東海大學管理學院

統計研究所

--- 碩士論文 ---

在半參數轉換馬可模型下探討多狀態存活分析

Multi-state Survival Analysis:

Semiparametric Transformation Markov Model

指導教授：沈葆聖 博士

研究生：郭玲毓

中華民國一〇五年七月

Multi-state Survival Analysis: Semiparametric Transformation Markov Model

Director: Pao-sheng Shen
Student: Ling-Yu Kuo
Department of Statistics
Tunghai University
Taichung, Taiwan 40704

東海大學碩士班研究生

論文口試委員審定書

統計學系碩士班郭玲毓君所提之論文

Multi-state Survival Analysis: Semiparametric Transformation Markov Model

---

經本委員會審議，認為符合碩士資格標準。

論文口試委員召集人 戴政 （簽章）

委員 沈存玲

黃愉閔

中華民國 105 年 07 月 21 日

謝誌

首先，我要感謝我的指導老師　沈葆聖教授，感謝沈老師這段時間的指導，面對五年一貫的我，要在短短的一年內完成論文並非一件容易的事，但沈老師很有耐心與熱心地傳授自身所學，讓我在這段時間學到許多，從沈老師身上我也看到了做學問的專注與熱誠，以及將知識傳承下去的使命感。我由衷地感謝沈老師在論文和課業上對我的幫助與教導，讓我能夠順利完成本篇碩士論文。

感謝　戴政教授和　黃愉閔教授在百忙之中能夠撥空擔任口試委員，給予我寶貴的指導與建議，讓我的論文能夠更加地充實與完善。

感謝東海統計系所有教授與助教們，由於大家的教導與幫助，除了熱心地傳授知識，給予成長的養分，讓我在這五年的求學過程中收穫豐富外，也分享了許多人生智慧，讓我受益良多。其中特別感謝王榮琮老師曾經對於我在學業上的指導與幫助。很榮幸能夠在東海大學這個美麗的校園完成我的碩士學位，未來我將會更精進專業，努力不懈。

感謝碩士班的所有同學們，這一年一路走來，大家一起同甘共苦，歷經了歡樂與悲傷，共創了許多回憶。很感謝大家在課業上對於我的幫助，當我遇到困難時，總是熱心地幫我解決，讓我能夠順利通過每次考試與每科修課課程，完成學業。

最後，我要感謝我的家人，因為有家人的的支持與資助，讓我能夠無後顧之憂的專注在學業上。我衷心地感謝這五年陪我一路走來，以及曾經幫助過我的所有人，老師、助教、同學、朋友以及家人們，感謝各位。

學生 郭玲毓 謹誌於
東海大學統計研究所
中華民國 105 年 7 月

Abstract

A multi-state model (MSM) is a model for a continuous time stochastic process allowing individuals to transit among a finite number of states. When there are covariate effects to be considered, Cox Markov regression model has been widely used for modelling transition intensities between the states. The Cox model, however, may not be suitable for describing transition rates. Semiparametric transformation models, which includes Cox's model as a special case, has been widely used in the analysis of survival data. The purpose of this article is to study semiparametric transformation model in a general finite-stat Markov process setting. Based on the product integral and the functional delta method, we present an estimator of the transition probability matrix and derive its large-sample theory. The proposed method is illustrated with bone marrow transplant data.

Keywords: Semiparametric Transformation Model; Markov regression model; Multi-state model; Cox proportional model; Transition Probability.

Contents

1. Introduction

Multi-state models (MSM) are models for stochastic processes which occupy one of a set of discrete states at any time. The MSM are well adapted for modeling complex event histories. They are useful in describing a process in which an individual moves through a series of states in continuous time and can provide a better understanding of the process of the failure, i.e. a better knowledge of the evolution of the disease/depressed over time. Based on MSM, one may estimate progression rates, assess the effects of risk factors and survival rates. In clinical and epidemiological follow-up studies, continuous-time MSM are widely used to describe disease processes in situations in which an individual is going through varying stages over time. The complexity of MSM depend on the number of states defined, as well as the transitions allowed among these states. A commonly-used model is the illness-death model, with three states representing health, illness and death. Transitions are permitted from health to illness, illness to death and health to death. Recovery from illness to health is sometimes also considered. A wide range of medical situations have been modelled using MSM, for example, a patient recovering from a bone marrow transplant (BMT) for leukaemia may fail therapy due to one of several terminal events, such as death in remission or relapse. As patients recover from their transplant, they may experience several intermediate events, which have influences on their eventual prognosis, such as the return of the platelets to a "normal" level, the development of various types of infections, the occurrence of acute or chronic graft-versus-host disease (GVHD), etc. This complex process can be described by using a multi-state model. Figure 1 shows a simplified diagram of the recovery process, where four events are taken into consideration, namely, acute GVHD (A), chronic GVHD (C), death in remission (D) and relapse (R). Since patients who relapse are typically considered as failures of the treatment, relapse is treated as an absorbing state. Thus, these four events can be modeled by a six-state model with two absorbing states, "5: D" and "6: R", and four transient states, "1: Tx (stands for transplantation)", "2: A", "3: C" and "4: AC (stands for both A and C)". Figure 1 shows a simplified diagram of the recovery process, where the arrows show which transitions are possible between states.

The inference in MSM is traditionally performed under a Markov assumption for which past and future are independent given its present state, i.e. the transition rates depend only on the current state of the patient and not on the patient's history. When there is no covariate, Aalen and Johansen (1978) demonstrated how counting process methods can be used to estimate transition probabilities. When there are covariates which may influence the rate of transition from one state to the next, it is necessary to accommodate covariates influence through the use of regression models, which leads to the so-called "Markov regression model". In literature, many Markov models have been proposed. For example, parametric models for the transition intensities have been considered by (see Begg and Larson (1982), Kalbfleisch and Lawless (1985), Marshall and Jones (1995), Alioum and Commenges (2001), Peŕeź-Ocon et al. (2001)). Andersen (1988) developed a semiparametric Markov regression model for multistate survival experiments where the intensity rate for each of the transitions among the states was modeled by a separate Cox (1972) proportional hazards regression model, i.e. the so-called "AG-Cox Markov model". Under AG-Cox Markov model, Andersen et al.

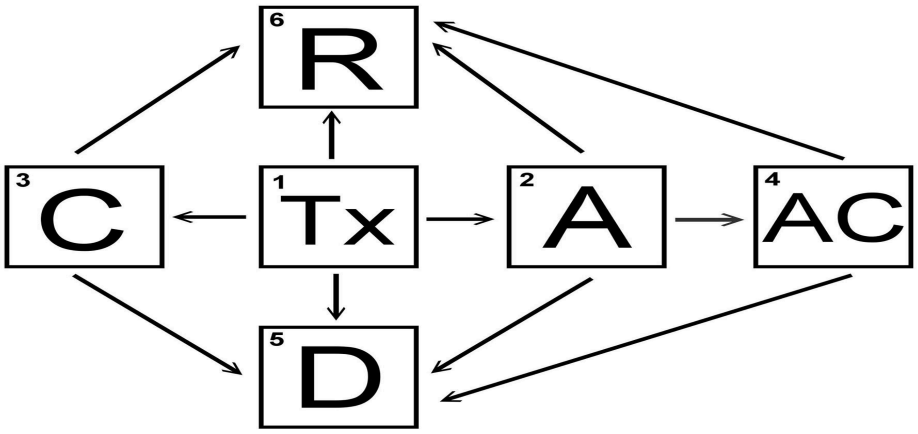

Figure 1. Schematic depiction of six-state model for BMT

(1991) (see also Andersen et al.(1993)) proposed estimators for transition probabilities and derived their large sample properties. Klein et al (1993) suggested an alternative approach to multistate modeling by fitting a Cox model to each of the events with time dependent covariates used to model the timing of the intermediate events that precede the event of interest. Shu and Klein (2005) studied two alternatives to Cox's, Aalen's (1989) nonparametric additive hazards model and Lin and Ying's (1994) semiparametric additive hazards model.

In some situations, the Cox model, however, may not be suitable for describing transition rates. The semiparametric transformation models have been proposed to allow various non-proportional hazards structures, such as proportional odds (Bennett, 1983; Pettitt, 1984). In Section 2, we study semiparametric transformation model in a general finite-stat Markov process setting. Based on the product integral and the functional delta method, we present an estimator of the transition probability matrix and derive its large-sample theory. In Section 3, the proposed method is illustrated with bone marrow transplant data.

2. Multi-state models

2.1. Markov Processes and The Product Integral

For individual k, a multi-state process is a stochastic process $X_k = \{X_k(t), t \in [0, \tau]\}$ with a finite state space $E = \{1, 2, \ldots, M\}$ and with right-continuous path: $X_k(t+) = X_k(t)$. For any t, the variable $X_k(t)$ has values in E, i.e. M states. We will assume that independent multi-state processes $\{X_k(t), 0 \leq t \leq \min(\tau, C_k); k = 1, \ldots, n\}$ are observed in continuous time, where C_k is the independent right-censoring time for individual k. Thus, the observations are subject to right censoring, i.e. we cannot observe the processes over an infinite time period since the observation of the process $X_k(t)$ is stopped at $\min(C_k, \tau)$. The data for individual k can then be represented as a multivariate counting process $N_{ijk}(t)$, $i, j \in E; i \neq j; t \leq C_k$, counting the number of direct $i \to j$ transitions observed for subject

k in $[0,t]$ (where some i,j combinations may not be possible). Associated with $X_k(t)$ is a counting process $N_{ijk}(t)$, which denotes the number of direct transition from state i to j in the interval $[0,t]$, i.e. $N_{ijk}(t) = \#\{s \leq t : X_k(s-) = i, X_k(s) = j\}$, $i \neq j$. The other process is the indicator $Y_{ik}(t) = I_{[X_k(t-)=i]}$, which denotes whether the process is in state i just before time t. Define the filtration or history process as

$$\mathcal{F}_t = \{N_{ijk}(u), Y_{ik}(u), 0 \leq u \leq t, k = 1, \ldots, n; i, j = 1, \ldots, M\}.$$

Notice that the history $\mathcal{F}_s$ of the process can also be generated by $\{X_k(u), u \leq s \ k = 1, \ldots, n\}$, i.e. $\mathcal{F}_s$ is an element of a 'filtration' and it can be understood intuitively as the trajectory of the process until time s. The law of multi-state processes can be specified by the transition probabilities

$$P_{ijk}(s,t,\mathcal{F}_s) = P(X_k(t) = j | X_k(s) = i, \mathcal{F}_{s-}) \ (i = 1, \ldots, M; j = 1, \ldots, M).$$

Under Markov model, given the state at time s, the whole history before s can be forgotten:

$$P_{ijk}(s,t,\mathcal{F}_{s-}) = P(X_k(t) = j | X_k(s) = i) = P_{ijk}(s,t).$$

The state occupation probabilities are $\pi_{jk}(t) = P(X_k(t) = j)$, $j \in E$ and, in particular, the initial distribution is $\pi_{jk}(0) = P(X_k(0) = j)$, $j \in E$. We may then write $\pi_{jk}(t) = \sum_{i \in E} \pi_{ik}(0) P_{ijk}(0,t)$.

Next, we consider transition intensities. Under a Markov assumption, if a randomly chosen individual k is in state i at time $t-$, the transition rate or intensity from i to j at time t is given by

$$d\Lambda_{ijk}(t) = P(X_k(t - +dt) = j | X_k(t-) = i) = P(X_k(t - +dt) = j | X_k(t-) = i),$$

which holds for all $X_k(u)$, $0 \leq u < t$ with $X_k(t-) = i$ and $i \neq j$. For convenience, define $d\Lambda_{iik}(t) = -\sum_{j \neq i} \Lambda_{ijk}(t)$ such that the row sums of the matrix $d\Lambda_{ijk}(t) = [d\Lambda_{ijk}]_{M \times M}$ are all equal to zero. For continuous case, we have $d\Lambda_{ijk}(t) = \lambda_{ijk}(t)dt$ for all $i \neq j$, where

$$\lambda_{ijk}(t) = \lim_{h \to 0} h^{-1} P(X_k(t - +h) = j | X_k(t-) = i).$$

Hence, $\lambda_{ijk}(t)$ is the intensity function for $i - to - j$ transition. Let $Y_{ik}(u) = I_{[X_k(u-)=i]}$. For $i \neq j; i, j \in E$, $M_{ijk}(t) = N_{ijk}(t) - \int_0^t Y_{ik}(u) \lambda_{ijk}(u) du$, are zero mean local square-integrable martingale with respect to $\mathcal{F}_t$. A state $h \in E$ is absorbing if for all $t \in [0, \tau]$, $j \neq h$, $\lambda_{hjk}(t) = 0$.

For homogenous population, $P_{ijk}(s,t) = P_{ij}(s,t)$ and $\lambda_{ijk}(t) = \lambda_{ij}(t)$ for all k. The transition probabilities can be estimated via the Aalen-Johansen (1978) estimator, which can be thought as the generalization of the Kaplan-Meier (1958) estimator for the simple mortality model (with states "alive" and "dead" and only one possible transition).

Next, we briefly describe nonparametric approach as follows. Let $\mathbf{I}$ be the identity matrix and $\mathbf{\Lambda}$ a matrix-valued function with element $\Lambda_{ij}(s) = \int_0^s \lambda_{ij}(u) du$, where $d\Lambda_{ii}(t) =$

$-\sum_{j\neq i} d\Lambda_{ij}(t)$. In the discrete case, there exists a set of times $\{t_k : k = 1, 2, \dots\}$, at which transition can occur and $\mathbf{P}_k = \mathbf{I} + d\mathbf{\Lambda}(t_k)$ is the usual one-step probability transition matrix of a nonhomogeneous Markov chain with element $P(X(t_k) = j|X(t_k-) = i)$. Let $\mathbf{P}^{(r)}$ denote the r-step transition probability with element $P(X(t_r) = j|X(0) = i)$, $r = 1, 2, \dots,$. It is well known that

$$\mathbf{P}^{(r)} = \prod_{i=1}^{r} \mathbf{P_i} = \mathbf{P_1 P_2 ... P_r},$$

where an empty product is interpreted as $\mathbf{I}$. Notice that the order of the multiplication matters here since in general $\mathbf{P}_k$ matrices does not commute.

In the continuous case, $d\Lambda_{ij}(t) = \lambda_{ij}(t)dt$ for all i, j, where $\lambda_{ij}(t)$ is the intensity function for $i - to - j$ transition and $\lambda_{ii}(t) = -\sum_{j\neq i} \lambda_{ij}(t)$. Similar to discrete case, we can write

$$\mathbf{P}(s, t) = \prod\nolimits_{(s,t]} \big(\mathbf{I} + \mathbf{\Lambda}(du)\big),$$

where $\prod_{(s,t]}$ is the product integral over the interval $(s, t]$ and can be defined as the limit of a product, refining the partition $s < s_1 < \cdots < s_{p+1} = t$ of $(s, t]$:

$$\lim_{\max |s_l - s_{l-1}| \to 0} \prod_l \big(\mathbf{I} + \mathbf{\Lambda}(s_l-) - \mathbf{\Lambda}(s_{l-1})\big).$$

The transition probability matrix $\mathbf{P}$ can, for a Markov process, be recovered from the Kolmogorov forward equations:

$$\mathbf{P}(s, s) = \mathbf{I} \text{ and } \frac{\partial}{\partial t}\mathbf{P}(s, t) = \mathbf{P}(s, t)\lambda(t).$$

This can also be written as follows:

$$\mathbf{P}(s, t) = \mathbf{P}(s, s) + \int_{u\in(s,t]} \frac{\partial}{\partial u}\mathbf{P}(s, u)du = \mathbf{I} + \int_{u\in(s,t]} \frac{\partial}{\partial u}\mathbf{P}(s, u)du.$$

Since $\lambda(\mathbf{u})du = \mathbf{\Lambda}(du)$, it follows from Volterra's equation that the unique solution to the above equation is $\mathbf{P}(s, t)$. The Aalen-Johansen estimator of $\mathbf{P}(s, t)$ is obtained by plugging the matrix of Nelson-Aalen estimated matrix, i.e.

$$\hat{\mathbf{P}}(s, t) = \prod\nolimits_{(s,t]} \big(\mathbf{I} + \hat{\mathbf{\Lambda}}(du)\big),$$

where $\hat{\mathbf{\Lambda}}$ is the Nelson-Aalen matrix with element $\hat{\Lambda}_{ij}(t) = \sum_{s\leq t} \hat{\Lambda}_{ij}(ds)$, where

$$\hat{\Lambda}_{ij}(t) = \int_0^t I_{[Y_i(u)>0]} \frac{dN_{ij}(u)}{Y_i(u)} = \sum_{t_l \leq t} \frac{dN_{ij}(t_l)}{Y_i(t_l)},$$

where $N_{ij}(t) = \sum_k N_{ijk}(t)$, $Y_i(t) = \sum_k Y_{ik}(t)$. t_l's are the observed times.

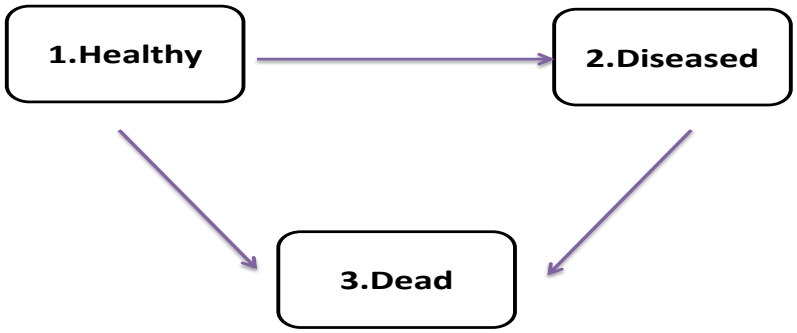

Figure 2. Schematic depiction of three-state model

A common model is the progressive three-state model (i.e. illness-death model) as shown in Figure 2.

Only three of them need to be estimated since the two other transition probabilities can be obtained from the following relations: $p_{11}(s,t) + p_{12}(s,t) + p_{13}(s,t) = 1$ and $p_{22}(s,t) + p_{23}(s,t) = 1$. Explicit formulae of the Aalen-Johansen estimator for the illness-death model are as follows:

$$\hat{p}_{11}(s,t) = \prod_{s<t_{(k)}\leq t}\left(1 - \frac{d_{12k}+d_{13k}}{n_{1k}}\right), \quad \hat{p}_{22}(s,t) = \prod_{s<t_{(k)}\leq t}\left(1 - \frac{d_{23k}}{n_{2k}}\right),$$

$$\hat{p}_{12}(s,t) = \sum_{s\leq t_{(k)}\leq t} \hat{p}_{11}(s,t_{(k-1)})\frac{d_{12k}}{n_{1k}}\hat{p}_{22}(t_{(k)},t),$$

where $t_{(1)} < t_{(2)} < \cdots < t_{(d)}$ are the event times for transitions (e.g. disease/death) arranged in increased order, n_{1k} and n_{2k} denote the number of subjects at states 1 and 2, respectively, just prior to the event time $t_{(k)}$, and d_{ijk} is the number of transition $i \to j$ at time $t_{(k)}$. Notice that the estimator $\hat{p}_{12}(s,t)$ is a plug-in estimator obtained from the following expression:

$$p_{12}(s,t) = \int_s^t p_{11}(s,u)\lambda_{12}(u)p_{22}(u,t)dt,$$

by replacing $p_{11}(s,u) = p_{11}(s,u-)$ by $\hat{p}_{11}(s,u)$, $p_{22}(u,t)$ by $\hat{p}_{22}(u,t)$ and $\lambda_{12}(u)$ by $d\hat{\Lambda}_{12}(u)$ the increment of the Nelson-Aalen estimator $\hat{\Lambda}_{12}(u) = \sum_{t_{(k)}\leq u} d_{12k}/n_{1k}$ of the cumulative disease intensity $\Lambda_{12}(t) = \int_0^t \lambda_{12}(u)du$.

2.2. Cox Markov Models (CMM)

Next, we consider heterogeneous population and for individual k there is a $p \times 1$ vector of possibly time-dependent covariates $Z_k = [Z_{1k}, \ldots, Z_{pk}]^T$. One important goal in multi-state modeling is to relate the individual characteristics to the intensity rates through a covariate vector Z_k. Several models have been used in literature. A common strategy is to decouple the whole process into various survival models, by fitting separate intensities to all permitted transitions based on some models while making appropriate adjustments to the risk set if necessary.

For individual k, let λ_{ijk} denote the intensity function for $i-to-j$ transition of individual k. Parametric or semiparametric models for λ_{ijk} can be specified, e.g. one may specify a parametric model depending on a vector of unknown parameters γ. Alternatively, one may consider the semiparametric model, e.g. Aalen's model (1980, 1989)) or Cox model (1972).

Andersen et al. (1991) developed the general theory of the "Cox Markov model" (CMM) where the intensities of the transitions from one state to the next are specified via Cox's (1972) proportional hazards regression models. Under CMM, the intensities depend only on time as measured from the origin (e.g., study entry) and not on the duration in a given state. Under CMM, given $Z_k(t)$, λ_{ijk} is written as

$$\lambda_{ijk}(t) = \lambda_{ij0}(t) \exp(Z_k^T \beta_{ij}), \tag{2.1}$$

for all i, j, k with $i \neq j$ and $t > 0$, where $\lambda_{ij0}(t)$ is an unknown baseline intensity function and β_{ij} is $p \times 1$ vector of regression parameters for $i-to-j$ transition. The CMM readily fits into the multiplicative intensity framework of Cox model. Consider a right-censored sample of n individuals from model (2.1) and define the filtration or history process as

$$\mathcal{F}_t = \{N_{ijk}(u), Z_k, Y_{ik}(u), 0 \leq u \leq t, k = 1, \ldots, n; i, j = 1, \ldots, M\},$$

where $N_{ijk}(t)$ is the right continuous process that counts the number observed direct $i-to-j$ transition for individual k and $Y_{ik}(t)$ is the corresponding at risk process, i.e. the indicator of individual k being at risk in state i just before time t. Suppose that the censoring is independent such that for all $i \neq j$, k, $\mathcal{F}_{t-}$ and $t > 0$,

$$P(dN_{ijk}(t) = 1|\mathcal{F}_{t-}) = Y_{ik}(t)\lambda_{ijk}(t),$$

where $Y_{ik}(t) = I_{[X_k(t-)=i]}$. Let $M_{ijk}(t) = N_{ijk}(t) - \int_0^t Y_{ik}(s)\lambda_{ij0}(s) \exp(Z_k(s)^T \beta_{ij}) ds$. Then, under model (2.1), $E[dM_{ijk}(t)|\mathcal{F}_{t-}] = 0$ and for $i \neq j; i, j \in E$, $M_{ijk}(t)$ are zero mean local square-integrable martingale with respect to $\mathcal{F}_t$.

Model (2.1) can be analyzed using partial likelihood arguments based on conditional probabilities of $dN_{ijk}(x)$, $k = 1, \ldots, n$ given $\{\mathcal{F}_{t-}, dN_{ij}(t), i, j \in [0, \ldots, M-1], i \neq j; t > 0\}$. Given a transition $i - to - j$ occurs at some $t \in [0, \tau]$, the contributing

$$P(dN_{ijk}(t) = 1|dN_{ij.}(t) = 1, \mathcal{F}_{t-}) = \frac{Y_{ik}(t) \exp(Z_k^T \beta_{ij})}{\sum_{l=1}^n Y_{il}(t) \exp(Z_l^T \beta_{ij})}.$$

The log partial likelihood is given by

$$\sum_{all\ i,j}\left\{\int_0^\tau \sum_{k=1}^n Z_k^T\beta_{ij} - \log\big(\sum_{l=1}^n Y_{il}(x)\exp(Z_l^T\beta_{ij})dN_{ij.}(t)\big)\right\}. \tag{2.2}$$

The parameters β_{ij}'s can be estimated by maximizing (2.2). Let $\hat{\beta}_{ij}$ denote the estimator. Since $E[dM_{ijk}(t)|\mathcal{F}_{t-}] = 0$, we have $E[dM_{ij.}(t)|\mathcal{F}_{t-}] = 0$, i.e.

$$E[dN_{ij.}(t)|\mathcal{F}_{t-}] = \lambda_{ij0}(t)\sum_{k=1}^n Y_{ik}(t)\exp(Z_k^T\beta_{ij}).$$

By letting $dN_{ij.}(t) - \lambda_{ij0}(t)\sum_{k=1}^n Y_{ik}(t)\exp(Z_k^T\beta_{ij}) = 0$, we obtain

$$\lambda_{ij0}(t) = \frac{dN_{ij.}(t)}{\sum_{k=1}^n Y_{ik}(t)\exp(Z_k^T\beta_{ij})}.$$

Given $\hat{\beta}_{ij}$, the baseline cumulative incidence function $\Lambda_{ij0}(t) = \int_0^t \lambda_{ij0}(u)du$ can be obtained using the Breslow estimator (1972,1974) $\hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, x) = \int_0^x d\hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, u)$, where

$$d\hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, t) = \frac{dN_{ij.}(t)}{\sum_{k=1}^n Y_{ik}(t)\exp(Z_k^T(t)\hat{\beta}_{ij})},$$

$dN_{ij.}(t) = \sum_{k=1}^n dN_{ijk}(t)$. Furthermore, $\Lambda_{ij}(t|Z_k)$ can be estimated by

$$\hat{\Lambda}_{ij}(t|Z_k) = \hat{\Lambda}_{ij0}(\hat{\beta}_{ij}, t)\exp(Z_k^T(t)\hat{\beta}_{ij}).$$

Given Z_k, the transition probability matrix $\mathbf{P}$ can be estimated by

$$\hat{\mathbf{P}}(s,t|Z_k) = \prod_{(s,t]}\big(\mathbf{I} + \hat{\mathbf{\Lambda}}(du|Z_k)\big),$$

where $\hat{\mathbf{\Lambda}}(u|Z_k)$ is the estimated matrix with elements $\hat{\Lambda}_{ij}(u|Z_k)$.

The asymptotic properties of the estimators $\hat{\beta}$ and $\hat{\Lambda}$ was established by Shu et al. (2007).

Remark 1:

For the bone marrow transplant example, let $\mathcal{T}$ denote the set of all possible transitions: $\mathcal{T} = \{12, 13, 15, 16, 24, 25, 26, 35, 36, 45, 46\}$. Then the 16 transition probability estimators are

$$\hat{P}_{hh}(s,t|Z_k) = \prod_{s<u\le t}\left(1 - \sum_{j>h, hj\in E} d\hat{\Lambda}_{hj}(u|Z_k)\right),\ (h = 1,2,3,4)$$

$$\hat{P}_{hj}(s,t|Z_k) = \int_s^t \hat{P}_{hh}(s, u-|Z_k)d\hat{\Lambda}_{hj}(u|Z_k),\ (hj = 35, 36, 45, 46),$$

$$\hat{P}_{2j}(s,t|Z_k)\int_s^t[\hat{P}_{22}(s,u-|Z_k)d\hat{\Lambda}_{2j}(u|Z_k)+d\hat{\Lambda}_{24}(u|Z_k)P_{4j}(s,u-|Z_k)],\ (j=5,6),$$

$$\hat{P}_{hj}(s,t|Z_k)=\int_s^t\hat{P}_{hh}(s,u-|Z_k)d\hat{\Lambda}_{hj}(u|Z_k)P_{jj}(u,t|Z_k),\ (hj=12,13,24),$$

$$\hat{P}_{14}(s,t|Z_k)=\int_s^t\hat{P}_{11}(s,u-|Z_k)d\hat{\Lambda}_{12}(u|Z_k)P_{24}(u,t|Z_k),$$

and

$$\hat{P}_{1j}(s,t|Z_k)\int_s^t[\hat{P}_{11}(s,u-|Z_k)d\hat{\Lambda}_{1j}(u|Z_k)+d\hat{\Lambda}_{12}(u|Z_k)P_{2j}(s,u-|Z_k)$$
$$+d\hat{\Lambda}_{13}(u|Z_k)P_{3j}(s,u-|Z_k)],\ (j=5,6).$$

2.3. Semiparametric Transformation Markov Models (STMM)

In some instances, the Cox model may not be suitable for describing transition rates. Semiparametric transformation models, which includes Cox's model as a special case, has been widely used in the analysis of survival data. A semiparametric transformation model (Zeng and Lin 2006) specifies the the cumulative hazard function for the survival time given Z_k takes the form

$$\Lambda_{ij}(t|Z_k)=G\Big\{\Lambda_{ij}(t)\exp(\beta_{ij}^TZ_k)\Big\},\tag{2.3}$$

where $\Lambda_{ij}(t)$ is an arbitrary baseline cumulative hazard function and G is a prespecified transformation function that is continuously differentiable and strictly increasing with $G(0)=0$ and $G(\infty)=\infty$. Cox's proportional hazards model and the proportional odds model are two special cases, corresponding to the specifications $G(t)=t$ and $G(t)=\log(1+t)$, respectively.

Note that model (2.3) is equivalent to the following model (Zeng et al. (2008)):

$$h_{ij}(T_{ijk})=-\beta_{ij}^TZ_k+\epsilon,\tag{2.4}$$

where T_{ijk} is a random variable describing a sojourn time of $X_k(t)$ in state i before transition to j, $h_{ij}(\cdot)=\log\Lambda_{ij}(t)$ is a completely unspecified strictly increasing function and the error ϵ is distributed with cumulative distribution function $P(\epsilon\le t)=F_\epsilon(t)=1-S_\epsilon(t)$, where $S_\epsilon(t)=e^{-G(e^t)}$ is a completely specified function with $\lim_{t\to-\infty}S_\epsilon(t)=1$ and $\lim_{t\to\infty}S_\epsilon(t)=0$. Note that when $S_\epsilon(t)=\exp\{-\exp(t)\}$, (2.4) gives the Cox proportional hazard model, and when $S_\epsilon(t)=(1+e^t)^{-1}$, it corresponds to the proportional odds model.

The equivalence between (2.3) and (2.4) is demonstrated as follows:

$$P(T_{ijk}>t)=P(h_{ij}(T_{ijk})>h_{ij}(t))=P(-\beta_{ij}^TZ_k+\epsilon>h_{ij}(t))$$

$$=P(\epsilon>h_{ij}(t)+\beta_{ij}^TZ_k)=S_\epsilon(h_{ij}(t)+\beta_{ij}^TZ_k)=e^{-G(e^{h_{ij}(t)+\beta_{ij}^TZ_k})}=e^{-G(\Lambda_{ij}(t)\exp(\beta_{ij}^TZ_k))}.$$

In literature, transformation models have received a lot of attention for right-censored data. Cheng et al. (1995) proposed a class of rank-based estimating equations for estimating regression parameter β_{ij}. Using martingale arguments, Chen et al. (2002) proposed an estimation procedure for estimating regression parameter β and unknown function $\log \Lambda_{ij}(\cdot)$. Zeng and Lin (2006) proposed efficient estimation of β_{ij} and $h_{ij}(t)$ using the nonparametric maximum likelihood method. Let

$$M_{ijk}(t) = N_{ijk}(t) - \int_0^t Y_{ik}(s) dG(\Lambda_{ij0}(s) \exp(Z_k^T \beta_{ij}))$$

$$= N_{ijk}(t) - \int_0^t Y_{ik}(s) d\Lambda_\epsilon(Z_k^T \beta_{ij} + h_{ij}(t)).$$

Then, under model (2.3), $E[dM_{ijk}(t)|\mathcal{F}_{t-}] = 0$ and for $i \neq j; i, j \in E$, $M_{ijk}(t)$ are zero mean local square-integrable martingale with respect to $\mathcal{F}_t$. Let $h_{ij}(t) = \log \Lambda_{ij}(t)$ and $\Lambda_\epsilon(t) = G(e^t)$. Similar to the approach of Chen et al. (2002), we consider the following two estimating equations for the estimation of β_{ij} and $\Lambda_{ij}(t)$:

$$U(\beta_{ij}, h_{ij}) = \sum_{k=1}^n \int_{-\infty}^\tau Z_k [dN_{ijk}(t) - Y_{ik}(t) d\Lambda_\epsilon(Z_k^T \beta_{ij} + h_{ij}(t))] = 0, \tag{2.5}$$

and

$$\sum_{k=1}^n [dN_{ijk}(t) - Y_{ik}(t) d\Lambda_\epsilon(Z_k^T \beta_{ij}^T + h_{ij}(t))] = 0. \tag{2.6}$$

Let $\hat{\beta}_{ij}$ and $\hat{h}_{ij}(t; \hat{\beta}_{ij})$ denote the EE estimator by solving (2.5) and (2.6). Note that $\hat{h}_{ij}(t; \hat{\beta}_{ij})$ is a step function in t that rises at the distinct jump points of $\{dN_{ijk}(t) = 1 \text{ for } k = 1, \ldots, n\}$.

Equations (2.5) and (2.6) suggest the following iterative algorithms for computing $\tilde{\beta}$ and $\hat{h}_{ij}(t; \hat{\beta}_{ij})$:

Step 0: Choose an initial value of β_{ij}, denoted by $\hat{\beta}_{ij}^{(0)}$.

Step 1: Let $t_1 < t_2 < \cdots < t_d < \tau$ denote the distinct jump points of $\{dN_{ijk}(t) = 1 \text{ for } k = 1, \ldots, n\}$. Obtain $\hat{h}_{ij}^{(0)}(t_1; \hat{\beta}_{ij}^{(0)})$ by solving

$$\sum_{k=1}^n Y_{ik}(t_1) \Lambda_\epsilon(Z_k^T \beta_{ij} + h_{ij}(t_1)) = 1,$$

with $\beta_{ij} = \hat{\beta}_{ij}^{(0)}$. Then, obtain $\hat{h}_{ij}(t_s)$ for $s = 2, \ldots, n_d$, one-by-one by solving the equation

$$\sum_{k=1}^n Y_{ik}(t_j) \Lambda_\epsilon(Z_k^T \beta_{ij} + h_{ij}(t_s)) = 1 + \sum_{k=1}^n Y_{ik}(t_s) \Lambda_\epsilon(Z_k^T \beta_{ij} + h_{ij}(t_s-)),$$

with $\beta_{ij} = \hat{\beta}_{ij}^{(0)}$.

Step 2: Obtain a new estimate of β_{ij} by solving (2.5) with $h_{ij}(t_s) = \hat{h}_{ij}^{(0)}(t_s; \hat{\beta}_{ij}^{(0)})$.

Step 3: Set $\hat{\beta}_{ij}^{(0)}$ to be the estimate obtained in Step 2 and repeat Steps 1 and 2 until prescribed convergence criteria are met.

Based on $\hat{\beta}_{ij}$ and $\hat{h}_{ij}(t; \hat{\beta}_{ij})$, we can obtain the estimated cumulative hazard function

$$\hat{\Lambda}_{G,ij}(t|Z_k) = G\Big\{\hat{\Lambda}_{ij}(t)\exp(\hat{\beta}_{ij}^T Z_k)\Big\},$$

where $\hat{\Lambda}_{ij}(t) = e^{\hat{h}_{ij}(t;\hat{\beta}_{ij})}$. Based on $\hat{\Lambda}_{G,ij}(t|Z_k)$, we can obtain the estimated transition matrix $\hat{\mathbf{P}}_G(s,t|Z_k)$.

$$\hat{\mathbf{P}}_G(s,t|Z_k) = \prod\nolimits_{(s,t]}\big(\mathbf{I} + \hat{\mathbf{\Lambda}}_G(du|Z_k)\big).$$

Next, we derive the asymptotic properties of $\hat{\mathbf{P}}_G(s,t|Z_k)$. For any vector x, let $x^{\otimes 2} = xx^T$. Let β_{ij0} and $h_{ij0}(t)$ be the true values of β_{ij} and $h_{ij}(t)$. Let $\mathcal{H}$ be the collection of all nondecreasing step functions on $[0, \tau_c]$ with $h(0) = -\infty$ and with jumps only at the observed failure times. For any two nondecreasing functions h_1 and h_2 on $[0,\tau]$ such that $h_1(0) = h_2(0) = -\infty$, define

$$d(h_1, h_2) = \sup(|\exp\{h_1(t)\} - \exp\{h_2(t)\}| : t \in [0,\tau]).$$

Similar to proposition of Chen et al. (2002), we have the following Theorem:

Theorem 1: Under model (2.3) and regularity conditions (Fleming and Harrington, 1991), we have (i) $d(\hat{h}_{ij}(t;\hat{\beta}_{ij}) - h_{ij}(t;\beta_0))$ converges almost surely to zero; (ii) $n^{\frac{1}{2}}(\hat{\beta}_{ij} - \beta_{ij0}) \to N(0, \Sigma_{\hat{\beta}_{ij}})$ in distribution, as $n \to \infty$, where $\Sigma_{\hat{\beta}_{ij}} = \Sigma_{ij2}^{-1}\Sigma_{ij1}(\Sigma_{ij2}^{-1})^T$

$$\Sigma_{ij1} = E\bigg[\int_0^\tau [Z_1 - \mu_z(t;\beta_{ij0})]^{\otimes 2}\lambda_\epsilon(h_{ij0}(t) + \beta_{ij0}^T)Y_{i1}(t)]dh_{ij0}(t)\bigg],$$

$$\Sigma_{ij2} = E\bigg[\int_0^\tau [Z_1 - \mu_z(t;\beta_{ij0})]Z_1^T\dot{\lambda}_\epsilon(h_{ij0}(t) + \beta_{ij0}^T)Y_{i1}(t)]dh_{ij0}(t)\bigg],$$

where $\lambda_\epsilon(x) = d\Lambda_\epsilon(x)/dx$,

$$\mu_z(t) = \frac{E[Z_1\lambda_\epsilon(h_{ij0}(X_{i1}) + Z_1^T\beta_{ij0}^T)Y_{i1}(t)B_{ij}(t;X_{i1})]}{E[\lambda_\epsilon(h_{ij0}(t) + Z_1^T\beta_{ij0})Y_{i1}(t)]}$$

and

$$B_{ij}(t,s) = \exp\bigg(\int_s^t \frac{E[\dot{\lambda}_\epsilon(h_{ij0}(x) + Z_1^T\beta_{ij0})Y_{i1}(x)]}{E[\lambda_\epsilon(h_{ij0}(x) + Z_1^T\beta_{ij0}Z_1)Y_{i1}(x)}dh_{ij0}(x)\bigg),$$

where $X_{ik} = \min(T_{ik}, C_k)$, The T_{ik} is sojourn time of individual k in state i.

Proof:

The proof is technical and not reported here.

Theorem 2. Let $[s, v] \in [0, \tau]$ with $s < v$. Under regularity conditions, the process $n^{1/2}(\hat{\mathbf{P}}_G(s, \cdot|Z_k) - \mathbf{P}_G(s, \cdot|Z_k))$ converges weakly on $[s, \tau]$ to a zero-mean Gaussian process.

Proof:

The proof is technical and not reported here.

2.4 Model Checking

The classes of semiparametric transformation models as shown in (2.3) require specification of the transformation function G. Misspecifying G can result in erroneous inference. For right censored data, Chen et al. (2012) introduced time-dependent martingale residuals for model (2.3) and used the cumulative sums of the residuals for model assessment. Similar to Chen et al.'s approach, we can also define martingale residual $M_{ijk}(t; \hat{\beta}_{ij}; \hat{h}_{ij})$ as follows: $M_{ijk}(t; \hat{\beta}_{ij}; \hat{h}_{ij}) = N_{ijk}(t) - \int_0^t Y_{ik}(s) d\Lambda_\epsilon(Z_k^T \hat{\beta}_{ij} + \hat{h}_{ij}(s))$. We can consider the cumulative sums of residuals over the linear predictor and the argument of the transformation function:

$$R(x; \hat{\beta}_{ij}; \hat{h}_{ij}) = n^{-1/2} \sum_{k=1}^{n} \int_0^\infty I_{[Z_k^T \hat{\beta}_{ij} \leq x]} dM_{ijk}(t; \hat{\beta}_{ij}; \hat{h}_{ij}).$$

For right censored data, Chen et al. (2012) showed that $R(x; \hat{\beta}_{ij}; \hat{h}_{ij})$ converges weakly to a zero-mean Gaussian process and calculate the p-value of a supremum test by using $r(\hat{\beta}_{ij}; \hat{h}_{ij}) = (\sup_x |R(x; \hat{\beta}_{ij}; \hat{h}_{ij})|$ and the Monte Carlo procedure.

3. Real Data Analysis

To see a real data application of the model (2.3) presented in Section 2, we consider a bone marrow transplantation (BMT) data set (Klein and Moeschberger (1997). A multicenter trial of patients prepared for transplantation with a radiation-free conditioning regimen. A total of 137 patients (99 with acute myeloctic leukemia (AML) and 38 with acute lymphoblastic leukemia (ALL)) were treated at one of four hospitals. The study consists of transplants conducted at these hospitals from March 1, 1984 to June 30, 1989. The maximum follow-up was 7 years. There were 42 patients who relapsed and 41 who died (i.e. transition $1-5$) while in remission (i.e. transition $1-6$). Twenty-six (26) patients had acute GVHD (i.e. transition $1-2$), Six-one (61) patients had chronic GVHD (i.e. transition $1-3$) and For each patient, several potential risk factors were measured at the time of transplantation. Table 1 lists all the risk factors considered. We consider four transitions: $1-to-2$, $1-to-3$ and $1-to-5$. For $1-to-2$ transition, there are 26 uncensored observations and 117 right-censored observations. For $1-to-3$ transition, there are 61 uncensored observations and 76 right-censored observations. For $1-to-5$ transition, there are 42 uncensored observations and 95 right-censored observations.

We consider the logarithm transformation $G(u) = \log(1+\rho u)/\rho, \rho \geq 0$. Notice that since

$\lim_{\rho\to 0} d\log(1+\rho u)/d\rho = \lim_{\rho\to 0} u/(1+\rho u) = u$, $\lim_{\rho\to 0}\log(1+\rho u)/\rho = u$. Thus, $\rho = 0$ corresponds to the Cox model. Furthermore, when $\rho = 1$, $\Lambda_{ij}(t|Z_k) = \log(1+\Lambda_{ij}(t)\exp(\beta_{ij}^T Z_k))$ and $S_{ij}(t|Z_k) = [1+\Lambda_{ij}(t)\exp(\beta_{ij}^T Z_k)]^{-1}$. Hence, $\rho = 1$ yields the proportional odds model. We include variables $Z_1 - Z_8$ in the model and consider the models with different values of ρ ($\rho = 0, 0.5, 1, 1.5, 2.0$). Tables 2 through 5 list the estimated coefficients for $1-to-2$, $1-to-3$, $1-to-5$ and $1-to-6$, respectively. Tables 2 through 5 indicate that the best choices for ρ are equal to 1.5, 2.0, 1.5 and 2.0, respectively for $1-to-2$, $1-to-3$, $1-to-5$ and $1-to-6$.

Table 1. Description of data from 2009 patients who underwent bone marrow transplantation

Variable Names	Description
T1	time (in days) to death or on study time
T2	disease-Free survival time (time to relapse, death or end of study)
delta1	death indicator; 1-Dead, 0-Alive
delta2	relapse indicator; 1-Relapsed, 0-Disease-Free
delta3	disease-Free survival indicator; 1-Dead or relapsed, 0-Alive disease-free
TA	time (in days) to acute GVHD
deltaA	acute GVHD indicator; 1-developed, 0-Never developed
TC	time (in days) to chronic GVHD
deltaC	chronic GVHD indicator; 1-Developed, 0-Never developed
TP	time (in days) to return of platelets to normal levels
deltaP	platelet recovery indicator; 1-returned to normal, 0-never returned to normal
Z1	disease group 1-ALL, 2-AML low-risk, 3-high-risk
Z2	patient age in years
Z3	donor age in years
Z4	patient sex; 1-Male, 2-Female
Z5	donor sex; 1-Male, 2-Female
Z6	patient CMV status; 1-CMV positive, 0-CMV negative
Z7	donor CMV status; 1-CMV positive, 0-CMV negative
Z8	FAB; 1-FAB Grade 4 or 5 and AML, 0-Otherwise
Z9	MTX used as a graft-versus-host-prophylactic; 1-Yes, 0-No

Table 2. The estimated coefficients for transition $1-to-2$

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	$r(\hat{\beta}_{ij};\hat{h}_{ij})$
$\rho=0.0$	-0.712	0.051	0.026	0.008	-0.375	-0.286	0.692	0.492	-0.524	17.85
$\rho=0.5$	-0.756	0.057	0.031	0.013	-0.405	-0.325	0.779	0.521	-0.646	15.67
$\rho=1.0$	-0.813	0.063	0.032	0.012	-0.428	-0.344	0.823	0.559	-0.697	13.52
$\rho=1.5$	-0.877	0.068	0.033	0.010	-0.457	-0.366	0.869	0.601	-0.751	11.36
$\rho=2.0$	-0.950	0.074	0.035	0.009	-0.490	-0.390	0.923	0.649	-0.815	12.08

Table 3. The estimated coefficients for transition $1-to-3$

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	$r(\hat{\beta}_{ij};\hat{h}_{ij})$
$\rho=0.0$	-0.079	-0.027	0.037	-0.038	-0.342	0.476	-0.271	-0.192	-0.013	11.56
$\rho=0.5$	-0.171	-0.033	0.039	-0.125	-0.449	0.577	-0.345	-0.233	-0.075	9.15
$\rho=1.0$	-0.199	-0.038	0.046	-0.156	-0.528	0.681	-0.406	-0.275	-0.096	8.52
$\rho=1.5$	-0.230	-0.045	0.053	-0.190	-0.615	0.792	-0.470	-0.320	-0.119	7.21
$\rho=2.0$	-0.264	-0.052	0.060	-0.228	-0.709	0.916	-0.541	-0.368	-0.146	6.86

Table 4. The estimated coefficients for transition $1-to-5$

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	$r(\hat{\beta}_{ij};\hat{h}_{ij})$
$\rho=0.0$	0.179	-0.017	0.035	-0.050	0.009	-0.015	0.046	0.569	0.453	27.21
$\rho=0.5$	0.145	-0.021	0.038	-0.127	-0.032	-0.016	0.033	0.706	0.523	23.59
$\rho=1.0$	0.162	-0.024	0.045	-0.162	-0.038	-0.029	0.035	0.854	0.630	20.34
$\rho=1.5$	0.179	-0.027	0.052	-0.200	-0.045	-0.043	0.038	1.014	0.746	16.65
$\rho=2.0$	0.197	-0.031	0.059	-0.241	-0.054	-0.057	0.043	1.190	0.872	18.51

Table 5. The estimated coefficients for transition $1-to-6$

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	$r(\hat{\beta}_{ij};\hat{h}_{ij})$
$\rho=0$	0.326	0.005	0.009	-0.146	0.411	0.193	0.097	0.908	0.497	24.72
$\rho=0.5$	0.219	-0.011	0.007	-0.280	0.315	0.251	0.067	1.021	0.512	19.03
$\rho=1.0$	0.237	-0.012	0.008	-0.313	0.344	0.277	0.071	1.149	0.579	17.62
$\rho=1.5$	0.255	-0.013	0.009	-0.349	0.375	0.306	0.073	1.288	0.649	15.91
$\rho=2.0$	0.274	-0.014	0.010	-0.388	0.411	0.338	0.077	1.445	0.728	14.59

4. Conclusions

In this article, we have pointed out that when there are covariate effects to be considered, Cox Markov regression model may not be suitable for describing transition rates. Instead, semiparametric transformation models, which includes Cox's model as a special case, can be used in the analysis of survival data. In some cases, failure times may be subject to interval censoring/truncation. Further research is required to extend semiparametric transformation Markov model to interval censored/truncated data.

References

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.

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting process. In:Klonecki, W., Kozek, A., Rosiski, J. (Eds). Lecture Notes on Mathematical Statistics and Probability, Vol. 2. Springer, New York, pp. 1-25.

Aalen, O. O. (1989). A linear regression model for analysis of life times. *Statistics in Medicine*, **8**, 907-925.

Aalen, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine*, **17**, 1569-1588.

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

Alioum, A. and Commenges, D. (2001). MKVPCI: a computer program for Markov models with piecewise constant intensities and covariates. *Computer Methodsand Programs in Biomedicine*, **64**, 109-119.

Andersen, P.K. (1988). Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine*, **7**, 661-670.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Methods Based on Counting Processes.* New York: Springer.

Andersen, P. K., Hansen, L. S. and Keiding, N. (1991). Non- and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process. *Scandinavian Journal of Statistics*, **18**, 153-167.

Andersen, P. K., Esbjerg, S. and Sorensen, T. I. A. (2000). Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine*, **19**, 587-599.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression. Chichester: Wiley.

Begg, C. B. and Larson, M. (1982). A study of the use of the probability-of-being-in-response function as a summary of tumor response data. *Biometrics*, **38**, 59-66.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, **2**, 273-277.

Breslow, N. E. (1972). Discussion following "Regression models and life tables" by D. R. Cox. *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.

Breslow, N. E. and Crowley, J. (1974). A large-sample study of the life table and product limit estimates under random censorship. *Annals of Statistics*, **2**, 437-454.

Cai, T., Cheng, S. C. AND Wei, L. J. (2002). Semiparametric mixed-effects models for clustered failure time data. *Journal of the American Statistical Association*, **97**, 514-522.

Cai, T. and Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika*, **91**, 277-290.

Chen, K., Jin, Z and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, **89**, 659-668.

Chen, L., Lin, D. Y. and Zeng, D. (2012). Checking semiparametric transformation models with censored data. *Biostatistics*, **13**, 18-31.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, **82**, 835-845.

Commenges D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis*, 315-327.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of Royal Statistical Society, Ser.B*, **34**, 187-220.

Fleming, T. R. and Harrington, D. P. (1991). Counting Processes and Survival Analysis. Wiley, New York

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

Hougaard P. (1999). Multi-state models: a review. *Lifetime Data Analysis*, 239-264.

Huffer, F. W. and McKeague, I.W . (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, **86**, 114-129.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observation. *Journal of the American Statistical Association*, **53**, 457-481.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, **80**, 863-871.

Keiding, N., Klein, J. P. and Horowitz, M. M. (2001). Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine*, **20**, 1871-1885.

Klein, J. P. and Moeschberger, M. L. (1997). Survival Analysis: Techniques for Censored and Truncated Data. New York: Springer-Verlag.

Klein, J. P. and Qian, C. (1996). Modelling multistate survival illustrated in bone marrow transplantation. In Proceeding Biometrics Section American Statistical Association, pp. 93-102. Alexandria, VA: American Statistical Association.

Klein, J. P., Keiding, N. and Copelan, E. A. (1993). Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine*, **12**, 2315-2332.

Lin, D. Y. and Ying, Zhiliang (1994). Semiparametric Analysis of the Additive Risk Model. *Biometrika*, **81**, 61-71.

Marshall, G and Jones, R. H. (1995). Multi-state models and diabetic retinopathy. *Statistics in Medicine*, **14**, 1975-1983.

Meira-Machado, L.; de de Uña-Álvarez, J.; Cadarso-Suárez, C. and Andersen, P.K. (2009). Multi-state models for the analysis of time to event data, *Statistical Methods in Medical Research*, **18**, 195-222.

Perez-Ocon, R., Ruiz-Castro, J.E., Gamiz-Perez, M.L. (2001). A piecewise Markov process for analysing survival from breast cancer in different risk groups. *Statistics in Medicine*, **20**, 109-122.

Pettitt, A. N. (1984). Proportional odds model for survival data and estimates using ranks. *Journal of the Royal Statistical Society*, *Series C* **33**, 169-175.

Shu, Y. and Klein J. P. (2005). Additive hazards Markov regression models illustrated with bone marrow transplant data. *Biometrika*, 283-301.

Shu, Y., Klein, J. P. and Zhang, M. J. (2007). Asymptotic theory for the Cox semi-Markov illness-death model. *Lifetime Data Analysis*, **13**, 91-117.

Yang, S and Prentice, R. (1999). Semiparametric inference in the proportional odds regression model. *Journal of the American Statistical Association*, **94**, 125-136.

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93**, 627-640.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression model with censored data. *Journal of the Royal Statistical Society*, *Series B*, **69**, 507-564.

Zeng, D., Lin, D. Y. and Lin, X. (2008). Semiparametirc transformation modes with random effects for clustered data. *Statis. Sinica*, **18**, 355-377.

Zeng D, Lin DY (2010). A general theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statistica Sinica*, **20**, 871-910.