The International Journal of Biostatistics

Volume 8, *Issue* 1 2012 *Article* 15

Estimation in a Semi-Markov Transformation Model

Dorota M. Dabrowska, *University of California*

Recommended Citation:
Dabrowska, Dorota M. (2012) "Estimation in a Semi-Markov Transformation Model," *The International Journal of Biostatistics*: Vol. 8: Iss. 1, Article 15.
DOI: 10.1515/1557-4679.1233

Estimation in a Semi-Markov Transformation Model

Dorota M. Dabrowska

Abstract

Semi-Markov and modulated renewal processes provide a large class of multi-state models which can be used for analysis of longitudinal failure time data. In biomedical applications, models of this kind are often used to describe evolution of a disease and assume that patient may move among a finite number of states representing different phases in the disease progression. Several authors proposed extensions of the proportional hazard model for regression analysis of these processes. In this paper, we consider a general class of censored semi-Markov and modulated renewal processes and propose use of transformation models for their analysis. Special cases include modulated renewal processes with interarrival times specified using transformation models, and semi-Markov processes with with one-step transition probabilities defined using copula-transformation models. We discuss estimation of finite and infinite dimensional parameters and develop an extension of the Gaussian multiplier method for setting confidence bands for transition probabilities and related parameters. A transplant outcome data set from the Center for International Blood and Marrow Transplant Research is used for illustrative purposes.

KEYWORDS: modulated renewal process, transformation models, M-estimation, U-processes

Author Notes: The data presented here were obtained from the Statistical Center of the Center for International Blood and Marrow Transplant Research (CIBMTR). The analysis has not been reviewed or approved by the Advisory or Scientific Committee of the CIBMTR. The CIBMTR is comprised of clinical and basic scientists who confidentially share data on their blood and marrow transplant patients with the CIBMTR Data Collection Center located in the Medical College, Wisconsin. The CIBMTR is a repository of information about results of transplant at more than 450 transplant centers worldwide. I thank Mei-Jie Zhang for preparation of the data and some discussions. I also appreciate comments of a reviewer and Editor Daniel Commenges. Research supported by the grant R01 AI067943 from the National Institute of Allergy and Infectious Diseases. The content is solely the responsibility of the author and does not necessarily represent the official views of NIAID, NIH or CIBMTR.

1 Introduction

We consider estimation in a semi-Markov regression model with a finite state space $\mathscr{J} = \{1,\ldots,r\}$. In the absence of covariates, the model can be described by a sequence $(T,J) = \{(T_n,J_n) : n \geq 0\}$, where $T_0 < T_1 < T_2 \ldots$ are consecutive times of entrances into the states $J_0,J_1,J_2,\ldots,J_n \in \mathscr{J} = \{1,\ldots,r\}$. The sequence $J = \{J_n : n \geq 0\}$ of states visited forms a Markov chain and given J, the sojourn times $T_1, T_2 - T_1, \ldots$ are independent with distributions depending on the adjoining states only. Alternatively, the distribution of the sojourn times $T_{n+1} - T_n, n \geq 0$ satisfies

$$\begin{aligned}
&P(T_{n+1} - T_n \leq x \;\; J_{n+1} = j | J_0, T_0, J_1, T_1, \ldots, J_n, T_n) \\
&= P(T_{n+1} - T_n \leq x \;\; J_{n+1} = j | J_n) \; .
\end{aligned}$$

Properties of semi-Markov processes were discussed in some detail in classical papers of Pyke (1961,ab), Pyke and Schaufele (1964,1966), and textbooks of Cinlar (1975), Daley and Vere-Jones (1988), Karr (1991), Last and Brandt (1995) and Limnios and Oprisan (2001). Numerous examples of applications to areas such as reliability, insurance and finance were provided by Janssen (1999), Janssen and Manca (2006,2007) and Janssen and Limnios (2001), for instance. In such studies, it is most common to consider estimation methods assuming that a single realization of a semi-Markov process is observed over a finite time interval $[0,\tau]$ whose length tends to infinity ($\tau \uparrow \infty$). Greenwood and Wefelmeyer (1996) and Greenwood, Müller and Wefelmeyer (2004) developed a general framework for analysis of non- and semi-parametric semi-Markov processes in this setting. In particular, they studied properties of classical estimators of the jump frequency and the proportion of visits to a given state, as well as Moore and Pyke's (1968) non-parametric estimator of the kernel of the process. Estimation of transition intensities and transition probabilities was considered by Ouhbi and Limnios (1996,1999).

In survival analysis, it is more common to consider estimation based on a large number of iid copies of a semi-Markov process observed over a deterministic or random time intervals. Lagakos, Sommer and Zelen (1978), Gill (1980), Voelkel and Crowley (1984) and Phelan (1999) developed nonparametric estimators of the semi-Markov kernel of the process in the presence of random censoring. Examples of applications of these processes to analysis of survival data can be found in Commenges (1986), Keiding (1986), Dabrowska *et al.* (1994), Chang *et al.* (1994, 1999,2000), Cook and Lawless (2007), among others.

In this paper, we assume that the evolution of the process $(T_m,J_m)_{m\geq 0}$ depends also on an R^d-valued covariate $(Z_m)_{m\geq 0}$, $Z_m = [Z_{jm} : j \in \mathscr{J}]$, which represents either a vector of time independent covariates, or a vector of time dependent covariates changing at the successive renewal times. As an extension of the semi-

Markov process to the regression setting, Cox (1973) proposed to consider a proportional hazards modulated renewal process. More precisely, let $\widetilde{N} = \{\widetilde{N}_j(t) : t \geq 0, j = (j_1, j_2) \in \mathscr{J} \times \mathscr{J}\}$ be the counting process registering transitions among adjoining states of the model,

$$\widetilde{N}_j(t) = \sum_{m \geq 0} 1(T_{m+1} \leq t, J_{m+1} = j_2, J_m = j_1).$$

Cox's model assumes that the compensator of this process, relative to the self-exciting filtration $\{\mathscr{F}_t\}_{t\geq 0}$, is given by $\Lambda_j(0) = 0$,

$$\Lambda_j(t) = \Lambda_j(T_m) + \int_0^{t-T_m} 1(J_m = j_1) e^{\beta^T Z_{j_1 m}} \Gamma_j(du)$$

for $t \in (T_m, T_{m+1}]$ and $j = (j_1, j_2) \in \mathscr{J} \times \mathscr{J}$. Here β is a regression coefficient and Γ_j in an unknown cumulative hazard function. If covariates are time independent and $\Gamma_j(x) = \gamma_j x$, the process reduces to a Markov chain regression model. In the general case, the modulated renewal process allows to incorporate dependence of the history on the sequence of states visited and the length of time spent in each state. As a result of this, it has a more flexible structure than Markov chains.

The purpose of this paper is to extend Cox's modulated process to a class of transformation models. In the case of single spell models, they provide a common alternative to the proportional hazard model. In particular, they may be more appropriate than the proportional hazard model if relative differences between covariates dissipate or diverge over time. As an extension to multistate models, we consider here a modulated renewal process assuming that the counting process $\widetilde{N}$ has compensator given by $\Lambda_j(0) = 0$,

$$\Lambda_j(t) = \Lambda_j(T_m) + \int_0^{t-T_m} 1(J_m = j_1) \alpha_j(\Gamma_{(j_1,.)}(u), \theta, Z_{j_1 m}) \Gamma_j(du) \tag{1.1}$$

for $t \in (T_m, T_{m+1}]$ and $j = (j_1, j_2) \in \mathscr{J} \times \mathscr{J}$. For any such pair $j = (j_1, j_2)$, α_j is a hazard function dependent on an unknown Euclidean parameter θ and a vector of unknown increasing functions $\Gamma_{(j_1,.)} = [\Gamma_j(x) = \int_0^x \gamma_j(u) du : j = (j_1, j_2) \in \mathscr{J} \times \mathscr{J}, x \geq 0]$. The components of $\Gamma_{(j_1,.)}$ depend on all states which can be reached from the state j_1 in one step. If covariates are time independent, then (1.1) includes as a special case renewal processes whose interarrival times satisfy common transformation models. Other choices include semi-Markov models with one-step transition probabilities defined using copula graphic models (e.g. Zheng and Klein (1995), Rivest and Wells (2001), Lo and Wilke (2010)) or extensions of the dynamic Cox-McFadden's model (Chintagunta and Prasad (1998)) combining transformation models and multinomial regression. These models are defined in more detail in Section 2, where covariates are also allowed to change at the renewal times of the process.

For purposes of estimation, we consider a modification of procedures studied by Bagdonovicius and Nikulin (1999,2004) and Dabrowska (2006) in the case of single spell transformation models. Section 3 provides properties of the estimates as well as an extension of the Gaussian multiplier method of Lin *et al.* (1994) for setting pointwise and simultaneous confidence bands for the unknown transformations and related parameters. In analogy to Cox's model, the counting process $\widetilde{N}$ has a compensator depending on the backwards recurrence time and as a result of this, it falls outside the class of multiplicative models studied by Andersen *et al.* (1993), for instance. In the case of Cox's modulated renewal process or nonparametric semi-Markov models, estimation of the cumulative hazards of one-step transitions leads to a time transformation which arranges observations according to the length of time spent in each state rather than calendar time. As a result of the rearrangement of the time scale, usual counting process methods for analysis of large sample properties of stochastic integrals do not apply (Gill (1980), Oakes (1981), Oakes and Cui (1994)). To alleviate these problems, we use Hoeffding's projection method and empirical processes in Section 5.

In Section 4, we consider a transplant outcome data set from the Center for International Blood and Marrow Transplant Research (CIBMTR). The example data set consists of patients who received HLA-identical sibling transplant from 1995 to 2004 for acute myelogenous leukemia (AML) or acute lymphoblastic leukemia (ALL). Multistate models for analysis of the bone marrow transplant recovery process have been proposed by several authors. The early work in this area focused on competing risk models and goes back Prentice *et al.* (1978) who discussed estimation of cause specific cumulative hazards in the proportional hazard model. More recent approaches towards analysis of leukemia transplant data are based on multistate models. They provide a convenient tool for evaluation of the impact of intermediate events in the transplant recovery process on the main outcome events corresponding to leukemia relapse and death in remission. However, analysis of multistate regression models leads to some difficulties in the interpretation of the results because there is no one-to-one correspondence between regression coefficients and transition probabilities. Each covariate may increase the risk of transition among some states of the model and at the same time decrease it among the others. Correspondingly, its overall impact on the outcome events is often not clear. To obviate difficulties, Arjas and Eerola (1993) and Eerola (1994) proposed a set of graphical tools which can be used for purposes of interpretation of regression analyzes based on multistate models. These included graphs of innovation gains and plots of the transition probabilities evaluated by conditioning on the follow-up history of a patient. The approach was illustrated using a proportional hazard model with time dependent covariates in Eerola (1994). Applications of these methods to proportional hazard Markov chain models were given in Klein *et al.* (1993) and Keiding *et al.* (2001) and Andersen and Parme (2008), and proportional hazard semi-Markov models in Dabrowska *et al.* (1993, 2006). Putter *et al.* (2007) discussed special cases of both models.

In this paper, we consider a data set involving patients who received either bone marrow (BMT) or peripheral blood stem cell transplant (PBSCT). Many clinical studies have reported that PBSCT may be beneficial during the early post-transplant period as it leads to faster engraftment and hematopoietic recovery than BMT (e.g. Flowers *et al.* 2002, Ringden *et al.* 2002). Several studies have also pointed out that differences between the two transplant types may dissipate over time (e.g. Friedrichs *et al.* 2010, Cutler *et al.* 2002ab). Such dissipating time effects are better captured by the proportional odds ratio model than the proportional hazard model, and in Section 5 we discuss an extension of it to semi-Markov models. In this section we also propose pointwise and simultaneous confidence bands for comparison of transition probabilities.

2 The model

Throughout the paper we assume that $(\Omega, \mathscr{F}, P)$ is a complete probability space and $(T_m, V_m)_{m \geq 0}$ is a marked point process defined on it with marks taking on values in a separable measure space $(E, \mathscr{E})$ and enlarged by the empty mark Δ. Thus $T_0 < T_1 < \dots T_m \dots$ is a sequence of random time points registering occurrence of some events in time such that T_m are almost surely distinct and $T_m \uparrow \infty$ P-a.s. At time T_m we observe a variable V_m such that $V_m \in E$ if $T_m < \infty$, and $V_m = \Delta$ if $T_m = \infty$.

For any $B \in \mathscr{E}$, let $\widetilde{N}(t, B) = \sum_{m \geq 0} 1(T_{m+1} \leq t, V_{m+1} \in B)$ be the process counting observations falling into the set $[0, t] \times B$. The internal history of the process, $\{\mathscr{F}_t^N\}_{t \geq 0}$, represents information collected on $\widetilde{N}$ until time t, and is given by $\mathscr{F}_t^N = \sigma(1(T_m \leq s, V_m \in B) : m \geq 1, s \leq t, B \in \mathscr{E}) \vee \sigma(V_0)$. Let $\mathscr{F}_t = \mathscr{N} \vee \mathscr{F}_t^N$ be the self-exciting filtration associated with the process $\widetilde{N}$, obtained by adjoining the P-null sets to the internal history of the process. The compensator of the process $\widetilde{N}$ with respect to $\mathscr{F}_t$ is given by

$$\widetilde{\Lambda}(t, B) = \widetilde{\Lambda}(T_m, B) + \int_{(T_m, t] \times B} \frac{P_m(d(s, v))}{P_m([s, \infty); E \cup \Delta)} \quad \text{for} \quad t \in (T_m, T_{m+1}] \, ,$$

where $P_m(d(s, v))$ is a version of a regular conditional distribution of (T_{m+1}, V_{m+1}) given $\mathscr{F}_{T_m}$ (Jacod (1975)).

In this paper we assume that the marks V_m have the form $V_m = (J_m, \widetilde{Z}_m)$, where $J_m \in \mathscr{J} = \{1, \dots, r\}$ is a discrete variable representing the type of the event occurring at time T_m and $\widetilde{Z}_m$ are covariates taking on value in R^d. The covariate $\widetilde{Z}_m$ may correspond to some measurements taken upon entrance into the state J_m. The process $\widetilde{N} = [\widetilde{N}_j, j = (j_1, j_2) \in \mathscr{J} \times \mathscr{J}]$,

$$\widetilde{N}_j(t,B) = \sum_{m \geq 0} 1(T_{m+1} \leq t, J_{m+1} = j_2, J_m = j_1, \widetilde{Z}_{m+1} \in B),$$

has compensator given by

$$\begin{aligned}\widetilde{\Lambda}_j(t,B) &= \widetilde{\Lambda}_j(T_m,B)\\ &+ \int_0^{t-T_m} \mu_{m+1}(B, u+T_m, j) 1(J_m = j_1) \alpha_j(\Gamma_{(j_1,.)}(u), \theta, Z_{j_1 m}) \Gamma_j(du),\end{aligned}$$

for $t \in (T_m, T_{m+1}]$. Here $\mu_{m+1}(B, T_{m+1}, J_m, J_{m+1})$ is the conditional probability of the event $\{\widetilde{Z}_{m+1} \in B\}$ given $\sigma(\mathscr{F}_{T_m}, T_{m+1}, J_{m+1})$.

Further, $Z_{j_1 m} = g_{j_1 m}(T_l, J_l, \widetilde{Z}_l : l = 0, \dots, m)$ is a fixed R^d valued function, measurable with respect to $\mathscr{F}_{T_m}$. Finally, α_j denotes a hazard rate dependent on a Euclidean parameter θ and a vector of unknown monotone increasing functions $\Gamma_{(j_1,.)} = [\Gamma_j : j = (j_1, j_2) \in \mathscr{J} \times \mathscr{J}]$. In particular, setting $B = R^d$ and using $\mu_{m+1}(R^d, T_{m+1}, J_m, J_{m+1}) 1(T_{m+1} < \infty) = 1$ P-a.s., $\widetilde{\Lambda}_j(t, R^d)$ reduces to (1.1) and represents the compensator of the “marginal” counting process

$$\widetilde{N}_j(t) = \widetilde{N}_j(t, R^d) = \sum_{m \geq 0} 1(T_{m+1} \leq t, J_{m+1} = j_2, J_m = j_1) \tag{2.1}$$

registering transitions among the adjoining states of the model.

To give examples of the model, we assume first that the covariates are time independent. If events are of a single type ($|\mathscr{J}| = 1$), then (1.1) represents compensator of a renewal regression model assuming that the interarrival times follow a transformation model. Thus in this case $\{\alpha(u, \theta, Z) : \theta \in \Theta\}$ is a parametric family of hazard rates, and the model stipulates that conditionally on Z, the interarrival times, X_{m+1} are independent and their conditional survival function has cumulative hazard function $A(\Gamma(x), \theta, Z)$.

Simple examples of multi-type processes are given by competing risk and semi-Markov regression models. In particular, a semi-Markov regression model assumes that one-step transition probabilities satisfy

$$P(X_{m+1} \leq x, J_{m+1} = j_2 | (T_\ell, J_\ell)_{\ell=0}^m, Z) = P(X_{m+1} \leq x, J_{m+1} = j_2 | J_m, Z).$$

The matrix $[F_j, j=(j_1,j_2)\in \mathscr{J}\times\mathscr{J}]$,

$$F_j(x|Z) = P(X_{m+1}\le x, J_{m+1}=j_2|J_m=j_1,Z),$$

forms the kernel of the process. One way to define it is to consider latent variable models. Specifically, suppose that transitions originating from the state j_1 have the same conditional distribution as the pair (U,V), where

$$\begin{aligned} U &= \min[U_j : j=(j_1,j_2)\in \mathscr{J}\times\mathscr{J}],\\ V &= [1(U=U_j) : j=(j_1,j_2)\in \mathscr{J}\times\mathscr{J}],\end{aligned}$$

and $[U_j : j=(j_1,j_2)\in \mathscr{J}\times\mathscr{J}]$ is a multivariate vector whose joint conditional survival function given Z is

$$S_{(j_1,.)}(u,\theta,z) = S^0_{(j_1,.)}([\Gamma_j(u_j)e^{\theta_j^T z} : j=(j_1,j_2)\in\mathscr{J}\times\mathscr{J}]).$$

Here $u=[u_j, j=(j_1,j_2)\in\mathscr{J}\times\mathscr{J}]$ and $S^0_{(j_1,.)}$ is a known multivariate survival function with a density with respect to Lebesgue measure supported on the entire upper orthant of $R^{q_{j_1}}$, $q_{j_1}=|\{j_2:(j_1,j_2)\in\mathscr{J}\times\mathscr{J}\}|$. The functions α_j in (1.1) are equal to

$$-\frac{\partial}{\partial y_j}\log S^0_{(j_1,.)}([y_j e^{\theta_j^T z} : j=(j_1,j_2)\in\mathscr{J}\times\mathscr{J}]).$$

With this choice the cumulative intensity (1.1) corresponds to a semi-Markov model whose kernel is given by

$$\begin{aligned} F_j(x|Z) &= P(X_{m+1}\le x, J_{m+1}=j_2|J_m=j_1,Z) \\ &= \int_0^x \overline{F}_{(j_1,.)}(u|Z)\alpha_j(\Gamma_{(j_1.)}(u),\theta,Z)\Gamma_j(du), \end{aligned} \tag{2.2}$$

where $j=(j_1,j_2)\in\mathscr{J}\times\mathscr{J}$ and $\overline{F}_{(j_1,.)}(x|z)$ is the survival function of the sojourn time in state j_1,

$$\begin{aligned} &\overline{F}_{(j_1,.)}(x|z) = P(X_{m+1}>x|J_m=j_1,z) = \\ &= \exp\big[-\sum_{j=(j_1,j_2)}\int_0^x \alpha_j(\Gamma_{(j_1,.)}(u),\theta,z)\Gamma_j(du)\big] \\ &= S_{(j_1,.)}(\Gamma_{(j_1,.)}(x),\theta,z). \end{aligned} \tag{2.3}$$

If the state space of the process consists of one ephemeral state ($J_0 = 1$, say) and $q-1$ absorbing states, $q \geq 3$, then the semi-Markov process reduces to a competing risk model. In this case transition probabilities (2.2) provide a regression analogue of copula-graphic models proposed for analysis of competing risks by Zhang and Klein (1995) and Rivest and Wells (2001). The special case of Archimedean copula models corresponds to the choice $S^{(0)}_{(j_1,.)}(y_{(j_1,.)}) = \overline{S}(\|y_{(j_1,.)}\|_1)$, where $\overline{S}$ is a known survival function with a density supported on the positive half-line and $\|\cdot\|_1$ is the ℓ_1-norm of a vector.

Another example of a semi-Markov model is provided by the dynamic Cox-McFadden model (Chintagunta and Prasad, 1998). In this case, the distribution of the sojourn time in state $j_1 \in \mathscr{J}$ is specified by means of a transformation model for univariate failure time data, i.e. the survival function (2.3) is of the form $\overline{F}_{(j_1,.)}(x|z) = \exp[-\widetilde{A}_{j_1}(\Gamma_{j_1}(x), \theta_1, z)]$ for some univariate cumulative hazard function $\widetilde{A}_{j_1}$. The kernel of the process is given by

$$F_j(x|z) = \int_0^x \pi_j(u, z, \theta_2) F_{(j_1,.)}(du|z),$$

where $F_{(j_1,.)}(\cdot|z) = 1 - \overline{F}_{(j_1,.)}(\cdot|z)$ and for $j = (j_1, j_2)$,

$$\pi_j(X_{m+1},, Z, \theta_2) = P(J_{m+1} = j_2 | X_{m+1}, J_m = j_1, Z) \tag{2.4}$$

are the one-step state transition probabilities. The state transition probabilities can be specified using multinomial regression models such as the logistic or probit model. If the state transition probabilities (2.4) do not depend on the length of the sojourn time X_{m+1}, the model reduces to a stationary process, i.e. conditionally on Z, the transition probabilities do not depend on m.

In practice, the assumptions of the semi-Markov process may be violated if transitions from a state j_1 to a state j_2 depend on the sequence or the time spent in states visited prior to the entrance into the state j_1. Both models can accommodate this problem by allowing the covariates to depend on the internal history of the process. The time dependent covariates may represent for instance the total number of events occurring prior to the entrance into the state j_2 or the length of time spent in states preceding entrance into the state j_1. The time dependent covariates may also represent changing treatment types or levels of drugs.

We further assume that the process is subject to censoring and times at which the process is observed is determined by a process $C(t) = \sum_{m \geq 1} 1(C_{m-1} < t \leq C_m)$, where $0 \leq C_0 \leq C_1 \leq \ldots \leq C_m \ldots$ is an increasing sequence such that $C_m \in [T_m, T_{m+1}]$ are stopping times with respect to a larger filtration $\{\mathscr{H}_t\}_{t \geq 0}, \mathscr{F}_t \subseteq \mathscr{H}_t$. If $T_m = C_m$ then no information is available on either the sojourn time $X_{m+1} = T_{m+1} - T_m$ or the marks (V_m, V_{m+1}). If $C_m = T_{m+1}$ then the sojourn time $X_{m+1} = T_{m+1} - T_m$ and the marks (V_m, V_{m+1}) are observable. Finally, if $T_m < C_m < T_{m+1}$ then the mark V_m is visible while the sojourn time X_{m+1} is only known to exceed $C_m - T_m$. Following Andersen *et al.* (1993), we assume that the compensator $\Lambda^{\mathscr{H}}$, of the marked point process $\widetilde{N}$, relative to the filtration $\{\mathscr{H}\}_{t\geq 0}$, satisfies $\Lambda^{\mathscr{H}} = \Lambda$, P-a.s. and that the censoring process and the compensator Λ depend on parameters which do not share components in common. We also make the assumption that the censoring process is monotone so that with probability 1, $T_m \leq C_m < T_{m+1} \Rightarrow C_{m'} = T_{m'}$ for all $m' > m$. This condition stipulates that the process terminates once censoring takes place.

These conditions are satisfied in two common applications. The first assumes that the process is subject to censoring by a univariate failure time T' such that T' is independent of the the sequence (T_m, V_m), conditionally on the initial state of the process, V_0. In this case, $C_m = T_m + \min(T' - T_m, X_{m+1})1(T' \geq T_m)$ and the augmented filtration is given by $\mathscr{H}_t = \mathscr{F}_t \vee \sigma(T')$.

The second example assumes that the state space of the process has an extra absorbing state corresponding censoring, say $\{c\}$, which can be reached in one step from each transient state $j_1 \in \mathscr{J}$. Time T till entrance into the censoring state forms then stopping time with respect to the filtration $\mathscr{H}_t = \mathscr{F}_t$. Consequently, there exist nonnegative variables U_m such that on the event $\{T \geq T_m\}$, we have $T \wedge T_{m+1} = (T_m + U_m) \wedge T_{m+1}$, and U_m is measurable with respect to $\mathscr{F}_{T_m}$. Correspondingly, $C_m = T_m + \min(U_m, X_{m+1})1(T \geq T_m)$. In this setting, the assumption of non-informative censoring means that the compensators of one-step transitions into the the censoring state depend on different parameters than the compensator of transitions among the remaining states of the model.

Let $\mathscr{J}_0 \subset \mathscr{J} \times \mathscr{J}$ be the set of pairs of adjacent states in the model, i.e. $j = (j_1, j_2) \in \mathscr{J}_0$ iff the subject may progress from state j_1 to state j_2 in one step. For $j = (j_1, j_2) \in \mathscr{J}_0$ and $m \geq 0$, let $N_{jm}(x) = 1(X_{m+1} \leq x, J_m = j_1, J_{m+1} = j_2, T_m = C_{m+1})$, $Y_{jm}(v) = 1(X_{m+1} \geq x, C_m - T_m \geq x, J_m = j_1)$ and set

$$
\begin{aligned}
M_{jm}(x,\theta) &= N_{jm}(x) - \Lambda_{jm}(x,\theta), \\
\Lambda_{jm}(x,\theta) &= \int_0^x Y_{jm}(u)\alpha_j(\Gamma_{j_1,\cdot}(u), Z_{j_1 m}, \theta)\Gamma_j(du) \,.
\end{aligned}
$$

The aggregate processes $N_{j.}, Y_{j.}$ and $M_{j.}$ are defined as $N_{j.} = \sum_m N_{jm}$, $Y_{j.} = \sum_m Y_{jm}$ and $M_{j.} = \sum_m M_{jm}$, respectively.

Note that the model depends on two parameters, θ and Γ, however, we suppress the dependence on Γ in the notation. In analogy to single spell models in Bagdonovicius and Nikulin (1999,2004) and Dabrowska (2006), under regularity conditions stated in Section 5, we can associate, with any $\theta \in \Theta$, a vector Γ_θ of locally bounded increasing functions. For this purpose, we shall require only that the processes $N_{j.}$ and $Y_{j.}$ have a finite expectation. To show asymptotic normality of estimates we shall require existence of the second moments of these processes. More precisely, we assume the following conditions.

Condition 2.1 For all $j \in \mathscr{J}_0$

(i) The functions $EY_{j.}(x)$ have at most a finite number of discontinuity points and $EY_{j.}(0)^2 < \infty$.
(ii) The functions $EN_{j.}(x)$ are continuous, $EN_{j.}(\tau)^2 < \infty$ and the point τ satisfies $\inf\{x : EN_{j.}(x) > 0\} < \tau < \tau_{j0}$, where $\tau_{j0} = \sup\{x : EY_{j.}(x) > 0\}$.
(iii) We have $P(|Z_{J(t-),\widetilde{N}_{..}(t-)}| \leq C) = 1$, where C is a finite constant, $J(t)$ is the state occupied by the process at time t and $\widetilde{N}_{..}(t) = \Sigma_j \widetilde{N}_{j.}(t)$ is the total number of events observed in the interval $[0,t]$.

Under the added assumption that the model corresponds to the censored modulated renewal process, and θ represents the true parameter, we have the following moment identities.

Lemma 2.1 Let $L(t) = t - T_{\widetilde{N}_{..}(t-)}$ be the backwards time of the process $\widetilde{N}$ and let $\{\varphi_m(x), m \geq 0, x \geq 0\}$ be a sequence of random functions such that the process $\varphi \circ L$, $\varphi \circ L(t) = \varphi_{\widetilde{N}_{..}(t-)}(t - T_{\widetilde{N}_{..}(t-)})$, is predictable with respect to the filtration $\{\mathscr{H}_t\}_{t \geq 0}$ and $E\int_0^\infty [\varphi \circ L]^2(s)\widetilde{\Lambda}_j(ds, \theta) < \infty$. Then

$$E\sum_m \int_0^\infty \varphi_m(u) N_{jm}(du) = E\sum_m \int_0^\infty \varphi_m(u)\Lambda_{jm}(du, \theta),$$
$$E[\sum_m \int_0^\infty \varphi_m(u) M_{jm}(du, \theta)]^2 = E\sum_m \int_0^\infty \varphi_m^2(u)\Lambda_{jm}(du, \theta).$$

In addition, if $\{\varphi_{1m} : m \geq 0\}$ and $\{\varphi_{2m} : m \geq 0\}$ are two such sequences, then

$$E[\sum_m \int_0^\infty \varphi_{1m}(u) M_{jm}(du, \theta)][\sum_m \int_0^\infty \varphi_{2m}(u) M_{j'm}(du, \theta)] = 0$$

for pairs $j \neq j'$, $j, j' \in \mathscr{J}_0$.

Similarly to Gill (1980), this lemma follows from the dominated convergence theorem, martingale properties of the processes $\widetilde{M}_j = \widetilde{N}_j(t) - \widetilde{\Lambda}_j(t)$, and the identities

$$\int_0^\infty [\varphi \circ L]^k(s) C(s) \widetilde{N}_j(ds) \quad = \quad \sum_{m \geq 0} \int_0^\infty \varphi_m^k(u) N_{jm}(du),$$

$$\int_0^\infty [\varphi \circ L]^k(s) C(s) \widetilde{\Lambda}_j(ds, \theta) = \sum_{m \geq 0} \int_0^\infty \varphi_m^k(u) \Lambda_{jm}(du, \theta).$$

The identities hold almost surely for $k = 1, 2$. We omit the details.

3 Estimation

Throughout the remainder of this paper, we assume that we have an iid sample of size n of the censored modulated renewal process and covariates. The subscript ”i" refers to the i-th subject under study and D_i represents the associated vector of observations. It corresponds to the sequence of states visited, duration of the time spent in each state, the initial covariate and its updates occurring at uncensored renewal times.

Further, let $q = |\mathscr{J}_0|$ be the total number of possible one-step transitions in the model. For each $j = 1, \ldots, q$, we let $(r(j), c(j)) = (j_1, j_2)$ if the pair $j \in \mathscr{J}_0$ corresponds to the one-step transition from state j_1 to the state j_2. For any such $j \in \mathscr{J}_0$, the covariate $Z_{j_1 m}$ is denoted as Z_{jm}. We shall also find it convenient to write $\Gamma = [\Gamma_1, \ldots, \Gamma_q]^T$ for the vector obtained by stacking the columns of the matrix $\Gamma = [\Gamma_j]_{j \in \mathscr{J} \times \mathscr{J}}$ on the top of each other and deleting all entries corresponding to the pairs $(j_1, j_2) \notin \mathscr{J}_0$. For the sake of convenience, we shall write $\alpha_j(y, \theta, z)$ for each $j \in \mathscr{J}_0$ and $y = (y_1, \ldots, y_q)^T, y_j \in R_+, j = 1, \ldots, q$. However, it is tacitly assumed here that for $j = (j_1, j_2) \in \mathscr{J}_0$, the function $\alpha_j(y, \theta, z)$ may depend only on y_k's such that $(r(k), c(k)) = (j_1, \ell)$ for some $(j_1, \ell) \in \mathscr{J}_0$.

Under assumptions stated in section 5, the parameter θ varies over a bounded open subset Θ of R^d and the functions $\ell_j(y, \theta, z) = \log \alpha_j(y, \theta, z)$, $y \in R^q$ are twice continuously differentiable with respect to (y, θ). We let $\ell'_j = (\ell_j^{(1)}, \ldots, \ell_j^{(q)})^T$ be a vector whose k-th component is equal to the partial derivative of $\ell_j(y, \theta, z)$ with respect to $y_k, k = 1, \ldots, q$. Likewise, $\dot{\ell}_j$ denotes the (column) vector of length d corresponding to the derivative of ℓ_j with respect to θ. We further set $S_j(y, \theta, x) = n^{-1} \sum_{i=1}^n \sum_m Y_{jmi}(x) \alpha_j(y, \theta, Z_{jmi}), y \in R^q$ and denote by $\dot{S}, S'$ the derivatives of these processes with respect to (y, θ). Here, $\dot{S}$ is a $d \times q$ matrix, whose j-th column is given by $\dot{S}_j(y, \theta, x)$, the derivative of S_j with respect to θ. Further $S' = [S_j^{(k)}]_{j,k=1,\ldots,q}$ is a $q \times q$ matrix, whose (k, j) entry is equal to the partial derivative $S_j^{(k)}(y, \theta, x)$ of $S_j(y, \theta, x)$ with respect to $y_k, k = 1, \ldots, q$. Let s and let $\dot{s}, s'$ be the matrices of expected $\dot{S}$ and S' processes. Finally, for each $j \in \mathscr{J}_0$, we let $N_{j\cdot\cdot}(x) = n^{-1} \sum_{i=1}^n \sum_m N_{jmi}(x)$ be the averaged process counting transitions from the state $j_1 = r(j)$ to the state $j_2 = c(j)$ and whose sojourn time in the state j_1 does not exceed x.

As an estimate of the unknown transformations $\Gamma = [\Gamma_1, \dots, \Gamma_q]^T$, we consider a vector valued analogue of the estimator proposed by Bagdonovicius and Nikulin (1999,2004) for analysis of single spell models. The estimator is given by

$$\Gamma_{jn\theta}(x) = \int_0^x \frac{N_{j..}(du)}{S_j(\Gamma_{n\theta}(u-), \theta, u)}, \qquad (3.1)$$
$$\Gamma_{jn\theta}(0-) = 0, \quad \theta \in \Theta, x \geq 0, j \in \mathscr{J}_0.$$

For fixed θ, (3.1) forms a sample analogue of the non-linear vector-valued Volterra equation

$$\Gamma_{j\theta}(x) = \int_0^x \frac{EN_{j..}(du)}{s_j(\Gamma_\theta(u-), \theta, u)}, \quad \Gamma_{j\theta}(0-) = 0, x \geq 0, j \in \mathscr{J}_0. \qquad (3.2)$$

Using arguments similar to Dabrowska (2006), we can show that under the regularity conditions stated in Section 5, the equation (3.2) has a unique solution $\Gamma_\theta = [\Gamma_{1\theta}, \dots, \Gamma_{q\theta}]^T$ and its estimator (3.1) is uniformly consistent. Further, the function $\Theta \ni \theta \to \{\Gamma_\theta(x) : x \in [0, \tau]\} \in C([0, \tau])^q$ is Frèchet differentiable with respect to θ. The derivative is a $d \times q$ matrix of continuous functions satisfying the matrix-valued linear Volterra equation

$$\dot{\Gamma}_\theta(x) = -\int_0^x \dot{s}(\Gamma_\theta(w-), \theta, w) C_\theta(dw) - \int_0^x \dot{\Gamma}_\theta(w-) Q_\theta(dw), \qquad (3.3)$$

where $C_\theta(x)$ is the diagonal $q \times q$ matrix $C_\theta(x) = \text{diag}\,[C_{1\theta}(x), \dots, C_{q\theta}(x)]$ with entries

$$C_{j\theta}(x) = \int_0^x \frac{EN_{j..}(du)}{s_j^2(\Gamma_\theta(u-), \theta, u)}$$

and

$$Q_\theta(x) = \int_0^x s'(\Gamma(w-), \theta, w) C_\theta(dw).$$

The solution to the Volterra equation is given by

$$\dot{\Gamma}_\theta(x) = -\int_0^x \dot{s}(\Gamma_\theta(w-), \theta, w) C_\theta(dw) \mathscr{P}_\theta(w, x). \qquad (3.4)$$

where $\mathscr{P}_\theta(w, x)$, $0 < w \leq x$ is the Peano series (Gill and Johansen, 1990)

$$\mathscr{P}_\theta(u, x) = I + \sum_{m=1}^{\infty} \int_{u < w_1 < \dots < w_m \leq x} (-1)^m Q_\theta(dw_1) \cdot \dots \cdot Q_\theta(dw_m). \qquad (3.5)$$

Here I is the $q\times q$ identity matrix. A uniformly consistent estimate of $\{\dot\Gamma_\theta(x): x\in[0,\tau], \theta\in\Theta\}$ can be obtained by substituting the processes $N_{j\cdot\cdot}$ and $S_j, S_j', \dot S_j$ into the preceding expressions.

To define the score equation for estimation of the Euclidean parameter, let

$$e_j[f_j](u,\theta) \;=\; \frac{E\sum_m Y_{jmi}(u)[f_j\alpha_j](\Gamma_\theta(u),\theta,Z_{jmi})}{E\sum_m Y_{jmi}(u)\alpha_j(\Gamma_\theta(u),\theta,Z_{jmi})},$$

where $f_j(y,\theta,Z_{jmi})$ is a function of covariates, jointly continuous with respect to (y,θ) and bounded on every compact set of $R^q\times\Theta$. Likewise, for any two vectors f_{1j} and f_{2j} of such functions, define

$$\mathrm{cov}_j[f_{1j},f_{2j}](u,\theta) = \left(e_j[(f_{1j}\otimes f_{2j})] - (e_j[f_{1j}]\otimes e_j[f_{2j}])\right)(u,\theta)$$

and set $\mathrm{var}_j[f_j](u,\theta) = \mathrm{cov}_j[f_j,f_j](u,\theta)$.

To estimate the parameter θ, we use a solution to the score equation $U_n(\theta) = U_{n\varphi_n}(\theta) = o_P(n^{-1/2})$, where

$$U_{n\varphi_n}(\theta) = \frac{1}{n}\sum_{i=1}^n\sum_j\sum_m\int_0^\tau \widehat{b}_{jmi}(\Gamma_{n\theta}(u),\theta,u)N_{jmi}(du), \tag{3.6}$$

$\widehat{b}_{jmi}(\Gamma_{n\theta}(u),u,\theta) = \widehat{b}_{jm1i}(\Gamma_{n\theta}(u),u,\theta) - \varphi_{n\theta}(u)\widehat{b}_{jm2i}(\Gamma_{n\theta}(u),u,\theta)$ and

$$\begin{aligned}\widehat{b}_{jm1i}(y,\theta,u) &= \dot\ell_j(y,\theta,Z_{jmi}) - [\dot S_j/S_j](y,\theta,u),\\ \widehat{b}_{jm2i}(y,\theta,u) &= \ell_j'(y,\theta,Z_{jmi}) - [S_j'/S_j](y,\theta,u).\end{aligned}$$

Here $\varphi_{n\theta}(x)$ is an estimate of a $d\times q$ matrix of bounded functions $\varphi_\theta(x)$, whose j-th column is absolutely continuous with respect to $\Gamma_{j\theta}$.

We further define matrices

$$\begin{aligned}\Sigma_0(\theta) &= \sum_j\int_0^\tau v_{j,\varphi}(u,\theta)EN_{j\cdot\cdot}(du),\\ \Sigma_1(\theta) &= \Sigma_0(\theta) + \sum_j\int_0^\tau \rho_{j\varphi}(u,\theta)EN_{j\cdot\cdot}(du)[\dot\Gamma_\theta(u)+\varphi_\theta(u)]^T,\\ \Sigma_2(\theta) &= \Sigma_0(\theta) + \int_0^\tau D_\varphi(u,\theta)^T C_\theta(du) D_\varphi(u,\theta),\end{aligned}$$

where $v_{j,\varphi}(u,\theta) = \text{var}_j[\dot{\ell}_j - \varphi_\theta \ell'_j](u,\theta)$, $\rho_{j,\varphi}(u,\theta) = \text{cov}_j[\dot{\ell}_j - \varphi_\theta \ell'_j, \ell'_j](u,\theta)$ and

$$D_\varphi(u,\theta) = \sum_j \int_u^\tau \mathscr{P}_\theta(u,w) EN_{j..}(dw) \rho_{j\varphi}(w,\theta)^T.$$

Proposition 3.1 Let $\varepsilon_n \downarrow 0$ be a sequence such that $\sqrt{n}\varepsilon_n \to \infty$ and let $\mathscr{B}(\theta_0,\varepsilon_n) = \{\theta : |\theta - \theta_0| \le \varepsilon_n\}$ be the ball of radius ε_n centered at θ_0. Suppose that the matrix $\Sigma_0(\theta_0)$ is positive definite and the matrix $\Sigma_1(\theta_0)$ is non-singular. Under conditions stated in Section 5, the score equation $U_{n\varphi_n}(\theta) = o_{P^*}(n^{-1/2})$ has a solution $\widehat{\theta}$ in the ball $\mathscr{B}(\theta_0,\varepsilon_n)$, with (inner) probability tending to 1. Further, let $\widehat{\Xi} = \sqrt{n}(\widehat{\theta} - \theta_0)$ and $\widehat{W}_0 = \sqrt{n}[(\Gamma_{n\widehat{\theta}} - \Gamma_{\theta_0})^T - (\widehat{\theta} - \theta_0)^T \dot{\Gamma}_{n\widehat{\theta}}]$. Then $[\widehat{\Xi}, \widehat{W}_0]$ converges weakly in $R^d \times \ell^\infty([0,\tau] \times \mathscr{J}_0)$ to a tight mean zero Gaussian process $[\Xi, W_0]$ with covariance

$$\begin{aligned}
&\text{cov}\Xi = \Sigma_1^{-1}(\theta_0)\Sigma_2(\theta_0)[\Sigma_1^{-1}(\theta_0)]^T,\\
&\text{cov}(W_0(x), W_0(x')) = K_{\theta_0}(x,x'),\\
&\text{cov}(\Xi, W_0(x)) = -\Sigma_1^{-1}(\theta_0)\sum_j \int_0^\tau \rho_{j,\varphi}(u,\theta_0) EN_{j..}(du) K_{\theta_0}(u,x),
\end{aligned}$$

where $K_\theta, \theta \in \Theta$ is a $q \times q$ matrix

$$K_\theta(x,x') = \int_0^{x \wedge x'} \mathscr{P}_\theta^T(u,x) C_\theta(du) \mathscr{P}_\theta(u,x'). \tag{3.7}$$

Here $\mathscr{X}_1 = \ell^\infty([0,\tau] \times \mathscr{J}_0)$ denotes the space of bounded functions mapping the set $[0,\tau] \times \mathscr{J}_0$ into R and equipped with uniform metric and Borel σ-field. The Borel σ-field $\mathscr{X} = R^d \times \mathscr{X}_1$ is generated by open sets in the product topology of the Euclidean space R^d and the space $\mathscr{X}_1$. It is equal to $\mathscr{B}(R^d) \otimes \mathscr{B}(\mathscr{X})$ because R^d is a complete separable metric space. The process $X = (\Xi, W_0)$ has a version whose almost all paths are in the separable subspace of $\mathscr{X}$ corresponding to $R^d \times C_b([0,\tau] \times \mathscr{J}_0)$, where $C_b([0,\tau] \times \mathscr{J}_0)$ is the space functions continuous with respect to the variance pseudo-metric. Weak convergence of the sequence $X_n = [\widehat{\Xi}, \widehat{W}_0]$ to (Ξ, W_0) means that for all bounded continuous functions f on $\mathscr{X}$, we have $E^* f(X_n) - E f(X) \to 0$, where E^* is the outer expectation. This implies that X_n is asymptotically measurable. In particular, we have $E^* f(X_n) - E_* f(X_n) \to 0$ for all bounded continuous functions f on $\mathscr{X}$, where $E_* f(X_n) = -E^*(-f(X_n))$ is the inner expected (van der Vaart and Wellner (1996), Dudley (1999)). We also note that the space $\mathscr{X}_1 = \ell^\infty(\mathscr{T} \times \mathscr{J}_0)$ is isometric to the product space $\mathscr{Y} = \ell^\infty([0,\tau])^q$ equipped with uniform metric $d_Y(x,y) = \max_j \sup_t |x_j(t) - y_j(t)|$ and product topology of $\mathscr{Y}$ coincides with the topology induced by metric d_Y. Under assumptions of section 5, the space $C_b([0,\tau]\times \mathscr{J}_0)$ is isometric to the space $C([0,\tau])^q$ and W_0 is a linear transformation of a vector of q independent time-transformed Brownian motions.

The M-estimator $\widehat{\theta}$ depends on the specification of the matrix φ_θ and its estimator $\varphi_{n\theta}$. Depending on the measurability properties of the estimator $\varphi_{n\theta}$, the solution to the score equation exists either with probability tending to 1, or with inner probability tending to 1 (Section 5). Two simple choices of the function φ_θ correspond to $\varphi_\theta \equiv 0$ and $\varphi_\theta = -\dot{\Gamma}_\theta$. In particular, with the latter choice, the estimate $\widehat{\theta}$ is an analogue of the pseudo- maximum likelihood estimators considered by Bagdonovicius and Nikulin (1999,2004) in the case of single spell models. Under regularity conditions, the optimal choice of this function corresponds to solution of a system of Sturm-Liouville equations and yields an asymptotically efficient estimate of the Euclidean component of the model. If the process registers only events of one type (i.e. $|\mathscr{J}_0| = 1$) then the form of φ_θ corresponding to the efficient estimate of θ is similar to the single spell version of this model and can be found in Bickel (1986) and Bickel and Ritov (1995) in the uncensored case, and in Dabrowska (2007) in the censored case. The estimate of the function φ_θ can be obtained in this case by inverting a simple tridiagonal band-symmetric matrix. The form of the information bound and efficient score function for the general case ($|\mathscr{J}_0| > 1$) is postponed to a separate paper, where we consider it under additional compatibility conditions.

To set confidence bands for the baseline Γ vector and related parameters, we consider Gaussian multiplier method of Lin, Fleming and Wei (1994). For this purpose, we shall need some additional notation.

(i) Let G_0 be a vector of independent $\mathscr{N}(0, I_{d\times d})$ variables. and let $G_i = (G_{mi} : m = 1,\ldots,K_i), i = 1,\ldots,n, K_i = Y_{..i}(0)$ be standard normal variables, independent of G_0 and mutually independent given the data $D_1,\ldots,D_n$.

(ii) For $j \in \mathscr{J}_0$, set

$$\widehat{V}_j^{\#}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{m} G_{mi}\int_0^x \frac{N_{jmi}(du)}{S_j(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)},$$

(iii) Put $\widehat{\Xi}^{\#} = \widehat{\Xi}_1^{\#} - \widehat{\Xi}_2^{\#}$, where $\widehat{\Xi}_1^{\#} = \widehat{\Sigma}_1^{-1}(\widehat{\theta})\widehat{\Sigma}_0(\widehat{\theta})^{1/2}G_0$ and

$$\widehat{\Xi}_2^{\#} = \widehat{\Sigma}_1^{-1}(\widehat{\theta})\sum_j \int_0^\tau \widehat{\rho}_{j,\varphi_n}(u,\widehat{\theta})N_{j..}(du)\widehat{W}_0^{\#}(u)^T,$$

$$\widehat{W}_0^{\#}(x) = \int_0^x \widehat{V}^{\#}(du)\widehat{\mathscr{P}}_{\widehat{\theta}}(u,x) = \widehat{V}^{\#}(x) - \int_0^x \widehat{W}_0^{\#}(u-)\widehat{Q}_{\widehat{\theta}}(du).$$

The estimates $\widehat{Q}_{\widehat{\theta}}$ and $\widehat{\mathscr{P}}_{\widehat{\theta}}$ are plug-in analogues of the matrices defined in (3.3)-(3.5).

Proposition 3.2 Suppose that the conditions of Proposition 3.1 are satisfied. Then, unconditionally, $(\widehat{\Xi}^{\#}, \widehat{W}_0^{\#})$, $\widehat{W}_0^{\#} = \{[\widehat{W}_j^{\#}(x) : x \in [0,\tau], j \in \mathscr{J}_0\}$ converges weakly in $R^d \times \ell^\infty([0,\tau] \times \mathscr{J}_0)$ to a mean zero Gaussian process $(\Xi^{\#}, W_0^{\#})$ with the same covariance function as (Ξ, W_0). Moreover, (Ξ, W_0) and $(\Xi^{\#}, W_0^{\#})$ are independent while $(\widehat{\Xi}, \widehat{W}_0)$ and $(\widehat{\Xi}^{\#}, \widehat{W}_0^{\#})$ are asymptotically independent. Conditionally, the process $(\widehat{\Xi}^{\#}, \widehat{W}_0^{\#})$ converges weakly to $(\Xi^{\#}, W_0^{\#})$, in probability.

As in van der Vaart and Wellner (1996, p. 181), conditional weak convergence means that $\sup_{h \in BL_1} |E_G h(\widehat{\Xi}^{\#}, \widehat{W}_0^{\#}) - Eh(\Xi^{\#}, W_0^{\#})| \to_{P^*} 0$, where E_G denotes expectation with respect to the G variables. Further, h varies over the class of bounded Lipschitz functions, and BL_1 is the set Lipschitz functions whose norm is bounded by 1.

This proposition can be further extended to approximate the distribution of functionals $\Phi(\theta, \Gamma)$. In sufficiently simple cases, functional delta method can be used for this purpose. In particular, we may consider estimation of the kernel F of a semi-Markov processes with a state space $\mathscr{J} = \{1, \dots, r\}$. In this case the covariates are time independent, and the entries of the matrix $F(x|z) = [F_j(x|z)]_{j \in \mathscr{J} \times \mathscr{J}}$ are specified by (2.2)-(2.3). Under the assumed differentiability conditions on the hazard functions α_j, the plug-in sample analogue $\widehat{F}$ of the matrix F has entries satisfying

$$\begin{aligned}
\widehat{W}_{F,j}(x|z) &= \sqrt{n}[\widehat{F}_j - F_j](x|z) = \qquad (3.8)\\
&= \widehat{\Xi}^T \int_0^x \dot{f}_j(\Gamma_{(j_1.)}(u), \theta_0, z)\Gamma_j(du) + \int_0^x \widetilde{W}_{(j_1.)}(u) f_j'(\Gamma_{(j_1.)}(u), \theta, z)\Gamma_j(du)\\
&+ \int_0^x f_j(\Gamma_{(j_1.)}(u), \theta_0, z)\widetilde{W}_{0j}(du) + o_{P^*}(1), j \in \mathscr{J}_0.
\end{aligned}$$

For any $j = (j_1, j_2) \in \mathscr{J}_0$, $\Gamma_{(j_1,.)}$ and $\widetilde{W}_{(j_1,.)}$ denote subvectors $\Gamma_{(j_1,.)} = \{\Gamma_{\theta_0 j} : j = (j_1, \ell) \in \mathscr{J}_0\}$ and $\widetilde{W}_{(j_1,.)} = \{\widetilde{W}_{0j} : j = (j_1, \ell) \in \mathscr{J}_0\}$, where

$$\widetilde{W}_0 = \{\sqrt{n}[\Gamma_{nj\widehat{\theta}} - \Gamma_{j\theta_0}] : j \in \mathscr{J}_0\} = \widehat{W}_0 + \widehat{\Xi}^T \dot{\Gamma}_{n\widehat{\theta}} + o_{P^*}(1). \qquad (3.9)$$

Denote by $\widehat{W}_F^{\#}$ the matrix obtained by replacing in (3.8)-(3.9) the process $(\widehat{\Xi}, \widehat{W}_0)$ by $(\widehat{\Xi}^{\#}, \widehat{W}_0^{\#})$ and the unknown parameters by their estimates $(\widehat{\theta}, \Gamma_{n\widehat{\theta}})$. Using integration by parts and Proposition 3.1 it is easy to verify that the process $\widehat{W}_F = [\widehat{W}_{F,j}(x|z) : x \le \tau, j \in \mathscr{J}_0]$ converges weakly to a mean zero Gaussian process W_F

in $\ell^\infty([0,\tau])^{|\mathscr{J}_0|}$. In addition, the conclusions of Proposition 3.2 carry over to the process $\widehat{W}_F^\# = [\widehat{W}_{F,j}^\#(x|z) : x \le \tau, j \in \mathscr{J}_0]$, i.e. unconditionally, $\widehat{W}_F^\#$ converges weakly to a mean zero Gaussian process $W_F^\#$ with the same covariance function as the process W_F and is independent of it. Conditionally, the process $\widehat{W}_F^\#$ converges weakly to $W_F^\#$ in probability.

Another example of a functional may correspond to the cumulative residual process arising in goodness-of-fit testing. In particular, suppose that covariates are partitioned into k disjoint categories, $I_1,\dots,I_k$. The cumulative residual process for the one-step transition between states $j_1 \to j_2$ is given by

$$\widehat{R}_j(x,\ell) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_m 1(Z_{jmi}\in I_\ell)\widehat{M}_{jmi}(x) =$$
$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_m \int_0^x \left[1(Z_{jmi}\in \ell) - \frac{S_{j\ell}}{S_j}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)\right] N_{jmi}(du),$$

where $S_{j\ell}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u) = \sum_{i=1}^n\sum_m Y_{jmi}(u)1(Z_{jmi}\in I_\ell)\alpha_j(\widehat{\Gamma}_{n\widehat{\theta}}(u-),\widehat{\theta},Z_{jmi})$ is the risk process corresponding to subjects in the group I_ℓ. Under the assumption that residuals are consistent with the model, the $\widehat{R} = \{\widehat{R}_j(t,\ell) : t \in [0,\tau], j \in \mathscr{J}_0, \ell = 1,\dots,k\}$ converges weakly to a mean zero Gaussian process and the Gaussian multiplier approximation to its distribution is given by

$$\widehat{R}_j^\#(x,\ell) = \frac{1}{\sqrt{n}}\int_0^x \left[1(Z_{jmi}\in I_\ell) - \frac{S_{j\ell}}{S_j}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)\right] G_{mi}N_{jmi}(du)$$
$$-(\widehat{\Xi}^\#)^T\int_0^x\left(\left[\frac{\dot{S}_{j\ell}}{S_{j\ell}} - \frac{\dot{S}_j}{S_j}\right]\frac{S_{j\ell}}{S_j}\right)(\Gamma_{n\widehat{\theta}},\widehat{\theta},u)N_{j\cdot\cdot}(du)$$
$$-\int_0^x \widetilde{W}^\#_{(j_1,\cdot)}(u)\left(\left[\frac{S'_{j\ell}}{S_{j\ell}} - \frac{S'_j}{S_j}\right]\frac{S_{j\ell}}{S_j}\right)(\Gamma_{n\widehat{\theta}},\widehat{\theta},u)N_{j\cdot\cdot}(du).$$

In analogy to Martinussen and Scheike (2006), the performance of residuals can be evaluated using Kolmogorov-Smirnov statistics such as $\sup_{x\in[\delta,\tau-\delta]}|\widehat{R}_j(x,\ell)|$ and the Guassian multiplier method can be used to obtain critical levels of tests. Alternate tests can be obtained by modifying chi-squared tests in Aalen *et al* (2008, p.144) or tests based on Schoenfeld residuals.

4 Example

We consider a transplant outcome data set from the Center for International Blood and Marrow Transplant Research (CIBMTR). The CIBMTR is comprised of clinical and basic scientists who confidentially share data on their blood and bone marrow transplant patients with CIBMTR Data Collection Center located at the Medical College of Wisconsin. The CIBMTR is a repository of information about results of transplants at more than 450 transplant centers worldwide. The example data set consists of patients who received HLA-identical sibling transplant from 1995 to 2004 for acute myelogenous leukemia (AML) or acute lymphoblastic leukemia (ALL) and transplanted in first remission. All patients received bone marrow transplantation or peripheral blood stem cell transplantation. Children under age 16 and all patients who received umbilical cord blood transplants were excluded as risk factors are likely to vary in this group.

Allogeneic stem cell transplantation (ASCT) is an accepted treatment for leukemia patients. Transplant candidates receive high doses of chemotherapy and radiation which destroy malignant cells in the bone marrow and elsewhere. Because stem cells in the normal bone marrow are destroyed in this process as well, patients subsequently receive a transplant from a suitably matched donor. The transplant can be followed by several complications. In this study, fatal complications correspond to relapse of leukemia or death in remission (hereafter referred to as death). The most important intermediate event in ASCT is graft-versus-host-disease (GVHD) in which transplanted immune cells recognize the recipient's body tissues as foreign. Acute and chronic GVHD (AGVHD and CGVHD) are two forms of this disease. AGVHD occurs during the early post-transplant period is defined here as moderate to severe using clinically established criteria. CGVHD occurs later in time and may be preceded by AGVHD.

The incidence of GVHD, leukemia relapse and death in remission depends on a number of variables characterizing the recipient, the donor and the transplant. The main variables considered in this paper include recipient's age, donor-recipient gender match, disease type and graft source. Bone marrow was the first source of stems cells used in used ASCT. Since 90'ies, peripheral-blood stem cell transplants have replaced bone marrow as the preferred source of stem cells because of a quicker hematologic recovery and relative ease of collection. Patients may receive also an infusion of both peripheral stem-cells and bone marrow. Several studies have shown that PBSCT recipients may be at a higher risk of GVHD than BMT patients. (e.g. Cutler *et al.* (2001), Flowers *et al.* (2002), Friedrichs *et al.* (2010)). A possible explanation of this phenomenon is that GVHD develops from the infusion of donor T cells and PBSCT recipients receive a significantly higher dose of T cells than BMT patients. As a result of the increased risk of GVHD, the patients who experience it may be at a higher risk of death in remission than BMT patients. GVHD is also more more common among older patients and among male recipients receiving transplants from female donors (Gale *et al.* 1987).

For purposes of modeling, we consider a five state modulated renewal model proposed for analysis of the transplant recovery process in Dabrowska *et al.* (1994). Table 1 collects some information about the type and number of the observed transitions, their range and median. The model assumes that a patient remains in the transplant state (tx, state 1) until the time of the first adverse event which may correspond to AGVHD (state 2), CGVHD (state 3), relapse (state 4) or death in remission (state 5). The model takes also in to the account that a patient who develops GVHD may subsequently relapse or die, and that CGVHD may be preceded by AGVHD. The observed model has an extra absorbing state corresponding to censoring (loss-to-follow-up). Further, age was categorized into 3 groups, each representing approximately one third of the patients. The baseline group corresponds to the age range $[29.5, 42.5]$. Transitions were also adjusted for the waiting time for transplant. Two continuous variables were used for this purpose: the length of time between leukemia diagnosis and first remission (DxCr) and the length of time between first remission and transplant (CrTx). Their medians and range were: median(DxCr)= 1.38, IQR(DxCr)=1.15, range(DxCr)=221.45 months and med(CrTx) = 3.06, IQR(CrTx)=2.5, range(CrTx)=46.74 months. To obviate skewness of the distribution, the log transformation of these variables is used in the regression analysis.

The modulated renewal process assumes that one-step transition probabilities are specified by means of a proportional odds ratio model. More precisely, hazard rates of one-step transitions originating from the transplant or AGVHD state are of the form

$$\alpha_j(\Gamma_{(j_1,.)}(x),\theta,Z)\gamma_j(x) = e^{\theta_j^T Z_j}[1+\sum_{k=j_1+1}^{5} 1(\ell=(j_1,k))\Gamma_\ell(x)e^{\theta_\ell^T Z_\ell}]^{-1}\gamma_j(x),$$

for $j=(j_1,j_2)$ such that $j_1 = 1$ or $j_1 = 2$ and $j_1+1 \leq j_2 \leq 5$, $\Gamma_j(x) = \int_0^x \gamma_j(u)du$. In the case of transition rates originating from the CGVHD state, we use covariate $Z_C = (Z, Z_A)$, where Z_A is a binary variable indicating by 1 whether AGVHD preceded onset of chronic graft versus host disease. The corresponding transition rates into the relapse and death states are given by

$$\alpha_j(\Gamma_{(3,.)}(x),\theta,Z_C)\gamma_j(x) = e^{\theta_j^T Z_{jC}}[1+\sum_{k=4}^{5} 1(\ell=(3,k))\Gamma_\ell(x)e^{\theta_\ell^T Z_{jC}}]^{-1}\gamma_j(x)$$

for $j=(3,j_2)$ and $j_2 = 4,5$. Here Z_j and Z_{jC}, $j=(j_1,j_2)$, represent transition specific covariates, which correspond to subvectors of Z and Z_C, respectively. Table 4 provides their entries as well as the estimates of the regression coefficients and standard errors. The estimates were obtained using Fisher scoring algorithm ap-

Table 1: Observed one-step transitions

	n	median (in months)	range (in months)
TX $\rightarrow$ AGVHD	491	.7	4.3
TX $\rightarrow$ CGVHD	372	5.5	106.4
TX $\rightarrow$ relapse	106	5.6	59.4
TX $\rightarrow$ death	179	2.9	131.9
TX $\rightarrow$ censoring	506	56.9	143.8
AGVHD $\rightarrow$ CGVHD	202	4.8	57.4
AGVHD $\rightarrow$ relapse	33	5.2	23.7
AGVHD $\rightarrow$ death	141	2.9	80.3
AGVHD $\rightarrow$ censoring	115	45.7	133.0
CGVHD $\rightarrow$ relapse	27	8.3	98.3
CGVHD $\rightarrow$ death	79	9.8	124.4
CGVHD $\rightarrow$ censoring	266	51.1	144.3
A+CGVHD $\rightarrow$ relapse	25	3.5	53.3
A+CGVHD $\rightarrow$ death	65	5.6	109.3
A+CGVHD $\rightarrow$ censoring	112	56.3	145.2

Table 2: Summary of covariates

Age	n	Graft source	n	Disease	n
< 30 (young)	550	[BMT]	842	[AML]	1168
$[30, 42.5]$	534	PB/PB+BMT	803	ALL	477
> 42.5 (old)	561				
Donor's Gender	**n**	**Gender-Match**	**n**		
F	890	FM	441		
$[M]$	755	[not FM]	1224		

Baseline groups are marked in brackets.
FM represents a female to male transplant

plied to the score process (3.6) with $\varphi_{n\theta} = -\dot{\Gamma}_{n\theta}$. Variable selection was based on backwards elimination and Wald testing. To asses adequacy of the model, we have used Kolmogorov-Smirnov tests described in Section 3. The results are summarized below and at in Table 5.

Table 3: One-step transition probability matrix

	tx	AGVH	CGVH	A+CGVH	rel	death
tx	0	F_{12}	F_{13}	0	F_{14}	F_{15}
AGVHD	0	0	0	F_{23}	F_{24}	F_{25}
CGVHD	0	0	0	0	F_{34}	F_{35}
A+CGVHD	0	0	0	0	$F_{\bar{3}4}$	$F_{\bar{3}5}$
rel	0	0	0	0	1	0
death	0	0	0	0	0	1

We note here that the transitions originating from the CGVHD state depend on whether or AGVHD was experienced prior to the entrance to the CGVHD state. This dependence violates the assumption that the sequence of states visited forms a Markov chain. However, this problem disappears if the state space of the process is enlarged to include an extra state A+CGVHD. This extra state is here denoted by $\overline{3}$. Conditionally on the time independent covariates, the resulting model has structure of a semi-Markov process with kernel $F(x|z) = [F_j(x|z)]$ specified in Table 3. The entries of the kernel matrix have a fairly explicit form. For transitions originating from the transplant (tx) or AGVHD state, we have

$$F_j(x|z) = \int_0^x e^{\theta_j^T Z_j}[1 + \sum_{k=j_1+1}^{5} 1(\ell = (j_1,k))\Gamma_\ell(u)e^{\theta_\ell^T Z_\ell}]^{-2}\Gamma_j(du)$$

for $j = (j_1, j_2), j_1 = 1,2$ and $j_2 = j_1 + 1 \leq j_2 \leq 5$. One-step transition probabilities originating from the CGVHD state are given by

$$F_j(x|z) = 1(Z_A = 0)\int_0^x [1 + \sum_{k=4}^{5} 1(\ell = (3,k))\Gamma_\ell(u)e^{\theta_\ell^T Z_{jC}}]^{-2} e^{\theta_j^T Z_{jC}}\Gamma_j(du)$$

for $j = (3, j_2)$ and $j_2 = 4,5$. One-step transition probabilities originating from the state A+CGVHD (labeled as "$\overline{3}$") have a similar form, with covariate covariate $Z_A = 1$.

We also consider multi-step probabilities of transitions into the absorbing states, i.e. probabilities of transition into the relapse and death states along any possible path of the model. Let $J(t)$ be the state occupied by the process at time

t and let e denote either relapse or death in remission. By noting that a patient may move into an absorbing state by first passing through the GVHD states, these probabilities are given by

$$H_e(t|z) = P(J(t) = e|z) = \sum_{k=1}^{4} H_e^{(k)}(t|z),$$

where

$$\begin{aligned} H_e^{(1)}(t|z) &= P(\{J(t) = e\} \cap A^c \cap C^c|z), \\ H_e^{(2)}(t|z) &= P(\{J(t) = e\} \cap A \cap C^c|z), \\ H_e^{(3)}(t|z) &= P(\{J(t) = e\} \cap A^c \cap C|z), \\ H_e^{(4)}(t|z) &= P(\{J(t) = e\} \cap A \cap C|z), \end{aligned} \tag{4.1}$$

and the events A and C represent

$$\begin{aligned} A &= \{\text{AGVHD occurs prior to the event e}\}, \\ C &= \{\text{CGVHD occurs prior to the event e}\}. \end{aligned}$$

The first of these probabilities corresponds to a move from the transplant to the state e in one step so that $H_e^{(1)}(t|z) = F_{1e}(t|z)$ for $e = 4,5$. The terms $H_e^{(2)}$ and $H_e^{(3)}$ provide the probabilities of transitions along the paths "tx $\to$ AGVHD $\to$ e" ($H_e^{(2)}$) and "tx $\to$ CGVHD $\to$ e" ($H_e^{(3)}$) and are given by $H_e^{(k)}(t|z) = (F_{1k} \star F_{ke})(t|z)$, $k = 2$ or 3, $e = 4$ or 5. Here for any two subdistribution functions F and F' on the positive half-line, $F \star F'$ is their convolution

$$(F \star F')(t) = \int_0^x F(t-u)F'(du) = \int_0^x F(du)F'(t-u).$$

Lastly, transition along the path "tx $\to$ AGVHD $\to$ A+CGVHD $\to$ e" ($H_e^{(4)}$) contributes to the sum $H_e^{(4)}(t|z) = (F_{12} \star F_{23} \star F_{\overline{3}e})(t|z)$.

The multi-step transition probabilities can be estimated using plug-in method. The estimates are consistent on time intervals $[0,\tau]$ strictly contained in the support of all sojourn time distributions. As an example, Figure 1 compares transition probabilities of hypothetical ALL patients receiving BMT and PBSCT transplant. The remaining covariates correspond to the age range 16-29.5 years and baseline subgroups specified in Table 2. The plots represent the four components of the multistep transition probabilities defined in (4.1). PBSCT seems to reduce one-step transition probabilities of both relapse and death ($H_e^{(1)}$, black curves), and

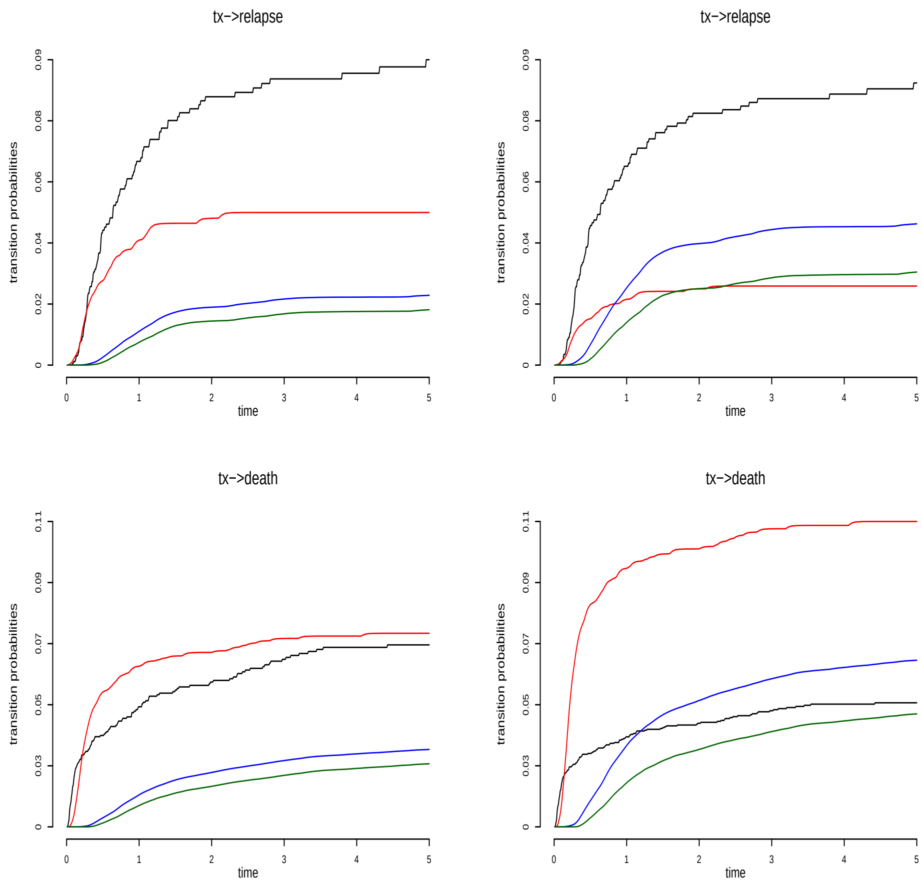

Figure 1: Transition probabilities of endpoint events of a young ALL patient receiving BMT (left panel) or PB (right panel). The remaining covariates correspond to the baseline. The curves represent one-step transitions tx $\rightarrow$ e (black), two-step transitions tx $\rightarrow$ AGVHD $\rightarrow$ e (red) and tx $\rightarrow$ CGVHD $\rightarrow$ e (blue), and three-step transitions tx $\rightarrow$ AGVHD $\rightarrow$ CGVHD $\rightarrow$ e (green).

the effect is more pronounced in the case of the tx $\rightarrow$ death transition. The graphs suggest also that PBSCT associates with a reduced probability of relapse preceded by AGVHD ($H_e^{(2)}$, red curves). At the same time, however, the probability of death in remission is higher than that of a BMT recipient. We also see an increase in the probability of relapse and death resulting from CGVHD without AGVHD ($H_e^{(3)}$, blue curves) and CGVHD with AGVHD ($H_e^{(4)}$, green curves).

To assess effects of covariates, we consider pointwise and simultaneous confidence bands for pairwise differences of one-step and multi-step transition probabilities. In the case of one-step transition probabilities, we consider functions

$$\Delta_j^F(t|z_1,z_2) = F_j(t|z_1) - F_j(t|z_2), \quad j \in \mathscr{J}_0,$$

where z_1 and z_2 are two covariate levels. We denote by $\widehat{\Delta}_j^F$ the corresponding sample analogue of the function Δ_j^F. Results of Section 5 imply that the normalized process $\widehat{W}_{j,\Delta}^F = \{\sqrt{n}[\widehat{\Delta}_j^F - \Delta_j^F](t|z_1,z_2) : t \in [0,\tau]\}$ converges weakly to a mean zero Gaussian process $W_{j\Delta}^F = \{W_j^F(t|z_1) - W_j^F(t|z_2) : t \in [0,\tau]\}$.

To construct confidence bands, we note that each Δ function forms a difference of two subdistributions functions. Correspondingly, it assumes values between -1 and 1. Direct application of the Gaussian approximation to the limiting distribution of the process $W_{j\Delta}^F$ may result in confidence intervals and confidence bands whose bounds may fall outside the interval $(-1,1)$. To circumvent this problem, we use transformation method.

Let $\Phi : R \to (-1,1)$ be strictly increasing differentiable function derivative φ satisfying $\varphi(x) > 0$ for all $x \in R$. By delta method,

$$\begin{aligned} &\sqrt{n}[\Phi^{-1}(\widehat{\Delta}_j^F(t|z_1,z_2)) - \Phi^{-1}(\Delta_j^F(t|z_1,z_2))] \Rightarrow \\ &\varphi(\Phi^{-1}(\Delta_j^F(t|z_1,z_2)))^{-1} W_{j,\Delta}^F(t|z_1,z_2), t \in [0,\tau]. \end{aligned}$$

Let $c_\alpha(t_1,t_2)$ be the upper α percentile of the distribution of

$$\sup_{t_1 \le t \le t_2} \left[\frac{|W_{j,\Delta}^F|}{\widehat{\sigma}_{\Delta_j^F}} \right] (t|z_1,z_2),$$

where $\widehat{\sigma}_{\Delta_j^F}(t|z_1,z_2)$ is an estimate if the standard deviation of $\Delta_j^F(t|z_1,z_2)$. Then, by the continuous mapping theorem, the $100 \times (1-\alpha)\%$ asymptotic confidence band for the Δ function has upper and lower bounds given by

$$\Phi\left(\Phi^{-1}(\widehat{\Delta}_j^F(t|z_1,z_2)) \pm c_\alpha(t_1,t_2) \frac{\widehat{\sigma}_{\Delta_j^F}(t|z_1,z_2)}{\varphi(\Phi^{-1}(\widehat{\Delta}_j^F(t|z_1,z_2)))}\right). \tag{4.2}$$

The corresponding pointwise confidence intervals can be obtained by replacing the constant $c_\alpha(t_1,t_2)$ by the upper $\alpha/2$ percentile of the standard normal distribution.

A possible choice of the Φ function may correspond to $\Phi(x) = 2G(x) - 1$, where G is a distribution function with density g supported on the whole real line. In analogy to the construction of the confidence bands for survival function in Andersen *et al.* (1993), we may consider the choice of the extreme value distribution $G(x) = 1 - \exp[-e^x]$. In this case $\Phi^{-1}(u) = \log[-\log[(1-u)/2]]$ and the bounds are given by

$$1 - 2\left[\frac{1-\widehat{\Delta}_j^F(t|z_1,z_2)}{2}\right]^{\exp[\pm c_\alpha(t_1,t_2)[h\widehat{\sigma}_{\Delta_j^F}](t|z_1,z_2)]}, \tag{4.3}$$
$$h(t|z_1,z_2) = [|\log[(1-\widehat{\Delta}_j^F(t|z_1,z_2))/2|](1-\widehat{\Delta}_J^F(t|z_1,z_2))]^{-1}.$$

Another possible choice may correspond to the logistic distribution, $G(x) = e^x/[1+e^x]$. We have $\Phi^{-1}(u) = \log([1+u]/[1-u])$, and the bounds assume form

$$1 - 2\left(1 + \frac{1+\widehat{\Delta}_j^F(t|z_1,z_2)}{1-\widehat{\Delta}_J^F(t|z_1,z_2)}\exp[\pm c_\alpha(t_1,t_2)[h\widehat{\sigma}_{\Delta_j^F}](t|z_1,z_2)]\right)^{-1}, \tag{4.4}$$
$$h(t|z_1,z_2) = 2[(\widehat{\Delta}_j^F(t|z_1,z_2)+1)(1-\widehat{\Delta}_j^F(t|z_1,z_2))]^{-1}.$$

A similar approach can be applied towards comparison of multi-step transition probabilities. For any two covariate levels, z_1 and z_2, we set

$$\Delta_j^H(t|z_1,z_2) = H_j(t|z_1) - H_j(t|z_2), j = 4,5.$$

The corresponding sample analogue is denoted by $\widehat{\Delta}_j^H$. It is easy to see that $\{\widehat{W}_{j,\Delta}^H(t|z_1,z_2) = \sqrt{n}[\widehat{\Delta}_j^H - \Delta_j^H](t|z_1,z_2) : t \in [0,\tau]\}$ converges weakly to a Gaussian process $W_{j\Delta}^H(t|z_1,z_2) = W_j^H(t|z_1) - W_j^H(t|z_2)$, where

$$W_j^H(t|z) = W_{1j}^F(t|z) + \sum_{i=2}^{3}[W_{1i}^F \star F_{ij} + F_{1i} \star W_{ij}](t|z)$$
$$+[W_{12}^F \star F_{23} \star F_{\overline{3}j} + F_{12} \star W_{23}^F \star F_{\overline{3}j} + F_{12} \star F_{23} \star W_{\overline{3}j}^F](t|z).$$

and the integrals are defined by means of the convolution formula.

In Figures 2-5, we compare one-step and multi-step transition probabilities of relapse and death in remission for patients with selected covariate profiles. To obtain the bands, we first used Gaussian multiplier method to estimate the approximate variance of the Δ function: the Monte Carlo variance of the Δ function was computed based on 5000 Monte Carlo experiments. A second application of the

Gaussian multiplier method was then used to obtain an approximation of the critical level $c_\alpha(t_1,t_2)$ based on 5000 Monte Carlo trials. The interval $[t_1,t_2]$ was chosen to correspond to $t_1 = 1.5$ and $t_2 = 60$ months. The bounds (4.2) and (4.3) showed a close numerical agreement and the resolution of the graphs does not allow to show the difference between the two choices. The difference between the upper/lower bounds did not exceed .07%, and the bands obtained using the logistic transformation were narrower.

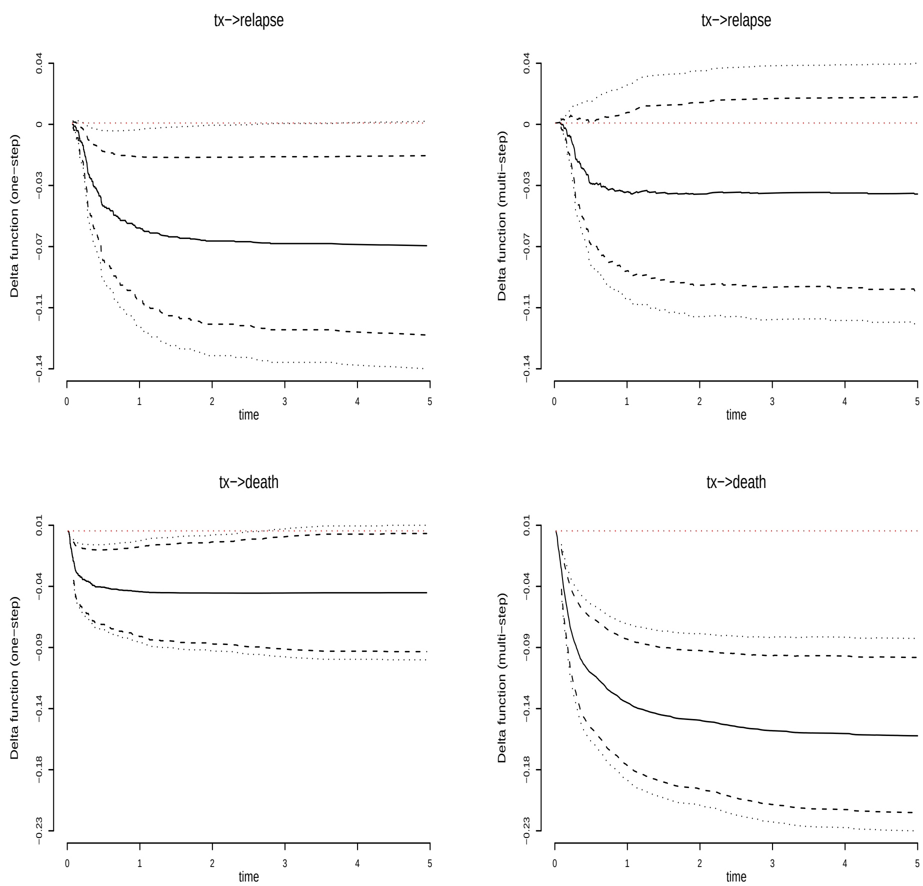

Figure 2: Pointwise and simultaneous confidence bands for the one-step and multi-step Δ functions of ALL patients receiving BMT. Covariates: age $z_1 \leq 29.5$ and $z_2 =$ baseline age. The remaining covariates correspond to the baseline.

Younger age associated with reduced probabilities of relapse and death of both AML and ALL patients. In Figure 2, we use Δ function to compare transition probabilities of hypothetical younger (z_1) and baseline age (z_2) ALL bone marrow transplant recipients. The remaining covariates correspond to baseline groups specified in Table 2 and median waiting times variables DxCr and CrTx. The plots show that younger age has "concordant" effect on endpoint probabilities, i.e. younger age associated with reduced probability of both relapse and death. In the case of one-step tx $\rightarrow$ relapse transitions, the pointwise bands suggest that the differences are significant but the wider simultaneous bands show that this is not the case. Examination of the four possible paths leading to the relapse state showed that although younger patients have lower one-step relapse transition probabilities, they are at a higher risk of relapse preceded by AGVHD than patients in the baseline age group. This accounts for marginal differences in the multistep relapse transition probabilities. Figure 2 shows also that multi-step transitions into the death state are significantly lower for a younger patient since the upper bounds of both pointwise and simultaneous bands are below the horizontal line passing through 0. While in the case of one-step transition probabilities there is a marginal difference during the early post-transplant period, patients in the baseline age group had higher probabilities death preceded by GVHD.

In Figure 3, we show the "discordant" effect of older age on the two endpoint probabilities. The graphs represent Δ function for hypothetical ALL patients receiving peripheral blood stem cell transplant. The covariate z_1 corresponds to the older age and z_2 to the baseline age group. The remaining covariates correspond to baseline (Table 2). Older age associated with lower transition probabilities into the relapse state. On the other hand, the role of the two covariates is reversed in the case of transitions into the death state. Plots of the four paths leading to the endpoint events showed that an older patient may have higher probabilities of death resulting from CGVHD (with or without AGVHD) while probability of transition along the path tx $\rightarrow$ AGVHD $\rightarrow$ death is comparable to that of a patient in the baseline age group.

In the next figure we show a "switching" treatment effect. Figure 4 compares two hypothetical young AML patients receiving PBSCT (z_1) and BMT (z_2). The one-step and multi-step relapse probabilities were lower in the case of the PBSCT but the differences were not significant. On the other hand, we see that PBSCT associates with a lower probability of one-step transition into the death state, while in the case of multi-step transitions the role of the two graft sources is reversed. This pattern is also seen in the case ALL young patients in Figure 1, but in the case of AML patients the differences in the multi-step transition probabilities were more pronounced.

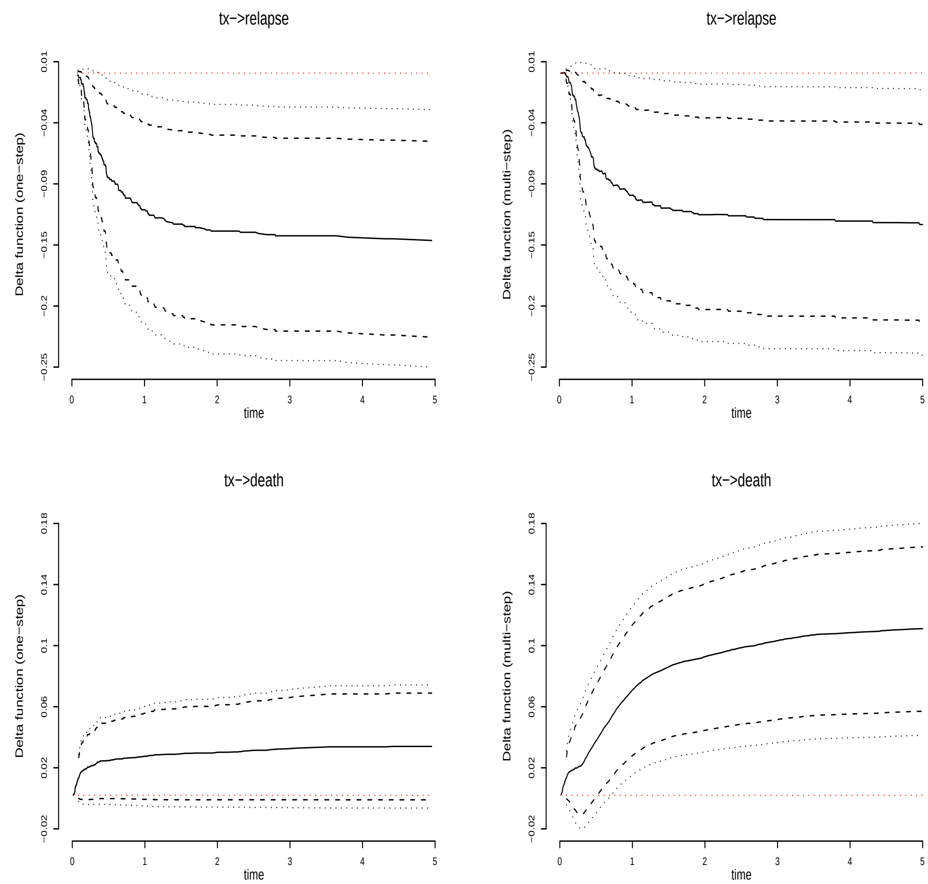

Figure 3: Pointwise and simultaneous confidence bands for the one-step and multi-step Δ functions of ALL patients receiving PBSCT.Covariates: $z_1 =$ age > 42.5 years, $z_2 =$ baseline age. The remaining covariates correspond to the baseline.

A similar approach can be applied to compare transition probabilities evaluated by conditioning on the follow-up history of a patient. In particular, Arjas and Eerola (1993) and Eerola (1994) have suggested the use of graphs of the conditional probabilities

$$P(J(t) = e|\mathscr{H}_s), \quad s < t \tag{4.5}$$

where $\mathscr{H}_s$ represents patient's history up-to time s. Examples of these graphs specialized to Markov chains and semi-Markov models were given in Klein *et al.*

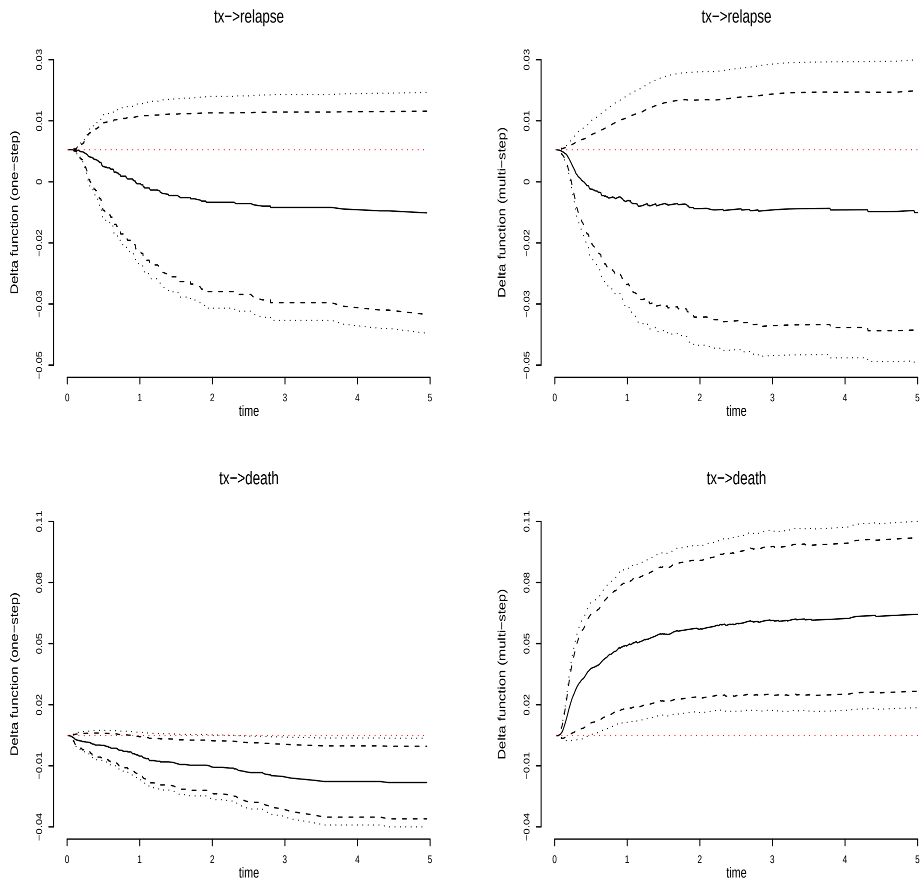

Figure 4: Pointwise and simultaneous confidence bands for the one-step and multi-step Δ functions of young AML patients. Covariates $z_1 = \text{PBSCT}$ $z_2 = \text{BMT}$. The remaining covariates correspond to the baseline.

(1993), Keiding *et al.* (2001), Dabrowska *et al.* (1994) and Putter *et al.* (2007). Here we note only that in the case of Markov chain regression models, the predictions depend only on the state occupied by the patient at time s and estimation of (4.5) reduces to estimation of the transition probability matrix because

$$P(J(t) = e|\mathscr{H}_s) = P(J(t) = e|J(s) = i, Z) \qquad \text{for} \quad s < t. \tag{4.6}$$

In the case of semi-Markov model, the conditional probabilities $P(J(t) = e|\mathscr{H}_s)$ are given by the transition probability matrix of a delayed Markov renewal process, with delay determined by the length of time spent on the state occupied at time s. On the other hand, the right-hand side of (4.6) depends also on the the initial state J_0, and all possible transitions leading to the state e and passing through the state i on or prior to time s. The two models coincide only if the sojourn times in each state are exponentially distributed.

In Table 5 we report results from analysis of residuals of the main variables in the model. We considered Martinussen and Scheike's Kolmogorov-Smirnov statistics for transitions between adjacent states of the model from each state. The test statistics were calculated in the range $t \in [1,90]$ months and the reported p-values were obtained using Gaussian multiplier method based on 5000 Monte Carlo samples. The results were also compared with a larger model, which included length of time spent in the transplant and AGVHD states as time dependent co-variates. The dependence on length of time spent in these states appeared to have marginal effect. In the case of the transitions originating from the CGVHD state, the latter may stem from a relatively small number of failures (relapse or death). On the other hand, AGVHD can occur only during the first 4 months and the state space of the process partially captures the dependence on the length of time spent in the transplant state. Although Table 5 shows an acceptable fit, there are several possible sources of departure from the model, In particular, they may be caused by calendar and center effects. For example, grading of acute and chronic GVHD is not uniform across centers. At the same time, the use of PBSCT in allogeneic transplants might have been more frequent towards the end of the study period than at its beginning. These factors were not taken into the account in this study as they identify patients in the population. Further, transplant may result in many other complications, including infections, pneumonia, as well secondary cancers, loss of vision and damage of other organs. We have not taken them into the account due to lack of data.

There has been very little work on variable and model selection problems in multistate models. Commenges *et al.* (2007) considered a flexible class of multi-state models which includes as special cases Markov chains and semi-Markov models. They extended the expected Kullback-Leibler (EKL) risk function to counting process models coarsened at random and proposed a leave-one-out cross-validation method for approximation of EKL based on penalized likelihoods. The approach was illustrated using a three state additive illness process, though the methodology applies to more complex situations as well. Another approach may be based on focused information criteria and model averaging of Hjort and Cleaskens (2003, 2006). Their approach is tailored towards selection of a model for given parameters of interest. In the case of single spell models, examples of such parameters include regression coefficients, quantiles, cumulative hazards or distribution functions evaluated at a fixed point or over a fixed interval. Extension of this method to multistate regression models may include one-step and multistep transition probabilities or other parameters arising in prediction problems.

Table 4: Regression coefficients

	1	2	3	4
ALL vs AML	.07 (.25)		1.32 (.36)	.50 (.23)
Age1	-.25 (.16)	-.68 (.28)	-.49 (.32)	-.45 (.26)
Age2				.27 (.20)
FM			-.20 (.28)	
PBSCT vs BMT		.09 (.22)	.01 (.29)	
ALLxPBSCT	.46 (.23)		.92 (.43)	
AMLxPBSCT				
AMLxBMT	-.30 (.22)			
DxCr	.12 (.08)		.45 (.16)	.22 (.12)
CrTx	-.21 (.07)	-.26 (.09)	-.33 (.15)	
Age1xBMT		-.57 (.33)		
Age2xPBSCT	-.28 (.25)			
Age2xBMT		-.37 (.24)		
Age0xPBSCT	-.27 (.26)			
AMLxPBSCTxAge2		.25 (.22)		
Age1xALL		.61 (.27)	-.87 (.48)	-.60 (.42)
Age2xALL		.57 (.30)	-.97 (.47)	
FMxALL		.63 (.27)		
FMxAML		.64 (.17)		
FMxPBSCT	.42 (.18)			.44 (.25)
FxALL	-.25 (.19)			
FxAML	-.19 (.14)			

Columns: 1 = Tx $\rightarrow$ AGVHD; 2 = Tx $\rightarrow$ CGVHD; 3 = Tx $\rightarrow$ Relapse; 4 = Tx $\rightarrow$ Death.

Rows: Age0: age in the $(29.5, 42, 5]$ range, Age1 = age $\leq$ 29.5 years, Age2 = age $>$ 42.5 years; F = female donor transplant; FM = female donor to male recipient transplant.

Table 4 (continued)

	5	6	7	8	9
ALL vs AML		.45 (.38)	.12 (.30)	.90 (.31)	.58 (.22)
Age2	.33 (.23)	.51 (.59)	.01 (.24)		.72 (.21)
FM	-.57 (.39)				
PBSCT vs BMT	.28 (.30)	-.12 (.66)	-.17 (.38)	.55 (.35)	
AMLxPBSCT	-.33 (.30)				
AMLxBMT			-.50 (.43)		
DxCr					.13 (.14)
CrTx				-.24 (.17)	-.10 (.12)
prioir AGVHD			.72 (.30)	.72 (.20)	
Age1xPBSCT			-.44 (.34)		
Age1xBMT		.70 (.60)	-.88 (.35)		-.64 (.34)
Age2xPBSCT				.13 (.37)	
Age2xBMT		.60 (.88)			
Age0xPBSCT	.40 (.31)				
FMxAML	.81 (.45)				

Columns: 5 = AGVHD $\rightarrow$ CGVHD; 6 = AGVHD $\rightarrow$ Relapse; 7 = AGVHD $\rightarrow$ Death; 8 = CGVHD $\rightarrow$ Relapse; 9 = CGVHD $\rightarrow$ Death.

Rows: Age0: age in the $(29.5, 42, 5]$ range, Age1 = age $\leq$ 29.5 years, Age2 = age $>$ 42.5 years; F = female donor transplant; FM = female donor to male recipient transplant.

Table 5: Kolmogorov-Smirnov residual statistics

	1	2	3	4	
AML	8.51 (.97)	6.35 (.96)	7.88 (.69)	5.87 (.86)	
ALL	10.22 (.89)	7.66 (.83)	7.65 (.69)	6.09 (.73)	
Age0	12.99 (.81)	6.52 (.93)	3.30 (.95)	6.50 (.63)	
Age1	6.42 (.98)	5.82 (.95)	3.20 (.94)	5.18 (.82)	
Age2	10.56 (.90)	7.64 (.91)	6.40 (.60)	8.11 (.72)	
BMT	7.84 (.97)	7.73 (.91)	7.16 (.69)	6.04 (.61)	
PBSCT	9.44 (.95)	9.07 (.90)	7.49 (.73)	5.75 (.82)	
non-FM	5.46 (.99)	7.84 (.94)	1.95 (.99)	9.82 (.66)	
FM	6.83 (.96)	9.36 (.82)	2.32 (.96)	9.52 (.43)	
M donor	5.48 (.99)	1.29 (.84)	2.96 (.98)	6.00 (.84)	
F donor	6.84 (.98)	11.95 (.80)	2.61 (.99)	5.83 (.86)	
	5	6	7	8	9
AML	4.40 (.96)	2.75 (.86)	4.14 (.97)	4.42 (.73)	6.70 (.82)
ALL	4.34 (.94)	2.66 (.85)	4.23 (.95)	4.49 (.71)	6.46 (.75)
Age0	4.55 (.92)	2.79 (.51)	4.37 (.93)	2.03 (.94)	8.17 (.47)
Age1	4.96 (.87)	3.66 (.71)	2.66 (.98)	2.04 (.95)	3.53 (.87)
Age2	6.35 (.83)	2.44 (.87)	2.92 (.99)	1.51 (.99)	8.07 (.76)
BMT	5.96 (.81)	1.36 (.99)	5.11 (.90)	2.94 (.79)	9.67 (.57)
PBSCT	6.66 (.88)	1.27 (.99)	4.84 (.95)	2.91 (.91)	9.39 (.69)
non-FM	4.22 (.97)	1.50 (.97)	4.57 (.96)	3.97 (.80)	4.20 (.95)
FM	5.06 (.88)	1.33 (.96)	5.07 (.84)	3.91 (.60)	4.37 (.88)
M donor	1.42 (.58)	2.02 (.94)	3.93 (.98)	2.06 (.97)	3.50 (.98)
F donor	11.50 (.53)	1.93 (.93)	4.43 (.95)	2.08 (.97)	3.47 (.97)

Each column provides test statistics and p-values determined based on 5000 re-sampling experiments.

Columns: 1 = Tx $\to$ AGVHD; 2 = Tx $\to$ CGVHD; 3 = Tx $\to$ Relapse; 4 = Tx $\to$ Death; 5 = AGVHD $\to$ CGVHD; 6 = AGVHD $\to$ Relapse; 7 = AGVHD $\to$ Death; 8 = CGVHD $\to$ Relapse ; 9 = CGVHD $\to$ Death.

Rows: Age0: age in the $(29.5, 42, 5]$ range, Age1 = age $\leq$ 29.5 years, Age2 = age $>$ 42.5 years. F = female donor transplant; FM = female donor to male recipient transplant.

Table 5 (continued)

	1	2	3	4	
DxCr-1	9.73 (.86)	16.86 (.42)	5.67 (.56)	7.88 (.49)	
DxCr-2	8.47 (.88)	11.67 (.59)	2.56 (.94)	2.86 (.98)	
DxCr-3	4.03 (1.00)	12.92 (.48)	4.34 (.71)	5.03 (.71)	
DxCr-4	11.80 (.83)	15.05 (.43)	4.01 (.89)	6.71 (.69)	
CrTx-1	14.69 (.74)	11.23 (.52)	5.19 (.66)	5.21 (.74)	
CrTx-2	9.92 (.85)	12.52 (.58)	3.96 (.81)	5.37 (.68)	
CrTx-3	12.08 (.76)	5.86 (.94)	5.96 (.62)	8.34 (.46)	
CrTx-4	8.84 (.87)	7.21 (.85)	2.94 (.92)	1.37 (.35)	
AMLxBMT	5.38 (.99)	9.84 (.76)	5.15 (.70)	1.62 (.41)	
AMLxPB	7.25 (.96)	6.30 (.97)	1.03 (.34)	5.04 (.80)	
ALLxBMT	7.62 (.85)	2.92 (.99)	4.44 (.74)	7.03 (.42)	
ALLxPBSCT	6.20 (.93)	6.39 (.69)	3.88 (.85)	3.01 (.86)	
	5	6	7	8	9
DxCr-1	4.92 (.85)	2.27 (.77)	5.17 (.81)	2.44 (.89)	5.34 (.74)
DxCr-2	6.80 (.61)	2.79 (.64)	5.93 (.73)	3.85 (.57)	4.72 (.74)
DxCr-3	9.88 (.33)	2.21 (.75)	4.27 (.88)	4.38 (.40)	11.23 (.30)
DxCr-4	6.15 (.71)	5.67 (.27)	12.11 (.43)	2.83 (.78)	3.52 (.94)
CrTx-1	6.53 (.74)	3.97 (.55)	7.24 (.76)	3.17 (.75)	2.43 (1.00)
CrTx-2	7.05 (.58)	2.61 (.66)	6.13 (.73)	3.49 (.67)	7.57 (.54)
CrTx-3	4.17 (.89)	2.52 (.67)	2.62 (.98)	2.34 (.86)	4.84 (.75)
CrTx-4	5.73 (.71)	4.21 (.37)	5.71 (.73)	3.59 (.57)	4.54 (.80)
AMLxBMT	5.12 (.82)	2.04 (.88)	5.46 (.78)	2.57 (.64)	7.96 (.50)
AMLxPBSCT	5.18 (.92)	1.84 (.87)	5.77 (.88)	3.18 (.84)	4.60 (.91)
ALLxBMT	2.44 (.97)	1.10 (.99)	2.79 (.95)	2.46 (.70)	3.22 (.85)
ALLxPB	4.62 (.82)	1.97 (.75)	3.24 (.95)	3.79 (.66)	5.92 (.62)

Each column provides test statistics and p-values determined based on 5000 re-sampling experiments.

Columns: 1 = Tx → AGVHD; 2 = Tx → CGVHD; 3 = Tx → Relapse;
4 = Tx → Death; 5 = AGVHD → CGVHD;
6 = AGVHD → Relapse; 7 = AGVHD → Death;
8 = CGVHD → Relapse ; 9 = CGVHD → Death.

Rows: DxCr-i and CrTX-i, i = 1,2,3,4: DxCr and CrTx variables grouped according to quartiles.

Table 5 (continued)

	1	2	3	4	
Age1xPB	14.92 (.71)	8.03 (.81)	6.58 (.63)	5.12 (.75)	
Age1xBMT	6.45 (.93)	6.12 (.84)	3.11 (.88)	7.13 (.51)	
Age2xPBSCT	7.31 (.94)	5.61 (.95)	7.69 (.34)	9.53 (.45)	
Age2xBMT	5.46 (.87)	6.96 (.68)	4.02 (.41)	4.58 (.76)	
Age0xPBSCT	4.29 (.98)	8.10 (.71)	4.73 (.67)	5.45 (.49)	
Age0xBMT	9.73 (.73)	9.06 (.59)	3.97 (.76)	8.60 (.22)	
Age1xAML	6.87 (.89)	5.47 (.87)	2.52 (.91)	2.49 (.98)	
Age1xALL	4.51 (.97)	2.73 (.99)	2.70 (.89)	3.19 (.74)	
Age2xAML	10.54 (.80)	7.95 (.83)	2.94 (.89)	3.69 (.88)	
Age2xALL	8.57 (.65)	5.01 (.60)	4.60 (.74)	3.34 (.74)	
Age0xAML	9.70 (.88)	9.63 (.80)	7.64 (.34)	8.95 (.58)	
Age0xALL	3.75 (.95)	2.97 (.94)	3.75 (.52)	2.10 (.94)	
	5	6	7	8	9
Age1xPBSCT	7.56 (.58)	1.63 (.84)	5.31 (.86)	3.28 (.71)	6.58 (.57)
Age1xBMT	4.22 (.78)	2.49 (.80)	2.01 (.96)	2.94 (.56)	3.09 (.76)
Age2 x PBSCT	4.36 (.93)	1.28 (.96)	3.74 (.95)	1.26 (1.00)	9.60 (.57)
Age2xBMT	3.66 (.77)	1.95 (.78)	2.37 (.93)	.93 (.91)	4.43 (.72)
Age0xPBSCT	5.34 (.74)	1.88 (.61)	2.46 (.98)	1.70 (.94)	3.58 (.75)
Age0xBMT	6.34 (.42)	1.76 (.58)	4.03 (.85)	3.31 (.24)	5.82 (.45)
Age1xAML	3.22 (.94)	3.98 (.34)	2.73 (.86)	2.39 (.59)	3.29 (.68)
Age1xALL	2.59 (.96)	2.18 (.82)	3.95 (.78)	3.88 (.50)	2.06 (.93)
Age2xAML	5.19 (.78)	3.98 (.17)	4.89 (.81)	1.78 (.92)	7.34 (.38)
Age2xALL	2.13 (.96)	1.74 (.29)	4.09 (.78)	1.94 (.67)	1.69 (.97)
Age0xAML	4.92 (.89)	3.30 (.63)	5.20 (.86)	3.54 (.67)	4.12 (.95)
Age0xALL	3.15 (.76)	1.27 (.86)	3.16 (.81)	2.32 (.67)	5.76 (.52)

Each column provides test statistics and p-values determined based on 5000 re-sampling experiments.

Columns: 1 = Tx $\rightarrow$ AGVHD; 2 = Tx $\rightarrow$ CGVHD; 3 = Tx $\rightarrow$ Relapse;
4 = Tx $\rightarrow$ Death; 5 = AGVHD $\rightarrow$ CGVHD;
6 = AGVHD $\rightarrow$ Relapse; 7 = AGVHD $\rightarrow$ Death;
8 = CGVHD $\rightarrow$ Relapse ; 9 = CGVHD $\rightarrow$ Death.

Rows: Age0: age in the $(29.5, 42, 5]$ range, Age1 = age $\leq$ 29.5 years,
Age2 = age $>$ 42.5 years.

5 Proofs

5.1 Assumptions and notation. We first recall that if $A = [a_{k\ell}]$ is a rectangular $d \times q$ matrix then its ℓ_1 and ℓ_∞ norms are given by

$$\|A\|_1 = \max_\ell \sum_{k=1}^d |a_{k\ell}| \quad \text{and} \quad \|A\|_\infty = \max_k \sum_{\ell=1}^q |a_{k\ell}|,$$

and we have $\|A\|_1 = \sup\{\mu^T A\lambda : \|\mu\|_\infty \leq 1, \|\lambda\|_1 \leq 1\} = \|A^T\|_\infty$, where $\mu = (\mu_1, \ldots, \mu_d)^T$ and $\lambda = (\lambda_1, \ldots, \lambda_q)^T$. If $A(s) = [a_{ij}(s)], s = (x, \theta)$ is a $d \times q$ matrix of functions defined on $\mathscr{T} = [0, \tau] \times \Theta$ then $\|A\| = \sup\{\|A(s)\|_1 : s \in \mathscr{T}\}$ is the corresponding supremum norm, and with some abuse of notations, we write $\|A\| = \sup\{\|A(s)\|_\infty : s \in \mathscr{T}\}$. We also use $\|\cdot\|$ to denote the supremum norm of scalar or vector-valued functions on $[0, \tau]$.

We shall assume the following regularity conditions on the hazard rates $\alpha_j(y, \theta, z), y \in R^q$, $j \in \mathscr{J}_0$.

Condition 5.1 (i) The parameter set $\Theta \subset R^d$ is bounded and open.

(ii) For fixed $z \in R^d$, the function $\ell_j(y, \theta, z) = \log \alpha_j(y, \theta, z)$, $j \in \mathscr{J}_0$ is twice continuously differentiable with respect to (y, θ). The derivatives with respect to y (denoted by primes) and with respect to θ (denoted by dots) satisfy $\|\ell_j'(y, \theta, z)\|_1 \leq \psi(\|y\|_1)$, $\|\ell_j''(y, \theta, z)\|_1 \leq \psi(\|y\|_1)$, $\|\dot{\ell}_j(y, \theta, z)\|_1 \leq \psi_1(\|y\|_1)$, $\|\ddot{\ell}_j(y, \theta, z)\|_1 \leq \psi_2(\|y\|_1)$ and $\|g(y, \theta, z) - g(y', \theta', z)\|_1 \leq \max(\psi_3(\|y\|_1), \psi_3(\|y'\|_1)) \times$ $\times [\|y - y'\|_1 + |\theta - \theta'|]$, where $g = \ddot{\ell}_j, \dot{\ell}_j'$ and ℓ_j''. Here ψ is a constant or a continuous bounded decreasing function. The functions $\psi_p, p = 1, 2, 3$ satisfy $\psi_p(0) < \infty$, are continuous and locally bounded.

(iii) For fixed $\theta \in \Theta$ and $y \in R^q$, the functions $\alpha_j(y, \theta, \cdot)$ and their logarithmic derivatives in (ii) are measurable with respect to the Borel σ-field of R^d.

(iv) We have either a) $m_1 < \alpha_j(y, \theta, z) < m_2$ for some $0 < m_1 < m_2 < \infty$ or b) $\alpha_j(y, \theta, z)$ is a bounded coordinate-wise decreasing function such that $\alpha_j(y_1, \ldots, y_k, \theta, z) \downarrow 0$ as $y_\ell \uparrow \infty$, $\ell = 1, \ldots, q$, and

$$m_1[1 + c_1\|y\|_1]^{-e_1} \leq \alpha_j(y, \theta, z) < m_2[1 + c_2 y_j]^{-e_2}, j = 1, \ldots, q$$

for some $c_1, c_2 > 0$, $e_1 \in (0, 1]$, $e_2 \in [0, 1]$ and $0 < m_1 < m_2 < \infty$.

The condition (ii) assumes that the function $\alpha(y, \theta, z)$ and its derivatives are jointly continuous in the arguments (y, θ). Together with the condition (iii), this implies that they are measurable with respect to the Borel σ-field of $\mathscr{B}(R^q) \otimes \mathscr{B}(\Theta) \otimes \mathscr{B}(R^d)$. The condition 5.1 (iii) serves to ensure that for each state $j_1 \in \mathscr{J}$, the Volterra equation corresponding to the transitions $j \in \mathscr{J}_0$ originating from the state j_1 has a non-explosive solution on the interval $[0, \tau_{j_1}] = \sup\{x : EY_j(x) > 0\}$.

Let P be a distribution satisfying assumptions 2.1 of Section 2. For any $j \in \mathscr{J}_0$ let

$$A_{jP}(x) = \int_0^x \frac{E_P N_{j.i}(du)}{E_P Y_{j.i}(u)}$$

and set $A_{.P} = \Sigma_{j \in \mathscr{J}_0} A_{jP}$. In analogy to the single spell models in Dabrowska (2006), we can show that the condition 5.1 (iii-a) implies that, the Volterra equation has a unique solution $\Gamma_\theta = [\Gamma_{1\theta}, \ldots, \Gamma_{q\theta}]^T$ such that $m_2^{-1} A_{jP}(x) \leq \Gamma_{j\theta}(x) \leq m_1^{-1} A_{jP}(x)$ for $x \in [0, \tau]$, $\theta \in \Theta$. In addition, there exist positive constants d_1, d_2, d_3 such that

$$\begin{aligned} \|\Gamma_\theta(x) - \Gamma_{\theta'}(x)\|_1 &\leq |\theta - \theta'| d_1 \exp[d_2 A_{.P}(x)], \\ |\Gamma_{j\theta}(x) - \Gamma_{j\theta}(x')| &\leq d_3 E_P N_{j.i}((x \wedge x', x \vee x']). \end{aligned} \tag{5.1}$$

Similar inequalities hold also for the left continuous version of $\Gamma_{j\theta}$. On the other hand, under the condition 5.1.(iii-b), we have $\Phi_2(A_{jP}(x)) \leq \Gamma_{j\theta}(x) \leq \Phi_1(A_{.P}(x))$, where $\Phi_q(u) = c_q^{-1}([1 + c_q u/m'_q]^{1/1-e_q} - 1)$ for $q = 1, 2$ and $m'_q = m_q/(1 - e_q)$ if $e_q \neq 1$, and $\Phi_q(u) = c_q^{-1}(e^{c_q u/m_q} - 1)$ if $e_q = 1$. The functions Φ_q are inverse cumulative hazards corresponding to the lower and upper bounds on hazard rates in the condition 5.1 (iii). The inequality (5.1) is in this case satisfied with the function $A_{.P}$ replaced by $\Phi_1(A_{.P})$.

5.2 Some measurability issues. In section 2, we assumed that the observations $D_1, \ldots, D_n$ of the censored modulated renewal process are defined on a common complete probability space $(\Omega, \mathscr{F}, P)$ and take on values in a separable measure space $(\mathbb{S}, \mathscr{S})$. A measure space is here called separable if its σ-field is countably generated and contains all singletons. Any such space is measurably isomorphic to a subspace of the real line equipped with its Borel σ-field (e.g. Dellacherie-Meyer, 1975, p.15). Let $(\mathbb{S}_n, \mathscr{S}_n)$ and $(\mathbb{S}_\infty, \mathscr{S}_\infty)$ be the corresponding n-fold and infinite product spaces and let P_n and P_∞ be the corresponding product measures on $\mathscr{S}_n$ and $\mathscr{S}_\infty$ induced by $(D_1, \ldots, D_n)$ and $D = (D_1, D_2, \ldots, D_n, \ldots)$, respectively. We denote by $\mathscr{S}_n^P$ the sigma-field of subsets $A \subset \mathbb{S}_n$ measurable in the completion of the product probability measure P_n and by $\mathscr{S}_n^u$ the universal sigma-field generated by $\mathscr{S}_n$, i.e. the sigma-field of subsets measurable in the completion of any probability measure Q on $\mathscr{S}_n$. We have $\mathscr{S}_n \subseteq \mathscr{S}_n^u \subseteq \mathscr{S}_n^P$. Whereas $\mathscr{S}_n$ is not complete with respect to the product measure P_n, any set $A \in \mathscr{S}_n^P$ satisfies $P_n^*(A) = P_{n,*}(A)$ and $g_n^{-1}(A) \in \mathscr{F}$ for $g_n = (D_1, \ldots, D_n)$. The sigma-fields $\mathscr{S}_\infty^P$ and $\mathscr{S}_\infty^u$ have similar property. Without much loss of generality, we can assume therefore that $(\Omega, \mathscr{F}) = (\mathbb{S}_\infty, \mathscr{S}_\infty)$ and, when necessary, require measurability with respect to these larger sigma-fields. With this choice the sequence D is the identity map on $\mathbb{S}_\infty$ and $(D_1,\ldots,D_n)$ are the corresponding coordinate projections on $(\mathbb{S}_n,\mathscr{S}_n)$.

Further, let $(\Omega_0,\mathscr{F}_0)$ be an arbitrary measure space let $\mathscr{Z}$ be a Polish space or a Borel subset of it. For any set $A\subset\Omega_0\times\mathscr{Z}$, its projection on Ω_0 is denoted by $\mathrm{proj}_{\Omega_0}(\mathrm{A})=\{\omega_0:(\omega_0,\mathrm{z})\in\mathrm{A}\text{ for some }\mathrm{z}\in\mathscr{Z}\}$. A multifunction (or correspondence) is a set-valued function assigning to each $\omega_0\in\Omega_0$ a subset of $\mathscr{Z}$. We shall write $H:\Omega_0\hookrightarrow\mathscr{Z}$ for such mappings to differentiate them from usual functions assigning to each ω_0 a single value ($h:\Omega_0\to\mathscr{Z}$). The domain and graph of a multifuction H are defined as

$$\mathrm{dom}H=\{\omega_0:H(\omega_0)\neq\emptyset\}\quad\text{and}\quad\mathrm{graph}H=\{(\omega_0,z):z\in H(\omega_0)\},$$

respectively. For any nonempty set $B\subseteq\mathscr{Z}$, the inverse image of H is given by

$$H^{-1}(B)=\{\omega_0:H(\omega_0)\cap B\neq\emptyset\}=\{\omega_0:z\in H(\omega_0)\text{ for some }z\in B\}$$

and the right side is equal to the projection $\mathrm{proj}_{\Omega_0}(\mathrm{graphH}\cap\Omega_0\times\mathrm{B})$. Finally, by a selector we mean a function $h:\Omega_0\to\mathscr{Z}\cup\{z^*\}$ such that $h(\omega_0)\in H(\omega_0)$ if $\mathrm{dom}H\neq\emptyset$ and $h(\omega_0)=z^*$, otherwise. (Here z^* is an extra point attached to $\mathscr{Z}$).

A set-valued mapping H is here called measurable if graph$H$ is jointly measurable with respect to $\mathscr{F}_0\otimes B(\mathscr{Z})$. By measurable projection theorems (e.g. Dellacherie and Meyer, 1975, p.252, Pollard, 1984, p. 196-197 or Dudley, 1999, Chapter 5), the joint measurability of graph$H$ entails that the inverse image $H^{-1}(B)$ of any Borel set $B\in\mathscr{B}(\mathscr{Z})$ belongs to the universal sigma field $\mathscr{F}_0^u$ generated by $\mathscr{F}_0$. Moreover, H admits at least one $\mathscr{F}_0^u$-measurable selector. If $\mathscr{F}_0$ is complete with respect to some probability measure then $\mathscr{F}_0^u=\mathscr{F}_0$. For alternative conditions for this equality we refer to Wagner (1976).

Further, let $\mathscr{T}$ be a Polish space and let $\{X_t:t\in\mathscr{T}\}$ be an R^k-valued random element defined on Ω_0. We refer to it as measurable if it forms a measurable stochastic process, i.e. the map $\Omega_0\times\mathscr{T}\ni(\omega,t)\to X(\omega,t)\in R^k$ is jointly measurable with respect to the σ-fields $\mathscr{F}_0\otimes\mathscr{B}(\mathscr{T})$ and $\mathscr{B}(R^k)$. Correspondingly, the set valued function $H:\Omega_0\hookrightarrow\mathscr{Z}=\mathscr{T}\times R^k$ given by $H(\omega_0)=\{(t,X(\omega_0,t):t\in\mathscr{T}\}$ has a measurable graph and for any Borel sets $B\in\mathscr{B}(R^k)$ and $C\in\mathscr{B}(\mathscr{T})$, we have $\{\omega_0:X(\omega_0,t)\in B\text{ for some }t\in C\}\in\mathscr{F}_0^u$. In section 5.3, we use that an R^k-valued process is measurable iff each of its components is measurable. Moreover, sums and products of such processes are measurable as well.

A class of scalar functions $\mathscr{G}=\{g_t(s):t\in\mathscr{T}\}$ defined on $\mathbb{S}_k, k\leq n$ is called here measurable if it forms a measurable process in the above sense. Following Nolan and Pollard (1987) and Pollard (1990), a measurable class of functions $\mathscr{G}$ is called Euclidean for an envelope G if $|g_t|(s)\leq G(s)$ for all $t\in\mathscr{T}$, and there exist constants A and V such that $N(\varepsilon\|G\|_{Q,r}, \mathscr{G}, \|\cdot\|_{Q,r}) \leq (A/\varepsilon)^V$ for all $\varepsilon \in (0,1)$ and all probability measures Q on $\mathscr{S}_k$ such that $\|G\|_{Q,r} < \infty$. Here $N(\eta, \mathscr{G}, \|\cdot\|_{Q,r})$ is the minimal number of $L_r(Q)$–balls of radius η covering the class $\mathscr{G}$ and $\|\cdot\|_{Q,r}$ is the $L_r(Q)$ norm. We use $r = 1,2$ in the sequel.

In our application the space $(\mathbb{S}, \mathscr{S})$ can be taken as the complete separable metric space $(\mathbb{S}, \mathscr{S}) = (E_0, \mathscr{B}(E_0)) \times (E_1 \times \mathscr{B}(E_0))^{\mathbb{N}}$, where $E_0 = \mathscr{J} \times R^d$ $E_1 = (\overline{R}^+ \times (\mathscr{J} \times R^d) \cup \Delta)^{\mathbb{N}}$. Here E_0 represents possible initial realizations of the mark $V_0 = (J_0, Z_0)$ and E_1 is the space of realizations of the censored modulated renewal process $(X_m, V_m = (J_m, Z_m))_{m\geq 1}$. Further, $\mathscr{T} = [0,\tau] \times \Theta$, where τ is a finite point on the positive half-line and Θ is a bounded open subset of a Euclidean space. Here $\mathscr{T}$ is a Polish space because $\mathscr{T}$ forms a $\mathbf{G}_\delta$ subset (a countable intersection of open sets) of a Polish space and Polishness is hereditary with respect to $\mathbf{G}_\delta$ sets. Finally, all classes $\mathscr{G} = \{g_t(s) : t = (x,\theta) \in \mathscr{T}\}$ correspond to cádlág (or cáglád) functions such that for $0 \leq x < x' \leq \tau$ and $\theta, \theta' \in \Theta$, we have

$$
\begin{aligned}
|g_{x\theta}(s) - g_{x'\theta}(s)| &\leq C_1[\widetilde{G}(x', s) - \widetilde{G}(x, s)], \\
|g_{x\theta}(s) - g_{x\theta'}(s)| &\leq C_2|\theta - \theta'|\widetilde{G}(\tau, s),
\end{aligned}
\tag{5.2}
$$

where $\widetilde{G}(s,x)$ is a nonnegative monotone increasing cádlág (respectively cáglád) function of x such that $\widetilde{G}(s,0) = 0$ and $\|\widetilde{G}(\tau,\cdot)\|_{Q,r} < \infty$. In this case, the Euclidean property is satisfied with envelope $G(s) = [C_1 + C_2 \mathrm{diam}\,\Theta]\widetilde{G}(\tau,s) + g_{x_0\theta_0}(s)$, where $g_{x_0\theta_0}(s)$ is an arbitrary function from the class $\mathscr{G}$.

To verify measurability of the estimates, we shall need some properties of Carathéodory integrands and cádlág or cáglád functions. If $\mathscr{T}$ and $\mathscr{Y}$ are Polish spaces then a function $f : \Omega_0 \times \mathscr{T} \to \mathscr{Y}$ is called a Carathéodory integrand if for fixed $t \in \mathscr{T}$, $f(\cdot,t) : \Omega_0 \to \mathscr{Y}$ is measurable, and for fixed $\omega_0 \in \Omega_0$, $f(\omega_0,\cdot) : \mathscr{T} \to \mathscr{Y}$ is continuous. Here $(\Omega_0, \mathscr{F}_0)$ is an arbitrary measure space and we have

Lemma 5.1 Let $f : \Omega_0 \times \mathscr{T} \to \mathscr{Y}$ be a Carathéodory mapping. Then

(i) f is measurable with respect to $\mathscr{F}_0 \times \mathscr{B}(\mathscr{T})$.

(ii) For any open set B of $\mathscr{Y}$, let $H(\omega_0) = \{t : f(\omega_0,t) \in B\}$. Then for any closed or open C set of $\mathscr{T}$, we have $H^{-1}(C) = \{\omega_0 : f(\omega_0,t) \in B \text{ for some } t \in C\} \in \mathscr{F}_0$.

(iii) If $g : \Omega_0 \to \mathscr{T}$ is measurable, then the composite mapping $f \circ g : \Omega_0 \to \mathscr{Y}$ given by $(f \circ g)(\omega_0) = f(\omega_0, g(\omega_0))$ is measurable.

(iv) Suppose that $\mathscr{Y}'$ is another Polish space and $h : \Omega_0 \times \mathscr{T} \times \mathscr{Y}' \to \mathscr{Y}$ is a Carathéodory integrand. Then the composite map $(h \circ f) : \Omega_0 \times \mathscr{T} \to \mathscr{Y}$ given by $(h \circ f)(\omega_0,t) = h(\omega_0,t,f(\omega_0,t))$ is a Carathéodory integrand.

Part (i) remains valid even if $\mathscr{Y}$ is replaced by a nonseparable metric (Kuratowski 1966 p. 378 , or Himmelberg, 1975). In part (ii), if C is a closed set then

$$H^{-1}(C) = \bigcup_{q \in C} \{\omega_0 : f(\omega_0, q) \in B\},$$

where the union is over a dense subset of C. If C is open then it can be represented as a countable increasing union of closed sets and part (ii) follows by noting that inverse images preserve unions of sets. Part (iii) follows from the definition of a measurable function and continuity of f with respect to t. Part (iv) follows from part (i) and (iii) and definition of a continuous function.

Part (i) of the lemma extends to functions f which are cádlág, cáglád, cád and cág in $t \in \mathscr{T}$, $\mathscr{T} = R_+$ or $\mathscr{T} = [0, \tau]$ and take on values in a complete separable metric space (e.g. Dellacherie and Meyer, 1975 p. 144). Any cádlág or cáglád function is also a pointwise limit of Carathéodory integrands.

Finally, suppose that $\mathscr{T} = [0, \tau] \times \Theta$ and f is a function such that (i) for fixed $(x, \theta) \in \mathscr{T}$, $f(\cdot, x, \theta)$ is the $\mathscr{F}_0$ measurable and (ii) for fixed $\omega_0 \in \Omega_0$, it is jointly cádlág with respect to (x, θ) and continuous with respect to θ. To see that f is jointly measurable, let $\{q_k : k \geq 1\}$ be a dense set in Θ and for given integer $m \geq 1$ let B_{mk} be a balls of radius $1/m$ centered at q_k covering Θ. Set $B'_{mk} = B_{mk} - \bigcup_{r=1}^{k-1} B_{mr}$ and

$$f_m(\omega_0, x, \theta) = \sum_{\ell=1}^{\infty} \sum_{k=1}^{\infty} f(\omega, \frac{\ell}{m} \wedge \tau, q_k) 1(\frac{\ell - 1}{m} \leq x < \frac{\ell}{m}) 1(\theta \in B'_{mk}).$$

Then f_m is joinly measurable and pointwise converges to f. Similarly, if f is jointly cáglád rather than cádlág function in (ii) then f is a jointly measurable with respect to $\mathscr{F}_0 \otimes \mathscr{B}(\mathscr{T})$. Similarly to the single parameter case, functions of this type are pointwise limits of Carathéodory integrands. Part (ii) of the lemma remains valid for sets of the form $C = I \times C'$, where C' is an open or closed subset of Θ and I is an interval contained in $[0, \tau]$. In particular, if f is a real valued cádlág function of this type then its supremum is $\mathscr{F}_0$ measurable.

5.3 Proof of Proposition 3.1. To show proposition 3.1, we shall first consider the process $\Gamma_{n\theta}(x), (x, \theta) \in [0, \tau] \times \Theta = \mathscr{T}$.

Lemma 5.2 (i) The process $\widehat{W} = \{\widehat{W}(t) = [\widehat{W}_j(t) : t = (x, \theta) \in \mathscr{T}, j \in \mathscr{J}_0\}$, $\widehat{W}_j(x, \theta) = \sqrt{n}[\Gamma_{nj\theta} - \Gamma_{j\theta}](x)$, converges weakly in $\ell^\infty(\mathscr{T} \times \mathscr{J}_0)$ to

$$W(x, \theta) = V(x, \theta) - \int_{[0,x]} V(u-, \theta) s'(\Gamma_\theta(u-), \theta, u) C_\theta(du) \mathscr{P}_\theta(u, x),$$

where $\{V(t) = [V_j(t) : t = (x,\theta) \in \mathscr{T}, j \in \mathscr{J}_0\}$ is a tight mean zero Gaussian process. Its covariance function is given by

$$\text{cov}(V_j(x,\theta), V_{j'}(x',\theta')) = E\sum_m\sum_{m'}\int_0^x\int_0^{x'} \frac{M_{jm}(du,\theta)M_{j'm'}(dv,\theta')}{s_j(\Gamma_\theta(u),\theta,u)s_{j'}(\Gamma_{\theta'}(v),\theta',v)}.$$

In addition, under the assumption that observations correspond to a censored modulated renewal process and $\theta = \theta_0$ is the true parameter, $\text{cov}(V_j(x,\theta), V_{j'}(x',\theta)) = 1(j = j')C_{j\theta}(x \wedge x')$.

(ii) Let θ_0 be an arbitrary point in Θ. If $\widehat{\theta}$ is a $\sqrt{n}$-consistent estimate of it, then the process $\widehat{W}_0 = \{\widehat{W}_0(x) : x \leq \tau\}$, $\widehat{W}_0 = \sqrt{n}[\Gamma_{n\widehat{\theta}} - \Gamma_{\theta_0} - (\widehat{\theta} - \theta_0)^T\dot{\Gamma}_{n\widehat{\theta}}]$ converges weakly in $\ell^\infty([0,\tau] \times \mathscr{J}_0)$ to $W_0 = W(\cdot,\theta_0)$.

Here the space $\mathscr{X} = \ell^\infty(\mathscr{T} \times \mathscr{J}_0)$ is equipped with uniform metric, $d_X(\widetilde{x},\widetilde{y}) = \sup_{t,j}|\widetilde{x}(t,j) - \widetilde{y}(t,j)|$ and is isometric to the space $\mathscr{Y} = \ell^\infty(\mathscr{T})^q$ equipped with metric $d_Y(x,y) = \max_j \sup_t |x_j(t) - y_j(t)|$. Apparently, the isometry is given by the mapping Φ assigning to each $\widetilde{x} \in \mathscr{X}$ the vector of coordinate functions, $\Phi(\widetilde{x}) = [\widetilde{x}(\cdot,1),\ldots,\widetilde{x}(\cdot,q)]^T$. Open sets of $\mathscr{X}$ can be represented as arbitrary unions of balls $\mathscr{B}_X(\widetilde{x},\varepsilon) = \{y : d_X(x,y) < \varepsilon\}$. On the other hand, the product topology of $\mathscr{Y}$ coincides with the topology induced by the metric d_Y so that any open set in the product topology is an arbitrary union of balls $\mathscr{B}_Y(x,\varepsilon)$, where $x = [x_1,\ldots,x_q]$.

Proof. To show part (i), define $V_n = [V_{jn} : j \in \mathscr{J}_0]$, where

$$V_{jn}(x,\theta) = \int_{(0,x]} \frac{N_{j..}(du)}{S_j(\Gamma_\theta(u-),\theta,u)} - \int_{(0,x]} \frac{EN_{j..}(du)}{s_j(\Gamma_\theta(u-),\theta,u)}.$$

Then $V_{jn} = V_{1jn} + \text{rem}_j$, where

$$V_{1jn}(x,\theta) = \frac{1}{n}\sum_{i=1}^n \int_{(0,x]} \left[\frac{N_{j.i}(du)}{s(\Gamma_\theta(u-),\theta,u)} - \frac{S_{ji}}{s_j^2}(\Gamma_\theta(u-),\theta,u)EN_{j..}(du)\right]$$

and $\text{rem}_j(x,\theta)$ is a remainder term. Lemma 5.3 gives its form and shows that $\|\text{rem}_j\| = o_P(n^{-1/2})$. Therefore the process $V_{1n} = [V_{1jn} : j \in \mathscr{J}_0]$ satisfies also $\|V_n - V_{1n}\| = o_P(n^{-1/2})$.

Using CLT and Cramer-Wold device, the finite dimensional distributions of $\sqrt{n}V_{1n}$ converge in distribution to finite dimensional distributions of V: for any distinct $t_1,\ldots,t_k \in \mathscr{T}$ and any numerical vector λ of length kq, the random variable $\lambda^T vec[V_{1n}(t_1),\ldots,V_{1n}(t_k)]$ converges in distribution to the corresponding linear combination of finite dimensional marginals of V.

For each $j \in \mathscr{J}_0$, the process V_{1jn} can be represented as $V_{1jn}(x,\theta) = [\mathbb{P}_n - P]g$, where g varies over a class $\mathscr{G}_j = \{g_{tj} : t = (x,\theta) \in \mathscr{T}\}$ consisting of cádlág functions such that each g_{tj} is a difference of two càdlàg functions, increasing in x and Lipschitz continuous with respect to θ. Setting $\widetilde{G}_j(D_i,x) = N_{j.i}(x) + \int_0^x Y_{j.i}(u)A_j(du)$, the condition (5.2) is satisfied with constants C_1 and C_2 determined by the functions ψ, ψ_1 of the condition 5.2 (ii) and $g_{t_0} \equiv g_{\tau,\theta_0}$, say. Correspondingly, the class $\mathscr{G}_j$ is Euclidean for a square integrable envelope G_j. From Pollard (1984,1990) it follows that the process $\sqrt{n}V_{1jn}$ converges weakly in $\ell^\infty(\mathscr{G}_j)$ to V_j, the j-th component of the process V because the class $\mathscr{G}_j$ is totally bounded and asymptotically uniformly equicontinuous with respect to the variance pseudo-metric $d_j(t,t') = sd(V_{1jn}(t) - V_{1jn}(t'))$, $t,t' \in \mathscr{T}$. Joint weak convergence of the process $\sqrt{n}V_n = \sqrt{n}(P_n - P)g$, $g \in \bigcup_j G_j$ follows from finite dimensional weak convergence and by noting that union of a finite number Euclidean classes of functions is also Euclidean (Pollard, 1990). In particular, the class $\mathscr{G}$ is totally bounded and asymptotically equicontinuous with respect to the variance pseudo-metric $d((t,j),(t',j')) = sd(V_{1nj}(t) - V_{1nj'}(t'))$. Denoting by V_n^- the left-continuous process (obtained by changing the integrals over $(0,x]$ to integrals over intervals $[0,x)$), the process $\sqrt{n}V_n^-$ converges weakly to V as well because the jumps of the process V_n are of the order $O_p(1/n)$ unifromly in $t \in \mathscr{T}$ and the functions EN_j are continous.

Finally, to show weak convergence of the standardized $\Gamma_{n\theta}$ process, we shall need bounds on the supremum of the norm of the vector V_n. Let $\mathscr{H}$ denote the class of functions $\mathscr{H} = \{h(\lambda,t) = \sum_{j=1}\lambda_j g_{tj} : g_{tj} \in \mathscr{G}_j, |\lambda_j| \le 1, j = 1,\dots,q\}$. Then $\mathscr{H}$ forms a Euclidean class for the envelope $H = \sum_j G_j$ and we have

$$E \sup_{t\in\mathscr{T}} \|\sqrt{n}V_{1n}(t)\|_1 = E \sup_{h\in\mathscr{H}} \sqrt{n}|\mathbb{P}_n - P|h = O(1).$$

Similarly, $E\sup_{t\in\mathscr{T}} \|\sqrt{n}V_{1n}(t)\|_\infty = O(1)$ and the left-continuous versions of the process satisfy similar bounds.

To show consistency of the estimate $\Gamma_{n\theta}$, we first assume the condition 5.1. (iii-a). Let A_{jn} be the Aalen-Nelson estimator. Let $A_{pjn} = m_p^{-1}A_{jn}, p = 1,2$. Then $A_{2jn}(x) \le \Gamma_{nj\theta}(x) \le A_{1jn}(x)$ for all $\theta \in \Theta$ and a similar algebra as in Dabrowska (2006) shows that

$$|\Gamma_{nj\theta}(x) - \Gamma_{j\theta}(x)| \le |V_{jn}(x,\theta)| + \int_{(0,x]} \|\Gamma_{n\theta} - \Gamma_\theta\|_1(u-)\rho_{jn}(du),$$

where $\rho_{jn} = \max(c_j,1)A_{1jn}$ for some constant c_j. Therefore $\|\Gamma_{n\theta} - \Gamma_\theta\|_1(x) \le \|V_n(x,\theta)\|_1 + \int_{(0,x]} \|\Gamma_{n\theta} - \Gamma_\theta\|_1(u-)\rho_n(du)$, where $\rho_n = \sum_j \rho_{nj}$. Gronwall's inequality (Beesack, 1973, Dabrowska, 2006) implies that $\sup_{x,\theta} \exp[-\rho_n(x)]\|\Gamma_{n\theta} - \Gamma_\theta\|_1(x) \to 0$ a.s., where the supremum is over $\theta \in \Theta$

and $x \in [0,\tau]$. In the case of the condition 5.1.(iii-b), the proof is the same, except that the function ρ_{jn} is replaced by $\rho_j = \max(c_j,1)\Phi_1(A_{.n})$, where $A_{.n} = \Sigma A_{jn}$. Note that Aalen-Nelson estimate is a measurable process, whereas measurability of the process $\Gamma_{n\theta}$ is verified below.

The process $\widehat{W}(x,\theta) = \sqrt{n}[\Gamma_{n\theta} - \Gamma_\theta]^T(x)$ satisfies

$$\widehat{W}(x,\theta) = \sqrt{n}V_n(x,\theta) - \int_{(0,x]} \widehat{W}(u-,\theta)\widetilde{b}_{n\theta}(u)\overline{N}(du),$$

where $\overline{N}(x)$ is the diagonal matrix $\overline{N}(x) = \text{diag}\,[N_{1..}(x),\ldots,N_{q..}(x)]$, and $\widetilde{b}_{n\theta}(u)$ is a $q\times q$ matrix with columns

$$\widetilde{b}_{jn\theta}(u) = \left[\int_0^1 \left(S_j'/S_j^2\right)(\theta,\Gamma_\theta(u-) + \lambda[\Gamma_{n\theta} - \Gamma_\theta](u-),u)d\lambda\right].$$

Let $b_\theta(u)$ be a $q\times q$ matrix with columns $b_{j\theta}(u) = [s_j'/s_j^2](\Gamma_\theta(u),\theta,u)$. Using consistency of $\Gamma_{n\theta}$ and Lemma 5.3, we have $[\widetilde{b}_{n\theta} - b_\theta](u) \to 0$ a.s. uniformly in $(u,\theta) \in \mathscr{T}$. Moreover, (5.1) and (5.2) imply also that $\|R_{1n}\| \to 0$ a.s., where

$$R_{1n}(x,\theta) = \int_{(0,x]} b_\theta(u)[\overline{N} - E\overline{N}](du).$$

Define

$$\widetilde{W}(x,\theta) = \sqrt{n}V_n(x,\theta) - \int_{(0,x]} \widetilde{W}(u-,\theta)b_\theta(u)E\overline{N}(du).$$

Then

$$\begin{aligned}\widetilde{W}(x,\theta) &= \sqrt{n}V_n(x,\theta) - \int_{(0,x]} \sqrt{n}V_n(u-,\theta)b_\theta(u)E\overline{N}(du)\mathscr{P}_\theta(u,x)\\ &= \int_{(0,x]} V_n(du,\theta)\mathscr{P}_\theta(u,x).\end{aligned}$$

and

$$\widehat{W}(x,\theta) - \widetilde{W}(x,\theta) = -\int_{(0,x]} [\widehat{W} - \widetilde{W}](u-,\theta)\widetilde{b}_{n\theta}(u)\overline{N}(du) + \text{rem}(x,\theta),$$

where

$$\text{rem}(x,\theta) = -\int_{(0,x]} \widetilde{W}(u-,\theta)[\widetilde{b}_{n\theta}(u)\overline{N}(du) - b_\theta(u)E\overline{N}(du)]. \tag{5.3}$$

Setting $v_n = \max(\|\sqrt{n}V_n\|_1, \|\sqrt{n}V_n^-\|_1) = O_P(1)$, we have

$$\max(\|\widetilde{W}_n\|, \|\widetilde{W}_n^-\|) \leq v_n \exp \sup_\theta \int_0^\tau \|b_\theta(u)\|_1 EN_{\ldots}(u) = O_p(1).$$

The process $\widetilde{W}$ is a sum of iid mean zero processes whose finite dimensional distributions are asymptotically normal and converge to the finite dimensional distributions of the process W in the statement of the proposition. Moreover, its components can be represented as empirical processes indexed by Euclidean classes of functions satisfying the condition (5.2). Therefore a similar argument as in the case of the process $\sqrt{n}V_{1n}$, shows that $\widetilde{W} \Rightarrow W$. The remainder term (5.3) is bounded by $\sum_{p=2}^4 R_{pn}(x,\theta)$, where

$$\begin{aligned}
R_{2n}(x,\theta) &= \int_{(0,x]} \sqrt{n}V_n(u-,\theta) R_{1n}(du,\theta), \\
R_{3n}(x,\theta) &= \int_{(0,x]} \sqrt{n}V_n(u-,\theta) b_\theta(du) E\overline{N}(du) J_n(u,x,\theta), \\
R_{4n}(x,\theta) &= \int_{(0,\tau]} \|[\widetilde{b}_{n\theta} - b_\theta](u)\|_1 N_{\ldots}(du) \|\widetilde{W}(u-,\theta)\|_1, \\
J_n(u,x,\theta) &= \int_{(u,x]} \mathscr{P}_\theta(u,w) R_{1n}(dw,\theta),
\end{aligned}$$

where $N_{\ldots} = \Sigma_q N_{q,\ldots}$. We have $\|R_{2n}\| = o_P(1)$, by a similar V-process expansion as in Lemma 5.4 below. Using Kolmogorov equations for matrix product integrals (Gill and Johansen, 1990), we also have

$$\begin{aligned}
&J_n(u,x,\theta) = R_{1n}(x,\theta) - R_{1n}(u,\theta) \\
&- \int_{(u,x]} \mathscr{P}_\theta(u,s-) b_\theta(s) E\overline{N}(ds) [R_{1n}(x,\theta) - R_{1n}(s,\theta)]
\end{aligned}$$

and

$$\begin{aligned}
&\|J_n(u,x,\theta)\|_1 \leq 2\|R_{1n}\| [1 + \int_{(u,x]} \|\mathscr{P}_\theta(u,s-)\|_1 \|b_\theta(s)\|_1 EN_{\ldots}(ds) \\
&\leq 2\|R_{1n}\| \exp \int_{(u,x]} \|b_\theta(s)\|_1 EN_{\ldots}(ds) \leq 2\|R_{1n}\| \exp \int_{(0,\tau]} \|b_\theta(s)\|_1 EN_{\ldots}(ds).
\end{aligned}$$

From this we also get $\|R_{3n}\| = o_P(1)$, because $b_\theta(u)$ is uniformly bounded. Finally, $\|R_{4n}\| = o_P(1)$. Combining, the right-hand side of (5.3) is of the order $o_P(1)$, uniformly in $(x,\theta)\in\mathscr{T}$. For fixed (x,θ), we also have

$$\|\widehat{W}(x,\theta)-\widetilde{W}(x,\theta)\|_1 \leq \|\mathrm{rem}(x,\theta)\|_1 + \int_{(0,x]} \|\widehat{W}-\widetilde{W}\|_1(u-,\theta)\rho_n(du)$$

and by uniform Gronwall's inequality (Beesack, 1993, Dabrowska, 2006), we have $\widehat{W}(x,\theta)=\widetilde{W}(x,\theta)+o_P(1)$ uniformly in $(x,\theta)\in\mathscr{T}$.

To complete the argument, we note that the processes V_{1n} V_n, $\widetilde{W}$ and the remainders R_{pn}, $p=1,\ldots,3$ satisfy measurability conditions of section 5.2, whereas to show that $\widehat{W}$ and R_{4n} have this property, it is enough to show that the process $\Gamma_{n\theta}$ is measurable. However, the aggregate process $N_{\ldots}(x)=\sum_{i=1}^n\sum_m N_{jmi}(x)$ is measurable since it is cádlág increasing with respect to x and measurable with respect to $\mathscr{S}_n$ for fixed x. For any integer k and $\omega_0=(s_1,\ldots,s_n)$, $T_k(\omega_0)=\inf\{x: N_{\ldots}(\omega_0,x)\geq k\}$ is a random variable because $\{\omega_0 : T_k(\omega_0)\leq x\}=\{\omega_0 : N_{\ldots}(x,\omega_0)\geq k\}\in\mathscr{S}_n$. Similarly, the censored data ranks $R_{im}=\sum_k 1(T_k\leq X_{im})$ are measurable. Define set valued mapping $H_n:\mathbb{S}_n\hookrightarrow R^q$ by setting $H_n(\omega_0)=\{(\theta,x):\Gamma_{n\theta}(\omega_0,x))\in B\}$ where B is an open set of R^q. Then $H_n(\omega_0)=\bigcup_{\ell\geq 0}H_{n\ell}(\omega_0)$ where

$$H_{n\ell}(\omega_0)=\{(\theta,x):\Gamma_{n\theta}(\omega_0,x)\in B \quad \text{and} \quad N_{\ldots}(\omega_0,\infty)=\ell\}.$$

On the set $A_l=\{\omega : N_{\ldots}(\omega_0,\tau)=\ell\}\in\mathscr{S}_n$, the process $\Gamma_{n\theta}$ is a weighted sum

$$\Gamma_{n\theta}(x,\omega_0)=\sum_{k=1}^{\ell} 1(T_k(\omega_0)\leq x)h_{n\theta}(\cdot,k,\omega_0)$$

and the weights form a finite composition of Carathéodory integrands. Suppressing dependence on ω_0, $h_{n\theta}(\cdot,k)$ is the k-th column of a $q\times\ell$ matrix h_n with entries

$$h_{n\theta}(j,k)=\frac{\sum_i\sum_m 1(R_{im}=k)1((J_{im},J_{im+1})=j)}{\sum_i\sum_m 1(R_{im}\geq k, J_{im}=j_1)\alpha_j(g_{n\theta}(\cdot,k-1),\theta,Z_{im})},$$

where $j=(j_1,j_2)\in\mathscr{J}_0$ and $g_{n\theta}$ is a $q\times\ell$ matrix with columns

$$g_{n\theta}(\cdot,0)=0 \quad g_{n\theta}(\cdot,k)=g_{n\theta}(\cdot,k-1)+h_{n\theta}(\cdot,k).$$

Alternatively, $g_{n\theta}=g_{n\theta}^{(\ell)}$, where $g_{n\theta}^{(0)}\equiv 0$ and for $r=1,\ldots,\ell$

$$g_{n\theta}^{(r)}(\cdot,k)=\sum_{p\leq k} h_{n\theta}^{(r)}(\cdot,p),$$

$$h_{n\theta}^{(r)}(j,k) = \frac{\sum_i \sum_m 1(R_{im} = k) 1((J_{im}, J_{im+1}) = j)}{\sum_i \sum_m 1(R_{im} \geq k, J_{im} = j_1) \alpha_j (g_{n\theta}^{(r-1)}(\cdot, k-1, \theta, Z_{im})}$$

for $j = (j_1, j_2) \in \mathscr{J}_0$. The indicators $1(T_k(\omega_0) \leq x)$ are jointly measurable with respect to $\mathscr{S}_n \otimes \mathscr{B}(\mathscr{T})$ and by Lemma 5.1, so are the weights $h_{n\theta}$ and $g_{n\theta}$. Therefore the graph of $H_{n\ell}$ is $\mathscr{S}_n \otimes \mathscr{B}(\mathscr{T})$ is measurable and

$$\{(\omega_0, x, \theta) : \Gamma_{n\theta}(\omega_0, x) \in B\} = \text{graph} H_n = \bigcup_{l \geq 0} \text{graph}(H_{n\ell}) \in \mathscr{S}_n \otimes \mathscr{B}(\mathscr{T}).$$

A similar argument can be used to show measurability of the process $\dot{\Gamma}_{n\theta}$ in part (ii). Using arguments analogous to Dabrowska (2006), $\|\Gamma_{n\theta_0+h_n} - \Gamma_{n\theta_0} - h_n \dot{\Gamma}_{n\theta_0}\| = O_P(\|h_n\|_1^2)$ and $\|\dot{\Gamma}_{n\theta_0+h_n} - \dot{\Gamma}_{n\theta_0}\| = O_P(\|h_n\|_1) = o_P(1)$ for any deterministic sequence $h_n \to 0$ or a random $\mathscr{S}_n^P$- measurable sequence $h_n \to_P 0$. Therefore if $\widehat{\theta}$ is an $\mathscr{S}_n^P$- measurable $\sqrt{n}$- consistent estimator of θ_0, then setting $h_n = \widehat{\theta} - \theta_0$, we have $\widehat{W}_0(x) = \widehat{W}(x, \theta_0) + \text{rem}_n(x)$, where $\text{rem}_n = \sqrt{n}[\Gamma_{n\widehat{\theta}} - \Gamma_{n\theta_0} - (\widehat{\theta} - \theta_0)\dot{\Gamma}_{n\widehat{\theta}}] = o_P(1)$. For non-measurable h_n and $\widehat{\theta}_n$, convergence is in outer probability. $\square$

Let us assume now that $f_j(y, \theta, z), j \in \mathscr{J}_0$ is a scalar Carathéodory integrands such that $|f_j(y, \theta, z)| \leq \widetilde{\psi}(\|y\|_1)$ and $|f_j(y, \theta', z) - f_j(y', \theta', z)| \leq [|\theta - \theta'| + \|y - y'\|_1] \max(\widetilde{\psi}'(\|y\|_1), \widetilde{\psi}'(\|y'\|_1)$, where $\widetilde{\psi} = \psi, \psi_1, \psi_2$ and $\widetilde{\psi}' = \psi_3$ satisfy conditions 5.1. Put $S_j[f_j](u, \theta) = n^{-1} \Sigma_{i=1}^n S_{j.i}[f_j](u, \theta)$, where $S_{j.i}[f_j](u, \theta) = \sum_m Y_{jmi}(u)(f_j \alpha_j)(\Gamma_\theta(u), \theta, Z_{jmi})$, and let $s_j[f_j] = ES_j[f_j]$. We write $S_j[1]$ and $s_j[1]$ when $f_j \equiv 1$, and set $\widehat{e}_j[f_j] = S_j[f_j]/S_j[1]$ and $e_j[f_j] = s_j[f_j]/s_j[1]$.

Lemma 5.3 We have $\|S_j[f_j]/s_j[1] - s_j[f_j]/s_j[1]\| \to 0$ a.s. for all $j \in \mathscr{J}_0$.

Proof. We have $([S_j[f_j]/s_j[1]])(x, \theta) = \mathbb{P}_n g_{x\theta}$, where

$$g_{x\theta}(D_i) = \frac{\sum_m Y_{jmi}(x)(f_j \alpha_j)(\Gamma_\theta(x), \theta, Z_{jmi})}{E \sum_m Y_{jmi}(x) \alpha_j(\Gamma_\theta(x), \theta, Z_{jmi})}.$$

The conditions 5.1 imply that there exist constants C_1 and C_2 (dependent on the functions $\widetilde{\psi}, \widetilde{\psi}'$) such that

$$\begin{aligned} &|g_{x\theta}(D_i) - g_{x'\theta}(D_i)| \leq C_1[|Y_{j.i}(x') - Y_{j.i}(x)| + \\ &Y_{j.i}(0)(|EN_{j.i}(x) - EN_{j.i}(x')| + |EY_{j.i}(x) - EY_{j.i}(x')|)], \\ &|g_{x\theta}(D_i) - g_{x\theta'}(D_i)| \leq |\theta - \theta'| C_2 Y_{j.i}(0)[1 + EN_{j.i}(\tau) + EY_{j.i}(0)]. \end{aligned}$$

Define $G(D_i) = Y_{j.i}(0)[C_2 \text{diam}\,\Theta + C_1][1 + EN_{j.i}(\tau) + EY_{j.i}(0)] + g_{x_0\theta_0}(D_i)$, where (x_0, θ_0) is an arbitrary point in $\Theta \times [0, \tau]$. Let $\theta_p, p = 1, \ldots, \ell = O(\text{diam}\,\Theta/\varepsilon)^d$ be centers of balls $B(\theta_p, \varepsilon)$ of radius ε covering the set Θ. By noting that $EN_{j.i}$ is an increasing continuous function and $EY_{j.i}$ is a decreasing cáglád function, we can construct a finite partition $0 = x_0 < x_1 < \ldots < x_k = \tau$ such that the intervals $I_r = [x_{r-1}, x_r], r = 1, \ldots, k$ satisfy $EN_{j.i}(I_q) \leq \varepsilon EN_{j.i}(\tau)$ and $E|Y_{j.i}(I_r)| \leq \varepsilon EY_{j.i}(0)$. Let x_q be the center of the interval I_r. Then for each $x \in I_r$ and $\theta \in B(\theta_p, \varepsilon)$, we have $\|g_{x\theta}(D_i) - g_{x_r\theta_p}(D_i)\|_{P,1} \leq \varepsilon \|G(D_i)\|_{P,1}$. It follows that the class of functions $\mathscr{G} = \{g_{x\theta} : x \in [0, \tau], \theta \in \Theta\}$ is Euclidean for the envelope $G(D_i)$ and Glivenko-Cantelli. $\square$

Lemma 5.4 For $j \in \mathscr{J}_0$, define $\text{rem}_j(x, \theta) = [V_{jn} - V_{1jn}](x, \theta)$ and

$$B_j(x, \theta) = \int_0^x [\widehat{e}_j[f_j] - e_j[f_j]](u, \theta) M_{j..}(du, \theta),$$

where f_j satisfies assumptions of Lemma 5.3. Then $\|\sqrt{n}\text{rem}_j\| = o_P(1)$ and $\|\sqrt{n}B_j\| = o_P(1)$.

Proof . For the sake of convenience write rem $=$ rem$_j$ and $B = B_j$. Put $\eta_j(u, \theta) = [S_j/s_j](\Gamma_\theta(u), \theta, u) - 1$. A little algebra shows that

$$\begin{aligned}\text{rem}(x, \theta) &= -\int_0^x \eta_j(u, \theta) \frac{[N_{j..} - EN_{j..}](du)}{s_j[1](u, \theta)} + \int_0^x \eta_j^2(u, \theta) \frac{N_{j..}(du)}{S_j[1](u, \theta)} \\ &= \text{rem}_1(x, \theta) + \text{rem}_2(x, \theta).\end{aligned}$$

We have $\text{rem}_2(x, \theta) = O_P(1)\text{rem}_3(\tau, \theta)$, where

$$\text{rem}_3(x, \theta) = \int_0^x \eta_j^2(u, \theta) \frac{[N_{j..} - EN_{j..}](du)}{s_j[1](u, \theta)} + \int_0^x \eta_j^2(u, \theta) \frac{EN_{j..}(du)}{s_j[1](u, \theta)}.$$

In addition,

$$\begin{aligned}B(x, \theta) &= \int_0^x \left(\frac{S_j[f_j] - S_j[1]e_j[f_j]}{s_j[1]} \right)(u, \theta)[N_{j..} - EN_{j..}](du) \\ &- \int_0^x \left[\left(\frac{S_j[f_j] - s_j[f_j]}{s_j[1]} \right) \eta_j \right](u, \theta)[N_{j..} - EN_{j..}](du) \\ &- \int_0^x \left[\left(\frac{S_j[f_j] - s_j[f_j]}{s_j[1]} \right) \eta_j \right](u, \theta) EN_{j..}(du)\end{aligned}$$

$$+\int_0^x S_j[f_j](u,\theta)\mathrm{rem}_2(du,\theta) = \sum_{p=1}^4 B_p(x,\theta).$$

We have $B_4(x,\theta) = O_P(1)B_5(\theta)$,

$$B_5(\theta) = \int_0^\tau (S_j[|f_j|] - s_j[|f_j|])(u,\theta)\mathrm{rem}_3(du,\theta) + \int_0^\tau s_j[|f_j|](u,\theta)\mathrm{rem}_3(du,\theta).$$

These expressions can be rewritten as V processes of degree $r+1$, $r \le 3$

$$\mathbb{V}_{n,r+1}(g) = \frac{1}{n^{r+1}}\sum_{\mathbf{i_{r+1}}} g(\mathbf{D}_{\mathbf{i_{r+1}}}), g \in \mathscr{G},$$

where the sum extends over sequences $r+1$-tuplets $\mathbf{D}_{\mathbf{i_{r+1}}} = (D_{i_1},\ldots,D_{i_{r+1}})$ $\mathbf{i_{r+1}} = (i_{r_1},\ldots,i_{r+1}), i_j \in 1,\ldots,n$. The kernels g vary over the class $\mathscr{G} = \{g_t : t \in \mathscr{T}\}$, where for $t = (x,\theta)$ we have

$$\begin{aligned} & g_t(\mathbf{D}_{\mathbf{i_{r+1}}}) = \\ & = \int_0^x \prod_{\ell=1}^r [h_\ell(D_{i_\ell},\theta,u) - Eh_\ell(D_{i_\ell},\theta,u)][N_{j.i_{r+1}} - EN_{j.i_{r+1}}](du) \end{aligned} \tag{5.4}$$

or

$$g_t(\mathbf{D}_{\mathbf{i_{r+1}}}) = \int_0^x \prod_{\ell=1}^{r+1}[h_\ell(D_{i_\ell},\theta,u) - Eh_\ell(D_{i_\ell},\theta,u)]EN_{j.}(du). \tag{5.5}$$

Here $h_\ell(D_{i\ell})$ are functions of the form $S_j[f_j]/s_j[1]$, $S_j[1]/s[1]$ and $(\sqrt{s_j[|f_j|]})S_j[1]/s[1]$. In all cases, there exists a constant C such that $h_\ell(D_i,\theta,u)$ $\le CY_{ji}(u)$ and $|h_\ell(D_i,\theta,u) - h_\ell(D_i,\theta',u)| \le |\theta-\theta'|CY_{j.i}(u)$. Therefore, for any sequence $\mathbf{D}_{\mathbf{i_{r+1}}} = (D_{i_1},\ldots,D_{i_{r+1}})$, we also have

$$\begin{aligned} |g_{x\theta} - g_{x'\theta}|(\mathbf{D}_{\mathbf{i_{r+1}}}) &\le |G(\mathbf{D}_{\mathbf{i_{r+1}}},x) - G(\mathbf{D}_{\mathbf{i_{r+1}}},x')|, \\ |g_{x\theta} - g_{x\theta'}|(\mathbf{D}_{\mathbf{i_{r+1}}}) &\le |\theta-\theta'|G(\mathbf{D}_{\mathbf{i_{r+1}}},\tau), \end{aligned}$$

where

$$G(\mathbf{D}_{\mathbf{i_{r+1}}},x) = \int_0^x \prod_{\ell=1}^r [H_\ell(D_{i_\ell},u) + EH_\ell(D_{i_\ell},u)][N_{j.i_{r+1}} + EN_{j.i_{r+1}}](du)$$

and $H_\ell(D_i,u) = CY_{j.i_\ell}(u), \ell = 1,\ldots,r$ for some constant C.

Let $\{\mathbb{U}_{r+1,n}(g_t) : t \in \mathscr{T}\}$ denote the U process associated with the kernels (5.4-5.5). It is easy to see that $\mathbb{U}_{r+1,n}(g_t)$ forms a canonical process. For

$\mathbf{D_{r+1}} = (D_1, \ldots, D_{r+1})$, we have $EG^p(\mathbf{D_{r+1}}) < \infty$ for $p = 1+1/(2r+1)$. Therefore, by Marcinkiewicz-Zygmund law in Teicher (1998) and Lemma A.1 in Dabrowska (2009), $\sqrt{n}\sup_t |\mathbb{U}_{r+1,n}(g_t)| \to_P 0$. By Marcinkiewicz-Zygmund theorem in de la Peña and Giné (1999), we also have $\sqrt{n}\sup_t |\mathbb{V}_{r+1,n}(g_t) - \mathbb{U}_{r+1,n}(g_t)| \to 0$ a.s. because

$$EG(\mathbf{D_{i_{r+1}}})^{2d(i_{r+1})/(2r+1)} < \infty,$$

where $\mathbf{i_{r+1}} = (i_1, \ldots, i_{r+1})$ and $d = d(\mathbf{i_{r+1}})$ is the number of distinct coefficients among $\{i_1, \ldots, i_{r+1}\}$, $d = 1, \ldots, r, r \leq 3$. $\square$

We denote now by $\|B\|_{\mathrm{v}}$ the variation norm of a $d \times q$-matrix of functions $B(x) = [b_{kl}(x)], x \in [0, \tau]$. For any interval $I \subseteq [0, \tau]$, $\|B\|_{\mathrm{v}}(I) = \sup \sum_{i=1}^m \|B(x_j) - B(x_{j-1})\|_1$, where the supremum is taken over finite partitions of I such that $x_i < x_j$.

Further, let $\mathscr{B}(\theta_0, \varepsilon_n)$ be a ball centered at θ_0 of radius $\varepsilon_n, \varepsilon_n \downarrow 0, \sqrt{n}\varepsilon_n \uparrow \infty$. Suppose that $\varphi_\theta(x)$ is a $d \times q$ matrix of functions, with columns of the form $\int_0^x g_{j\theta} d\Gamma_{\theta,j}$ such that $\|\varphi_{\theta_0}\|_{\mathrm{v}} = O(1)$. Let $\varphi_{n\theta}$ be a sequence of consistent estimators such that

- (i) $\varphi_{n\theta}(x)$ is a càdlàg or càglàd function (jointly in (x, θ)), continuous with respect to θ;
- (ii) $\limsup_n \sup\{\|\varphi_{n\theta}\|_{\mathrm{v}} : \theta \in \mathscr{B}(\theta_0, \varepsilon_n)\} = O_P(1)$;
- (iii) $\sup\{\|\varphi_{n\theta} - \varphi_{\theta_0}\|_\infty : \theta \in \mathscr{B}(\theta_0, \varepsilon_n)\} = o_P(1)$ or
- (iii') $\varphi_{n\theta} - \varphi_{n\theta'} = (\theta - \theta')\psi_{n\theta,\theta'}$ where $\limsup_n \sup\{\|\psi_{n\theta\theta'}\|_{\mathrm{v}} : \theta, \theta' \in \mathscr{B}(\theta_0, \varepsilon_n)\} = O_P(1)$.

If $\varphi_{n\theta}$ is a jointly $\mathscr{S}_n^P \otimes \mathscr{B}(\mathscr{T})$ measurable estimator then conditions (ii)-(iii) are assumed to hold in probability. If this is not the case then the conditions (ii)-(iii) are taken to hold in outer probability.

Lemma 5.5 (i) If $\varphi_{n\theta}(x)$ is a measurable process satisfying (i)-(ii) and (iii) or (iii') then with probability tending to 1, the equation $U_{n\varphi_n}(\theta) = 0$ has a consistent root $\widehat{\theta}$ in the ball $\mathscr{B}(\theta_0, \varepsilon_n)$. In addition, under the condition (iii'), the score equation has a unique root in $\mathscr{B}(\theta_0, \varepsilon_n)$, with probability tending to 1.

(ii) If $\varphi_{n\theta}$ is not measurable, then statements in part (1) hold with inner probability tending to 1.

(iii) If $\widetilde{\theta}$ is an arbitrary consistent estimator of θ_0, then the equation $U_{n\widetilde{\varphi}_n}(\theta) = 0$, where $\widetilde{\varphi}_n(x) = \varphi_{n\widetilde{\theta}}(x)$ has a unique solution $\widehat{\theta}$, with (inner) probability tending to 1, and $U_{n\varphi_n}(\widehat{\theta}) = o_{p^*}(n^{-1/2})$.

In all three cases, $\widehat{\Xi} = \sqrt{n}(\widehat{\theta} - \theta_0)$ and the process $\widehat{W}_0 = \{\sqrt{n}[\Gamma_{n\widehat{\theta}} - \Gamma_{\theta_0} - (\widehat{\theta} - \theta_0)^T \dot{\Gamma}_{n\widehat{\theta}}](x) : x \leq \tau\}$ converge weakly in $R^d \times \ell^\infty([0, \tau] \times \mathscr{J}_0)$ to a mean zero Gaussian process defined in the statement of Proposition 3.1.

Proof . Case (1). Write $U_n(\theta) = U_{n\varphi_n}(\theta)$ for short. Set $\widetilde{b}_{jmi}(\Gamma_\theta(u), \theta, u) = = \widetilde{b}_{jmi1}(\Gamma_\theta(u), \theta, u) - \varphi_{\theta_0}(u)\widetilde{b}_{jmi2}(\Gamma_\theta(u), \theta, u)$ where

$$\widetilde{b}_{jmi1}(\Gamma_\theta(u), \theta, u) = \dot{\ell}_j(\Gamma_\theta(u), \theta, Z_{jmi}) - e_j[\dot{\ell}_j](u, \theta),$$
$$\widetilde{b}_{jmi2}(\Gamma_\theta(u), \theta, u) = \ell'_j(\Gamma_\theta(u), \theta, Z_{jmi}) - e_j[\ell'_j](u, \theta).$$

Define $\overline{b}_{jmi}(\Gamma_\theta(u), \theta, u), \overline{b}_{jmi1}(\Gamma_\theta(u)\theta, u)$ and $\overline{b}_{jmi2}(\Gamma_\theta(u), \theta, u)$ using similar expressions with $e_j[\dot{\ell}_j]$ and $e_j[\ell'_j]$ replaced by $\widehat{e}_j[\dot{\ell}_j]$ and $\widehat{e}_j[\ell'_j]$. We have $U_n(\theta) = \sum_{p=1}^4 U_{np}(\theta)$, where

$$U_{1n}(\theta) = \frac{1}{n}\sum_{i=1}^n \sum_j \sum_m \int_0^x \widetilde{b}_{jmi}(\Gamma_\theta(u), \theta, u) M_{jmi}(du, \theta),$$
$$U_{2n}(\theta) = \sum_{q=1}^2 \int_0^\tau r_{nq}(du, \theta)[\Gamma_{n\theta} - \Gamma_\theta]^T(u) = \sum_{q=1}^2 U_{2n;q}(\theta),$$
$$U_{3n}(\theta) =$$
$$-\sum_j \int_0^\tau \left[(\widehat{e}_j[\dot{\ell}_j] - e_j[\dot{\ell}_j])(u, \theta) - \varphi_{\theta_0}(u)(\widehat{e}_j[\ell'_j] - e_j[\ell'_j])(u, \theta)\right] M_{j\cdot\cdot}(du, \theta),$$
$$U_{4n}(\theta) = -\frac{1}{n}\sum_{i=1}^n \sum_j \sum_m \int_0^\tau [\varphi_{n\theta} - \varphi_{\theta_0}](u)]\widehat{b}_{jmi2}(\Gamma_{n\theta}(u), \theta, u) N_{jmi}(du),$$

and

$$r_{n1}(x, \theta) = \frac{1}{n}\sum_{i=1}^n \sum_j \sum_m \int_0^x \overline{b}'_{jmi}(\Gamma_\theta(u), \theta, u) N_{jmi}(du),$$
$$r_{n2}(x, \theta) = \frac{1}{n}\sum_{i=1}^n \sum_j \sum_m \int_0^1 \int_0^x [\overline{b}'_{jmi}(\Gamma^\lambda_{n\theta}(u), \theta, u) N_{jmi}(du) d\lambda - r_{n1}(x, \theta).$$

Here $\Gamma^\lambda_{n\theta} = \Gamma_\theta + \lambda(\Gamma_{n\theta} - \Gamma_\theta)$ for $\lambda \in (0,1)$. We have $U_{2n;2}(\theta_0) = \int_0^\tau O_P(\|\Gamma_{n\theta_0} - \Gamma_{\theta_0}\|^2) \sum_j N_{j\cdot\cdot}(du) = o_P(n^{-1/2})$. Moreover, $r_{1n}(x, \theta_0)$ converges almost surely to

$$r(x, \theta_0) = \sum_j \int_0^x [\mathrm{cov}_j(\ell'_j, \dot{\ell}_j)(u, \theta_0) - \varphi_{\theta_0}(u)\mathrm{cov}_j(\ell'_j, \ell'_j)(u, \theta_0)] E N_{j\cdot\cdot}(du)$$

uniformly in $x, x \le \tau$. Lemma 5.2 and integration by parts imply that the terms $[\sqrt{n}U_{1n}(\theta_0), \sqrt{n}U_{2n;1}(\theta_0)]$ converge weakly to a pair of independent normal variables with mean zero and covariances $\Sigma_0(\theta_0)$ and $\Sigma_2(\theta_0) - \Sigma_0(\theta_0)$, respectively.

By Lemma 5.3-4, we also have $U_{3n}(\theta_0) = o_P(n^{-1/2})$. Finally,

$$U_{4n}(\theta) = -\sum_{p=1}^{3}\int_0^{\tau}[\varphi_{n\theta} - \varphi_{\theta_0}](u)]B_{pn}(du,\theta) = \sum_{p=1}^{3} U_{4n;p}(\theta),$$

where

$$B_{1n}(x,\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_j\sum_m\int_0^x[\widehat{b}_{2jmi}(\Gamma_{n\theta}(u),\theta,u) - \widehat{b}_{2jmi}(\Gamma_{\theta}(u),\theta,u)]N_{jmi}(du),$$
$$B_{2n}(x,\theta) = -\sum_j\int_0^x(\widehat{e}_j[\ell'_j] - e_j[\ell'_j])(u,\theta)M_{j\cdot\cdot}(du,\theta),$$
$$B_{3n}(x,\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_j\sum_m\int_0^x\widetilde{b}_{2jmi}(\Gamma_{\theta}(u),\theta,u)M_{jmi}(du,\theta).$$

By Lemmas 5.2-5.4, we have $\sqrt{n}U_{4n;2}(\theta) = o_P(1)$ and $\sqrt{n}U_{4n;1}(\theta) = \sum_j\int_0^{\tau}O_P(\sqrt{n}\|\Gamma_{n\theta} - \Gamma_{\theta}\|_1(u)\|\varphi_{n\theta} - \varphi_{\theta_0}\|_1(u)N_{j\cdot\cdot}(du) = o_P(1)$, uniformly in $\theta \in \mathscr{B}(\theta_0,\varepsilon_n)$. On the other hand, at $\theta = \theta_0$, $\{\sqrt{n}B_{3n}(x,\theta_0) : x \leq \tau\}$ is a sum of iid mean zero processes. The finite dimensional distributions are mean zero variables with finite variance-covariance matrix and converge weakly to mean zero Gaussian variables. Each component of $B_{3n}(x,\theta_0)$ is a measurable process which can be represented as a finite linear combination of càdlàg monotone functions of x with a square integrable envelope satisfying (5.2). The same argument as in Lemma 5.2 implies that the process is $\sqrt{n}B_{3n}(x,\theta_0)$ converges weakly to a mean zero Gaussian process with sample paths continuous with respect to the variance semi-metric. The space of functions continuous with respect to the variance semi-metric is isometric to the space $C([0,\tau])^q$. By almost sure representation theorem and a similar integration by parts argument as in Bilias et al (1997) we have $\sqrt{n}U_{4n;3}(\theta_0) = o_P(1)$.

Set $\widehat{U}_n(\theta) = \sum_{j=1}^{3}U_{jn}(\theta)$. Some elementary algebra shows that for θ, $\theta' \in \mathscr{B}(\theta_0,\varepsilon_n)$, we have $\widehat{U}_n(\theta) = \widehat{U}_n(\theta') + (\Sigma_n(\theta_0) + \text{rem}_{0n}(\theta,\theta'))(\theta - \theta')$, where $\Sigma_n(\theta_0)$ is a matrix which converges in probability $-\Sigma_1(\theta_0)$. The matrix $\Sigma_1(\theta)$ is defined in Section 3 and is non-singular. Further, $U_{4n}(\theta) - U_{4n}(\theta') = \text{rem}_{2n}(\theta,\theta')$ $(\theta - \theta') + \text{rem}_{3n}(\theta,\theta')$ $+O(|\theta - \theta_0| \vee |\theta' - \theta_0|)\text{rem}_{4n}(\theta,\theta')$. Setting $\text{rem}_{1n}(\theta,\theta')$ $= I + \Sigma_1^{-1}(\theta_0)[\Sigma_n(\theta_0) + \text{rem}_{0n}(\theta,\theta')]$, and $b_{qn} = \sup\{|\text{rem}_{qn}(\theta,\theta')| : \theta,$ $\theta' \in \mathscr{B}(\theta_0,\varepsilon_n)\}$, q = 1, …,4, we have $b_{1n} = o_P(1)$, $b_{2n} = o_P(1)$. Under the condition (iii'), $\text{rem}_{nq} \equiv 0 \equiv b_{qn}, q = 3,4$, while under the condition (iii), $b_{3n} = o_P(n^{-1/2})$ and $b_{4n} = o_P(1)$.

Put $a_n = b_{1n} + b_{2n} + b_{4n}$ and $A_n = b_{5n} + b_{3n}$, where $b_{5n} = |\Sigma(\theta_0)^{-1}\widehat{U}_n(\theta_0)| = O_P(n^{-1/2})$. Let $0 < \eta < 1/2$ and $0 < \eta' < 1$ be given. By asymptotic tightness of A_n, we can find a compact set $K = K(\eta)$ and n_0 such that for all $n \geq n_0$ and all open sets G containing K, we have $P_n(\sqrt{n}A_n \notin G) < \eta$ and $P_n(a_n > \eta') < \eta$. Therefore, we also have $P_n(\sqrt{n}A_n > M(1-\eta')) < \eta$ for all finite $M \geq M_0$, where $M_0 = M_0(\eta)$ is a large enough finite nonnegative constant. Since $\sqrt{n}\varepsilon_n \uparrow \infty$ and $\varepsilon_n \downarrow 0$, by eventually increasing n_0, we can assume that for $n \geq n_0$, we have $\mathscr{B}(\theta_0, \varepsilon_n) \subset \Theta$ and $M < \sqrt{n}\varepsilon_n$. Consequently, the set $E_n \subset \mathbb{S}_n$ given by $E_n = \{\omega_0 : A_n(\omega_0)/(1-a_n(\omega_0) < \varepsilon_n, a_n(\omega_0) \leq \eta'\}$ satisfies $P_n(E_n) \geq 1-2\eta$ for all $n \geq n_0$.

For $n \geq n_0$, consider the set-valued mapping $H_n : \mathbb{S}_n \hookrightarrow R^d$ given by

$$\begin{aligned} H_n(\omega_0) &= \overline{\mathscr{B}}(\theta_0, \frac{A_n(\omega_0)}{1-a_n(\omega_0)}) = \{\theta : |\theta-\theta_0| \leq \frac{A_n(\omega_0)}{1-a_n(\omega_0}\} \quad \text{if} \quad \omega_0 \in E_n, \\ &= \emptyset \qquad \text{if} \quad \omega_0 \notin E_n. \end{aligned}$$

The graph of H_n, graph$H_n = \{(\omega_0, \theta) : \theta \in H_n(\omega_0)\}$ is $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$-measurable and dom$H_n = E_n \in \mathscr{S}_n^P$. Further, let $g_n(\omega_0, \theta) = \theta + \Sigma_1^{-1}(\theta_0)U_n(\omega_0, \theta)$. Then g_n is $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$ measurable, because it is continuous with respect to θ for fixed ω_0 and $\mathscr{S}_n^P$-measurable for fixed θ. It follows that the set valued mapping

$$\begin{aligned} C_n(\omega_0) &= \{\theta : g_n(\omega_0, \theta) = 0 \quad \text{and} \quad \theta \in H_n(\omega_0)\} \quad \text{for} \quad \omega_0 \in E_n, \\ &= \emptyset \qquad \text{for} \quad \omega_0 \notin E_n \end{aligned}$$

is closed-valued and has an $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$- measurable graph. We have dom$C_n = E_n$: for fixed $\omega_0 \in E_n$, $H_n(\omega_0)$ is a closed ball, $g_n(\omega_0, \theta)$ is continuous and maps $H_n(\omega_0)$ into itself. By Brouwer's fixed point theorem, $C_n(\omega_0) \neq \emptyset$. Thus $E_n \subseteq$ domC_n, while the reversed inclusion is obvious.

Further, for any root $\widehat{\theta}$ in domC_n, we have $|\sqrt{n}(\widehat{\theta}-\theta_0)|^* \leq A_n/(1-a_n) = O_P(1)$, and $\sqrt{n}(\widehat{\theta}-\theta_0) = \Sigma(\theta_0)^{-1}\sqrt{n}\widehat{U}_n(\theta_0) + o_{P^*}(n^{-1/2})$ so that $\sqrt{n}(\widehat{\theta}-\theta_0)$ converges in law to the normal distribution given in Section 3. An argument similar to Bickel *et al.* (1993, p.517) shows also that under the condition (iii'), $g_n(\omega_0, \theta)$ is a contraction on $H_n(\omega_0), \omega_0 \in E_n$, with contraction coefficient $a_n(\omega_0)$. Thus in this case, the root is unique: $C_n(\omega_0) = \{\widehat{\theta}(\omega_0)\}$ for $\omega_0 \in E_n$ and $n \geq n_0$.

Case (2). If $\varphi_{n\theta}$ estimators are not $\mathscr{S}_n^P \otimes \mathscr{B}(\mathscr{T})$ measurable, then the score function splits into two parts: $U_n(\theta) = \widehat{U}_n(\theta) + U_{4n}(\theta)$. The term $\widehat{U}_n(\theta)$ remains $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$ measurable, while the second term is not. However, $b_{3n} = o_{p*}(n^{-1/2})$, $a_n = o_{P*}(1)$ while $b_{5n} = |\Sigma(\theta_0)^{-1}\widehat{U}_n(\theta_0)| = O_p(n^{-1/2})$. In this case, the set E_n satisfies $\liminf_n P_{n,*}(E_n) \geq 1-2\eta$ and the closed ball $\overline{\mathscr{B}}(\theta_0, A_n/1-a_n)$ is contained in $\mathscr{B}(\theta_0, \varepsilon_n)$ with inner probability tending to 1.

Case (3). We write $\widetilde{U}_n(\theta)$ for the modified score function obtained by substituting in $\widetilde{\varphi}_n(x) = \varphi_{n\widetilde{\theta}}(x)$ in place of $\varphi_{n\theta}$. Suppose that $\widetilde{\theta}$ is $\mathscr{S}_n^P$-measurable and $\varphi_{n\theta}(x)$ is $\mathscr{S}_n^P \otimes \mathscr{B}(\mathscr{T})$ measurable. Then the plug-in estimator $\varphi_{n\widetilde{\theta}}(x)$ is $\mathscr{S}_n^P \otimes$

$\mathscr{B}([0,\tau])$ measurable and the modified score process $\widetilde{U}_n(\theta)$ is $\mathscr{S}_n^P \otimes \mathscr{B}(\Theta)$ measurable. Moreover, we have $\widetilde{U}_n(\theta) = \widehat{U}_n(\theta) + \widetilde{U}_{n4}(\theta)$, where the remainder $\widetilde{U}_{4n}(\theta)$ satisfies $\sqrt{n}[U_{n4}(\theta) - \widetilde{U}_{n4}(\theta)] = o_P(1+\sqrt{n}|\theta-\theta_0|)$, uniformly in $\theta \in B(\theta_0, \varepsilon_0)$. We also have $\widetilde{U}_{4n}(\theta) - \widetilde{U}_{4n}(\theta') = (\theta-\theta')\widetilde{\mathrm{rem}}_{2n}(\theta,\theta')$, $\sup\{\widetilde{\mathrm{rem}}_{2n}(\theta,\theta') : \theta,\theta' \in \mathscr{B}(\theta_0,\varepsilon_n)\} = o_P(1)$. With probability tending to 1, the modified equation has a unique root $\widehat{\theta}$ in a compact random ball contained in $B(\theta_0,\varepsilon_n)$ and $U_n(\widehat{\theta}) = o_{P^*}(n^{-1/2})$. On the other hand, if either $\widetilde{\theta}$ or $\varphi_{n\theta}$ are not measurable, then this remains to hold, except that the modified equation has a unique solution with inner probability tending to one. □

Under assumptions of part (1), measurable selection theorems (Wagner, 1976) ensure that there exists at least one function $\widehat{\widehat{\theta}} : \mathbb{S}_n \to R^d$ such that $\widehat{\widehat{\theta}}(\omega_0) \in C_n(\omega_0)$ whenever $\omega_0 \in E_n$ and $\widehat{\widehat{\theta}}$ is measurable with respect to $\mathscr{S}_n^p$. This also applies to part (3), provided $\widetilde{\theta}$ and $\varphi_{n\theta}$ are $\mathscr{S}_n^P$- measurable.

5.4 Proof of Proposition 3.2. With some abuse of notation, set $V = [V_j, j \in \mathscr{J}_0]$ where $V(x) = V(x,\theta_0)$ and $V(x,\theta)$ is the Gaussian process of Lemma 5.1. Under the assumption that θ_0 is the true parameter of the modulated renewal process, the process V corresponds to a vector of independent time-transformed Brownian motions with covariance

$$\mathrm{cov}(V_j(x),V_j(y)) = C_j(x\wedge y) \quad \text{and} \quad \mathrm{cov}(V_j(x),V_\ell(y)) = 0 \quad \text{if} \quad j \neq \ell.$$

Similarly, let $\check{V} = [\check{V}_j : j \in \mathscr{J}_0]$ be equal to $\check{V}(x) = \sqrt{n}V_{1n}(x,\theta_0)$ where $V_{1n}(x,\theta)$ is defined as in Lemma 5.1. Thus the j-th component of $\check{V}$ is

$$\check{V}_j(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_m \int_0^x \frac{M_{jmi}(du)}{s_j(\Gamma_{\theta_0}(u-),\theta_0,u)}.$$

Put $\check{V}^\# = [\check{V}_j^\# : j = 1,\ldots,q]$,

$$\check{V}_j^\#(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_m G_{mi}\int_0^x \frac{N_{jmi}(du)}{s_j(\Gamma_{\theta_0}(u-),\theta_0,u)}.$$

Finally, let G_0 be a $\mathscr{N}(0,I_{d\times d})$ variable, independent of (D_i,G_i)'s. Set $\Xi_1^\# = \Sigma_1^{-1}(\theta_0)\Sigma_0(\theta_0)^{1/2}G_0$ and $\widehat{\Xi}_1^\# = \widehat{\Sigma}_1^{-1}(\widehat{\theta})\widehat{\Sigma}_0(\widehat{\theta})^{1/2}G_0$. We have $E\check{V}_j^\#(x) = 0 = E\check{V}_j(x)$,

$$\begin{aligned} &\mathrm{cov}(\check{V}_j^\#(x),\check{V}_\ell^\#(x')) = \mathrm{cov}(\check{V}_j(x),\check{V}_\ell(x')) = \delta_{jl}C_{j\theta_0}(x\wedge x'), \\ &\mathrm{cov}(\check{V}_j^\#(x),\check{V}_\ell(x')) = 0. \end{aligned} \tag{5.6}$$

Moreover, $\Xi_1^\#$ is independent of $D_1,\ldots,D_n$. This also means that it is independent of $(\check{V}^\#,\check{V})$.

We consider first unconditional weak convergence. By central limit theorem and strong law of large numbers, the finite dimensional distributions of the processes $(\check{V},\check{V}^\#)$ converge weakly to finite dimensional distributions of $(V,V^\#)$, two independent vectors of Brownian motions with variance functions C_{j,θ_0}, $j=1,\ldots,q$.

For each $j=1,\ldots,q$, the process $\check{V}_j^\#$ can be represented as $\check{V}_j^\#(x) = n^{-1/2}\Sigma_{i=1}^n f_x^{(j)}(G_i,D_i)$, where

$$f_x^{(j)}(G_i,D_i) = \sum_m G_{mi}\int_0^x \frac{N_{jmi}(du)}{s_j(\Gamma_{\theta_0}(u-),\theta_0,u)}.$$

The class of functions $\mathscr{F}_j = \{f_x^{(j)}(G_i,D_i) : x\in[0,\tau]\}$ has a square integrable envelope

$$F_j(G_i,D_i) = \sum_{m=1}|G_{mi}|\int_0^\tau \frac{N_{jmi}(du)}{s_j(\Gamma_{\theta_0}(u-),\theta_0,u)}$$

and is Euclidean for this envelope because each $f_x^{(j)}\in\mathscr{F}_j$ is a difference of two functions increasing in x and bounded by $F_j(G_i,D_i)$. Thus $\mathscr{F}_j$ forms a Donsker class of functions. The union of these classes, $\mathscr{F}=\bigcup_j\mathscr{F}_j$ is Donsker as well. From Lemma 1, the process $\check{V}=\{\check{V}_j(x) : x\in[0,\tau], j\in\mathscr{J}\}$ can be also represented as an empirical process over a Euclidean class of functions $\mathscr{G}$ and the union $\mathscr{F}\cup\mathscr{G}$ forms a Donsker class. Using consistency of the estimates $(\widehat{\theta},\Gamma_{n\widehat{\theta}})$, Lemma 5.5 and a couple of lines integration by parts yields also $\|\widehat{V}^\# - \check{V}^\#\| = o_P(1)$ in outer probability.

Write $\check{V}^\#$ as the empirical process $\check{V}^\# = \mathbb{P}_n f$, $f\in\mathscr{F}$. Further, let BL_1 be the collection of Lipschitz functions h from $R^d\times\ell^\infty(\mathscr{F})$ into $[0,1]$, such that $|h(r,w)-h(r',w')|\le|r-r'|+\|w-w'\|$ for $r,r'\in R^d$ and $w,w'\in\ell^\infty(\mathscr{F})$. The set $\mathscr{F}$ is totally bounded with respect to the variance pseudo-metric d. Therefore, for fixed $\delta>0$, it can be covered by a finite number of d-balls of radius δ, say $\mathscr{B}(f_l,\delta)$ $\ell=1,\ldots,k=k(\delta)$. Set $V^\#\circ\pi_\delta = \mathbb{P}_n\pi_\delta(f)$, where $\pi_\delta(f)=f_\ell$ for $f\in\mathscr{B}(f_\ell,\delta)$ (pick one f_ℓ for each $f\in\mathscr{F}$). By triangular inequality, we have

$$\sup_{h\in BL_1}|E_G h(\widehat{\Xi}_1^\#,\widehat{V}^\#) - Eh(\Xi_1^\#,V^\#)| \le \sum_{r=1}^4 I_4(\delta),$$

$$
\begin{aligned}
I_1(\delta) &= \sup_{h\in BL_1} |Eh(\Xi_1^\#, V^\# \circ \pi_\delta) - Eh(\Xi_1^\#, V^\#)|, \\
I_2(\delta) &= \sup_{h\in BL_1} |Eh(\Xi_1^\#, V^\# \circ \pi_\delta) - E_G h(\Xi_1^\#, \check{V}^\# \circ \pi_\delta)|, \\
I_3(\delta) &= \sup_{h\in BL_1} |E_G h(\Xi_1^\#, \check{V}^\# \circ \pi_\delta) - E_G h(\Xi_1^\#, \check{V}^\#)|, \\
I_4(\delta) &= \sup_{h\in BL_1} |E_G h(\widehat{\Xi}_1^\#, \widehat{V}^\#) - E_G h(\Xi_1^\#, \check{V}^\#)|.
\end{aligned}
$$

For given $\varepsilon > 0$, we can choose δ_0 so that $I_1(\delta) < \varepsilon$ for all $\delta < \delta_0$. The second term converges in outer probability to 0, for any δ. This follows from weak convergence of finite dimensional distributions of $\check{V}^\#$ and the same argument as in Van der Vaart and Wellner (1996, p. 182), except that in our setting, the Lindeberg condition of their Lemma 2.9.5 is not needed to verify conditional weak convergence of finite dimensional distributions. We also have $I_3(\delta) \le E_G^* \|\check{V}^\# \circ \pi_\delta - \check{V}^\#\|_{\mathscr{F}_\delta} \le \Sigma E_G^* \|\check{V}^\#\|_{\mathscr{F}_\delta}$ where $\mathscr{F}_\delta = \{f - f' : f, f' \in \mathscr{F} : d(f - f') < \delta\}$. Since $\mathscr{F}$ forms a Euclidean class of functions with a square integrable envelope, we have $\lim_{\delta\downarrow 0} \limsup_n E^* I_3(\delta) \le \lim_{\delta\downarrow 0} \limsup_n E^* E_G^* \|\check{V}^\#\|_{\mathscr{F}_\delta} = 0$. Finally, the term $I_4(\delta)$ does not depend on δ, and we have $I_4(\delta) \le 2P_G^*(|\widehat{\Xi}_1 - \Xi_1| + \|\widehat{V}^\# - \check{V}^\#\| > \varepsilon) + \varepsilon$. By unconditional convergence, we have $I_4(\delta) \to 0$ in outer probability.

Finally, set $\Psi(\widehat{\Xi}_1^\#, \widehat{V}^\#) = [\check{\Xi}^\#, \check{W}_0^\#]$, where

$$
\begin{aligned}
&\check{\Xi}^\# = \widehat{\Xi}_1^\# - \Sigma_1^{-1}(\theta_0) \sum_j \int_0^\tau \rho_{j,\varphi}(u,\theta_0) EN_{j\cdot\cdot}(du) \check{W}_0^\#(u)^T, \\
&\check{W}_0^\#(x) = \int_0^x \widehat{V}^\#(du) \mathscr{P}_{\theta_0}(u,x) \\
&= \widehat{V}^\#(x) - \int_0^x \widehat{V}^\#(u-) Q_{\theta_0}(du) \mathscr{P}_{\theta_0}(u,x).
\end{aligned}
$$

The estimates $[\widehat{\Xi}^\#, \widehat{W}_0^\#]$ defined in Section 4 are $[\widehat{\Xi}^\#, \widehat{W}_0^\#] = \widehat{\Psi}(\widehat{\Xi}_1^\#, \widehat{V}^\#)$, where $\widehat{\Psi}$ is the sample analogue of Ψ obtained by plugging in the estimates $\widehat{\mathscr{P}}_{\widehat{\theta}}, \widehat{Q}_{\widehat{\theta}}$, $\rho_{j,\varphi_n}(\cdot, \widehat{\theta}_0)$. By the continuous mapping theorem, unconditionally, $\Psi(\widehat{\Xi}_1^\#, \widehat{V}^\#) \Rightarrow \Psi(\Xi_1^\#, V^\#) = (\Xi^\#, W_0^\#)$. By triangular inequality one more time, we have $\sup_{h\in BL_1} |E_G h(\widehat{\Xi}^\#, \widehat{W}_0^\#) - Eh(\Xi^\#, W_0^\#)| \le J_1 + J_2$, where

$$
J_1 = \sup_{h\in BL_1} |E_G h(\widehat{\Xi}^\#, \widehat{W}_0^\#) - E_G h(\check{\Xi}^\#, \check{W}_0^\#)|,
$$

$$J_2 = \sup_{h \in BL_1} |E_G h(\check{\Xi}^{\#}, \check{W}_0^{\#}) - E h(\Xi^{\#}, W_0^{\#})|.$$

For any Lipschitz continuous function $h \in BL_1$, $h \circ \Psi \in BL_c$ for some constant c. Therefore the preceding implies that J_2 tends to 0 in outer probability. This also holds for the term J_1, because $\|\check{\Xi}^{\#} - \widehat{\Xi}^{\#}\| \to_{P^*} 0$ and $\|\widehat{W}_0^{\#} - \check{W}_0^{\#}\|_1 \to_{P^*} 0$, by consistency of the estimates $(\widehat{\theta}, \Gamma_{n\widehat{\theta}})$ and integration by parts. □

References

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

Arjas, E. and Eerola, M. (1993). On predictive causality in longitudinal studies. *J. Statist. Planning and Inference,* **34** 361-386.

Bagdonovicius, V. and Nikulin, M. (1999). Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Analysis,* **5** 329-350.

Bagdonovicius, M. Hafdi, M. A. and Nikulin, M. (2004). Analysis of survival data with cross-effects of survival functions. *Biostatistics,* **5** 415-425.

Beesack, P. R. (1973). *Gronwall Inequalities.* Carlton Math. Lecture Notes **11** Carlton University, Ottawa.

Bickel, P. J. (1986). Efficient testing in a class of transformation models. In *Proceedings of the 45th Session of the International Statistical Institute* 23.3.63-23.3.81. ISI, Amsterdam.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993) Efficient and adaptive estimation in semi-parametric models. Johns Hopkins University Press, Baltimore.

Bickel, P. J. and Ritov, Y. (1995). Local asymptotic normality ranks and covariates in transformation models. In *Festschrift for L. LeCam* (Pollard, D. and Yang, G.,eds). Springer, New York.

Bilias, Y., Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.,* **25** 662-683.

Chang, I-S. and Hsiung, C. A. (1994). Information and asymptotic efficiency in some generalized proportional hazard models for counting processes. *Ann. Statist.,* **22** 1275-1298.

Chang, I-S, Chuang Y-C and Hsiung, C.A. (1999). A class of nonparametric k-sample tests for semi-Markov processes. *Statistica Sinica,* **9** 211-277.

Chang, I-S, Hsiung, C.A. and Wu, S-M. (2000). Estimation in a proportional hazard model for semi-Markov counting process. *Statistica Sinica,* **10** 1257-1266.

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

Chintagunta, P. and Prasad, A.R. (1998). An empirical investigation of the "Dynamic McFadden" model of purchase timing and brand choice: implications for market structure. *J. Business and Economic Statist.*, **16** 2-12.

Cinlar, E. (1975). *Introduction to Stochastic Processes.* Prentice-Hall, New Jersey.

Cook, R. J. and Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Events*. Springer, New York.

Commenges, D. Semi-Markov and non-homogeneous Markov models in medical studies. (1986) . In *Semi-Markov models.* (Edited by J. Janssen) 411-422. Plenum Press, New York, 411-422.

Commenges, D., Joly, P., Gégout-Petit, A. and Liquet, B. (2007). Choice between semi-parametric estimators for Markov and non-Markov multistate models from coarsened observations. *Scand. J. Statist.,* **34** 33-52.

Cox, D. R. The statistical analysis of dependencies in point processes. (1973). *Symposium on Point Processes* (Lewis, P. A. W., Ed.). Wiley, New York.

Cutler, C. and Antin, J.H. (2001) Peripheral blood stem cells for allogeneic transplantation: a review. *Stem Cells*, **19** 108-117.

Cutler, C., Giri, S., Jeyapalan, S., Paniagua, D., Viswanathan, A., and Antin, J. H. (2001). Acute and chronic graft-versus-host disease after allogeneic peripheral blood stem-cell and bone marrow transplantation: a meta analysis. *J. Clin. Oncol.* **19** 3685-3691.

Dabrowska, D. M., Sun, G. and Horowitz, M. M. (1994). Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. *J. Amer. Statist. Assoc.,* **89** 867–877.

Dabrowska, D. M. (1995). Estimation of transition probabilities and bootstrap in a semi–parametric Markov renewal model. *J. Nonparametric Statist.,* **5** 237–259.

Dabrowska, D.M. (2006). Estimation in a class of semi-parametric transformation models. In *"Second Erich L. Lehmann Symposium - Optimality"* (J. Rojo, Ed.) Institute of Mathematical Statistics, Lecture Notes and Monograph Series, **49** 166-216.

Dabrowska, D.M. (2007). Information bounds and efficient estimation in a class of censored transformation models. *Acta Applicandae Mathematicae,* **96** 177-201.

Dabrowska, D.M. (2009). Estimation in a semi-parametric two-stage renewal regression model. *Statistica Sinica*, **19**, 981-996.

Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.

Dellacherie, C. and Meyer, P.A. (1975) *Probabilities and Potentiel*, Hermann, Paris.

de la Peña, V. and Giné, H. (1999). *Decoupling: From Dependence to Independence*. Springer, New York.

Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

Eerola, M. (1994). *Probabilistic causality in longitudinal studies*. Springer, New York.

Flowers, M. E. D., Parker, P.M., Johnston, L.J., Matos, A.V., Storer, B., Bensinger, W. I., Storb, R., Appelbaum, F. R., Forman, S. J., Blume, K. G. and Martin, P. J. (2002). Comparison of chronic graft-versus-host disease after transplantation of peripheral blood stem cells versus bone marrow in allogeneic recipients: long-term follow-up of a randomized trial. *Blood,* **100** 415-419.

Friedrichs, B., Tichelli, A., Bacigalupo, A. Russel,N.H., Ruutu, T., Shapira, Y.M., Beksac, M., Hasenclever, D., Socié, G. and Schmitz, N. (2001) Long-term outcome and late effects in patients transplanted with mobilised blood or bone marrow: a randomised trial. *Lancet Oncology,* **11**, 331-338.

Gale, R. P., Bortin, M. M., van Bekkum, D. W., Biggs, J.C., Dicke, K.A., Gluckman, E., Good, R.A., Hoffman, R.G., Key, H. E. M., Kersey, J.H., Marmont, A., Masaoka, T., Rimm, A.A., van Rood, J.J.and Zwaan, F.E. (1987). Risk factors for acute graft-versus-host disease. *Br. J. Haematol.,*, **67**, 397-406.

Gill, R. D. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrscheinlichkeitstheorie verv. Gebiete,* **53** 97-116.

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

Greenwood, P. and Wefelmeyer W . Empirical estimators for semi-Markov processes. (1996). *Math. Meth. Statist.,* **5** 299-315.

Greenwood, P., Müller, U.U. and Wefelmeyer W . Semi-Markov processes and their applications. (2004). *Commun. Stat. Theory Methods,* **33** 419-435.

Himmelberg, C. J. (1975). Measurable relations. *Fund. Math.* **87** 53-72.

Hjort, N. L. and Cleaskens, G. (2003) Frequentist model average estimators. *J. Amer.Statist.Assoc.,* **98** 938-945.

Hjort, N. L. and Cleaskens, G. (2006) Focused information criteria and model averaging for Cox's hazard regression model. *J. Amer.Statist. Assoc.,* **101** 1449-1464.

Jacod, J. (1975) . Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verv. Gebiete*, **31** 235-254.

Janssen, J. (1999) *Semi-Markov Models: Theory and Applications.* Springer, New York.

Janssen, J. and Manca, R. (2007) *Semi-Markov Risk Models for Finance, Insurance and Reliability*. Springer, New York.

Janssen, J. and Manca, R. (2006) *Applied Semi-Markov Processes* Springer, New York.

Janssen, J., Limnios, N. (2001) *International Symposium on Semi-Markov Models: Theory and Applications*. Kluwer, Academic Press.

Jones, M. P. and Crowley, J.J. (1992). Nonparametric tests of the Markov model for survival data. *Biometrika*, **79** 513-522.

Kalbfleisch, J. D. and Prentice, R. L. (1981). *Statistical Analysis of Failure Time Data*. Wiley

Karr, A.F. (1991). *Point Processes and their Statistical Inference,* Marcel Dekker, New York.

Keiding, N. (1986). Statistical analysis of semi-Markov models based on the theory of counting processes. In *Semi-Markov models. Theory and Applications*, (J. Janssen, Ed.) Plenum Press, 301-315.

Keiding, N., Klein, JP and Horowitz,MM. (2001). Multistate models and outcome prediction in bone marrow transplantation. *Statist. Med.*, **20** 1871-1885.

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. *Statist. Med.*, **12** 2315-2332.

Kosorok, M. R. , Lee B. L. and Fine J. P. (2004). Robust inference for univariate proportional hazard models. *Ann. Statist.*, **32** 1448-1491.

Kuratowski, K. (1966). *Topology*. Academic Press.

Lagakos, S. W., Sommer, C. J. and Zelen, M. (1978). Semi-Markov models for censored data. *Biometrika*, **65** 311-317.

Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line: the Dynamic Approach.* Springer, New York.

Limnios, N and Oprisan. (2001). *Semi-Markov Processe and Reliability*. Springer.

Lin, D. Y., Fleming, T. R., Wei, L.J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika*, **81** 73-81.

Lo, S.M.S. and Wilke, R.A. (2010). A copula model for dependent competing risks. *Appl. Statist.*, **59**, 359-376.

Martinussen, T. and Scheike,T. (2006). *Dynamic Regression Models for Survival Data*. Springer, New York.

Moore,E. M. and Pyke,R. (1968). Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Stat. Math.*, **20** 411-468.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.,* **15** 780-799.

Oakes, D. (1981). Survival analysis: aspects of partial likelihood (with discussion). *Int. Statist. Rev., **49*** 235-264.

Oakes, D. and Cui, L. (1994) On semi-parametric inference for modulated renewal processes. *Biometrika*, **81** 83-91

Ouhbi, L. and Limnios, N. (1996). Nonparametric estimation for semi-Markov kernels with applications to reliability analysis. *Appl. Stochastic Models and Data Analysis.,* **12** 209-220.

Ouhbi, L. and Limnios, N. (1999). Nonparametric estimation for semi-Markov processes based on its hazard rate functions. *Stat. Inference Stoch. Processes,* **2** 151-173.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Verlag, New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Inst. Math. Statist., Hayward.

Phelan, M. F. (1999). Bayes estimation from a Markov renewal process. *Ann. Statist.*, **18** 603-616.

Putter, H., Fiocco, M. and Geskus, R.B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statist. Med.,* **26** 2389-2430.

Pyke, R. (1961,a). Markov renewal processes: definitions and preliminary properties. *Ann. Math. Statist.,* **32** 1231–1242.

Pyke, R. (1961,b) Markov renewal processes with finitely many states. *Ann. Math. Statist.,* **32** 1243–1259.

Pyke, R. and Schaufele, R. (1964). Limit theorems for Markov renewal processes. *Ann. Math. Statist.,* **35** 1746-1764.

Pyke, R. and Schaufele, R. (1966). The existence and uniqueness of stationary measures for Markov renewal processes. *Ann. Math. Statist.,* **37** 1439-1462.

Ringden,O. , Labopin, M., Bacigalupo, A., Arcese, W., Schaefer, U.W., Willemze, R., Koc, H., Bunjes, D. Gluckman,E., Rocha, V. Schattenberg, A. and Frassoni, F. (2002). Transplantation of peripheral blood stem cell as compared with bone marrow from HLA-identical siblings in adult patients with acute myeloid leukemia and acute lymphoblastic leukemia. *J. Clin. Oncol.* **20(24)** 4655-4664.

Rivest, L. P. and Wells, M. T. (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *J. Multiv. Analysis,* **79**, 138-155.

Teicher , H. (1998). On the Marcinkiewicz-Zygmund strong law for U-statistics. *J. Theoret. Probab.,* **11** 279-288.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and Empirical Processes with Applications to Statistics*. Springer, New York.

Voelkel, J. G. and Crowley, J. J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations. *Ann. Statist.,* **12** 142-160.

Wagner, D. H. (1977). Survey of measurable selection theorems. *SIAM, J. Control and Optimization*, **15**, 859-903.

Weiss, G.H. and Zelen, M. (1965). A semi-Markov model for clinical trials. *J. Appl. Probab.,* **2** 269-285.

Zheng, M. and Klein, J. P. (1995). estimates of marginal survival for dependent competing risks based on an assumed copula model. *Biometrika*, **82** 127-138.