jk}(t; Z), \quad j = 1, 2. \tag{3.28}$$

We then estimate $P(s, t; Z)$ in (2.9) by

$$\hat{P}(s, t; Z) = \mathbf{P}_{(s,t]} \{ \mathrm{I} + d\hat{A}(u; Z) \}. \tag{3.29}$$

The estimate (3.29) is meaningful as long as $1 + d\hat{A}_{jj}(t; Z) \geq 0$, $j = 1, 2$.

The each estimate $\hat{A}_{jk}(t; Z)$ is a step-function with a finite number of jumps in $(s, t]$. Thus estimate in (3.29) is a finite product of matrices. If one or more transitions are observed a at time u (allowing for ties), then the contribution to (3.29) from this time point

is a matrix $\mathbf{I}+\Delta\hat{A}(u;Z)$, where $\Delta\hat{A}(u;Z)$ is the 4×4 matrix with entries (j,k) equal to $\Delta\hat{A}_{jk}(u;Z) = \Delta\hat{A}_{jk0}(u;\beta)\exp(\hat{\beta}^T Z_{jk})$, $(j,k)\in I$ and entries (j,j) equal to $\Delta\hat{A}_{jj}(u;Z) = -\sum_{k\neq j}\Delta\hat{A}_{jk}(u;Z)$, $j=1,2$, where

$$\Delta\hat{A}_{jk0}(u;\beta) = \frac{\Delta N_{jk}(t)}{S_{jk}^{(0)}(\beta,t)} . \tag{3.30}$$

For the computation of (3.29), let $t_1<t_2<\cdots<t_m$ be the times of all observed transitions between s and t. Then $P(s,t;Z)$ is estimated by

$$\hat{P}(s,t;Z) = \prod_{i=1}^{m}\{\mathbf{I}+\Delta\hat{A}(t_i;Z)\}, \tag{3.31}$$

with

$$\mathbf{I}+\Delta\hat{A}(t_i;Z) = \begin{pmatrix} 1+\Delta\hat{A}_{11}(t_i;Z) & \Delta\hat{A}_{12}(t_i;Z) & \Delta\hat{A}_{13}(t_i;Z) & 0 \\ 0 & 1+\Delta\hat{A}_{22}(t_i;Z) & \Delta\hat{A}_{23}(t_i;Z) & \Delta\hat{A}_{24}(t_i;Z) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{3.32}$$

The expression (3.31) provides a simple algorithm for the computation of (3.29). It may, nevertheless, be illustrative to give the explicit expressions for the elements of $\hat{P}(s,t;Z)$ by substituting $\Delta\hat{A}_{jk}(u;Z)$ into $dA_{jk}(u;Z)$ in the forms of (2.17) - (2.23). Specifically we have the following forms:

$$\hat{P}_{11}(s,t;Z) = \prod_{i=1}^{m}\left\{1-\frac{\exp(\hat{\beta}^T Z_{12})\Delta N_{12}(t_i)}{S_{12}^{(0)}(\hat{\beta},t_i)}-\frac{\exp(\hat{\beta}^T Z_{13})\Delta N_{13}(t_i)}{S_{13}^{(0)}(\hat{\beta},t_i)}\right\}, \tag{3.33}$$

$$\hat{P}_{22}(s,t;Z) = \prod_{i=1}^{m}\left\{1-\frac{\exp(\hat{\beta}^T Z_{23})\Delta N_{23}(t_i)}{S_{23}^{(0)}(\hat{\beta},t_i)}-\frac{\exp(\hat{\beta}^T Z_{24})\Delta N_{24}(t_i)}{S_{24}^{(0)}(\hat{\beta},t_i)}\right\}, \tag{3.34}$$

$$\begin{aligned}\hat{P}_{12}(s,t;Z) = \sum_{i=1}^{m}\Bigg\{ & \prod_{h=1}^{i-1}\left(1-\frac{\exp(\hat{\beta}^T Z_{12})\Delta N_{12}(t_h)}{S_{12}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{13})\Delta N_{13}(t_h)}{S_{13}^{(0)}(\hat{\beta},t_h)}\right) \\ & \times\frac{\exp(\hat{\beta}^T Z_{12})\Delta N_{12}(t_i)}{S_{12}^{(0)}(\hat{\beta},t_i)} \\ & \times\prod_{h=i+1}^{m}\left(1-\frac{\exp(\hat{\beta}^T Z_{23})\Delta N_{23}(t_h)}{S_{23}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{24})\Delta N_{24}(t_h)}{S_{24}^{(0)}(\hat{\beta},t_h)}\right)\Bigg\}\end{aligned} \tag{3.35}$$

$$\hat{P}_{13}(s,t;Z) = \hat{P}_{13}^{(1)}(s,t;Z) + \hat{P}_{13}^{(2)}(s,t;Z), \tag{3.36}$$

where

$$\hat{P}_{13}^{(1)}(s,t;Z) = \sum_{i=1}^{m}\left\{\prod_{h=1}^{i-1}\left(1-\frac{\exp(\hat{\beta}^T Z_{12})\Delta N_{12}(t_h)}{S_{12}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{13})\Delta N_{13}(t_h)}{S_{13}^{(0)}(\hat{\beta},t_h)}\right)\times\frac{\exp(\hat{\beta}^T Z_{13})\Delta N_{13}(t_i)}{S_{13}^{(0)}(\hat{\beta},t_i)}\right\}, \tag{3.36a}$$

$$\begin{aligned}\hat{P}_{13}^{(2)}(s,t;Z) = \sum_{i=1}^{m}\Bigg[\sum_{l=1}^{i-1}\Bigg\{\prod_{h=1}^{l-1}\left(1-\frac{\exp(\hat{\beta}^T Z_{12})\Delta N_{12}(t_h)}{S_{12}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{13})\Delta N_{13}(t_h)}{S_{13}^{(0)}(\hat{\beta},t_h)}\right)\\ \times\frac{\exp(\hat{\beta}^T Z_{12})\Delta N_{12}(t_l)}{S_{12}^{(0)}(\hat{\beta},t_l)}\\ \times\prod_{h=l+1}^{i-1}\left(1-\frac{\exp(\hat{\beta}^T Z_{23})\Delta N_{23}(t_h)}{S_{23}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{24})\Delta N_{24}(t_h)}{S_{24}^{(0)}(\hat{\beta},t_h)}\right)\Bigg\}\\ \times\frac{\exp(\hat{\beta}^T Z_{23})\Delta N_{23}(t_i)}{S_{23}^{(0)}(\hat{\beta},t_i)}\Bigg],\end{aligned} \tag{3.36b}$$

$$\begin{aligned}\hat{P}_{14}(s,t;Z) = \sum_{i=1}^{m}\Bigg[\sum_{l=1}^{i-1}\Bigg\{\prod_{h=1}^{l-1}\left(1-\frac{\exp(\hat{\beta}^T Z_{12})\Delta N_{12}(t_h)}{S_{12}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{13})\Delta N_{13}(t_h)}{S_{13}^{(0)}(\hat{\beta},t_h)}\right)\\ \times\frac{\exp(\hat{\beta}^T Z_{12})\Delta N_{12}(t_l)}{S_{12}^{(0)}(\hat{\beta},t_l)}\\ \times\prod_{h=l+1}^{i-1}\left(1-\frac{\exp(\hat{\beta}^T Z_{23})\Delta N_{23}(t_h)}{S_{23}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{24})\Delta N_{24}(t_h)}{S_{24}^{(0)}(\hat{\beta},t_h)}\right)\Bigg\}\\ \times\frac{\exp(\hat{\beta}^T Z_{24})\Delta N_{24}(t_i)}{S_{24}^{(0)}(\hat{\beta},t_i)}\Bigg],\end{aligned} \tag{3.37}$$

$$\hat{P}_{23}(s,t;Z) = \sum_{i=1}^{m}\left\{\prod_{h=1}^{i-1}\left(1-\frac{\exp(\hat{\beta}^T Z_{23})\Delta N_{23}(t_h)}{S_{23}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{24})\Delta N_{24}(t_h)}{S_{24}^{(0)}(\hat{\beta},t_h)}\right)\times\frac{\exp(\hat{\beta}^T Z_{23})\Delta N_{23}(t_i)}{S_{23}^{(0)}(\hat{\beta},t_i)}\right\}, \tag{3.38}$$

$$\hat{P}_{24}(s,t;Z) = \sum_{i=1}^{m}\left\{\prod_{h=1}^{i-1}\left(1-\frac{\exp(\hat{\beta}^T Z_{23})\Delta N_{23}(t_h)}{S_{23}^{(0)}(\hat{\beta},t_h)}-\frac{\exp(\hat{\beta}^T Z_{24})\Delta N_{24}(t_h)}{S_{24}^{(0)}(\hat{\beta},t_h)}\right)\times\frac{\exp(\hat{\beta}^T Z_{24})\Delta N_{24}(t_i)}{S_{24}(\hat{\beta},t_i)}\right\}. \tag{3.39}$$

The above expressions (3.33)-(3.39) may be used as an alternative to the direct application

of the form (3.31).

4. Asymptotic Properties

For the asymptotic properties of $\hat{\beta}$ and $\hat{A}_{jk0}(t, \hat{\beta})$, we introduce the following notations as in Andersen and Gill(1982). Let

$$S_{jk}^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} S_{jk}^{(0)}(\beta, t), \quad (j, k) \in I \tag{4.1}$$

and

$$S_{jk}^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta^2} S_{jk}^{(0)}(\beta, t), \quad (j, k) \in I \tag{4.2}$$

Then we have the p-vectors

$$S_{jk}^{(1)}(\beta, t) = \sum_{i=1}^{n} Z_{jki} \exp(\beta^T Z_{jki}) Y_{ji}(t), \quad (j, k) \in I, \tag{4.3}$$

and the $p \times p$ matrices

$$S_{jk}^{(2)}(\beta, t) = \sum_{i=1}^{n} Z_{jki}^{\otimes 2} \exp(\beta^T Z_{jki}) Y_{ji}(t), \quad (j, k) \in I, \tag{4.4}$$

where $a^{\otimes 2} = a a^T$ for any vector a. We also define the p-vector

$$E_{jk}(\beta, t) = \frac{S_{jk}^{(1)}(\beta, t)}{S_{jk}^{(0)}(\beta, t)}, \tag{4.5}$$

and the $p \times p$ matrix

$$\begin{aligned} V_{jk}(\beta, t) &= \{V_{jkhl}(\beta, t), h, l = 1, \cdots, p\} \\ &= \frac{S_{jk}^{(2)}(\beta, t)}{S_{jk}^{(0)}(\beta, t)} - E_{jk}(\beta, t)^{\otimes 2}. \end{aligned} \tag{4.6}$$

With the above definitions, the vector of score statistics $U(\beta, \tau) = (U_1(\beta, \tau), \cdots, U_p(\beta, \tau))$, where $U_h(\beta, \tau) = \frac{\partial}{\partial \beta_h} C(\beta, \tau)$, $h = 1, \cdots, p$ can be written as

$$U(\beta, \tau) = \sum_{(j,k) \in I} \left\{ \sum_{i=1}^{n} \int_0^\tau Z_{jki} dN_{jki}(t) - \int_0^\tau E_{jk}(\beta, t) dN_{jk}(t) \right\}. \tag{4.8}$$

And the matrix of second-order partial derivatives of $C(\beta, \tau)$ is $-I(\beta, \tau)$, where

$$I(\beta, \tau) = \sum_{(j,k)\in\Gamma} \int_0^\tau V_{jk}(\beta, t) dN_{jk}(t). \tag{4.9}$$

Then, from the standard likelihood theory, we find that under certain regularity conditions (e.g. Andersen and Gill(1982, §1105, §1110), $\sqrt{n}(\hat{\beta} - \beta_0)$, where β_0 is the vector of true parameters, is asymptotically multivariate normally distributed with mean zero vector and a covariance matrix $nI^{-1}(\beta, \tau)$ which may be estimated consistently by $nI^{-1}(\hat{\beta}, \tau)$. From this result, it follows that for a simple hypothesis $H_0: \beta = \beta_0$, the Wald test statistic

$$Q_1 = (\hat{\beta} - \beta_0)^T I(\hat{\beta}, \tau)(\hat{\beta} - \beta_0) \tag{4.10}$$

is asymptotically χ^2-distribution with p degrees of freedom. Inferences can also be based on the score test statistic and the Cox partial likelihood ratio test statistic, i.e. under the same regularity conditions as in the Wald test statistic, for testing $H_0: \beta = \beta_0$,

$$Q_2 = U(\beta_0, \tau)^T I^{-1}(\beta_0, \tau)\, U(\beta_0, \tau) \tag{4.11}$$

and

$$Q_3 = 2\big(C(\hat{\beta}, \tau) - C(\beta_0, \tau)\big) \tag{4.12}$$

have the same asymptotic χ^2-distribution with p degrees of freedom.

Next, for the asymptotic property of $\hat{A}_{jk0}(t, \hat{\beta})$, under the same regularity conditions as in Andersen and Gill(1982, §1110), each process $\sqrt{n}\big(\hat{A}_{jk0}(t, \hat{\beta}) - A_{jk0}(t)\big)$ is asymptotically distributed as a Gaussian process with mean zero, independent increments and the estimated variance function

$$n\left\{\int_0^t \frac{dN_{jk}(u)}{S_{jk}^{(0)}(\hat{\beta}, u)^2} + H_{jk}(\hat{\beta}, t)^T I^{-1}(\hat{\beta}, \tau) H_{jk}(\hat{\beta}, t)\right\}, \tag{4.13}$$

where

$$H_{jk}(\hat{\beta}, t) = -\int_0^t \frac{S_{jk}^{(1)}(\hat{\beta}, u)}{\{S_{jk}^{(0)}(\hat{\beta}, u)\}^2} dN_{jk}(u). \tag{4.14}$$

On the other hand, for the asymptotic property of the each transition probabilities $P_{jk}(s, t; Z)$, $(j, k) \in E$, we find that from the asymptotic properties of $\big(\hat{\beta}, \hat{A}_{jk0}(t, \hat{\beta})\big)$ via the functional delta method, $\sqrt{n}\big(\hat{P}_{jk}(s, t; Z) - P_{jk}(s, t, ; Z)\big)$ is asymptotically distributed as a Gaussian process with mean zero and the estimated variance function $\hat{V}_1 + \hat{V}_2$, where

$$\hat{V}_1 = \sum_{h \neq l} \int_s^t \hat{P}_{jh}(s,u;Z)^2 \{\hat{P}_{lk}(u,t;Z) - \hat{P}_{hk}(u,t;Z)\}^2 \times J_h(u) \frac{\{\exp(Z_{jl}\hat{\beta})\}^2}{S_{hl}^{(0)}(\hat{\beta},u)^2} dN_{hl}(u), \tag{4.15}$$

and

$$\hat{V}_2 = Q^T(s,t,\hat{\beta};Z)\, I^{-1}(\hat{\beta},\tau)\, Q(s,t,\hat{\beta};Z), \tag{4.16}$$

and where

$$Q = \int_s^t \sum_{h,l} \hat{P}_{jk}(s,u;Z)\, d\,W_{hl}(u)\, \hat{P}_{lk}(u,t;Z), \tag{4.17}$$

with $$W_{hl}(t) = \exp(\hat{\beta}^T Z_{hl}) \int_0^t \frac{(Z_{hl} - E_{hl}(\hat{\beta},u))J_h(u)}{S_{hl}^{(0)}(\hat{\beta},u)} dN_{hl}(u), \text{ for } h \neq l, \tag{4.18}$$

and $$W_{hh}(t) = -\sum_{h \neq l} W_{hl}$$ (e.g. Aalen and Johansen(1978) and Andersen et al.(1991)).

Therefore, we may use the above result to set approximate pointwise confidence limits for each transition probability.

5. An application

In this section, we illustrate the results of applying the illness-death model(with two alive states and two death states) to the survival of patients from a heart transplant study. The data used here came from the Stanford Heart Transplantation Program and were described in detail in Clark et al.(1971).

In summary, patients are admitted to the program after it is decided that the patient is not likely to respond to other forms of therapy. Then a donor heart, matched on blood type, is sought for a patient. This search has taken from a few days to almost a year. Thus, some patients die before a heart is found, while who receive a transplant may die from the rejection of the new heart from other causes. Sometimes, several patients are matched on blood type for a given donor heart. Then other criteria such as tissue typing or absence of pulmonary infection are used to choose the recipient of the heart. In general, no serious bias were found from this selection procedure. All patients(except two who are lost) were followed from their entrance into the program to their death or to the end of the study on April 1, 1974. The first patient was entered the program on November 15, 1967, and the last on March 22, 1974. During this period, 103 patients were entered the program. Of these, four patients have improved enough while waiting for transplant, and were then 'deselected' and followed as long as possible(two have been lost to follow-up and two have died). These four patients are treated like other patients who were not transplanted. Also one patient failed the first

transplant at 50 days but was immediately retransplanted and survived another 15 days more. For the purpose of present analysis, this patient is considered to have died by rejection at 50 days.

On the other hand, four of the transplanted patients were not tissue typed and are excluded from the analysis. Thus, the survival of 99 patients has been analyzed here. Of these 99 patients, the number of patients corresponding to each transition is as follows:

Transition	Number	Transition	Number
$1 \rightarrow 1$	4	$1 \rightarrow 2 \rightarrow 2$	24
$1 \rightarrow 3$	30	$1 \rightarrow 2 \rightarrow 3$	12
		$1 \rightarrow 2 \rightarrow 4$	29

Many authors have analyzed the survival data of patients in the Stanford Heart Transplantation Program: among these are Gail(1972), Turnbull et al.(1974), Crowley and Hu(1977). Beck(1979), Kalbfleisch and Prentice (1980, Ch.5), and Aitkin et al.(1983). In particular, Crowley and Hu(1977) evaluated the effects of several covariates on survival using the regression model of Cox to see for which variables, if any, transplantation is likely to prolong survive. Beck applied a similar stochastic survival model to the present work, but he restricted assumption to the case where the underlying intensity functions are constant. Aitkin et al. suggested various parametric models to examine the effects of covariates on survival. They considered pretransplant and post transplant survival separately. However, the effect of competing risks for the transplanted patients was not addressed in their work. That is, the intensity function and the effect of the covariates may differ depending on whether the patient dies from a rejected transplant of from other causes. They also mentioned residuals in connection with parametric model checking but indicated that they did not find them to be particularly useful.

In order to evaluate this survival process of the heart patients, the illness-death model in Figure 1 is used, where in this situation, the states are: 1, accepted into the program (nontransplant); 2, received a new heart (transplant); 3, died from rejection of the donor heart (rejection); and 4, died from any other causes. We note that this model allows to avoid the problem of changing covariate values at a random point in time (as opposed to Crowley and Hu(1977)), and to analyzed the data to be studied in one analysis that permits the death of transplanted patients to be examined by cause of death (as opposed to Aitkin et al.(1983)), and to let the underlying intensity functions need not to be specified (as opposed to Beck(1979), important in the present case because of difficulty in finding a parametric hazard with an adequate fit to the data.

The model used here has the underlying intensity functions $\alpha_{jk0}(t)$, $(j,k) \in \Gamma$. Also associated with each patient is a vector $\boldsymbol{Z}_i$ of covariates. The covariates used on survival of

patients not transplanted are age (in years) at acceptance into the program, previous open-heart surgery (1=yes, 0=no). The covariates for transplanted patients are age at transplant, previous open-heart surgery, calendar time at transplant (in days from 10/1/1967), waiting time (in days) to transplant, and three measures of the degree of the match to which donor and recipient are mismatched for tissue type. The first of these is the number of mismatches which is the number of donor alleles with no match in the recipient (1 through 4). The second is a dichotomous mismatch score on the antigen HLA-A2 (1=donor has HLA-A2 and recipient has neither HLA-A2 nor the similar HLA-A28, 0=otherwise). The third is a continuous mismatch score derived from antibody response of pregnant women by Charles Bieber of Stanford University which ranges between and 0 and 3.05 in this data.

An initial model has been fitted for which all covariates are included. Solving $\boldsymbol{U}(\boldsymbol{\beta}, \tau) = \mathbf{0}$ in (4.8) and using (4.9), the estimates of the covariates coefficients and their estimated standard errors are given in Table 1(a). It is clear that several of the covariates have relatively little effect on survival of patients and should be eliminated. After removing insignificant covariates from the initial model, the final model has been fitted and is given in Table 1(b). From this table, we see that the age at acceptance into the program is only shown to be significant covariate for both 1→2 and 1→3 transitions. For 2→3 transition, calendar time of transplant is shown to be the only significant covariate. For 2→4 transition, age at transplant and the continuous mismatch score are shown to be significant covariates. In order to check the model fit, the residual plots for the final model were shown in Figure 2 and indicates that the model fits data well overall sense.

On the other hand, the average values of the covariates of all individuals were age at acceptance into the program, 44.63 years; age at transplant, 45.63 years; calendar time of transplant, 1346 days; and mismatch score, 1.1646. Using these values in the final model the estimated survival functions from the time (in days) of acceptance into the program for an individual with and with no previous surgery were shown in Figure 3, respectively. For example, the survival probabilities to 500 and 1,000 days for an individual with previous surgery are 0.315 and 0.271, respectively. And the corresponding survival probabilities for an individual with no previous surgery are 0.445 and 0.307, respectively. Thus, we see that the individuals with no previous surgery have slightly the higher survival probability than the individuals with previous surgery, although the previous surgery does not seem to be a significant covariate in Table 1. The estimated transition probabilities of 1→3(directly), 1→3 (via state 2) and 1→4 based on Model 2 with no previous surgery are shown in Figure 4. From this figure, we see that during initial period, 1→3(directly) is the most critical transition for the death of a patient, but after 1,000 days the main death of the patients is due to 1→4 transition.

The competing risks probabilities can also be calculated. The net probability is given by eliminating the risk of death from rejection. Using the average values of covariates, the estimated net and crude probabilities are shown in Figure 5. For example, the net probability

to 500 days for a patient is 0.08 and the total crude probability via state 2 is 0.22. Therefore, if death by rejection from receiving a new heart can be reduced, the death probability of a patient may be reduced considerably provided that the independent risks of death is assumed and the model fits.

Table 1. Covariates Estimates and Standard Errors in the Models

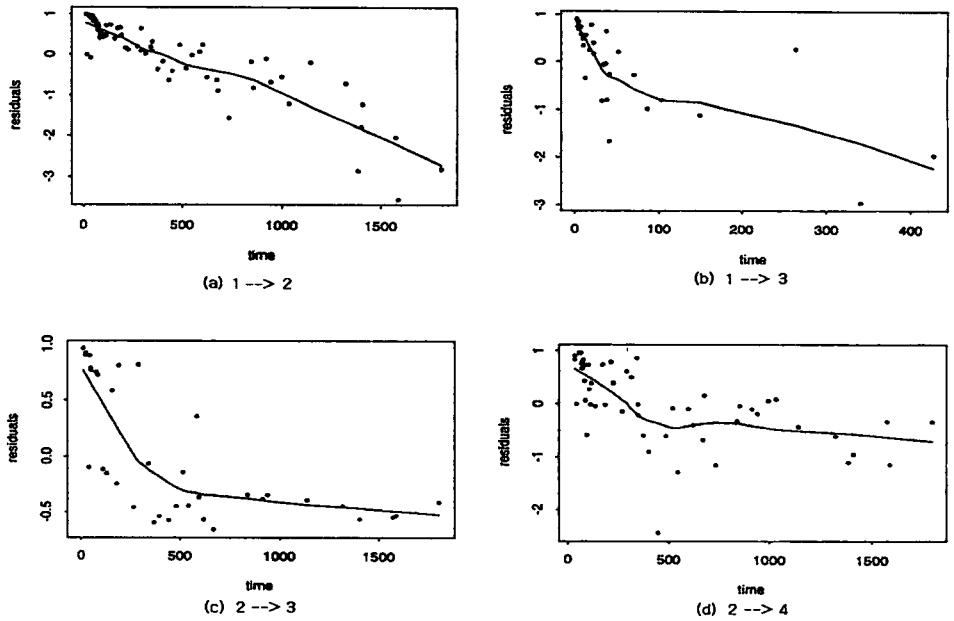
(a) Model 1

Transition	Covariates	Estimates	S.E	P-value
1→2	Age at acceptance	0.0324	0.0158	0.0408
	Prior surgery	-0.4581	0.3245	0.1580
	Likelihood ratio test	= 6.22 on	2 df,	0.0446
	Efficient score test	= 6.12 on	2 df,	0.047
1→3	Age at acceptance	0.0475	0.0205	0.0205
	Prior surgery	0.9333	0.6285	0.1376
	Likelihood ratio test	= 8.22 on	2 df	0.0164
	Efficient score test	= 7.52 on	2 df	0.0233
2→3	Age at transplant	0.015213	0.03712	0.682
	Calendar time	-0.000823	0.00054	0.127
	Waiting time	-0.001365	0.00698	0.845
	Prior surgery	-0.427719	0.82431	0.604
	# of mismatches	-0.824335	0.53987	0.127
	HL-A2	-0.800574	1.19206	0.502
	Mismatch score	1.026609	0.71197	0.149
	Likelihood ratio test	= 11.6 on	7 df	0.114
	Efficient score test	= 12.6 on	7 df	0.0832
2→4	Age at transplant	9.62e-02	0.03188	0.00254
	Calendar time	-7.35e-05	0.00038	0.84661
	Waiting time	-8.86e-03	0.00761	0.24432
	Prior surgery	-9.00e-01	0.63374	0.15540
	# of mismatches	-4.81e-02	0.22884	0.83354
	HL-A2	-3.17e-02	0.57098	0.95568
	Mismatch score	1.15e+00	0.47838	0.01631
	Likelihood ratio test	= 26.9 on	7 df	0.000341
	Efficient score test	= 23.7 on	7 df	0.0013

(b) Model 2

Transition	Covariates	Estimates	S.E	P-value
1→2	Age at acceptance	0.0324	0.015.8	0.0408
	Prior surgery	-0.4581	0.3245	0.1580
	Likelihood ratio test	= 6.22 on	2 df	0.0446
	Efficient score test	= 6.12 on	2 df	0.047
1→3	Age at acceptance	0.0475	0.0205	0.0205
	Prior surgery	0.9333	0.6285	0.1376
	Likelihood ratio test	= 8.22 on	2 df	0.0164
	Efficient score test	= 7.52 on	2 df	0.0233
2→3	Calendar time	-0.00106	0.000493	0.031
	# of mismatches	-0.86151	0.550916	0.118
	Mismatch score	1.01946	0.711751	0.152
	Likelihood ratio test	= 10.6 on	3 df	0.0141
	Efficient score test	= 11.6 on	3 df	0.00872
2→4	Age at transplant	0.101	0.0322	0.00168
	Prior surgery	-0.935	0.6218	0.13285
	Mismatch score	1.164	0.3999	0.00361
	Likelihood ratio test	= 25 on	3 df	1.55e-05
	Efficient score test	= 22.7 on	3 df	4.68e-05

Figure 2. Residual plot



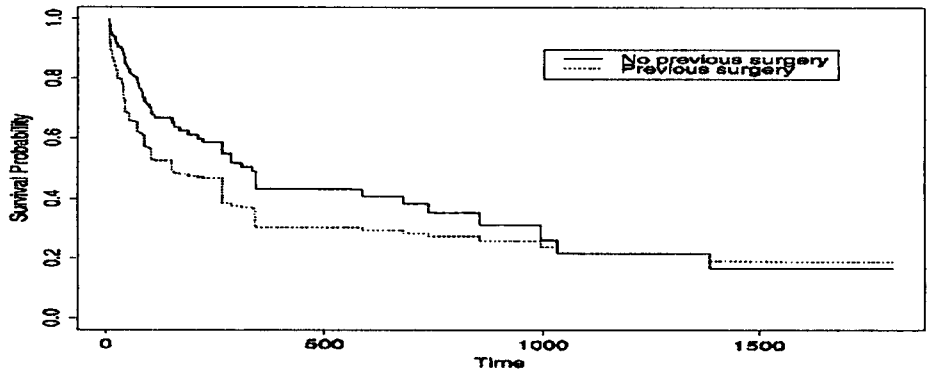


Figure 3. Estimated survival probabilities

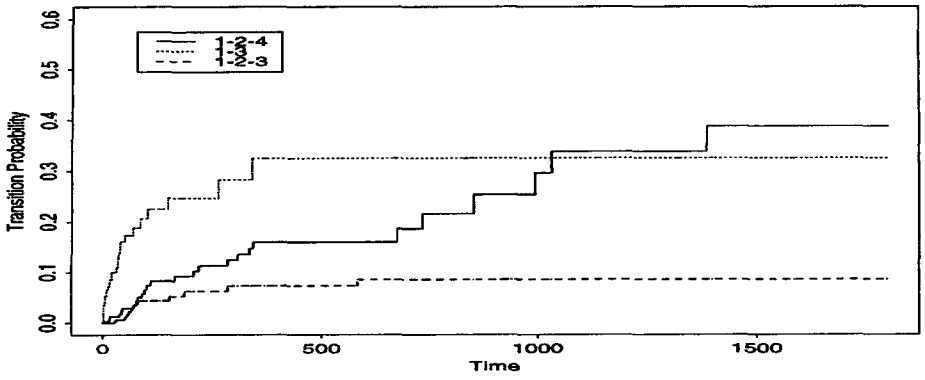


Figure 4. Three different transition probabilities

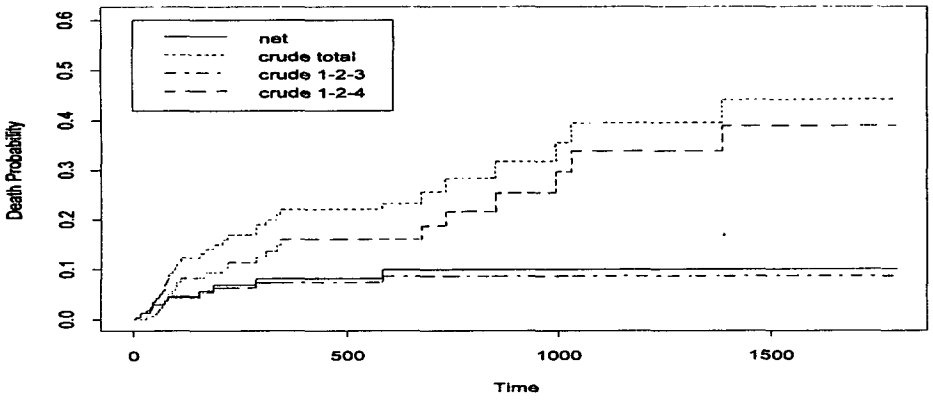


Figure 5. Competing risks probabilities

6. Concluding Remarks

In this paper we have used a Cox regression model to incorporate the covariates in a multistate stochastic survival model. We have seen that the multistate Cox regression model provide a flexible frame work for the study of the effects of various covariates on several transition rates and the important biological insight may be gained from the analysis of such a model which may have been overlooked by merely considering simple survival models.

However, in this paper we have only concerned with the non-homogeneous Markov process. In fact, we may envisage a Cox regression model for the semi-Markov process, and then we may want to develope methods of distinguishing between the Markov and the semi-Markov

processes.

In conclusion, the method of analysis given in this paper should be looked upon as an advantage of how the results of Aalen and Johansen (1978) and the generalization of these to the regression context(cf., Andersen(1988); Andersen et al. (1991)) can be carried out to evaluate the survival, transition and competing risks probabilities when the intensity function of an individual for each transition depends on the covariates. Finally we hope that the results in this paper would be useful to a refined analysis of data arising from a clinical trial such as the Stanford Heart Transplantation Program considered.

Acknowledgement

This paper was prepared during my visit from 1996 to 1997 to Division of Statistics, University of California, Davis. I wish to express my gratitude to Professor George Roussas for his hospitality. I also want to thank Sam Chung for his computer assistance.

References

[1] Aalen, O. O. (1975). Statistical Inference for a Family of Counting Processes, Ph.D. Dissertation, Department of Statistics, University of California, Berkeley.

[2] Aalen, O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations, *Scandinavian Journal of Statististics*, Vol 5, 141-150.

[3] Aitkin et al. (1983). A Reanalysis of the Stanford Heart Transplant Data, *Journal of the American Statistical Association*, Vol. 78, 264-292.

[4] Andersen, P. K. and Gill, R. D. (1982). Cox's Regression Model for Counting Processes: a Large Sample Study, *Annals of Statistics*, Vol. 10, 1100-1120.

[5] Andersen, P. K. and Borgan, (1985). Counting Process Models for Life History Data: a Review (with discussion). *Scandinavian Journal of Statistics*, Vol. 12, 97-158.

[6] Andersen, P. K. et al. (1988). Censoring, truncation and filtering in statistical models based on counting processes, *Contemp. Math.* Vol. 80, 19-60.

[7] Andersen, P. K. et al. (1991). Non- and semi-parametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process, *Scandinavian Journal of Statistics*. Vol. 18, 153-167.

[8] Beck, G. J. (1979). Stochastic Survival Models with Competing Risks and Covariates, *Biometrics*, Vol. 35, 427-438.

[9] Chiang, C. L. (1961a). A Stochastic Study of the Life Table and its Applications, III. The follow-up study with the Consideration of Competing Risks, *Biometrics*, Vol. 17, 57-78.

[10] Chiang, C. L. (1961b). On the probability of death from specific causes in the presence of competing risks, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, (J. Neyman, Ed.), Univ. of Calif. Press, Vol. IV, 168-180.

[11] Chiang, C. L. (1968). *Introduction to Stochastic Process in Biostatistics*, Wiley, New York, 1968.

[12] Clark, D. A., Stinson, E. B., Griepp, R. B., Schroeder, J. S., Shumway, N. E., and Harrison, D. C. (1971). Cardiac Transplantation in Man, VI. Prognosis of Patients Selected for Cardiac Transplantation, *Annals of Internal Medicine*, Vol. 75, 15-21.

[13] Cox, D. R. (1972). Regression Models and Life Tables (with discussion), *Journal of the Royal Statistical Society*, B, Vol. 34, 187-220.

[14] Crowley, J. and Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data, *Journal of the American Statistical Association*, Vol. 72, 27-36.

[15] Dewanji, A. and Kalbfleisch, J. D. (1986). Nonparametric Methods for Survival/Sacrifice Experiments, *Biometrics*, Vol. 42, 325-341.

[16] Dinse, G. E. and Lagakos, S. W. (1982). Nonparametric Estimation of Lifetime and disease onset distribution from incomplete observations, *Biometrics*, Vol. 38, 921-932.

[17] Fix, E. and Neyman, J. (1951). A simple stochastic model of recovery, relapse, death, and loss of patients, *Hum. Biol.* Vol. 23, 205-241.

[18] Gail, M. H. (1972). Does Cardiac Transplantation Prolong Life? A Reassessment, *Annals of Internal Medicine*, Vol. 76, 815-817.

[19] Gill. R. D. and Johansen, S. (1990). A survey of product-integration with a view towards application in survival analysis, *Annals of Statistics*, Vol. 18, 1501-1555.

[20] Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.

[21] Kodell, R. and Nelson, C. (1980). An illness-death model for the study of the carcinogenic process using survival/sacrifice data. *Biometrics*, Vol. 36, 267-277.

[22] Kodell, R., Shaw, G. and Johnson, A. (1982). Nonparametric joint estimators for disease resistance and survival functions in survival/sacrifice experiments. *Biometrics*, Vol. 38, 43-58.

[23] McKnight, B. and Crowley, J. (1984). Tests for differences in tumor incidence based on animal carcinogenesis experiments, *Journal of American Statistical Association*, Vol 79, 639-648.

[24] Sacks, S. T. and Chiang, C. L. (1977). A transition-probability model for the study of chronic diseases, *Mathematical Biosciences*, Vol. 34. 325-346.

[25] Tsiatis, A. (1975). A nonidentifiability Aspect of the problem of Competing risks, *Proc. Nat. Acad. Sci.*, Vol. 72, 20-22.

[26] Turnbull, B. W., Brown, B. W., Jr., and Hu, M. (1974). Survivorship Analysis of Heart

Transplant Data, *Journal of the American Statistical Association*, Vol. 69, 74-80.

[27] Turnbull, B. W. and Mitchell, T. (1984). Nonparametric estimation of the distribution of time to onset for specific disease in survival/sacrifice experiments, *Biometrics*, Vol. 40, 41-50.