Additive hazards Markov regression models illustrated with bone marrow transplant data

BY YOUYI SHU
Centocor, Inc., 200 Great Valley Parkway, Mailstop C-4-1, Malvern, Pennsylvania 19355, U.S.A.
yshu@cntus.jnj.com

AND JOHN P. KLEIN
Division of Biostatistics, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, U.S.A.
klein@mcw.edu

SUMMARY

When there are covariate effects to be considered, multistate survival analysis is dominated either by parametric Markov regression models or by semiparametric Markov regression models using Cox's (1972) proportional hazards models for transition intensities between the states. The purpose of this research work is to study alternatives to Cox's model in a general finite-state Markov process setting. We shall look at two alternative models, Aalen's (1989) nonparametric additive hazards model and Lin & Ying's (1994) semiparametric additive hazards model. The former allows the effects of covariates to vary freely over time, while the latter assumes that the regression coefficients are constant over time. With the basic tools of the product integral and the functional delta-method, we present an estimator of the transition probability matrix and develop the large sample theory for the estimator under each of these two models. Data on 1459 HLA identical sibling transplants for acute leukaemia from the International Bone Marrow Transplant Registry serve as illustration.

Some key words: Additive hazards; Additive risk; Bone marrow transplant; Markov regres-

sion model; Multistate model; Multistate survival analysis; Product integral.

1. Introduction

In many clinical and epidemiological follow-up studies, subjects may experience events of multiple types. For example, a patient recovering from a bone marrow transplant for leukaemia may fail therapy because of one of several terminal events, such as death in remission, relapse or simply death. As the patient recovers, intermediate events may occur that have an influence on the eventual prognosis, such as the return of the patient's platelets to a 'normal' level, the development of various types of infection and the occurrence of acute or chronic graft-versus-host disease.

A natural way to model complex experiments of this kind is to use a multistate model. Figure 1 shows a simplified diagram of the recovery process as previously explored by Keiding et al. (2001). Here, for illustrative purposes, we only consider four events: acute graft-versus-host disease (A), chronic graft-versus-host disease (C), death in remission (D) and relapse (R). These four events are modelled by a six-state model with two absorbing states, D and R, and four transient states, Tx, which stands for transplantation, A, C and AC, which stands for both A and C. Note that relapse is treated as an absorbing state since patients who relapse are typically considered as failures of the treatment, and that chronic prior to acute graft-versus-host disease is not biologically possible. Our interest centres on predicting the probability that a patient will be in one of the six states at some time after transplant.

[Fig. 1 about here.]

Multistate models have traditionally been represented by Markov models which assume that the transition rates depend only on the current state of the patient and not on the patient's history. When there is no covariate, Aalen & Johansen (1978) showed how counting process methods can be used to estimate transition probabilities.

When there are covariates which may affect the rate of transition from one state to the next, a number of Markov models have been proposed in the literature. These include parametric models for the transition intensities (Begg & Larson, 1982; Kalbfleisch & Lawless, 1985; Marshall & Jones, 1995; Alioum & Commenges, 2001; Pérez-Ocón et al., 2001) or semiparametric Markov regression models where transition intensities are modelled by the Cox (1972) proportional hazards regression models (Andersen, 1988; Andersen et al., 1991; Klein et al., 1993; Klein & Qian, 1996; Andersen et al., 2000). In this paper, we focus on the semiparametric case. For convenience, the semiparametric proportional hazards Markov regression model will be called the 'Cox Markov model' in the sequel.

Two versions of the Cox Markov model have been suggested. The first, due to Andersen & Gill (1982) and Andersen et al. (1991), fits a distinct Cox model to each of the transition rates. Hereafter we shall refer to this version as the 'Andersen-Gill Cox Markov model'. In the bone marrow transplant example, this model entails eleven separate Cox regression analyses. These regression analyses are then synthesized to obtain estimates of the transition probabilities.

Klein et al. (1993) suggested fitting a Cox model to each of the events with indicator-type time-dependent covariates used to model the timing of the intermediate events that precede the event of interest. Their model, which essentially assumes proportional hazards for all transitions into the same event, is a special case of the Andersen-Gill Cox Markov model and will be referred to as the 'Klein-Keiding Cox Markov model'. In the bone marrow transplant example this approach requires the fitting of four Cox models. The transition probability estimators and their asymptotics for this model can be found in Klein & Qian (1996).

Recently, Aalen et al. (2001) suggested an alternative to the Cox model in this setting; see also the 2001 unpublished Medical College of Wisconsin Ph.D. dissertation of Y. Shu. They suggested that the transition intensities be modelled by Aalen's (1989) additive hazards regression models rather than Cox's proportional hazards regression models. They briefly touched upon the asymptotics of the estimators of the transition probabilities.

In this paper, we discuss the use of Lin & Ying's (1994) additive hazards regression model for the transition intensities. This model, as opposed to Aalen's (1989) model, assumes that the regression coefficients are constant over time. It is flexible enough to allow analogues of both the Andersen-Gill and Klein-Keiding models in the additive regression settings. It has a further advantage over both the Cox and Aalen (1989) models in that the estimates of the regression coefficients have a closed-form solution.

The remainder of the paper proceeds as follows. In § 2, we describe Markov processes and introduce the powerful tool of the product integral. In § 3, we present three additive hazards Markov regression models, namely, the Aalen Markov model, the Andersen-Gill Lin-Ying Markov model and the Klein-Keiding Lin-Ying Markov model. The first two are analogues of the Andersen-Gill Cox Markov model, and the third one is an analogue of the Klein-Keiding Cox Markov model. For each model, we present the synthesized estimators of the transition probabilities and the asymptotics of the estimators. Proofs of the asymptotics are found in the appendices. The results of fitting various Markov models to the bone marrow transplant data are reported in § 4. We conclude with some discussion in § 5.

2. Markov Processes and the Product Integral

Suppose that observations are made on a group of individuals who independently move among k states, denoted by $1, \ldots, k$, according to a nonhomogeneous, time-continuous Markov process $\{\Gamma(t),\, t \geq 0\}$ having $k \times k$ transition intensity matrix $\alpha(t) = \{\alpha_{hj}(t);\, h, j = 1, \ldots, k\}$ with diagonal elements $\alpha_{hh}(t) = -\sum_{j \neq h} \alpha_{hj}(t)$, and having transition probability matrix $P(s,t) = \{P_{hj}(s,t);\, h, j = 1, \ldots, k\}$ with (h, j)th element

$$P_{hj}(s,t) = \mathrm{pr}\{\Gamma(t) = j | \Gamma(s) = h\}, \;\; s \leq t.$$

Conditioning on the initial states, for n conditionally independent replications of this process, subject to quite general censoring patterns, the multivariate counting process $N(t) = \{N_{hj}(t); h \neq j\}$, with $N_{hj}(t)$ counting the number of observed direct transitions from state h to state j in $[0, t]$, has intensity process $\lambda(t) = \{\lambda_{hj}(t); h \neq j\}$ of the multiplicative form $\lambda_{hj}(t) = Y_h(t)\alpha_{hj}(t)$. Here $Y_h(t) \leq n$ is the number of sample paths observed to be in state h just prior to time t.

We define $A_{hj}(t) = \int_0^t \alpha_{hj}(u)du$ for all h, j. The cumulative transition intensities $A_{hj}(t)$ $(h \neq j)$ can be estimated by the well-known Nelson-Aalen estimators

$$\hat{A}_{hj}(t) = \int_0^t \frac{J_h(u)dN_{hj}(u)}{Y_h(u)},$$

where $J_h(t) = I\{Y_h(t) > 0\}$ with $I(\cdot)$ being the indicator function. The transition probabilities $P_{hj}(s, t)$ depend on the transition intensities α_{hj} through the Kolmogorov forward differential equations, whose solution may be represented as the matrix product integral (Gill & Johansen, 1990)

$$P(s, t) = \prod_{(s,t]} \Big\{I + dA(u)\Big\},$$

where $A(u) = \{A_{hj}(u)\}$ and I is the $k \times k$ identity matrix. Aalen & Johansen (1978) exploited this relationship to estimate the transition probability matrix by

$$\hat{P}(s, t) = \prod_{(s,t]} \Big\{I + d\hat{A}(u)\Big\}, \tag{1}$$

where $\hat{A}(u) = \{\hat{A}_{hj}(u)\}$ with diagonal elements $\hat{A}_{hh}(u) = -\sum_{j \neq h} \hat{A}_{hj}(u)$.

When there are covariates to be adjusted for, we assume that each transition intensity $\alpha_{hj}(t; Z_i)$ $(h, j = 1, \ldots, k,\ h \neq j)$ from state h to state j in a Markov process for individual i $(i = 1, \ldots, n)$ with a vector of fixed-time covariates Z_i follows a given regression model. Our interest lies in predicting the transition probabilities $P_{hj}(s, t; Z_0)$ for individuals with a given vector of fixed-time covariates, Z_0. These estimates are obtained by replacing $\hat{A}(t)$ in (1) with $\hat{A}(t; Z_0)$, an estimate of $A(t; Z_0)$, obtained from the regression model.

The large sample properties of the transition probability matrix estimator $\hat{P}(s,t;Z_0) = \{\hat{P}_{hj}(s,t;Z_0);\ h,j = 1,\ldots,k\}$ can be found for Markov regression models with the aid of the following lemma.

LEMMA 1. *Let τ be a model-dependent upper bound of the time interval for which asymptotic results are desired. Let $s, v \in [0,\tau)$ with $s < v$. Assume that the matrix-valued estimator $\hat{A}(\cdot;Z_0)$ is of uniformly bounded total variation, with probability tending to 1, over $[s,v]$, and that $n^{1/2}\int_s^{\cdot} d\{\hat{A}(u;Z_0) - A(u;Z_0)\}$ converges weakly on $[s,v]$ to a limiting process $U(\cdot;Z_0)$. Define, for $t \in [s,v]$,*

$$P(s,t;Z_0) = \prod_{(s,t]}\Big\{I + dA(u;Z_0)\Big\}, \quad \hat{P}(s,t;Z_0) = \prod_{(s,t]}\Big\{I + d\hat{A}(u;Z_0)\Big\}.$$

Then $n^{1/2}\{\hat{P}(s,\cdot;Z_0) - P(s,\cdot;Z_0)\}$ converges weakly on $[s,v]$ to the process

$$\int_s^{\cdot} P(s,u-;Z_0)\, dU(u;Z_0) P(u,\cdot;Z_0).$$

Moreover, we have

$$n^{1/2}\{\hat{P}(s,t;Z_0) - P(s,t;Z_0)\} \overset{a}{=} \int_s^t P(s,u-;Z_0)\, d\Big\{n^{1/2}(\hat{A} - A)(u;Z_0)\Big\} P(u,t;Z_0),$$

where $\overset{a}{=}$ denotes 'asymptotically equivalent'; that is, convergence in probability to zero of the supremum norm of the difference.

Proof. This is easily proved by the compact differentiability of the product integral (Gill & Johansen, 1990, Theorem 8) and by applying the functional delta-method (Andersen et al., 1993, Theorem II.8.1). □

The rather condensed notation of matrix product integral may not be so illuminating. In many cases it is possible to write out explicit expressions for the transition probability estimates $\hat{P}_{hj}(s,t;Z_0)$. This is the case with our six-state bone marrow transplant model, see Fig. 1, as demonstrated in the following example.

Example 1. For the bone marrow transplant example, the estimated cumulative transition intensity matrix is, with t and Z_0 suppressed for ease of exposition,

$$\hat{A} = \begin{bmatrix} -\hat{A}_{12}-\hat{A}_{13}-\hat{A}_{15}-\hat{A}_{16} & \hat{A}_{12} & \hat{A}_{13} & 0 & \hat{A}_{15} & \hat{A}_{16} \\ 0 & -\hat{A}_{24}-\hat{A}_{25}-\hat{A}_{26} & 0 & \hat{A}_{24} & \hat{A}_{25} & \hat{A}_{26} \\ 0 & 0 & -\hat{A}_{35}-\hat{A}_{36} & 0 & \hat{A}_{35} & \hat{A}_{36} \\ 0 & 0 & 0 & -\hat{A}_{45}-\hat{A}_{46} & \hat{A}_{45} & \hat{A}_{46} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let $\mathcal{J}$ denote the set of all possible transitions, that is,

$$\mathcal{J} = \{12, 13, 15, 16, 24, 25, 26, 35, 36, 45, 46\}.$$

Then the sixteen nonzero transition probability estimates are

$$\hat{P}_{hh}(s,t;Z_0) = \prod_{(s,t]}\Big\{1 - \sum_{j:\, j>h,\, hj\in\mathcal{J}} d\hat{A}_{hj}(u;Z_0)\Big\}, \quad h = 1,2,3,4;$$

$$\hat{P}_{hj}(s,t;Z_0) = \int_s^t \hat{P}_{hh}(s,u-;Z_0)\, d\hat{A}_{hj}(u;Z_0), \quad hj = 35, 36, 45, 46;$$

$$\hat{P}_{2j}(s,t;Z_0) = \int_s^t \hat{P}_{22}(s,u-;Z_0)\left\{d\hat{A}_{2j}(u;Z_0) + \hat{P}_{4j}(u,t;Z_0)\, d\hat{A}_{24}(u;Z_0)\right\}, \quad j = 5,6;$$

$$\hat{P}_{hj}(s,t;Z_0) = \int_s^t \hat{P}_{hh}(s,u-;Z_0)\hat{P}_{jj}(u,t;Z_0)\, d\hat{A}_{hj}(u;Z_0), \quad hj = 12, 13, 24;$$

$$\hat{P}_{14}(s,t;Z_0) = \int_s^t \hat{P}_{11}(s,u-;Z_0)\hat{P}_{24}(u,t;Z_0)\, d\hat{A}_{12}(u;Z_0);$$

$$\hat{P}_{1j}(s,t;Z_0) = \int_s^t \hat{P}_{11}(s,u-;Z_0)\left\{d\hat{A}_{1j}(u;Z_0) + \hat{P}_{2j}(u,t;Z_0)\, d\hat{A}_{12}(u;Z_0) + \hat{P}_{3j}(u,t;Z_0)\, d\hat{A}_{13}(u;Z_0)\right\}, \quad j = 5,6.$$

At this point, a few words about these transition probability estimates are in order, for some of these quantities are of particular interest and will be illustrated in § 4. For example, given that individuals with fixed covariates Z_0 were initially in state 1 at time 0, $\hat{P}_{1j}(0,t;Z_0)$ $(j = 1,\ldots,6)$ is the estimated probability of being in state j at time t. Here, in particular, $\hat{P}_{15}(0,t;Z_0)$ and $\hat{P}_{16}(0,t;Z_0)$ are the estimated probability of death in remission at time t and the estimated probability of relapse at time t, respectively. The estimated leukaemia-free survival function is given by $1 - \hat{P}_{15}(0,t;Z_0) - \hat{P}_{16}(0,t;Z_0)$.

3. The Models

3·1. *The Aalen Markov Model*

Consider a group of n individuals indexed by $i = 1, \ldots, n$, each with a $(p+1) \times 1$ time-fixed covariate vector $Z_i = (1, Z_{i1}, \ldots, Z_{ip})^\top$. Define the counting process $N_{hji}(t)$ $(h, j = 1, \ldots, k,\ h \neq j)$ to be the number of direct transitions from state h to state j observed for individual i in the time interval $[0, t]$. Let $\alpha_{hj}(t; Z_i)$ be the transition intensity from state h to state j for individual i, and let

$$Y_{hi}(t) = I(\text{the } i\text{th individual is observed to be in state } h \text{ just prior to time } t).$$

Then, under independent censoring, $N_{hji}(t)$ can be uniquely decomposed as

$$N_{hji}(t) = \int_0^t Y_{hi}(u)\alpha_{hj}(u; Z_i)du + M_{hji}(t),$$

where $M_{hji}(t)$ is a local square integrable martingale.

The Aalen Markov model assumes that each transition intensity $\alpha_{hj}(t; Z_i)$ follows an Aalen (1989) additive model:

$$\alpha_{hj}(t; Z_i) = \beta_{hj0}(t) + \beta_{hj1}(t)Z_{i1} + \cdots + \beta_{hjp}(t)Z_{ip},$$
$$h, j = 1, \ldots, k,\ h \neq j,\ i = 1, \ldots, n,$$

where $\beta_{hjw}(t)$, $w = 0, 1, \ldots, p$, are unknown regression functions; note that if a covariate has no effect on the intensity of the $h \to j$ transition then the regression function for that covariate is set to a constant value of 0. This formulation of the model was initially presented in Aalen et al. (2001), and had been independently studied in Y. Shu's dissertation.

Let $N_{hj}(t) = (N_{hj1}(t), \ldots, N_{hjn}(t))^\top$, $\beta_{hj}(t) = (\beta_{hj0}(t), \beta_{hj1}(t), \ldots, \beta_{hjp}(t))^\top$. Let $Y_h(t)$ be the $n \times (p+1)$ matrix with ith row, $i = 1, \ldots, n$, given by

$$Y_{hi}(t)(1, Z_{i1}, \ldots, Z_{ip}) = (Y_{hiw}(t),\ w = 0, 1, \ldots, p).$$

Aalen (1989) showed that the vector of cumulative regression functions $B_{hj}(t) = \int_0^t \beta_{hj}(u)du$ may be estimated by

$$\hat{B}_{hj}(t) = \int_0^t J_h(u) \left\{ Y_h^\top(u)\, Y_h(u) \right\}^{-1} Y_h^\top(u) dN_{hj}(u), \quad h \neq j,$$

where $J_h(t)$ is the predictable indicator of $Y_h(t)$ having full column rank, with the assumption that $p + 1 \leq n$.

For individuals with given fixed covariate vector $Z_0 = (1, Z_{01}, \ldots, Z_{0p})^\top$, their cumulative intensities for the $h \to j$ transition are estimated by

$$\hat{A}_{hj}(t; Z_0) = Z_0^\top \hat{B}_{hj}(t), \quad h \neq j. \tag{2}$$

Thus, if we define

$$\hat{A}_{hh}(t; Z_0) = -\sum_{j \neq h} \hat{A}_{hj}(t; Z_0), \quad h = 1, \ldots, k,$$

then the transition probability matrix $P(s, t; Z_0) = \{P_{hj}(s, t; Z_0);\, h, j = 1, \ldots, k\}$ is estimated by the product integral, cf. § 2,

$$\hat{P}(s, t; Z_0) = \prod_{(s,t]} \left\{ I + d\hat{A}(u; Z_0) \right\}, \tag{3}$$

where $\hat{A}(t; Z_0) = \{\hat{A}_{hj}(t; Z_0)\}$.

The asymptotic properties of the estimator $\hat{P}(s, t; Z_0) = \{\hat{P}_{hj}(s, t; Z_0);\, h, j = 1, \ldots, k\}$ are given by the following theorem.

Theorem 1 (*The Aalen Markov model*). *Let*

$$\tau = \sup \left\{ u : \int_0^u |\beta_{hjw}(\tilde{u})| d\tilde{u} < \infty,\, h, j = 1, \ldots, k,\, h \neq j,\, w = 0, 1, \ldots, p \right\}$$

and let $s, v \in [0, \tau)$ *with* $s < v$. *Then, under regularity conditions (see Appendix 1), we have the following:*

(i) the process $n^{1/2}\{\hat{P}(s, \cdot; Z_0) - P(s, \cdot; Z_0)\}$ *converges weakly on* $[s, v]$ *to a zero-mean Gaussian process;*

(ii) $\text{cov}\{\hat{P}_{hj}(s,t;Z_0),\ \hat{P}_{mr}(s,t;Z_0)\}$ $(s \le t \le v)$ *can be estimated uniformly consistently by*

$$
\begin{aligned}
&\widehat{\text{cov}}\{\hat{P}_{hj}(s,t;Z_0),\ \hat{P}_{mr}(s,t;Z_0)\} \\
&= \sum_{g=1}^{k}\sum_{l\neq g}\int_s^t \hat{F}_{gl}^{(hj)}(u;s,t,Z_0)\hat{F}_{gl}^{(mr)}(u;s,t,Z_0) \\
&\qquad \times Z_0^\top J_g(u)\left\{Y_g^\top(u)\,Y_g(u)\right\}^{-1} Y_g^\top(u)\,\text{diag}\{dN_{gl}(u)\}\,Y_g(u)\left\{Y_g^\top(u)\,Y_g(u)\right\}^{-1} Z_0, \quad (4)
\end{aligned}
$$

where $\hat{F}_{gl}^{(hj)}(u;s,t,Z_0) = \hat{P}_{hg}(s,u-;Z_0)\{\hat{P}_{lj}(u,t;Z_0) - \hat{P}_{gj}(u,t;Z_0)\}$, *and* $\text{diag}(\rho)$ *for an* n*-vector* ρ *is the* $n \times n$ *diagonal matrix with the elements of* ρ *on the diagonal.*

3·2. *The Andersen-Gill Lin-Ying Markov model*

An alternative to Aalen's (1989) additive hazards model is Lin & Ying's (1994) additive hazards model. This model assumes that the regression coefficients do not depend on time, so that the transition intensities are modelled by

$$\alpha_{hj}(t;Z_i) = \alpha_{hj0}(t) + \beta_{hj}^\top Z_i, \quad h,j = 1,\ldots,k,\ h \neq j,\ i = 1,\ldots,n,$$

where $Z_i = (Z_{i1},\ldots,Z_{ip})^\top$ is a p-vector of time-fixed covariates, $\alpha_{hj0}(t)$ is an unspecified baseline intensity for the $h \to j$ transition, and β_{hj} is a p-vector of unknown regression parameters for the $h \to j$ transition.

Let $N_{hj}(t) = \sum_{i=1}^n N_{hji}(t)$, $Y_h(t) = \sum_{i=1}^n Y_{hi}(t)$, $\bar{Z}_h(t) = \sum_{i=1}^n Y_{hi}(t)Z_i/Y_h(t)$. For a p-vector a, let $a^{\otimes 2}$ denote the $p \times p$ matrix $aa^\top$. Lin & Ying (1994) showed that the estimator of β_{hj} can be written explicitly as

$$\hat{\beta}_{hj} = \left[\sum_{i=1}^n \int_0^\infty Y_{hi}(u)\{Z_i - \bar{Z}_h(u)\}^{\otimes 2}\,du\right]^{-1}\left[\sum_{i=1}^n \int_0^\infty \{Z_i - \bar{Z}_h(u)\}\,dN_{hji}(u)\right]$$

with variance-covariance matrix consistently estimated by $\widehat{\text{cov}}(\hat{\beta}_{hj}) = \dfrac{1}{n}\hat{\Omega}_h^{-1}\hat{V}_{hj}\hat{\Omega}_h^{-1}$, where

$$\hat{\Omega}_h = \frac{1}{n}\sum_{i=1}^n \int_0^\infty Y_{hi}(u)\{Z_i - \bar{Z}_h(u)\}^{\otimes 2}\,du,$$

$$\hat{V}_{hj} = \frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \{Z_i - \bar{Z}_h(u)\}^{\otimes 2}\, dN_{hji}(u).$$

Furthermore, the cumulative baseline hazard function, $A_{hj0}(t) = \int_0^t \alpha_{hj0}(u)du$, can be estimated uniformly consistently by

$$\hat{A}_{hj0}(t, \hat{\beta}_{hj}) = \int_0^t \frac{dN_{hj}(u)}{Y_h(u)} - \left\{\int_0^t \bar{Z}_h(u)du\right\}^\top \hat{\beta}_{hj}. \tag{5}$$

To estimate the transition probability matrix $P(s,t;Z_0)$ for individuals with given fixed covariate vector $Z_0 = (Z_{01}, \ldots, Z_{0p})^\top$, we let

$$\hat{A}_{hj}(t; Z_0) = \hat{A}_{hj0}(t, \hat{\beta}_{hj}) + t\hat{\beta}_{hj}^\top Z_0, \quad h \neq j.$$

Then $P(s,t;Z_0)$ is again estimated by the product integral in the manner of (3).

The asymptotic properties of the estimated transition probabilities are found in the following theorem.

Theorem 2 (*The Andersen-Gill Lin-Ying Markov model*). *Let*

$$\tau = \sup\left\{u : \int_0^u \alpha_{hj0}(\tilde{u})d\tilde{u} < \infty,\ h,j = 1,\ldots,k,\ h \neq j\right\}$$

and let $s, v \in [0,\tau)$ with $s < v$. Then, under regularity conditions (see Appendix 2), we have the following:

(i) the process $n^{1/2}\{\hat{P}(s,\cdot;Z_0) - P(s,\cdot;Z_0)\}$ converges weakly on $[s,v]$ to a zero-mean Gaussian process;

(ii) $\mathrm{cov}\{\hat{P}_{hj}(s,t;Z_0),\ \hat{P}_{mr}(s,t;Z_0)\}$ $(s \leq t \leq v)$ *can be estimated uniformly consistently by*

$$\begin{aligned} &\hat{\mathrm{cov}}\{\hat{P}_{hj}(s,t;Z_0),\ \hat{P}_{mr}(s,t;Z_0)\} \\ &= \sum_{g=1}^{k}\sum_{l\neq g}\{\hat{Q}_{gl}^{(hj)}(s,t;Z_0)\}^\top \hat{\mathrm{cov}}(\hat{\beta}_{gl})\hat{Q}_{gl}^{(mr)}(s,t;Z_0) \\ &\quad + \sum_{g=1}^{k}\sum_{l\neq g}\int_s^t \hat{F}_{gl}^{(hj)}(u;s,t,Z_0)\hat{F}_{gl}^{(mr)}(u;s,t,Z_0)\frac{dN_{gl}(u)}{\{Y_g(u)\}^2} \end{aligned}$$

$$+\frac{1}{n}\sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\hat{Q}_{gl}^{(hj)}(s,t;Z_0)\}^\top\hat{\Omega}_g^{-1}\int_s^t \frac{Z_i-\bar{Z}_g(u)}{Y_g(u)}\hat{F}_{gl}^{(mr)}(u;s,t,Z_0)\,dN_{gli}(u)$$

$$+\frac{1}{n}\sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\hat{Q}_{gl}^{(mr)}(s,t;Z_0)\}^\top\hat{\Omega}_g^{-1}\int_s^t \frac{Z_i-\bar{Z}_g(u)}{Y_g(u)}\hat{F}_{gl}^{(hj)}(u;s,t,Z_0)\,dN_{gli}(u), \qquad (6)$$

where

$$\hat{F}_{gl}^{(hj)}(u;s,t,Z_0)=\hat{P}_{hg}(s,u-;Z_0)\{\hat{P}_{lj}(u,t;Z_0)-\hat{P}_{gj}(u,t;Z_0)\},$$

$$\hat{Q}_{gl}^{(hj)}(s,t;Z_0)=\int_s^t \hat{F}_{gl}^{(hj)}(u;s,t,Z_0)\{Z_0-\bar{Z}_g(u)\}\,du.$$

3·3. *The Klein-Keiding Lin-Ying Markov Model*

The two additive hazards Markov regression models presented in the previous subsections require that an additive hazards regression model be fitted to each of the transition intensities. An alternative approach, in the spirit of Klein et al. (1993), is to fit a distinct additive regression model to each event with time-dependent indicators for the intermediate events. In the bone marrow transplant example, see Fig. 1, this approach fits a model to acute graft-versus-host disease, a model to chronic graft-versus-host disease with a time-dependent indicator for acute graft-versus-host disease, and models to relapse and death in remission with time-dependent indicators for acute and chronic graft-versus-host disease. The model for each of these regressions is taken to be a Lin & Ying (1994) additive model with fixed regression coefficients. Here, we will not consider the Aalen (1989) model which turns out to produce estimates identical to those of the Aalen Markov model of § 3.1.

To formulate the general finite-state Klein-Keiding Lin-Ying Markov model, we assume that an individual is at the risk of having any one of the events in a set $\mathcal{E}$. This set consists of both the intermediate and terminal events. In our bone marrow transplant example, $\mathcal{E}=\{A,C,D,R\}$.

As in Example 1 of § 2, we let $\mathcal{J}$ denote the set of all possible transitions. In the bone marrow transplant example, $\mathcal{J}=\{12,13,15,16,24,25,26,35,36,45,46\}$ has eleven elements. For any event $X \in \mathcal{E}$, we define $\mathcal{J}(X)$ as the set of transitions into event X that are possible. We call the transitions in $\mathcal{J}(X)$ 'X-transitions'. In our example, $\mathcal{J}(A) = \{12\}$, $\mathcal{J}(C) = \{13, 24\}$, $\mathcal{J}(D) = \{15, 25, 35, 45\}$ and $\mathcal{J}(R) = \{16, 26, 36, 46\}$. Obviously, we have $\mathcal{J} = \cup_{X \in \mathcal{E}} \mathcal{J}(X)$.

For any event $X \in \mathcal{E}$, we define the ancestor set $\mathcal{A}(X)$ as the set of intermediate events that may happen prior to the occurrence of the event X. In our example, we have $\mathcal{A}(A) = \emptyset$, the empty set, $\mathcal{A}(C) = \{A\}$ and $\mathcal{A}(D) = \mathcal{A}(R) = \{A, C\}$.

Let $Z_i = (Z_{i1}, \ldots, Z_{ip})^\top$ be the vector of fixed-time covariates that may have an influence on any event in $\mathcal{E}$ for individual i and let T_{Xi} be the occurrence time of event X $(X \in \mathcal{E})$ for individual i. For simplicity and for notational ease, we assume, for the moment, that there is no interaction effect. Then

$$Z_{Xi}(t) = \left(Z_i^\top, \{I(T_{X'i} < t), X' \in \mathcal{A}(X)\}\right)^\top \tag{7}$$

will be the ith individual's full covariate vector used in modelling the hazard rate for event X. Define the counting process $N_{Xi}(t)$ to be the number of X events observed for individual i in the time interval $[0, t]$. Let $\alpha_X\{t; Z_{Xi}(t)\}$ be the hazard rate of the time to event X for individual i, and let

$$Y_{Xi}(t) = I(\text{the } i\text{th individual is at risk of event } X \text{ just prior to time } t).$$

Then, under independent censoring, $N_{Xi}(t)$ can be uniquely decomposed as

$$N_{Xi}(t) = \int_0^t Y_{Xi}(u)\alpha_X\{u; Z_{Xi}(u)\}\, du + M_{Xi}(t),$$

where $M_{Xi}(t)$ is a local square integrable martingale.

The Klein-Keiding Lin-Ying Markov model assumes that, for each event X, the hazard rate $\alpha_X\{t; Z_{Xi}(t)\}$ follows a Lin & Ying (1994) additive model:

$$\alpha_X\{t; Z_{Xi}(t)\} = \alpha_{X0}(t) + \beta_X^\top Z_{Xi}(t), \quad X \in \mathcal{E},\ i = 1, \ldots, n, \tag{8}$$

where $\alpha_{X0}(t)$ is an arbitrary baseline hazard function for event X and β_X is a vector of unknown regression parameters for event X. Conforming to (7), we may write $\beta_X = \left(\beta_{FX}^\top, \{\beta_{X'X}, X' \in \mathcal{A}(X)\}\right)^\top$, where β_{FX} is a vector of risk coefficients for the fixed-time covariates Z_i for the event X; note that if a fixed-time covariate has no effect on the timing of event X then the risk coefficient for that covariate is set to 0. Also, $\beta_{X'X}$ is the risk coefficient for the effect of the occurrence of event X' on the time to event X. Then (8) can be rewritten more explicitly as

$$\begin{aligned} &\alpha_X\left(t; Z_i, \{I(T_{X'i} < t), X' \in \mathcal{A}(X)\}\right) \\ &= \alpha_{X0}(t) + \beta_{FX}^\top Z_i + \sum_{X' \in \mathcal{A}(X)} \beta_{X'X} I(T_{X'i} < t), \quad X \in \mathcal{E},\ i = 1, \ldots, n. \end{aligned} \tag{9}$$

Since, for each $gl \in \mathcal{J}(X)$, at any time t, state g determines the value of all the indicator functions $\{I(T_{X'i} < t), X' \in \mathcal{A}(X)\}$ by definition, we can define the hazard rate for the X-transition gl by

$$\alpha_{gl}(t; Z_i) = \alpha_{X0}(t) + \beta_X^\top Z_{gli},$$

where $Z_{gli} = \left(Z_i^\top, \{I(T_{X'i} < t), X' \in \mathcal{A}(X)\}\right)^\top$, a fixed X-transition-specific covariate vector for individual i composed of the fixed covariates plus the vector $\{I(T_{X'i} < t), X' \in \mathcal{A}(X)\}$ of 0's and 1's determined by state g.

Example 2. In the bone marrow transplant example, $\mathcal{E} = \{A, C, D, R\}$, so we fit four separate Lin & Ying (1994) additive hazards models, one for each of the four events A, C, D and R. For illustrative purposes, let us look at the event D only. Then model (9) means

$$\alpha_D\{t; Z_i, I(T_{Ai} < t), I(T_{Ci} < t)\} = \alpha_{D0}(t) + \beta_{FD}^\top Z_i + \beta_{AD} I(T_{Ai} < t) + \beta_{CD} I(T_{Ci} < t), \tag{10}$$

which implies that

$$\begin{cases} \alpha_{15}(t; Z_i) = \alpha_{D0}(t) + \beta_{FD}^\top Z_i \\ \alpha_{25}(t; Z_i) = \alpha_{D0}(t) + \beta_{FD}^\top Z_i + \beta_{AD} \\ \alpha_{35}(t; Z_i) = \alpha_{D0}(t) + \beta_{FD}^\top Z_i + \beta_{CD} \\ \alpha_{45}(t; Z_i) = \alpha_{D0}(t) + \beta_{FD}^\top Z_i + \beta_{AD} + \beta_{CD}. \end{cases} \tag{11}$$

Model (9) assumes that the fixed-time covariate vector Z_i has the same regression coefficients β_{FX} for all the X-transitions. This condition may be relaxed by including in the model interaction terms between the fixed-time covariates and the time-dependent indicator covariates.

Example 3. In Example 2, the four D-transitions, $1\to 5$, $2\to 5$, $3\to 5$ and $4\to 5$, have the same covariate effects for Z_i, cf. (11). This assumption could be easily relaxed by fitting a model for the event D with interaction effects as follows. Define $Z_{Ai}(t) = I(T_{Ai} < t,\, T_{Ci} \geq t)$, $Z_{Ci}(t) = I(T_{Ai} \geq t,\, T_{Ci} < t)$ and $Z_{ACi}(t) = I(T_{Ai} < t,\, T_{Ci} < t)$. Then an alternative model to (10) is

$$\begin{aligned}\alpha_D\{t; Z_i, Z_{Ai}(t), Z_{Ci}(t), Z_{ACi}(t)\} &= \alpha_{D0}(t) + \beta_{15}^\top Z_i\, \{1 - Z_{Ai}(t) - Z_{Ci}(t) - Z_{ACi}(t)\} \\ &\quad + \beta_{25}^\top Z_i Z_{Ai}(t) + \beta_{35}^\top Z_i Z_{Ci}(t) + \beta_{45}^\top Z_i Z_{ACi}(t) \\ &\quad + \gamma_{AD} Z_{Ai}(t) + \gamma_{CD} Z_{Ci}(t) + \gamma_{AC,D} Z_{ACi}(t),\end{aligned}$$

which leads to

$$\begin{cases} \alpha_{15}(t; Z_i) = \alpha_{D0}(t) + \beta_{15}^\top Z_i \\ \alpha_{25}(t; Z_i) = \alpha_{D0}(t) + \beta_{25}^\top Z_i + \gamma_{AD} \\ \alpha_{35}(t; Z_i) = \alpha_{D0}(t) + \beta_{35}^\top Z_i + \gamma_{CD} \\ \alpha_{45}(t; Z_i) = \alpha_{D0}(t) + \beta_{45}^\top Z_i + \gamma_{AC,D}. \end{cases}$$

Let $N_X(t) = \sum_{i=1}^n N_{Xi}(t)$, $Y_X(t) = \sum_{i=1}^n Y_{Xi}(t)$, $\bar{Z}_X(t) = \sum_{i=1}^n Y_{Xi}(t) Z_{Xi}(t)/Y_X(t)$. Lin & Ying (1994) showed that the estimator for β_X, the vector of regression parameters in model (8), can be expressed explicitly as

$$\hat\beta_X = \left[\sum_{i=1}^n \int_0^\infty Y_{Xi}(u)\{Z_{Xi}(u) - \bar{Z}_X(u)\}^{\otimes 2}\, du\right]^{-1} \left[\sum_{i=1}^n \int_0^\infty \{Z_{Xi}(u) - \bar{Z}_X(u)\}\, dN_{Xi}(u)\right],$$

the variance-covariance matrix of which is consistently estimated by $\widehat{\mathrm{cov}}(\hat\beta_X) = \dfrac{1}{n}\hat\Omega_X^{-1}\hat V_X \hat\Omega_X^{-1}$ with

$$\hat\Omega_X = \frac{1}{n}\sum_{i=1}^n \int_0^\infty Y_{Xi}(u)\{Z_{Xi}(u) - \bar{Z}_X(u)\}^{\otimes 2}\, du,$$

$$\hat{V}_X = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty}\{Z_{Xi}(u) - \bar{Z}_X(u)\}^{\otimes 2}\, dN_{Xi}(u).$$

Furthermore, the cumulative baseline hazard rate for event X, $A_{X0}(t) = \int_0^t \alpha_{X0}(u)du$, can be estimated uniformly consistently by

$$\hat{A}_{X0}(t, \hat{\beta}_X) = \int_0^t \frac{dN_X(u)}{Y_X(u)} - \left\{\int_0^t \bar{Z}_X(u)du\right\}^{\top} \hat{\beta}_X. \tag{12}$$

For individuals with given fixed covariate vector $Z_0 = (Z_{01}, \ldots, Z_{0p})^{\top}$, their estimated cumulative intensities from state g to l, $\hat{A}_{gl}(t; Z_0)$, are given by

$$\hat{A}_{gl}(t; Z_0) = \begin{cases} \hat{A}_{X0}(t, \hat{\beta}_X) + t\hat{\beta}_X^{\top} Z_{gl0}, & \text{if } g \neq l \text{ and } gl \in \mathcal{J}(X) \text{ for some } X \in \mathcal{E}; \\ 0, & \text{if } g \neq l \text{ and } gl \notin \mathcal{J}. \end{cases}$$

These estimators are again substituted into (3) to obtain the estimator for $P(s, t; Z_0)$.

We have the following result about the asymptotic distribution of the estimated transition probabilities.

Theorem 3 (*The Klein-Keiding Lin-Ying Markov model*). *Let*

$$\tau = \sup\left\{u : \int_0^u \alpha_{X0}(\tilde{u})d\tilde{u} < \infty,\ X \in \mathcal{E}\right\}$$

and let $s, v \in [0, \tau)$ with $s < v$. Then, under regularity conditions (see Appendix 3), we have the following:

(i) the process $n^{1/2}\{\hat{P}(s, \cdot; Z_0) - P(s, \cdot; Z_0)\}$ converges weakly on $[s, v]$ to a zero-mean Gaussian process;

(ii) $\mathrm{cov}\{\hat{P}_{hj}(s, t; Z_0),\ \hat{P}_{mr}(s, t; Z_0)\}$ $(s \leq t \leq v)$ can be estimated uniformly consistently by

$$\begin{aligned} &\widehat{\mathrm{cov}}\{\hat{P}_{hj}(s, t; Z_0),\ \hat{P}_{mr}(s, t; Z_0)\} \\ &= \sum_{X \in \mathcal{E}} \{\hat{Q}_X^{(hj)}(s, t; Z_0)\}^{\top} \widehat{\mathrm{cov}}(\hat{\beta}_X) \hat{Q}_X^{(mr)}(s, t; Z_0) \\ &\quad + \sum_{X \in \mathcal{E}} \int_s^t \hat{F}_X^{(hj)}(u; s, t, Z_0) \hat{F}_X^{(mr)}(u; s, t, Z_0) \frac{dN_X(u)}{\{Y_X(u)\}^2} \end{aligned}$$

$$+ \frac{1}{n} \sum_{X \in \mathcal{E}} \sum_{i=1}^{n} \{ \hat{Q}_X^{(hj)}(s,t;Z_0) \}^\top \hat{\Omega}_X^{-1} \int_s^t \frac{Z_{Xi}(u) - \bar{Z}_X(u)}{Y_X(u)} \hat{F}_X^{(mr)}(u;s,t,Z_0)\, dN_{Xi}(u)$$

$$+ \frac{1}{n} \sum_{X \in \mathcal{E}} \sum_{i=1}^{n} \{ \hat{Q}_X^{(mr)}(s,t;Z_0) \}^\top \hat{\Omega}_X^{-1} \int_s^t \frac{Z_{Xi}(u) - \bar{Z}_X(u)}{Y_X(u)} \hat{F}_X^{(hj)}(u;s,t,Z_0)\, dN_{Xi}(u), \qquad (13)$$

where

$$\hat{F}_X^{(hj)}(u;s,t,Z_0) = \sum_{gl \in \mathcal{J}(X)} \hat{H}_{gl}^{(hj)}(u;s,t,Z_0),$$

$$\hat{Q}_X^{(hj)}(s,t;Z_0) = \int_s^t \sum_{gl \in \mathcal{J}(X)} \left[\hat{H}_{gl}^{(hj)}(u;s,t,Z_0) \{ Z_{gl0} - \bar{Z}_X(u) \} \right] du,$$

$$\hat{H}_{gl}^{(hj)}(u;s,t,Z_0) = \hat{P}_{hg}(s,u-;Z_0) \{ \hat{P}_{lj}(u,t;Z_0) - \hat{P}_{gj}(u,t;Z_0) \}.$$

4. Example

We return to the bone marrow transplant example. A dataset of 1459 patients receiving an HLA-identical sibling bone marrow transplant between 1988 and 1996 was drawn from the International Bone Marrow Transplant Registry data base. Of these, 1081 patients were treated for acute myeloid leukaemia and 378 for acute lymphoblastic leukaemia. All patients were transplanted in the first remission of their disease, 267 (18.3%) died in remission and 217 (14.9%) relapsed. A total of 418 (28.7%) patients developed acute graft-versus-host disease and 407 (27.9%) developed chronic graft-versus-host disease. For each patient, a number of fixed-time covariates were recorded, such as the patient's and the donor's age and sex, Karnofsky score and the time from diagnosis to transplantation.

The purpose of this example is to illustrate the methodology we have developed, and not to attempt a thorough analysis of the data. Here we shall consider only two covariates, AGE and SEX-MATCH, of which AGE is a 0–1 binary variable with 1 indicating that the patient is more than 28 years old. The four combinations of recipient-donor SEX-MATCH are represented by three indicator variables for the instances of 'male patient, female donor', 'female patient, male donor' and 'male patient, male donor'. We fitted the three additive hazards Markov regression models of § 3, the Aalen Markov model, the Andersen-Gill Lin-Ying Markov model and the Klein-Keiding Lin-Ying Markov model with full interaction effects, as exemplified in Example 3 of § 3. For comparison, we also fitted the two proportional hazards Markov regression models, the Andersen-Gill Cox Markov model and the Klein-Keiding Cox Markov model, also with full interaction effects.

We focus on predicting outcome for a male patient aged over 28 years with a female donor. Table 1 reports the estimates and associated standard errors of the chance the patient will be in one of the six states at 100 days, 6 months, 1 year and 2 years post-transplant for each of the five models.

[Table 1 about here.]

In Fig. 2(a), we compare the predicted leukaemia-free survival function, i.e., the chance of not being in state 5 or 6, for the five models. Figure 2(b) shows the standard errors from the five models.

[Fig. 2 about here.]

We see from Table 1 and Fig. 2(a) that the probability estimates under the two Lin-Ying Markov models are pretty similar, as are the probability estimates under the two Cox Markov models. In fact, all the five models yield reasonably comparable results except at right-hand tails, where there are fewer data and greater variabilities. However, away from the tails, the precision of the probability estimates based on the two Lin-Ying Markov models seems to be higher than that based on the other three models, as is evident from Table 1 and Fig. 2(b).

Remark. In order not to make asymptotic arguments unwieldy, Theorems 1–3 impose slightly stronger sufficient conditions, such as condition (c) in Appendix 1 and the condition $\text{pr}\{Y_{h1}(u) = 1 \text{ for all } u \in [s, v]\} > 0$ in Appendix 2. For the bone marrow transplant data, such conditions are not satisfied for states $h \neq 1$ when $s = 0$. However, we believe that the asymptotic distributions for $\hat{P}_{1j}(s, t; Z_0)$ given in our theorems will continue to hold even when $s = 0$.

5. Discussion

For Aalen's (1989) model, the effects of the covariates are functions of time rather than single parameter values as in the Cox model and the Lin & Ying (1994) model. Although being capable of providing detailed information concerning the temporal influence of each covariate, Aalen's (1989) model is more limited in the number of covariates it can handle in practical data analysis. Note that, as opposed to a step function from a Cox model or an Aalen (1989) model, the estimated cumulative baseline hazard function from a Lin & Ying (1994) model is piecewise continuous, or, to be more precise, piecewise linear with a slope not necessarily of zero between any two adjacent follow-up times if there are no time-dependent covariates or only indicator-type time-dependent covariates when present, with jumps at event times only; cf. (5) and (12). Thus the Lin-Ying Markov model, in either of its two forms, usually gives a much smoother fit to the data than do the Cox Markov models and the Aalen Markov model. This explains why the probability estimates based on the two Lin-Ying Markov models tend to have smaller standard errors, as found in § 4.

The estimated survival curves from the three additive hazards Markov regression models shown in Fig. 2(a) are not necessarily monotonically decreasing, as it is obviously the case with the curve from the Aalen Markov model. This is a consequence of a drawback of the additive hazards models: in neither Aalen's (1989) model nor Lin & Ying's (1994) model is the hazard rate constrained to be positive. Nevertheless, in practice, this is seldom a serious disadvantage, since the monotonicity property is usually only slightly violated in a small neighbourhood of some time points. A simple remedy for this problem is to do monotone smoothing using the so-called 'pool-adjacent-violators algorithm' (Barlow et al., 1972, pp. 13–5). In our bone marrow transplant example, we could have easily done that, but we leave it as it is, to show that the lack of monotonicity is minimal. Indeed, an estimated survival curve with substantially increasing sections would probably indicate a lack of fit, since in theory the estimator is uniformly consistent for the true survival function.

McKeague & Sasieni (1994) proposed a more versatile additive hazards model in which the influence of only a few covariates varies nonparametrically over time, and that of the remaining covariates is unchanging. In a finite-state Markov process, if we assume each transition intensity follows a McKeague & Sasieni (1994) model, then we add to our multistate survival models family another new member: the 'McKeague-Sasieni Markov model'. It is emphasized that our unifying approach to Markov regression models using the basic tools of the product integral and the functional delta-method is easily extended to this model.

Finally, note that the regression modelling assumptions made in these models are testable assumptions. For example, Shen & Cheng (1999) provided graphical methods for assessing the appropriateness of Lin & Ying's (1994) regression model. A number of techniques are also available for checking the fit of the Cox model or the fit of the Aalen (1989) additive model (Klein & Moeschberger, 1997). Klein & Moeschberger (1997) showed how one can distinguish between the Andersen-Gill and Klein-Keiding models in the proportional hazards framework. Y. Shu's dissertation shows how their methods can be adapted to the additive hazards framework.

Acknowledgement

This research was supported by Grant R01-CA54706-09 from the National Cancer Institute. The authors thank the editor and two anonymous referees for their insightful comments which led to a significantly improved version of the manuscript. The authors are also indebted to Dr. Per Kragh Andersen for his helpful suggestions about the proofs of the asymptotics, and to Dr. Meijie Zhang for a general discussion on martingales.

Appendix 1

The Aalen Markov model

To study the asymptotics of the Aalen Markov model, we assume regularity conditions similar to those given by Example VII.4.4 of Andersen et al. (1993), with slight modifications to suit our multistate modelling setting. In particular, we assume that

$$E\left\{\sup_{u\in[s,v]} Y_{h1}(u)|Z_{1w}^3|\right\} < \infty, \quad h = 1,\dots,k,\ w = 1,\dots,p,$$

and that the covariates are linearly independent.

Recall from § 3.1 that $Y_{hi0}(u) = Y_{hi}(u)$ and $Y_{hiw}(u) = Y_{hi}(u)Z_{iw}$ for $w = 1,\dots,p$. Then, by the same arguments as in Example VII.4.4 of Andersen et al. (1993), the following conditions similar to those for Theorem VII.4.1 of Andersen et al. (1993) hold true.

(a) For $h = 1,\dots,k$ and $w, w', w'' = 0,1,\dots,p$, there exist continuous functions $R_{hw}^{(1)}$, $R_{hww'}^{(2)}$ and $R_{hww'w''}^{(3)}$ defined on $[s,v]$ such that as $n \to \infty$

$$\sup_{u\in[s,v]}\left|\frac{1}{n}\sum_{i=1}^n Y_{hiw}(u) - R_{hw}^{(1)}(u)\right| \to 0 \ \text{ in probability},$$

$$\sup_{u\in[s,v]}\left|\frac{1}{n}\sum_{i=1}^n Y_{hiw}(u)Y_{hiw'}(u) - R_{hww'}^{(2)}(u)\right| \to 0 \ \text{ in probability},$$

$$\sup_{u\in[s,v]}\left|\frac{1}{n}\sum_{i=1}^n Y_{hiw}(u)Y_{hiw'}(u)Y_{hiw''}(u) - R_{hww'w''}^{(3)}(u)\right| \to 0 \ \text{ in probability}.$$

(b) For $h = 1,\dots,k$ and $w = 0,1,\dots,p$,

$$n^{-1/2}\sup_{\substack{i = 1,\dots,n\\ u\in[s,v]}} |Y_{hiw}(u)| \to 0 \ \text{ in probability}, \quad \text{as } n \to \infty.$$

(c) For each $h = 1, \ldots, k$, the matrix $R_h^{(2)}(u) = \left\{R_{hww'}^{(2)}(u);\ w, w' = 0, 1, \ldots, p\right\}$ is nonsingular for all $u \in [s, v]$.

Before moving to the proof of Theorem 1, we first review in a lemma some key asymptotic results of Aalen's (1989) additive hazards model for classical two-state survival analysis.

Lemma A1. *Let $g, l \in \{1, \ldots, k\}$ and $g \neq l$. Define $M_{gl}(u) = (M_{gl1}(u), \ldots, M_{gln}(u))^\top$. Then, under conditions* (a)–(c) *above, we have the following:*

(i) $n^{1/2}\int_s^t d\{\hat{B}_{gl}(u) - B_{gl}(u)\} \stackrel{a}{=} n^{-1/2}\int_s^t J_g(u)\{R_g^{(2)}(u)\}^{-1} Y_g^\top(u) dM_{gl}(u)$, for $t \in [s, v]$;

(ii) $n^{1/2}\int_s^\cdot d\{\hat{B}_{gl}(u) - B_{gl}(u)\}$ converges weakly on $[s, v]$ to a zero-mean $(p+1)$-variate Gaussian martingale, the variance-covariance matrix of which can be estimated uniformly consistently by

$$n\int_s^\cdot J_g(u)\left\{Y_g^\top(u)\, Y_g(u)\right\}^{-1} Y_g^\top(u)\,\mathrm{diag}\{dN_{gl}(u)\}\, Y_g(u)\left\{Y_g^\top(u)\, Y_g(u)\right\}^{-1};$$

(iii) $n^{1/2}\int_s^\cdot d\{\hat{A}_{gl}(u; Z_0) - A_{gl}(u; Z_0)\}$ converges weakly on $[s, v]$ to a zero-mean Gaussian process, $U_{gl}(\cdot; Z_0)$ say, the variance function of which can be estimated uniformly consistently by

$$n\int_s^\cdot Z_0^\top J_g(u)\left\{Y_g^\top(u)\, Y_g(u)\right\}^{-1} Y_g^\top(u)\,\mathrm{diag}\{dN_{gl}(u)\}\, Y_g(u)\left\{Y_g^\top(u)\, Y_g(u)\right\}^{-1} Z_0.$$

Proof. Parts (i) and (ii) follow straightforwardly from the proof of Theorem VII.4.1 in Andersen et al. (1993, pp. 576–7), whose proof is based on that of Huffer and McKeague (1991). Part (iii) is trivial in view of Part (ii) and equation (2). □

Proof of Theorem 1. (i) Let $U(\cdot; Z_0) = \{U_{gl}(\cdot; Z_0);\ g, l = 1, \ldots, k\}$ be a $k \times k$ matrix-valued process, where, for $g \neq l$, the $U_{gl}(\cdot; Z_0)$ are independent zero-mean Gaussian processes as given by Lemma A1(iii), and $U_{gg}(\cdot; Z_0) = -\sum_{l \neq g} U_{gl}(\cdot; Z_0)$. Then $n^{1/2}\int_s^\cdot d\{\hat{A}(u; Z_0) -$

$A(u;Z_0)\}$ converges weakly on $[s,v]$ to $U(\cdot;Z_0)$. It follows from Lemma 1 that $n^{1/2}\{\hat{P}(s,\cdot;Z_0)-P(s,\cdot;Z_0)\}$ converges weakly on $[s,v]$ to the zero-mean Gaussian process

$$\int_s^{\cdot} P(s,u-;Z_0)\,dU(u;Z_0)P(u,\cdot;Z_0).$$

(ii) From Lemma 1, we also have

$$n^{1/2}(\hat{P}(s,t;Z_0)-P(s,t;Z_0)) \stackrel{a}{=} \int_s^t P(s,u-;Z_0)\,d\Big\{n^{1/2}(\hat{A}-A)(u;Z_0)\Big\}P(u,t;Z_0).$$

Multiplying both sides on the right by the inverse of $P(s,t;Z_0)=P(s,u;Z_0)P(u,t;Z_0)$, we obtain

$$n^{1/2}\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1}-I\right\} \stackrel{a}{=} \int_s^t P(s,u-;Z_0)\,d\Big\{n^{1/2}(\hat{A}-A)(u;Z_0)\Big\}P(s,u;Z_0)^{-1}. \tag{A1}$$

Let $P(s,u;Z_0)^{-1}=\Big\{P^{hj}(s,u;Z_0);\, h,j=1,\ldots,k\Big\}$ and let

$$\tilde{F}_{gl}^{(hj)}(u;s,Z_0)=P_{hg}(s,u-;Z_0)\{P^{lj}(s,u;Z_0)-P^{gj}(s,u;Z_0)\}.$$

Denote the (h,j)th entry of $n^{1/2}\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1}-I\right\}$ by $\xi_{hj}^{(n)}(s,t;Z_0)$. Then, from (A1), by applying Lemma A1(i), we have

$$\begin{aligned}
\xi_{hj}^{(n)}(s,t;Z_0) &\stackrel{a}{=} \sum_{g=1}^{k}\sum_{l=1}^{k}\int_s^t P_{hg}(s,u-;Z_0)d\left\{n^{1/2}(\hat{A}_{gl}-A_{gl})(u;Z_0)\right\}P^{lj}(s,u;Z_0)\\
&= \sum_{g=1}^{k}\sum_{l\neq g}\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)d\left\{n^{1/2}(\hat{A}_{gl}-A_{gl})(u;Z_0)\right\}\\
&= \sum_{g=1}^{k}\sum_{l\neq g}\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)d\left[n^{1/2}Z_0^\top\{\hat{B}_{gl}(u)-B_{gl}(u)\}\right]\\
&\stackrel{a}{=} \sum_{g=1}^{k}\sum_{l\neq g}\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)Z_0^\top n^{-1/2}J_g(u)\left\{R_g^{(2)}(u)\right\}^{-1}Y_g^\top(u)dM_{gl}(u).
\end{aligned} \tag{A2}$$

Note that the expression on the right-hand side of (A2) is a martingale integral, where the $M_{gl}(\cdot)$ are orthogonal n-variate martingales with optional variation processes given by $[M_{gl}](\cdot)=\text{diag}\{N_{gl}(\cdot)\}$. Therefore,

$$\text{cov}\left\{\xi_{hj}^{(n)}(s,t;Z_0),\,\xi_{mr}^{(n)}(s,t;Z_0)\right\}$$

$$
\begin{aligned}
&\stackrel{a}{=} \sum_{g=1}^{k}\sum_{l\neq g} \operatorname{cov}\Big[\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)Z_0^\top n^{-1/2} J_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Y_g^\top(u) dM_{gl}(u), \\
&\qquad\qquad \int_s^t \tilde{F}_{gl}^{(mr)}(u;s,Z_0)Z_0^\top n^{-1/2} J_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Y_g^\top(u) dM_{gl}(u)\Big] \\
&= \sum_{g=1}^{k}\sum_{l\neq g} E\Big(\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)\tilde{F}_{gl}^{(mr)}(u;s,Z_0) \\
&\qquad\qquad \times Z_0^\top \frac{1}{n} J_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Y_g^\top(u)\, d[M_{gl}](u)\, Y_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Z_0\Big) \\
&= \sum_{g=1}^{k}\sum_{l\neq g} E\Big[\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)\tilde{F}_{gl}^{(mr)}(u;s,Z_0) \\
&\qquad\qquad \times Z_0^\top \frac{1}{n} J_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Y_g^\top(u)\operatorname{diag}\left\{dN_{gl}(u)\right\} Y_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Z_0\Big]. \quad \text{(A3)}
\end{aligned}
$$

For $g \neq l$, let C_{gl} denote the $k \times k$ matrix with element (g,l) equal to 1, element (g,g) equal to -1, and the rest equal to zero. Then $\tilde{F}_{gl}^{(hj)}(u;s,Z_0)$ is the (h,j)th entry of $P(s,u-;Z_0)C_{gl}P(s,u;Z_0)^{-1}$. It follows from (A3) that

$$
\begin{aligned}
&\operatorname{cov}\left[n^{1/2}\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1} - I\right\}\right] \\
&\stackrel{a}{=} \sum_{g=1}^{k}\sum_{l\neq g} E\Big[\int_s^t \operatorname{vec}\left\{P(s,u-;Z_0)C_{gl}P(s,u;Z_0)^{-1}\right\} \\
&\qquad\qquad \times Z_0^\top \frac{1}{n} J_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Y_g^\top(u)\operatorname{diag}\left\{dN_{gl}(u)\right\} Y_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Z_0 \\
&\qquad\qquad \times \operatorname{vec}^\top\left\{P(s,u-;Z_0)C_{gl}P(s,u;Z_0)^{-1}\right\}\Big]. \quad \text{(A4)}
\end{aligned}
$$

Here, in line with Andersen et al. (1993, § IV.4.1.3), $\operatorname{vec}\{\Psi\}$ for a $k \times k$ matrix Ψ stacks the columns of Ψ on the top of each other into a $k^2 \times 1$ vector, and we define the covariance matrix of a $k \times k$ matrix-valued random variable W as the ordinary covariance matrix of $\operatorname{vec}\{W\}$, i.e., as the $k^2 \times k^2$ matrix

$$
\operatorname{cov}(W) = E\left([\operatorname{vec}\{W\} - \operatorname{vec}\{E(W)\}]\,[\operatorname{vec}\{W\} - \operatorname{vec}\{E(W)\}]^\top\right).
$$

Using equation (4.4.11) of Andersen et al. (1993, p. 291), (A4), and elementary properties of the vec-operator and Kronecker products of matrices, we find that

$$
\operatorname{cov}\left\{\hat{P}(s,t;Z_0)\right\}
$$

$$
\begin{aligned}
&= \operatorname{cov}\left[I\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1}\right\}P(s,t;Z_0)\right] \\
&= \left\{P(s,t;Z_0)^\top \otimes I\right\} \operatorname{cov}\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1}\right\}\left\{P(s,t;Z_0)\otimes I\right\} \\
&= \left\{P(s,t;Z_0)^\top \otimes I\right\} \frac{1}{n}\operatorname{cov}\left[n^{1/2}\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1} - I\right\}\right]\left\{P(s,t;Z_0)\otimes I\right\} \qquad \text{(A5)} \\
&\stackrel{a}{=} \sum_{g=1}^{k}\sum_{l\neq g} E\Big[\int_s^t \operatorname{vec}\left\{P(s,u-;Z_0)C_{gl}P(u,t;Z_0)\right\} \\
&\qquad\qquad \times Z_0^\top \frac{1}{n^2} J_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Y_g^\top(u)\operatorname{diag}\left\{dN_{gl}(u)\right\} Y_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Z_0 \\
&\qquad\qquad \times \operatorname{vec}^\top\left\{P(s,u-;Z_0)C_{gl}P(u,t;Z_0)\right\}\Big]. \qquad \text{(A6)}
\end{aligned}
$$

Now let $F_{gl}^{(hj)}(u;s,t,Z_0) = P_{hg}(s,u-;Z_0)\{P_{lj}(u,t;Z_0) - P_{gj}(u,t;Z_0)\}$. Then the (h,j)th entry of $P(s,u-;Z_0)C_{gl}P(u,t;Z_0)$ is $F_{gl}^{(hj)}(u;s,t,Z_0)$. It follows from (A6) that

$$
\begin{aligned}
&\operatorname{cov}\left\{\hat{P}_{hj}(s,t;Z_0),\ \hat{P}_{mr}(s,t;Z_0)\right\} \\
&\stackrel{a}{=} \sum_{g=1}^{k}\sum_{l\neq g} E\Big[\int_s^t F_{gl}^{(hj)}(u;s,t,Z_0)F_{gl}^{(mr)}(u;s,t,Z_0) \\
&\qquad\qquad \times Z_0^\top \frac{1}{n^2} J_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Y_g^\top(u)\operatorname{diag}\left\{dN_{gl}(u)\right\} Y_g(u)\left\{R_g^{(2)}(u)\right\}^{-1} Z_0\Big],
\end{aligned}
$$

which can be estimated uniformly consistently by (4). □

Appendix 2

The Andersen-Gill Lin-Ying Markov model

We assume that regularity conditions similar to those given by Kulich & Lin (2000, Appendix 1) hold. In particular, $\operatorname{pr}\{Y_{h1}(u) = 1 \text{ for all } u \in [s,v]\} > 0$,

$$
E\left\{\sup_{u\in[s,v]}\left|Y_{h1}(u)Z_1^{\otimes 2}(\beta_{hj}^\top Z_1)^2\right|\right\} < \infty,
$$

and $V_{hj} := E\left[\int_0^\infty \{Z_1 - \bar{Z}_h(u)\}^{\otimes 2}\, dN_{hj1}(u)\right]$ is positive definite, for $h,j = 1,\ldots,k,\ h \neq j$. Then by applying functional forms of the strong law of large numbers (Andersen & Gill, 1982, Appendix III), we have

$$
\frac{1}{n}Y_h(u) \to \pi_h^{(0)}(u), \quad \frac{1}{n}\sum_{i=1}^{n} Y_{hi}(u)Z_i \to \pi_h^{(1)}(u), \quad \bar{Z}_h(u) \to e_h(u),
$$

all uniformly in $u \in [s, v]$ in probability, where

$$\pi_h^{(0)}(u) = E\{Y_{h1}(u)\}, \quad \pi_h^{(1)}(u) = E\{Y_{h1}(u) Z_1\}, \quad e_h(u) = \pi_h^{(1)}(u)/\pi_h^{(0)}(u);$$

see Kulich & Lin (2000, Appendix 1) for a similar argument.

To prove Theorem 2, we need the following lemma.

LEMMA A2. *Let $g, l \in \{1, \ldots, k\}$ and $g \neq l$. Define*

$$\Omega_g = E\left[\int_0^\infty Y_{g1}(u)\{Z_1 - \bar{Z}_g(u)\}^{\otimes 2}\, du\right], \quad M_{gl}(u) = \sum_{i=1}^n M_{gli}(u).$$

Then, under the regularity conditions stated above, we have the following:

(i) $n^{1/2}(\hat{\beta}_{gl} - \beta_{gl}) \overset{a}{=} \Omega_g^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\infty \{Z_i - \bar{Z}_g(u)\}\, dM_{gli}(u)$;

(ii)
$$n^{1/2} \int_s^t d\left\{\hat{A}_{gl0}(u, \hat{\beta}_{gl}) - A_{gl0}(u)\right\} \overset{a}{=} -\int_s^t e_g^\top(u)\, du\, n^{1/2}(\hat{\beta}_{gl} - \beta_{gl}) + n^{1/2} \int_s^t \frac{dM_{gl}(u)}{Y_g(u)}, \quad \text{for } t \in [s, v];$$

(iii)
$$n^{1/2} \int_s^t d\left\{\hat{A}_{gl}(u; Z_0) - A_{gl}(u; Z_0)\right\} \overset{a}{=} \int_s^t \{Z_0 - e_g(u)\}^\top du\, n^{1/2}(\hat{\beta}_{gl} - \beta_{gl}) + n^{1/2} \int_s^t \frac{dM_{gl}(u)}{Y_g(u)}, \quad \text{for } t \in [s, v];$$

(iv) $n^{1/2} \int_s^\cdot d\{\hat{A}_{gl}(u; Z_0) - A_{gl}(u; Z_0)\}$ *converges weakly on $[s, v]$ to a zero-mean Gaussian process, $U_{gl}(\cdot; Z_0)$ say, the variance function of which can be estimated uniformly consistently by*

$$\int_s^\cdot \frac{n\, dN_{gl}(u)}{\{Y_g(u)\}^2} + \int_s^\cdot \{Z_0 - \bar{Z}_g(u)\}^\top du\, \hat{\Omega}_g^{-1} \hat{V}_{gl} \hat{\Omega}_g^{-1} \int_s^\cdot \{Z_0 - \bar{Z}_g(u)\}\, du + 2\int_s^\cdot \{Z_0 - \bar{Z}_g(u)\}^\top du\, \hat{\Omega}_g^{-1} \sum_{i=1}^n \int_s^\cdot \frac{\{Z_i - \bar{Z}_g(u)\}\, dN_{gli}(u)}{Y_g(u)}.$$

Proof. These are the basic asymptotic results that were either implicit or embedded in Lin & Ying (1994). Full derivations of Parts (i) and (ii) are straightforward along the lines of Kulich & Lin (2000, Appendix 2). Part (iii) is a direct consequence of Part (ii). Part (iv) follows immediately from Part (iii) and Rebolledo's (1980) martingale central limit theorem. □

Proof of Theorem 2. (i) Let $U(\cdot; Z_0) = \{U_{gl}(\cdot; Z_0);\, g, l = 1, \ldots, k\}$ be a $k \times k$ matrix-valued process, where, for $g \neq l$, the $U_{gl}(\cdot; Z_0)$ are independent zero-mean Gaussian processes as given by Lemma A2(iv), and $U_{gg}(\cdot; Z_0) = -\sum_{l \neq g} U_{gl}(\cdot; Z_0)$. Then $n^{1/2} \int_s^{\cdot} d\{\hat{A}(u; Z_0) - A(u; Z_0)\}$ converges weakly on $[s, v]$ to $U(\cdot; Z_0)$. It follows from Lemma 1 that $n^{1/2}\{\hat{P}(s, \cdot; Z_0) - P(s, \cdot; Z_0)\}$ converges weakly on $[s, v]$ to the zero-mean Gaussian process

$$\int_s^{\cdot} P(s, u-; Z_0)\, dU(u; Z_0) P(u, \cdot; Z_0).$$

(ii) As in the beginning of the proof of Theorem 1(ii) in Appendix 1, from Lemma 1, we may also derive (A1) in the current context.

Let $P(s, u; Z_0)^{-1} = \left\{P^{hj}(s, u; Z_0);\, h, j = 1, \ldots, k\right\}$ and let

$$\tilde{F}_{gl}^{(hj)}(u; s, Z_0) = P_{hg}(s, u-; Z_0)\{P^{lj}(s, u; Z_0) - P^{gj}(s, u; Z_0)\},$$

$$\tilde{Q}_{gl}^{(hj)}(s, t; Z_0) = \int_s^t \tilde{F}_{gl}^{(hj)}(u; s, Z_0)\{Z_0 - e_g(u)\}\, du. \tag{A7}$$

Denote the (h, j)th entry of $n^{1/2}\left\{\hat{P}(s, t; Z_0) P(s, t; Z_0)^{-1} - I\right\}$ by $\xi_{hj}^{(n)}(s, t; Z_0)$. Then, from (A1), by applying Lemma A2 (i) and (iii), we have

$$\begin{aligned}
\xi_{hj}^{(n)}(s, t; Z_0) &\overset{a}{=} \sum_{g=1}^{k} \sum_{l=1}^{k} \int_s^t P_{hg}(s, u-; Z_0) d\left\{n^{1/2}(\hat{A}_{gl} - A_{gl})(u; Z_0)\right\} P^{lj}(s, u; Z_0) \\
&= \sum_{g=1}^{k} \sum_{l \neq g} \int_s^t \tilde{F}_{gl}^{(hj)}(u; s, Z_0) d\left\{n^{1/2}(\hat{A}_{gl} - A_{gl})(u; Z_0)\right\} \\
&\overset{a}{=} \sum_{g=1}^{k} \sum_{l \neq g} \{\tilde{Q}_{gl}^{(hj)}(s, t; Z_0)\}^{\top} n^{1/2}(\hat{\beta}_{gl} - \beta_{gl}) \\
&\quad + \sum_{g=1}^{k} \sum_{l \neq g} \int_s^t \tilde{F}_{gl}^{(hj)}(u; s, Z_0) n^{1/2} \frac{dM_{gl}(u)}{Y_g(u)} \\
&\overset{a}{=} \sum_{g=1}^{k} \sum_{l \neq g} \sum_{i=1}^{n} \{\tilde{Q}_{gl}^{(hj)}(s, t; Z_0)\}^{\top} \Omega_g^{-1} n^{-1/2} \int_0^{\infty} \{Z_i - \bar{Z}_g(u)\}\, dM_{gli}(u) \\
&\quad + \sum_{g=1}^{k} \sum_{l \neq g} \sum_{i=1}^{n} \int_s^t \tilde{F}_{gl}^{(hj)}(u; s, Z_0) n^{1/2} \frac{dM_{gli}(u)}{Y_g(u)}.
\end{aligned} \tag{A8}$$

Notice that in (A8), the $M_{gli}(\cdot)$ are orthogonal martingales with optional variation processes given by $[M_{gli}](\cdot) = N_{gli}(\cdot)$. Therefore,

$$
\begin{aligned}
&\operatorname{cov}\left\{\xi_{hj}^{(n)}(s,t;Z_0),\ \xi_{mr}^{(n)}(s,t;Z_0)\right\} \\
&\overset{a}{=} \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\tilde{Q}_{gl}^{(hj)}(s,t;Z_0)\}^{\top}\Omega_g^{-1}\frac{1}{n}\operatorname{cov}\left[\int_0^{\infty}\{Z_i-\bar{Z}_g(u)\}\,dM_{gli}(u)\right]\Omega_g^{-1}\tilde{Q}_{gl}^{(mr)}(s,t;Z_0) \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\operatorname{cov}\left\{\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)n^{1/2}\,\frac{dM_{gli}(u)}{Y_g(u)},\ \int_s^t \tilde{F}_{gl}^{(mr)}(u;s,Z_0)n^{1/2}\,\frac{dM_{gli}(u)}{Y_g(u)}\right\} \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\tilde{Q}_{gl}^{(hj)}(s,t;Z_0)\}^{\top}\Omega_g^{-1} \\
&\qquad\qquad \times \operatorname{cov}\left[\int_0^{\infty}\{Z_i-\bar{Z}_g(u)\}\,dM_{gli}(u),\ \int_s^t \tilde{F}_{gl}^{(mr)}(u;s,Z_0)\frac{dM_{gli}(u)}{Y_g(u)}\right] \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\tilde{Q}_{gl}^{(mr)}(s,t;Z_0)\}^{\top}\Omega_g^{-1} \\
&\qquad\qquad \times \operatorname{cov}\left[\int_0^{\infty}\{Z_i-\bar{Z}_g(u)\}\,dM_{gli}(u),\ \int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)\frac{dM_{gli}(u)}{Y_g(u)}\right] \\
&= \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\tilde{Q}_{gl}^{(hj)}(s,t;Z_0)\}^{\top}\Omega_g^{-1}\frac{1}{n}E\left(\int_0^{\infty}\{Z_i-\bar{Z}_g(u)\}^{\otimes 2}d[M_{gli}](u)\right)\Omega_g^{-1}\tilde{Q}_{gl}^{(mr)}(s,t;Z_0) \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}E\left(\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)\tilde{F}_{gl}^{(mr)}(u;s,Z_0)\frac{n\,d[M_{gli}](u)}{\{Y_g(u)\}^2}\right) \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\tilde{Q}_{gl}^{(hj)}(s,t;Z_0)\}^{\top}\Omega_g^{-1}E\left\{\int_s^t \frac{Z_i-\bar{Z}_g(u)}{Y_g(u)}\tilde{F}_{gl}^{(mr)}(u;s,Z_0)d[M_{gli}](u)\right\} \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\tilde{Q}_{gl}^{(mr)}(s,t;Z_0)\}^{\top}\Omega_g^{-1}E\left\{\int_s^t \frac{Z_i-\bar{Z}_g(u)}{Y_g(u)}\tilde{F}_{gl}^{(hj)}(u;s,Z_0)d[M_{gli}](u)\right\} \\
&= \sum_{g=1}^{k}\sum_{l\neq g}\{\tilde{Q}_{gl}^{(hj)}(s,t;Z_0)\}^{\top}\Omega_g^{-1}V_{gl}\Omega_g^{-1}\tilde{Q}_{gl}^{(mr)}(s,t;Z_0) \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}E\left[\int_s^t \tilde{F}_{gl}^{(hj)}(u;s,Z_0)\tilde{F}_{gl}^{(mr)}(u;s,Z_0)\frac{n\,dN_{gl}(u)}{\{Y_g(u)\}^2}\right] \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\tilde{Q}_{gl}^{(hj)}(s,t;Z_0)\}^{\top}\Omega_g^{-1}E\left\{\int_s^t \frac{Z_i-\bar{Z}_g(u)}{Y_g(u)}\tilde{F}_{gl}^{(mr)}(u;s,Z_0)dN_{gli}(u)\right\} \\
&\quad + \sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{\tilde{Q}_{gl}^{(mr)}(s,t;Z_0)\}^{\top}\Omega_g^{-1}E\left\{\int_s^t \frac{Z_i-\bar{Z}_g(u)}{Y_g(u)}\tilde{F}_{gl}^{(hj)}(u;s,Z_0)dN_{gli}(u)\right\}. \qquad \text{(A9)}
\end{aligned}
$$

Let C_{gl} be defined as in Appendix 1. Then $\tilde{F}_{gl}^{(hj)}(u;s,Z_0)$ is the (h,j)th entry of $P(s,u-;Z_0)C_{gl}P(s,u;Z_0)^{-1}$. It follows from (A7) and (A9) that

$$\operatorname{cov}\left[n^{1/2}\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1}-I\right\}\right]$$

$$
\begin{aligned}
&\overset{a}{=} \sum_{g=1}^{k}\sum_{l\neq g}\int_s^t \mathrm{vec}\{P(s,u-;Z_0)\,C_{gl}P(s,u;Z_0)^{-1}\}\{Z_0-e_g(u)\}^\top du\,\Omega_g^{-1}\,V_{gl}\Omega_g^{-1}\\
&\qquad\times\int_s^t\{Z_0-e_g(u)\}\,\mathrm{vec}^\top\{P(s,u-;Z_0)\,C_{gl}P(s,u;Z_0)^{-1}\}\,du\\
&\quad+\sum_{g=1}^{k}\sum_{l\neq g}E\biggl[\int_s^t \mathrm{vec}\{P(s,u-;Z_0)\,C_{gl}P(s,u;Z_0)^{-1}\}\\
&\qquad\qquad\times\mathrm{vec}^\top\{P(s,u-;Z_0)\,C_{gl}P(s,u;Z_0)^{-1}\}\,\frac{n\,dN_{gl}(u)}{\{Y_g(u)\}^2}\biggr]\\
&\quad+2\sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\int_s^t \mathrm{vec}\{P(s,u-;Z_0)\,C_{gl}P(s,u;Z_0)^{-1}\}\{Z_0-e_g(u)\}^\top du\,\Omega_g^{-1}\\
&\qquad\qquad\times E\biggl[\int_s^t\frac{Z_i-\bar Z_g(u)}{Y_g(u)}\,\mathrm{vec}^\top\{P(s,u-;Z_0)\,C_{gl}P(s,u;Z_0)^{-1}\}\,dN_{gli}(u)\biggr].
\end{aligned}
$$

Using this result, (A5), and elementary properties of the vec-operator and Kronecker products of matrices, we find that

$$
\begin{aligned}
&\mathrm{cov}\left\{\hat P(s,t;Z_0)\right\}\\
&\overset{a}{=} \sum_{g=1}^{k}\sum_{l\neq g}\int_s^t \mathrm{vec}\{P(s,u-;Z_0)\,C_{gl}P(u,t;Z_0)\}\{Z_0-e_g(u)\}^\top du\,\frac{1}{n}\Omega_g^{-1}\,V_{gl}\Omega_g^{-1}\\
&\qquad\times\int_s^t\{Z_0-e_g(u)\}\,\mathrm{vec}^\top\{P(s,u-;Z_0)\,C_{gl}P(u,t;Z_0)\}\,du\\
&\quad+\sum_{g=1}^{k}\sum_{l\neq g}E\biggl[\int_s^t \mathrm{vec}\{P(s,u-;Z_0)\,C_{gl}P(u,t;Z_0)\}\\
&\qquad\qquad\times\mathrm{vec}^\top\{P(s,u-;Z_0)\,C_{gl}P(u,t;Z_0)\}\,\frac{dN_{gl}(u)}{\{Y_g(u)\}^2}\biggr]\\
&\quad+\frac{2}{n}\sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\int_s^t \mathrm{vec}\{P(s,u-;Z_0)\,C_{gl}P(u,t;Z_0)\}\{Z_0-e_g(u)\}^\top du\,\Omega_g^{-1}\\
&\qquad\qquad\times E\biggl[\int_s^t\frac{Z_i-\bar Z_g(u)}{Y_g(u)}\,\mathrm{vec}^\top\{P(s,u-;Z_0)\,C_{gl}P(u,t;Z_0)\}\,dN_{gli}(u)\biggr]. \qquad \text{(A10)}
\end{aligned}
$$

Now let

$$
F_{gl}^{(hj)}(u;s,t,Z_0)=P_{hg}(s,u-;Z_0)\{P_{lj}(u,t;Z_0)-P_{gj}(u,t;Z_0)\},
$$

$$
Q_{gl}^{(hj)}(s,t;Z_0)=\int_s^t F_{gl}^{(hj)}(u;s,t,Z_0)\{Z_0-e_g(u)\}\,du.
$$

Then the (h,j)th entry of $P(s,u-;Z_0)\,C_{gl}P(u,t;Z_0)$ is $F_{gl}^{(hj)}(u;s,t,Z_0)$. It follows from (A10) that

$$
\begin{aligned}
&\operatorname{cov}\left\{\hat{P}_{hj}(s,t;Z_0),\ \hat{P}_{mr}(s,t;Z_0)\right\}\\
\stackrel{a}{=}&\sum_{g=1}^{k}\sum_{l\neq g}\{Q_{gl}^{(hj)}(s,t;Z_0)\}^\top\frac{1}{n}\Omega_g^{-1}V_{gl}\Omega_g^{-1}\,Q_{gl}^{(mr)}(s,t;Z_0)\\
&+\sum_{g=1}^{k}\sum_{l\neq g}E\left[\int_s^t F_{gl}^{(hj)}(u;s,t,Z_0)F_{gl}^{(mr)}(u;s,t,Z_0)\frac{dN_{gl}(u)}{\{Y_g(u)\}^2}\right]\\
&+\frac{1}{n}\sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{Q_{gl}^{(hj)}(s,t;Z_0)\}^\top\Omega_g^{-1}E\left\{\int_s^t\frac{Z_i-\bar{Z}_g(u)}{Y_g(u)}F_{gl}^{(mr)}(u;s,t,Z_0)\,dN_{gli}(u)\right\}\\
&+\frac{1}{n}\sum_{g=1}^{k}\sum_{l\neq g}\sum_{i=1}^{n}\{Q_{gl}^{(mr)}(s,t;Z_0)\}^\top\Omega_g^{-1}E\left\{\int_s^t\frac{Z_i-\bar{Z}_g(u)}{Y_g(u)}F_{gl}^{(hj)}(u;s,t,Z_0)\,dN_{gli}(u)\right\},
\end{aligned}
$$

which can be estimated uniformly consistently by (6). □

Appendix 3

The Klein-Keiding Lin-Ying Markov model

We assume that regularity conditions like those given at the beginning of Appendix 2 hold. In particular, $\mathrm{pr}\{Y_{X1}(u)=1 \text{ for all } u\in[s,v]\}>0$,

$$
E\left(\sup_{u\in[s,v]}\left|Y_{X1}(u)Z_{X1}(u)^{\otimes 2}\{\beta_X^\top Z_{X1}(u)\}^2\right|\right)<\infty,
$$

and $V_X := E\left[\int_0^\infty\{Z_{X1}(u)-\bar{Z}_X(u)\}^{\otimes 2}\,dN_{X1}(u)\right]$ is positive definite, for $X\in\mathcal{E}$. Then by the same arguments as in Appendix 2, we have

$$
\frac{1}{n}Y_X(u)\to\pi_X^{(0)}(u),\quad \frac{1}{n}\sum_{i=1}^{n}Y_{Xi}(u)Z_{Xi}(u)\to\pi_X^{(1)}(u),\quad \bar{Z}_X(u)\to e_X(u),
$$

all uniformly in $u\in[s,v]$ in probability, where

$$
\pi_X^{(0)}(u)=E\{Y_{X1}(u)\},\quad \pi_X^{(1)}(u)=E\{Y_{X1}(u)Z_{X1}(u)\},\quad e_X(u)=\pi_X^{(1)}(u)/\pi_X^{(0)}(u).
$$

To establish Theorem 3, we start with the following lemma.

Lemma A3. *Let $X \in \mathcal{E}$ and let $gl \in \mathcal{J}(X)$. Define*

$$\Omega_X = E\left[\int_0^\infty Y_{X1}(u)\{Z_{X1}(u) - \bar{Z}_X(u)\}^{\otimes 2}\, du\right], \quad M_X(t) = \sum_{i=1}^n M_{Xi}(t).$$

Then, under the regularity conditions stated above, we have the following:

(i) $n^{1/2}(\hat{\beta}_X - \beta_X) \overset{a}{=} \Omega_X^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\infty \{Z_{Xi}(u) - \bar{Z}_X(u)\}\, dM_{Xi}(u)$;

(ii) $$n^{1/2}\int_s^t d\left\{\hat{A}_{X0}(u,\hat{\beta}_X) - A_{X0}(u)\right\} \overset{a}{=} -\int_s^t e_X^\top(u)\, du\, n^{1/2}(\hat{\beta}_X - \beta_X) + n^{1/2}\int_s^t \frac{dM_X(u)}{Y_X(u)}, \quad \text{for } t \in [s,v];$$

(iii) $$n^{1/2}\int_s^t d\left\{\hat{A}_{gl}(u;Z_0) - A_{gl}(u;Z_0)\right\} \overset{a}{=} \int_s^t \{Z_{gl0} - e_X(u)\}^\top du\, n^{1/2}(\hat{\beta}_X - \beta_X) + n^{1/2}\int_s^t \frac{dM_X(u)}{Y_X(u)}, \quad \text{for } t \in [s,v];$$

(iv) $n^{1/2}\int_s^\cdot d\{\hat{A}_{gl}(u;Z_0) - A_{gl}(u;Z_0)\}$ *converges weakly on* $[s,v]$ *to a zero-mean Gaussian process,* $U_{gl}(\cdot;Z_0)$ *say, the variance function of which can be estimated uniformly consistently by*

$$\int_s^\cdot \frac{n\, dN_X(u)}{\{Y_X(u)\}^2} + \int_s^\cdot \{Z_{gl0} - \bar{Z}_X(u)\}^\top du\, \hat{\Omega}_X^{-1}\hat{V}_X\hat{\Omega}_X^{-1}\int_s^\cdot \{Z_{gl0} - \bar{Z}_X(u)\}\, du$$
$$+ 2\int_s^\cdot \{Z_{gl0} - \bar{Z}_X(u)\}^\top du\, \hat{\Omega}_X^{-1}\sum_{i=1}^n \int_s^\cdot \frac{\{Z_{Xi}(u) - \bar{Z}_X(u)\}\, dN_{Xi}(u)}{Y_X(u)}.$$

Proof. Same arguments as in the proof of Lemma A2. □

Proof of Theorem 3. (i) Let $U(\cdot;Z_0) = \{U_{gl}(\cdot;Z_0);\, g,l = 1,\ldots,k\}$ be a $k\times k$ matrix-valued process. Here, for $g \neq l$ and $gl \in \cup_{X\in\mathcal{E}}\mathcal{J}(X)$, the $U_{gl}(\cdot;Z_0)$ are independent zero-mean Gaussian processes as given by Lemma A3(iv); for $g \neq l$ and $gl \notin \mathcal{J}$, $U_{gl}(\cdot;Z_0) = 0$; and $U_{gg}(\cdot;Z_0) = -\sum_{l\neq g} U_{gl}(\cdot;Z_0)$. Then $n^{1/2}\int_s^\cdot d\{\hat{A}(u;Z_0) - A(u;Z_0)\}$ converges weakly on $[s,v]$ to $U(\cdot;Z_0)$. It follows from Lemma 1 that $n^{1/2}\{\hat{P}(s,\cdot;Z_0) - P(s,\cdot;Z_0)\}$ converges weakly on $[s,v]$ to the zero-mean Gaussian process

$$\int_s^\cdot P(s,u-;Z_0)\, dU(u;Z_0)P(u,\cdot;Z_0).$$

(ii) Again, as in the beginning of the proof of Theorem 1(ii) in Appendix 1, from Lemma 1, we may also derive (A1) in the current context.

Let $P(s,u;Z_0)^{-1} = \left\{P^{hj}(s,u;Z_0);\, h,j=1,\ldots,k\right\}$ and let

$$\tilde{H}_{gl}^{(hj)}(u;s,Z_0) = P_{hg}(s,u-;Z_0)\{P^{lj}(s,u;Z_0) - P^{gj}(s,u;Z_0)\},$$

$$\tilde{F}_X^{(hj)}(u;s,Z_0) = \sum_{gl\in\mathcal{J}(X)} \tilde{H}_{gl}^{(hj)}(u;s,Z_0), \tag{A11}$$

$$\tilde{Q}_X^{(hj)}(s,t;Z_0) = \int_s^t \sum_{gl\in\mathcal{J}(X)} \left[\tilde{H}_{gl}^{(hj)}(u;s,Z_0)\{Z_{gl0} - e_X(u)\}\right] du. \tag{A12}$$

Denote the (h,j)th entry of $n^{1/2}\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1} - I\right\}$ by $\xi_{hj}^{(n)}(s,t;Z_0)$. Then, from (A1), by applying Lemma A3 (i) and (iii), we have

$$\begin{aligned}
\xi_{hj}^{(n)}(s,t;Z_0) &\stackrel{a}{=} \sum_{g=1}^{k}\sum_{l=1}^{k}\int_s^t P_{hg}(s,u-;Z_0)d\left\{n^{1/2}(\hat{A}_{gl} - A_{gl})(u;Z_0)\right\}P^{lj}(s,u;Z_0)\\
&= \sum_{g=1}^{k}\sum_{l\neq g}\int_s^t \tilde{H}_{gl}^{(hj)}(u;s,Z_0)d\left\{n^{1/2}(\hat{A}_{gl} - A_{gl})(u;Z_0)\right\}\\
&= \left(\sum_{X\in\mathcal{E}}\sum_{gl\in\mathcal{J}(X)} + \sum_{g\neq l,\, gl\notin\mathcal{J}}\right)\int_s^t \tilde{H}_{gl}^{(hj)}(u;s,Z_0)d\left\{n^{1/2}(\hat{A}_{gl} - A_{gl})(u;Z_0)\right\}\\
&\stackrel{a}{=} \sum_{X\in\mathcal{E}}\sum_{gl\in\mathcal{J}(X)}\int_s^t \tilde{H}_{gl}^{(hj)}(u;s,Z_0)\{Z_{gl0} - e_X(u)\}^\top du\, n^{1/2}(\hat{\beta}_X - \beta_X)\\
&\quad + \sum_{X\in\mathcal{E}}\sum_{gl\in\mathcal{J}(X)}\int_s^t \tilde{H}_{gl}^{(hj)}(u;s,Z_0)n^{1/2}\,\frac{dM_X(u)}{Y_X(u)}\\
&= \sum_{X\in\mathcal{E}}\{\tilde{Q}_X^{(hj)}(s,t;Z_0)\}^\top n^{1/2}(\hat{\beta}_X - \beta_X)\\
&\quad + \sum_{X\in\mathcal{E}}\int_s^t \tilde{F}_X^{(hj)}(u;s,Z_0)n^{1/2}\,\frac{dM_X(u)}{Y_X(u)}\\
&\stackrel{a}{=} \sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(hj)}(s,t;Z_0)\}^\top\Omega_X^{-1}n^{-1/2}\int_0^\infty\{Z_{Xi}(u) - \bar{Z}_X(u)\}\,dM_{Xi}(u)\\
&\quad + \sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\int_s^t \tilde{F}_X^{(hj)}(u;s,Z_0)n^{1/2}\,\frac{dM_{Xi}(u)}{Y_X(u)}.
\end{aligned} \tag{A13}$$

Notice that in (A13), the $M_{Xi}(\cdot)$ are orthogonal martingales with optional variation processes given by $[M_{Xi}](\cdot) = N_{Xi}(\cdot)$. Therefore,

$$\mathrm{cov}\left\{\xi_{hj}^{(n)}(s,t;Z_0),\, \xi_{mr}^{(n)}(s,t;Z_0)\right\}$$

$$
\begin{aligned}
&\overset{a}{=} \sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(hj)}(s,t;Z_0)\}^\top\Omega_X^{-1}\frac{1}{n}\operatorname{cov}\left[\int_0^\infty\{Z_{Xi}(u)-\bar{Z}_X(u)\}\,dM_{Xi}(u)\right]\Omega_X^{-1}\tilde{Q}_X^{(mr)}(s,t;Z_0)\\
&\quad+\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\operatorname{cov}\left\{\int_s^t\tilde{F}_X^{(hj)}(u;s,Z_0)n^{1/2}\frac{dM_{Xi}(u)}{Y_X(u)},\int_s^t\tilde{F}_X^{(mr)}(u;s,Z_0)n^{1/2}\frac{dM_{Xi}(u)}{Y_X(u)}\right\}\\
&\quad+\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(hj)}(s,t;Z_0)\}^\top\Omega_X^{-1}\\
&\qquad\qquad\times\operatorname{cov}\left[\int_0^\infty\{Z_{Xi}(u)-\bar{Z}_X(u)\}\,dM_{Xi}(u),\int_s^t\tilde{F}_X^{(mr)}(u;s,Z_0)\frac{dM_{Xi}(u)}{Y_X(u)}\right]\\
&\quad+\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(mr)}(s,t;Z_0)\}^\top\Omega_X^{-1}\\
&\qquad\qquad\times\operatorname{cov}\left[\int_0^\infty\{Z_{Xi}(u)-\bar{Z}_X(u)\}\,dM_{Xi}(u),\int_s^t\tilde{F}_X^{(hj)}(u;s,Z_0)\frac{dM_{Xi}(u)}{Y_X(u)}\right]\\
&=\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(hj)}(s,t;Z_0)\}^\top\Omega_X^{-1}\frac{1}{n}E\left(\int_0^\infty\{Z_{Xi}(u)-\bar{Z}_X(u)\}^{\otimes2}d[M_{Xi}](u)\right)\\
&\qquad\qquad\times\Omega_X^{-1}\tilde{Q}_X^{(mr)}(s,t;Z_0)\\
&\quad+\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}E\left(\int_s^t\tilde{F}_X^{(hj)}(u;s,Z_0)\tilde{F}_X^{(mr)}(u;s,Z_0)\frac{n\,d[M_{Xi}](u)}{\{Y_X(u)\}^2}\right)\\
&\quad+\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(hj)}(s,t;Z_0)\}^\top\Omega_X^{-1}E\left\{\int_s^t\frac{Z_{Xi}(u)-\bar{Z}_X(u)}{Y_X(u)}\tilde{F}_X^{(mr)}(u;s,Z_0)d[M_{Xi}](u)\right\}\\
&\quad+\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(mr)}(s,t;Z_0)\}^\top\Omega_X^{-1}E\left\{\int_s^t\frac{Z_{Xi}(u)-\bar{Z}_X(u)}{Y_X(u)}\tilde{F}_X^{(hj)}(u;s,Z_0)d[M_{Xi}](u)\right\}\\
&=\sum_{X\in\mathcal{E}}\{\tilde{Q}_X^{(hj)}(s,t;Z_0)\}^\top\Omega_X^{-1}V_X\Omega_X^{-1}\tilde{Q}_X^{(mr)}(s,t;Z_0)\\
&\quad+\sum_{X\in\mathcal{E}}E\left[\int_s^t\tilde{F}_X^{(hj)}(u;s,Z_0)\tilde{F}_X^{(mr)}(u;s,Z_0)\frac{n\,dN_X(u)}{\{Y_X(u)\}^2}\right]\\
&\quad+\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(hj)}(s,t;Z_0)\}^\top\Omega_X^{-1}E\left\{\int_s^t\frac{Z_{Xi}(u)-\bar{Z}_X(u)}{Y_X(u)}\tilde{F}_X^{(mr)}(u;s,Z_0)dN_{Xi}(u)\right\}\\
&\quad+\sum_{X\in\mathcal{E}}\sum_{i=1}^{n}\{\tilde{Q}_X^{(mr)}(s,t;Z_0)\}^\top\Omega_X^{-1}E\left\{\int_s^t\frac{Z_{Xi}(u)-\bar{Z}_X(u)}{Y_X(u)}\tilde{F}_X^{(hj)}(u;s,Z_0)dN_{Xi}(u)\right\}. \qquad \text{(A14)}
\end{aligned}
$$

Let C_{gl} be defined as in Appendix 1. Then $\tilde{H}_{gl}^{(hj)}(u;s,Z_0)$ is the (h,j)th entry of $P(s,u-;Z_0)C_{gl}P(s,u;Z_0)^{-1}$. It follows from (A11), (A12) and (A14) that

$$
\begin{aligned}
&\operatorname{cov}\left[n^{1/2}\left\{\hat{P}(s,t;Z_0)P(s,t;Z_0)^{-1}-I\right\}\right]\\
&\overset{a}{=}\sum_{X\in\mathcal{E}}\int_s^t\sum_{gl\in\mathcal{J}(X)}\left[\operatorname{vec}\{P(s,u-;Z_0)C_{gl}P(s,u;Z_0)^{-1}\}\{Z_{gl0}-e_X(u)\}^\top\right]du\,\Omega_X^{-1}V_X\Omega_X^{-1}
\end{aligned}
$$

$$\times \int_s^t \sum_{gl \in \mathcal{J}(X)} \Big[\{Z_{gl0} - e_X(u)\} \operatorname{vec}^\top \{P(s, u-; Z_0) C_{gl} P(s, u; Z_0)^{-1}\} \Big] \, du$$

$$+ \sum_{X \in \mathcal{E}} E \Bigg[\int_s^t \sum_{gl \in \mathcal{J}(X)} \operatorname{vec}\{P(s, u-; Z_0) C_{gl} P(s, u; Z_0)^{-1}\}$$

$$\times \sum_{gl \in \mathcal{J}(X)} \operatorname{vec}^\top \{P(s, u-; Z_0) C_{gl} P(s, u; Z_0)^{-1}\} \frac{n \, dN_X(u)}{\{Y_X(u)\}^2} \Bigg]$$

$$+ 2 \sum_{X \in \mathcal{E}} \sum_{i=1}^n \int_s^t \sum_{gl \in \mathcal{J}(X)} \Big[\operatorname{vec}\{P(s, u-; Z_0) C_{gl} P(s, u; Z_0)^{-1}\} \{Z_{gl0} - e_X(u)\}^\top \Big] \, du \, \Omega_X^{-1}$$

$$\times E \Bigg[\int_s^t \frac{Z_{Xi}(u) - \bar{Z}_X(u)}{Y_X(u)} \sum_{gl \in \mathcal{J}(X)} \operatorname{vec}^\top \{P(s, u-; Z_0) C_{gl} P(s, u; Z_0)^{-1}\} \, dN_{Xi}(u) \Bigg].$$

Using this result, (A5), and elementary properties of the vec-operator and Kronecker products of matrices, we find that

$$\operatorname{cov}\Big\{\hat{P}(s, t; Z_0)\Big\}$$

$$\overset{a}{=} \sum_{X \in \mathcal{E}} \int_s^t \sum_{gl \in \mathcal{J}(X)} \Big[\operatorname{vec}\{P(s, u-; Z_0) C_{gl} P(u, t; Z_0)\} \{Z_{gl0} - e_X(u)\}^\top \Big] \, du \, \frac{1}{n} \Omega_X^{-1} V_X \Omega_X^{-1}$$

$$\times \int_s^t \sum_{gl \in \mathcal{J}(X)} \Big[\{Z_{gl0} - e_X(u)\} \operatorname{vec}^\top \{P(s, u-; Z_0) C_{gl} P(u, t; Z_0)\} \Big] \, du$$

$$+ \sum_{X \in \mathcal{E}} E \Bigg[\int_s^t \sum_{gl \in \mathcal{J}(X)} \operatorname{vec}\{P(s, u-; Z_0) C_{gl} P(u, t; Z_0)\}$$

$$\times \sum_{gl \in \mathcal{J}(X)} \operatorname{vec}^\top \{P(s, u-; Z_0) C_{gl} P(u, t; Z_0)\} \frac{dN_X(u)}{\{Y_X(u)\}^2} \Bigg]$$

$$+ \frac{2}{n} \sum_{X \in \mathcal{E}} \sum_{i=1}^n \int_s^t \sum_{gl \in \mathcal{J}(X)} \Big[\operatorname{vec}\{P(s, u-; Z_0) C_{gl} P(u, t; Z_0)\} \{Z_{gl0} - e_X(u)\}^\top \Big] \, du \, \Omega_X^{-1}$$

$$\times E \Bigg[\int_s^t \frac{Z_{Xi}(u) - \bar{Z}_X(u)}{Y_X(u)} \sum_{gl \in \mathcal{J}(X)} \operatorname{vec}^\top \{P(s, u-; Z_0) C_{gl} P(u, t; Z_0)\} \, dN_{Xi}(u) \Bigg]. \quad \text{(A15)}$$

Now let

$$H_{gl}^{(hj)}(u; s, t, Z_0) = P_{hg}(s, u-; Z_0) \{P_{lj}(u, t; Z_0) - P_{gj}(u, t; Z_0)\},$$

$$F_X^{(hj)}(u; s, t, Z_0) = \sum_{gl \in \mathcal{J}(X)} H_{gl}^{(hj)}(u; s, t, Z_0),$$

$$Q_X^{(hj)}(s, t; Z_0) = \int_s^t \sum_{gl \in \mathcal{J}(X)} \Big[H_{gl}^{(hj)}(u; s, t, Z_0) \{Z_{gl0} - e_X(u)\} \Big] \, du.$$

Then the (h,j)th entry of $P(s,u-;Z_0)C_{gl}P(u,t;Z_0)$ is $H_{gl}^{(hj)}(u;s,t,Z_0)$. It follows from (A15) that

$$
\begin{aligned}
&\operatorname{cov}\left\{\hat{P}_{hj}(s,t;Z_0),\ \hat{P}_{mr}(s,t;Z_0)\right\} \\
&\overset{a}{=} \sum_{X\in\mathcal{E}}\{Q_X^{(hj)}(s,t;Z_0)\}^\top \frac{1}{n}\Omega_X^{-1}V_X\Omega_X^{-1}Q_X^{(mr)}(s,t;Z_0) \\
&\quad + \sum_{X\in\mathcal{E}} E\left[\int_s^t F_X^{(hj)}(u;s,t,Z_0)F_X^{(mr)}(u;s,t,Z_0)\frac{dN_X(u)}{\{Y_X(u)\}^2}\right] \\
&\quad + \frac{1}{n}\sum_{X\in\mathcal{E}}\sum_{i=1}^n\{Q_X^{(hj)}(s,t;Z_0)\}^\top\Omega_X^{-1}E\left\{\int_s^t \frac{Z_{Xi}(u)-\bar{Z}_X(u)}{Y_X(u)}F_X^{(mr)}(u;s,t,Z_0)\,dN_{Xi}(u)\right\} \\
&\quad + \frac{1}{n}\sum_{X\in\mathcal{E}}\sum_{i=1}^n\{Q_X^{(mr)}(s,t;Z_0)\}^\top\Omega_X^{-1}E\left\{\int_s^t \frac{Z_{Xi}(u)-\bar{Z}_X(u)}{Y_X(u)}F_X^{(hj)}(u;s,t,Z_0)\,dN_{Xi}(u)\right\},
\end{aligned}
$$

which can be estimated uniformly consistently by (13). □

References

Aalen, O.O. (1989). A linear regression model for the analysis of life times. *Statist. Med.* **8**, 907–25.

Aalen, O.O., Borgan, Ø. & Fekjær, H. (2001). Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics* **57**, 993–1001.

Aalen, O.O. & Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5**, 141–50.

Alioum, A. & Commenges, D. (2001). MKVPCI: a computer program for Markov models with piecewise constant intensities and covariates. *Comp. Meth. Prog. Biomed.* **64**, 109–19.

Andersen, P.K. (1988). Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statist. Med.* **7**, 661–70.

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

Andersen, P.K., Esbjerg, S. & Sørensen, T.I.A. (2000). Multistate models for bleeding episodes and mortality in liver cirrhosis. *Statist. Med.* **19**, 587–99.

Andersen, P.K. & Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–20.

Andersen, P.K., Hansen, L.S. & Keiding, N. (1991). Non- and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process. *Scand. J. Statist.* **18**, 153–67.

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

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

Cox, D.R. (1972). Regression models and life-tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–220.

Gill, R.D. & Johansen, S. (1990). A survey of product-integration with a view towards application in survival analysis. *Ann. Statist.* **18**, 1501–55.

Huffer, F.W. & McKeague, I.W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Am. Statist. Assoc.* **86**, 114–29.

Kalbfleisch J.D. & Lawless J.F. (1985). The analysis of panel data under a Markov assumption. *J. Am. Statist. Assoc.* **80**, 863–71.

Keiding, N., Klein, J.P. & Horowitz, M.M. (2001). Multistate models and outcome prediction in bone marrow transplantation. *Statist. Med.* **20**, 1871–85.

Klein, J.P., Keiding, N. & 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**, 2315–32.

Klein, J.P. & Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and*

Truncated Data. New York: Springer-Verlag.

Klein, J.P. & Qian, C. (1996). Modelling multistate survival illustrated in bone marrow transplantation. In *Proc. Biometrics Section Am. Statist. Assoc.*, pp. 93–102. Alexandria, VA: Am. Statist. Assoc.

Kulich, M. & Lin, D.Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika* **87**, 73–87.

Lin, D.Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

Marshall G. & Jones, R.H. (1995). Multistate models and diabetic retinopathy. *Statist. Med.* **14**, 1975–83.

McKeague I.W. & Sasieni, P.D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501-14.

Pérez-Ocón, R., Ruiz-Castro, J.E. & Gámiz-Pérez, M.L. (2001). Nonhomogeneous Markov models in the analysis of survival after breast cancer. *Appl. Statist.* **50**, 111–24.

Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Geb.* **51**, 269–86.

Shen, Y. & Cheng, S.C. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics* **55**, 1093–100.

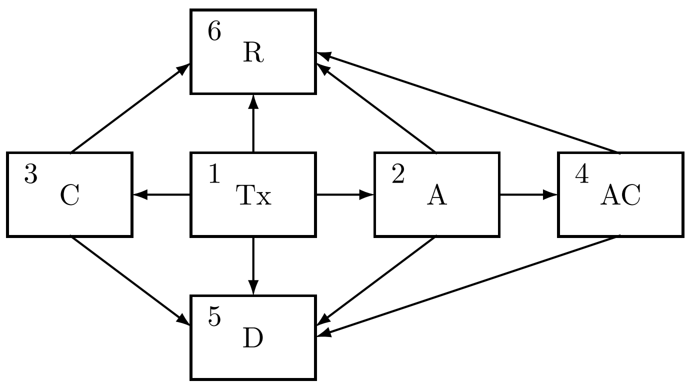

Fig. 1. A six-state bone marrow transplant model.

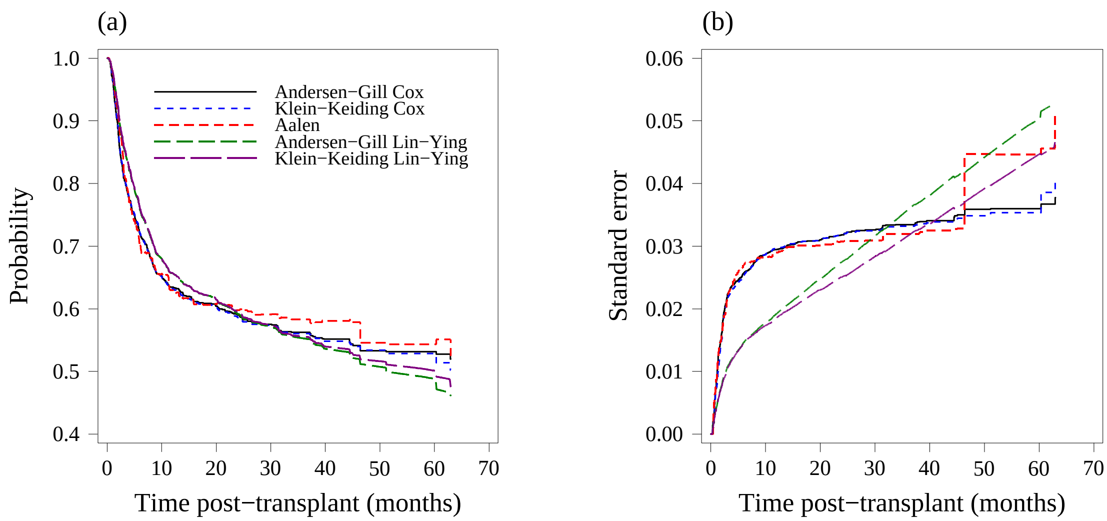

Fig. 2. The results of fitting the five Markov regression models to the bone marrow transplant data. (a) Predicted probabilities of leukaemia-free survival, (b) standard errors of predicted probabilities of leukaemia-free survival.

Table 1: *Estimated probabilities of being in states 1 to 6 and the associated standard errors at 100 days, 6 months, 1 year and 2 years post-transplant under each of the five Markov regression models, for a male patient aged over 28 years with a female donor, given that the patient was initially in state 1 at time 0.*

State	Time Post-transplant	AG Cox Est. (SE)	KK Cox Est. (SE)	Aalen Est. (SE)	AG-LY Est. (SE)	KK-LY Est. (SE)
1	100 days	0.539 (0.029)	0.533 (0.028)	0.543 (0.028)	0.590 (0.015)	0.566 (0.014)
(Tx)	6 months	0.423 (0.028)	0.420 (0.028)	0.414 (0.028)	0.443 (0.017)	0.421 (0.014)
	1 year	0.339 (0.029)	0.336 (0.029)	0.321 (0.027)	0.330 (0.020)	0.325 (0.017)
	2 years	0.307 (0.029)	0.307 (0.029)	0.290 (0.027)	0.261 (0.026)	0.296 (0.025)
2	100 days	0.192 (0.022)	0.204 (0.022)	0.178 (0.022)	0.181 (0.012)	0.210 (0.011)
(A)	6 months	0.108 (0.018)	0.120 (0.019)	0.114 (0.019)	0.106 (0.011)	0.139 (0.011)
	1 year	0.071 (0.016)	0.074 (0.016)	0.073 (0.016)	0.078 (0.013)	0.083 (0.012)
	2 years	0.060 (0.015)	0.060 (0.014)	0.059 (0.015)	0.080 (0.021)	0.048 (0.013)
3	100 days	0.033 (0.007)	0.035 (0.006)	0.040 (0.010)	0.041 (0.007)	0.045 (0.006)
(C)	6 months	0.103 (0.017)	0.101 (0.016)	0.113 (0.018)	0.124 (0.011)	0.130 (0.009)
	1 year	0.142 (0.022)	0.136 (0.021)	0.149 (0.021)	0.162 (0.014)	0.164 (0.012)
	2 years	0.148 (0.023)	0.142 (0.022)	0.164 (0.023)	0.175 (0.020)	0.164 (0.019)
4	100 days	0.040 (0.009)	0.033 (0.007)	0.048 (0.013)	0.039 (0.007)	0.032 (0.005)
(AC)	6 months	0.085 (0.016)	0.081 (0.015)	0.058 (0.015)	0.085 (0.009)	0.070 (0.007)
	1 year	0.083 (0.018)	0.086 (0.017)	0.086 (0.018)	0.087 (0.011)	0.085 (0.011)
	2 years	0.075 (0.018)	0.077 (0.018)	0.085 (0.018)	0.075 (0.016)	0.086 (0.016)
5	100 days	0.178 (0.023)	0.175 (0.022)	0.173 (0.021)	0.130 (0.010)	0.126 (0.010)
(D)	6 months	0.208 (0.025)	0.205 (0.024)	0.214 (0.024)	0.163 (0.013)	0.159 (0.012)
	1 year	0.241 (0.027)	0.243 (0.027)	0.240 (0.025)	0.211 (0.017)	0.207 (0.016)
	2 years	0.265 (0.029)	0.268 (0.029)	0.255 (0.026)	0.254 (0.024)	0.247 (0.022)
6	100 days	0.019 (0.005)	0.020 (0.005)	0.019 (0.009)	0.019 (0.005)	0.020 (0.005)
(R)	6 months	0.072 (0.013)	0.074 (0.013)	0.086 (0.017)	0.079 (0.009)	0.081 (0.009)
	1 year	0.124 (0.019)	0.125 (0.019)	0.131 (0.021)	0.133 (0.014)	0.136 (0.013)
	2 years	0.145 (0.022)	0.146 (0.022)	0.148 (0.023)	0.154 (0.020)	0.158 (0.018)

AG, Andersen-Gill; KK, Klein-Keiding; LY, Lin-Ying; Est., estimated probability; SE, standard error.