Semiparametric Regression Analysis of Interval-Censored Multi-State Data with An Absorbing State

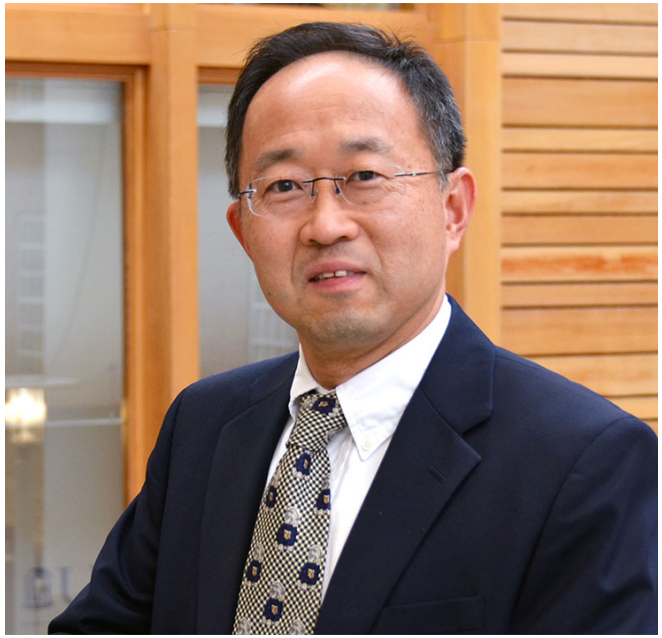
Yu Gu

Department of Biostatistics
University of North Carolina at Chapel Hill

July 1, 2023

Acknowledgments

Dr. Donglin Zeng



Dr. Danyu Lin

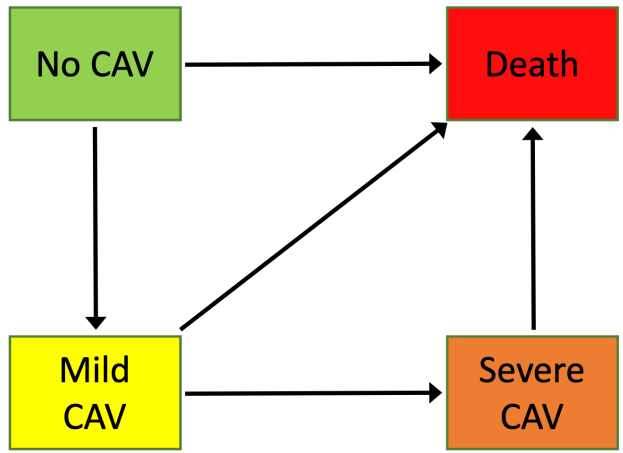
Gu, Y., Zeng, D., Lin, D. Y. Semiparametric Regression Analysis of Interval-Censored Multi-State Data with An Absorbing State. Under review.

Outline

Outline

Multi-State Data

- Multi-state data arise frequently in studies of chronic diseases.
- Health status can be characterized by a finite number of disease states.
- ***Transition***: change from one state to another.



Cardiac allograft vasculopathy

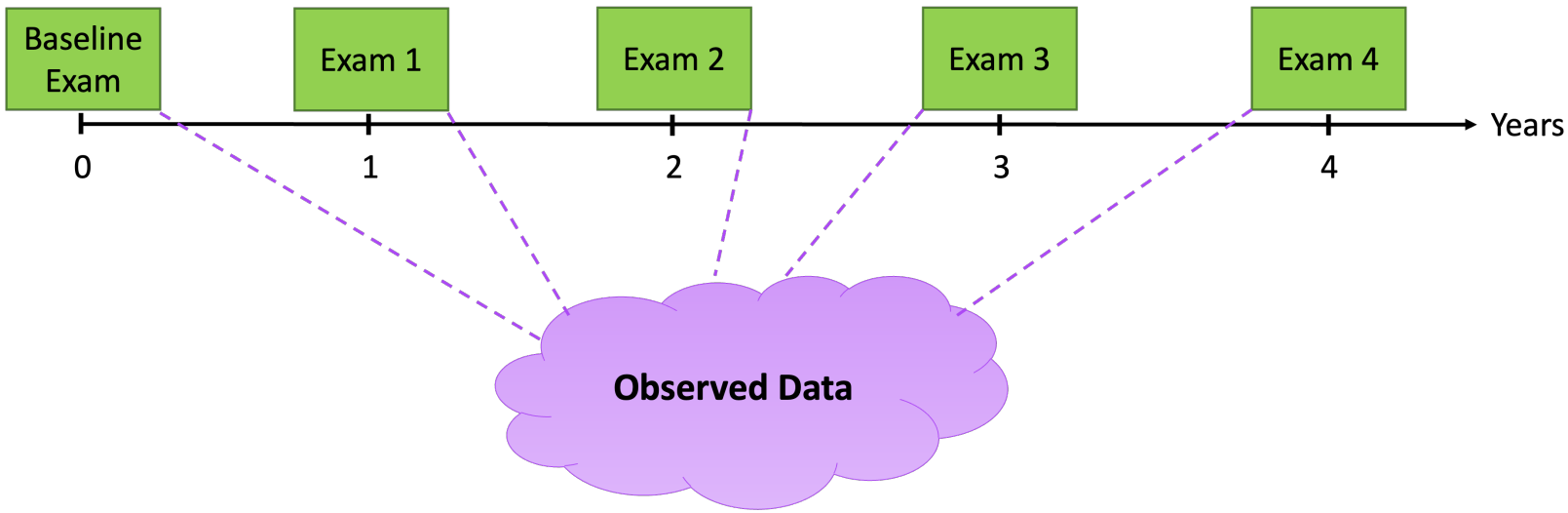
Significance

Analysis of multi-state data allows us to

- understand how a subject transitions from one state to another over time
- study the associations between risk factors and the disease process
- predict future disease progression using the disease history

Interval Censoring

- For economic and logistical reasons, subjects can only be examined periodically, such that transitions are only known to occur between two successive examinations.



- Such data are called ***interval-censored multi-state data***.

Absorbing State and Right Censoring

- In many applications, there is an ***absorbing state*** (e.g., death) which terminates the disease process.
- A common situation is that the time of entering the absorbing state is observed exactly or right-censored.
- However, the ***transient state*** from which a subject enters the absorbing state is still unknown.

Analysis Challenges

- None of the transition times among transient states are directly observed.
- Trajectory of transitions from one examination to the next is unknown.
- Dependence among transitions from the same subject.
- A mixture of interval- and right-censored transition times.
- Missing data on the transient state right before the absorbing state.

Existing Methods

Nonhomogeneous

Time-homogeneous Markov models

- Kalbfleisch & Lawless (1985); Satten (1999); Cook et al. (2002, 2004)
- simple; parametric; implemented in the msm package
- homogeneous assumption is unrealistic

Piecewise constant transition intensities

Nonparametric

- Gentleman et al. (1994); Saint-Pierre et al. (2003); Jackson (2011); Lawless & Nazeri Rad (2015)
- relatively simple; parametric
- restrictive; sensitive to the choice of change points

Spline-based intensities + penalized likelihood

- Joly & Commenges (1999); Machado & van den Hout (2018); Machado et al. (2021)
- more flexible; semiparametric
- tuning parameters (e.g., knots); inconsistent estimators

Overview of this work

- We provide a new framework to study semiparametric regression models for general interval-censored multi-state data with an absorbing state whose entry time is exactly known or right-censored.
- Our models use random effects to capture the dependence among transitions and accommodate time-dependent covariates.
- We combine ***nonparametric maximum likelihood estimation*** (NPMLE) and ***sieve estimation*** for inference.
- We devise a stable EM algorithm to compute the estimators.
- We leverage random effects to dynamically predict future process using the evolving disease history.

Outline

Notation

- We consider a random sample of n subjects and K states, with state K being absorbing and all others being transient.
- Let $N_{ijk}(t)$ denote the number of transitions $j \to k$ the ith subject has experienced by time t.
- For the ith subject, let $\boldsymbol{X}_i(t)$ denote a set of potentially time-dependent covariates and $\boldsymbol{b}_i$ denote a d-vector of random effects.

Semiparametric Regression Model

We specify that the transition intensity of $N_{ijk}(t)$ is related to $\boldsymbol{X}_i(t)$ and $\boldsymbol{b}_i$ through the proportional intensity model:

$$\lambda_{ijk}(t; \boldsymbol{X}_i, \boldsymbol{b}_i) = \lambda_{jk}(t) \exp\left\{\boldsymbol{\beta}_{jk}^{\mathsf{T}} \boldsymbol{X}_i(t) + \boldsymbol{b}_i^{\mathsf{T}} \boldsymbol{Z}_i(t)\right\} \tag{1}$$

- $\lambda_{jk}(\cdot)$: arbitrary baseline intensity function
- $\boldsymbol{\beta}_{jk}$: unknown regression parameters
- $\boldsymbol{b}_i \sim N_d(\boldsymbol{0}, \boldsymbol{\Sigma}(\gamma))$
- $\boldsymbol{Z}_i(\cdot)$: consists of 1 and covariates that are part of $\boldsymbol{X}_i(\cdot)$

Observed Data

- Examination times: $(U_{i0}, U_{i1}, \ldots, U_{iM_i})$
- Observed transient states: $(S_{i0}, S_{i1}, \ldots, S_{iM_i})$
- Time and status of entering state K: (Y_i, Δ_i)
  - $Y_i = \min(T_i, C_i)$
  - $\Delta_i = I(T_i \leqslant C_i)$
  - T_i is actual event time, C_i is censoring time
- Covariates: $\boldsymbol{X}_i(t)$

Likelihood

Under the noninformative censoring and conditional Markov assumptions, the likelihood is proportional to

$$\prod_{i=1}^{n} \int_{\boldsymbol{b}_i} \prod_{l=1}^{M_i} \boldsymbol{P}(U_{i,l-1}, U_{il}; \boldsymbol{X}_i, \boldsymbol{b}_i)^{(S_{i,l-1}, S_{il})} \times \left\{ \sum_{j \neq K} \boldsymbol{P}(U_{iM_i}, Y_i; \boldsymbol{X}_i, \boldsymbol{b}_i)^{(S_{iM_i}, j)} \right\}^{1-\Delta_i} \times \left\{ \sum_{j \in \mathcal{D}_K} \boldsymbol{P}(U_{iM_i}, Y_i; \boldsymbol{X}_i, \boldsymbol{b}_i)^{(S_{iM_i}, j)} \lambda_{ijK}(Y_i) \right\}^{\Delta_i} \phi(\boldsymbol{b}_i; \boldsymbol{\gamma}) d\boldsymbol{b}_i \tag{2}$$

- $\boldsymbol{P}(u, v; \boldsymbol{X}_i, \boldsymbol{b}_i)$ denotes the transition probability matrix between times u and v for the ith subject.
- $\phi(\boldsymbol{b}; \boldsymbol{\gamma})$ denotes the density function of $N_d(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\gamma}))$.

Compute Transition Probability

- The transition probability matrix is given by

$$\boldsymbol{P}(u, v; \boldsymbol{X}_i, \boldsymbol{b}_i) = \prod_{u<t\leqslant v} \{\boldsymbol{I}_K + d\boldsymbol{A}(t; \boldsymbol{X}_i, \boldsymbol{b}_i)\}$$

- $\boldsymbol{A}(t; \boldsymbol{X}_i, \boldsymbol{b}_i)$ is the cumulative transition intensity matrix, with

$$\boldsymbol{A}(t; \boldsymbol{X}_i, \boldsymbol{b}_i)^{(j,k)} = \int_0^t \exp\{\beta_{jk}^{\mathsf{T}} \boldsymbol{X}_i(s) + \boldsymbol{b}_i^{\mathsf{T}} \boldsymbol{Z}_i(s)\} d\Lambda_{jk}(s).$$

- $\Lambda_{jk}(t) = \int_0^t \lambda_{jk}(s) ds$.

NPMLE + Sieve Estimation

- We adopt NPMLE approach for Λ_{jk} of
  1. all transitions among states $1, 2, \ldots, K-1$
  2. transition from state 1 to state $K \Rightarrow$ reference Λ_a
- We use B-splines to approximate the log transition intensity ratio $\psi_j = \log(\lambda_{jK}/\lambda_a)$
- Why not NPMLE for all Λ_{jk}? Inconsistent estimators!
  - unknown state right before entering state K
  - Ma & Wang (2012); Wang et al. (2012)

More on NPMLE

- We treat Λ_{jk} as step functions with nonnegative jumps at all examination times and Y_i's.
- We treat Λ_a as a step function that jumps only at those Y_i with $\Delta_i = 1$.
- Then the transition probability matrix

$$\boldsymbol{P}_i(u, v) = \prod_{u<t\leqslant v} \{\boldsymbol{I}_K + d\boldsymbol{A}_i(t)\}$$

$$\Downarrow$$

$$\widetilde{\boldsymbol{P}}_i(u, v) = \prod_{u<t_q\leqslant v} \{\boldsymbol{I}_K + \delta\boldsymbol{A}_i(t_q)\}.$$

- $\delta\boldsymbol{A}_i(t_q)$ involves the jump sizes λ_{jkq} and λ_{aq}.

Remarks

- Advantages of NPMLE:
  - uses all information told by the data
  - minimal assumption about Λ's compared to splines
  - no tuning parameters
- The combination of NPMLE and B-splines ensures estimation consistency while achieving the maximal model flexibility.
- Challenges of computing sieve NPMLE:
  - high-dimensional parameters λ_{jkq} and λ_{aq}
  - lack of analytical expressions

Poissonization

- We introduce independent latent Poisson random variables W_{ijkq} with means $\{\delta \boldsymbol{A}_i(t_q)\}^{(j,k)}$.

- The key fact is that the transition probability $\widetilde{\boldsymbol{P}}_i(u,v)^{(s_0,s_r)}$ is equal to the probability of the event

$$\bigcup_{\text{traj}(s_0,s_1,\ldots,s_r)} \Big\{ W_{ijkq} > 0 \text{ if there's a transition } j \to k \text{ at time } t_q \text{ and } W_{ijkq} = 0 \text{ otherwise} \Big\}.$$

- Thus, maximizing the original likelihood is tantamount to maximizing the likelihood arising from the events of W's.

EM Algorithm

- We can treat W_{ijkq}'s and $\boldsymbol{b}_i$'s as missing data and apply the EM algorithm for maximizing the likelihood function.

- The complete-data log-likelihood is

$$\begin{aligned}
&\sum_{i=1}^{n}\Bigg\{\sum_{(j,k)\in\mathcal{D}^*}\sum_{q=1}^{m} I(t_q \leqslant Y_i)\Big[W_{ijkq}\left\{\log\left(\lambda_{jkq}\right)+\beta_{jk}^{\mathsf{T}}\boldsymbol{X}_{iq}+\boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{Z}_{iq}\right\}\\
&\qquad -\lambda_{jkq}\exp\left\{\beta_{jk}^{\mathsf{T}}\boldsymbol{X}_{iq}+\boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{Z}_{iq}\right\}-\log(W_{ijkq}!)\Big]\\
&+\sum_{j\in\mathcal{D}_K}\sum_{q=1}^{m} I(t_q \leqslant Y_i)\Big[W_{ijKq}\left\{\log\left(\lambda_{aq}\right)+\boldsymbol{\alpha}_j^{\mathsf{T}}\boldsymbol{B}_q+\beta_{jK}^{\mathsf{T}}\boldsymbol{X}_{iq}+\boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{Z}_{iq}\right\} \qquad (3)\\
&\qquad -\lambda_{aq}\exp\left\{\boldsymbol{\alpha}_j^{\mathsf{T}}\boldsymbol{B}_q+\beta_{jK}^{\mathsf{T}}\boldsymbol{X}_{iq}+\boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{Z}_{iq}\right\}-\log(W_{ijKq}!)\Big]\\
&\qquad -\frac{d_2}{2}\log(2\pi)-\frac{1}{2}\log|\boldsymbol{\Sigma}(\gamma)|-\frac{1}{2}\boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{\Sigma}(\gamma)^{-1}\boldsymbol{b}_i\Bigg\}.
\end{aligned}$$

E-Step

- We evaluate the conditional expectations of W_{ijkq}'s and functions of $\boldsymbol{b}_i$'s given the observed data.
- All conditional expectations have explicit expressions.
- Integrals of $\boldsymbol{b}_i$ can be approximated using Gaussian-Hermite quadratures.

M-Step

- The regression and spline parameters can be updated by solving their respective score equations with the one-step Newton-Raphson approach.

- The biggest advantage of the EM algorithm is that the jump sizes can be updated explicitly:

$$\lambda_{jkq} = \frac{\sum_{i=1}^{n} I(t_q \leqslant Y_i)\widetilde{E}(W_{ijkq})}{\sum_{i=1}^{n} I(t_q \leqslant Y_i)\widetilde{E}\{\exp(\beta_{jk}^{\mathsf{T}}\boldsymbol{X}_{iq} + \boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{Z}_{iq})\}},$$

$$\lambda_{aq} = \frac{\sum_{i=1}^{n} I(t_q \leqslant Y_i)\sum_{j\in\mathcal{D}_K}\widetilde{E}(W_{ijKq})}{\sum_{i=1}^{n} I(t_q \leqslant Y_i)\sum_{j\in\mathcal{D}_K}\widetilde{E}\{\exp(\boldsymbol{\alpha}_j^{\mathsf{T}}\boldsymbol{B}_q + \beta_{jK}^{\mathsf{T}}\boldsymbol{X}_{iq} + \boldsymbol{b}_i^{\mathsf{T}}\boldsymbol{Z}_{iq})\}}.$$

- Therefore, the EM algorithm is immune to the high-dimensional parameters in NPMLE.

Asymptotic Properties

- (Mixed rates of convergence) Under some regularity conditions,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2 + \sum_{(j,k)\in\mathcal{D}^*} \|\widehat{\Lambda}_{jk} - \Lambda_{0jk}\|_{L_2}^2 + \|\widehat{\Lambda}_a - \Lambda_{0a}\|_{L_2}^2 + \sum_{j\in\mathcal{D}_K^*} \|\widehat{\psi}_j - \psi_{0j}\|_{L_2}^2 = o_p(n^{-1/2}).$$

- (Asymptotic normality) $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal vector with mean zero and a covariance matrix that attains the semiparametric efficiency bound.

- (Variance estimation) The limiting covariance matrix of $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ can be consistently estimated by the inverse of

$$n^{-1} \sum_{i=1}^{n} \left\{ \nabla pl_i(\widehat{\boldsymbol{\theta}}) \right\}^{\otimes 2}.$$

Dynamic Prediction

- The key is to update the posterior density of random effects given the disease history $\mathcal{H}(t_0)$, which is proportional to

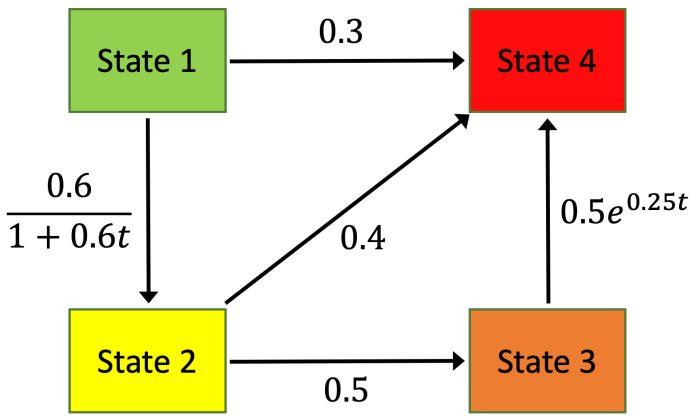
$$\prod_{l=1}^{L(t_0)} \widehat{\boldsymbol{P}}(U_{l-1}, U_l)^{(S_{l-1}, S_l)} \times \left\{ \sum_{j \neq K} \widehat{\boldsymbol{P}}(U_{L(t_0)}, t_0)^{(S_{L(t_0)}, j)} \right\} \phi(\boldsymbol{b}; \widehat{\gamma}).$$

- We then plug in this posterior density to estimate
  1. conditional probability of $S(t) = k$ given $\mathcal{H}(t_0)$, $k = 1, \ldots, K-1$
  2. conditional probability of reaching state K at time t given $\mathcal{H}(t_0)$

Outline

Simulation Setting

- $X_1 \sim \text{Bernoulli}(0.5)$ and $X_2 \sim \text{Unif}(0, 1)$
- $b \sim N(0, 0.5)$
- Six potential examination times separated by $0.05 + \text{Unif}(0, 1)$; no examinations after reaching state 4 or beyond $\tau = 3$.



Details of B-Splines

- We used cubic spline basis functions.
- We placed the boundary knots at 0 and τ, and placed two internal knots at the first and second tertiles of the observed time points.
- The estimation results are not sensitive to these choices.

Simulation Results

Table 1: Estimation of regression parameters

	$n = 500$				$n = 1000$				$n = 2000$			
	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
$\beta_{121} = 0.5$	0.026	0.272	0.254	94	0.014	0.178	0.180	96	0.003	0.134	0.127	95
$\beta_{122} = -0.5$	0.010	0.482	0.445	93	−0.005	0.322	0.312	94	−0.010	0.224	0.219	95
$\beta_{231} = 0.4$	−0.004	0.259	0.229	92	0.003	0.171	0.163	94	0.004	0.116	0.115	95
$\beta_{232} = 0.2$	0.022	0.467	0.398	91	0.016	0.296	0.283	94	0.012	0.208	0.199	94
$\beta_{141} = 0.3$	−0.007	0.309	0.256	91	0.000	0.193	0.180	94	−0.005	0.133	0.127	95
$\beta_{142} = 0.3$	−0.019	0.510	0.447	93	0.009	0.326	0.312	94	−0.002	0.219	0.220	95
$\beta_{241} = 0.3$	0.041	0.349	0.251	88	0.019	0.198	0.176	92	0.004	0.129	0.125	95
$\beta_{242} = 0.5$	0.001	0.570	0.426	86	−0.002	0.330	0.302	94	−0.006	0.226	0.214	95
$\beta_{341} = -0.2$	−0.030	0.272	0.220	90	−0.009	0.170	0.158	93	−0.003	0.114	0.113	95
$\beta_{342} = 0.5$	0.027	0.456	0.381	89	0.009	0.304	0.276	92	0.009	0.212	0.196	92
$\sigma^2 = 0.5$	−0.000	0.354	0.294	90	0.012	0.209	0.207	95	−0.002	0.130	0.141	97

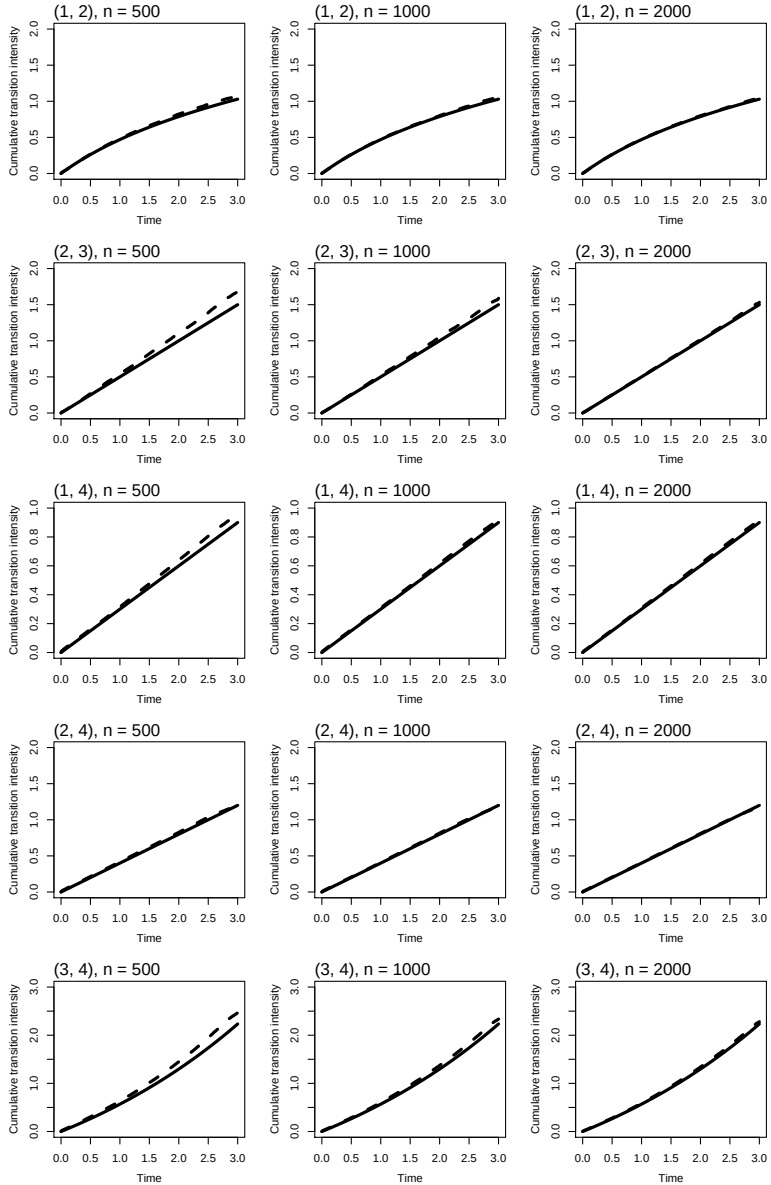


Figure 1: Estimation of cumulative baseline transition intensity functions.

Outline

CAV Study

- CAV is a deterioration of the arterial walls and is a common cause of death after heart transplantation.
- An important goal of the CAV study (Sharples et al., 2003) was to assess the effects of risk factors on CAV onset, progression, and survival.
- Starting from August 1979, a total of 622 heart transplant recipients underwent approximately yearly angiographic examinations and were classified as having no CAV, mild CAV, or severe CAV.
- Each patient was followed for up to 20 years, until death or the end of follow-up.
- The median follow-up time was 5 years; $\sim 40\%$ of the patients died during follow-up.

Analysis Specifics

- We fit a four-state random effects model.
- Tuning parameters in B-splines were determined based on AIC.
- For comparison, we also fit a homogeneous Markov model using the msm package.

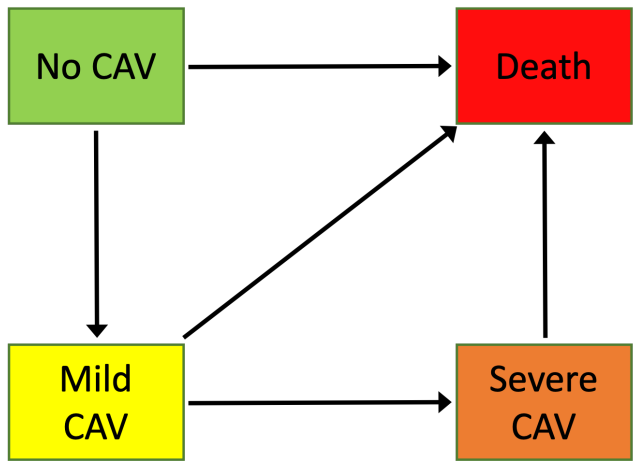


Table 2: Estimation results of regression parameters

	Proposed methods			msm package		
Covariate	Estimate	St error	p-value	Estimate	St error	p-value
No CAV to mild CAV						
Recipient age	−0.009	0.009	0.296	−0.013	0.007	0.071
Donor age	0.039	0.008	<0.001	0.027	0.006	<0.001
IHD	0.693	0.176	<0.001	0.549	0.151	<0.001
Acute rejection	0.163	0.041	<0.001	0.157	0.031	<0.001
Mild CAV to severe CAV						
Recipient age	−0.027	0.012	0.030	0.003	0.014	0.838
Donor age	−0.019	0.011	0.091	−0.009	0.011	0.378
IHD	0.335	0.269	0.213	0.081	0.227	0.722
Acute rejection	0.091	0.066	0.170	−0.015	0.055	0.783
No CAV to death						
Recipient age	0.074	0.015	<0.001	0.054	0.016	0.001
Donor age	0.032	0.010	0.002	0.020	0.009	0.032
IHD	0.144	0.255	0.573	0.085	0.223	0.703
Acute rejection	0.045	0.103	0.664	−0.058	0.085	0.498
Mild CAV to death						
Recipient age	0.074	0.031	0.017	−0.016	0.057	0.786
Donor age	0.025	0.027	0.349	−0.017	0.050	0.739
IHD	0.567	0.672	0.399	0.021	0.859	0.981
Acute rejection	0.101	0.123	0.412	0.250	0.123	0.043
Severe CAV to death						
Recipient age	−0.021	0.014	0.136	0.004	0.014	0.777
Donor age	−0.017	0.014	0.239	−0.008	0.012	0.518
IHD	−0.088	0.310	0.776	−0.232	0.225	0.303
Acute rejection	0.058	0.052	0.260	0.015	0.045	0.740

Key Findings

- Both donor age at transplant and cumulative number of acute rejection episodes are positively associated with the risk of mild CAV.
- Patients who received heart transplant due to ischemic heart disease have significantly higher risk of mild CAV.
- Older recipient and donor ages at transplant increase the risk of transition from no CAV to death.
- The variance of the random effect is estimated at 0.880 with a standard error estimator of 0.234, suggesting strong dependence among transitions.

Test Dynamic Prediction

- We randomly divide the 622 patients into training and testing sets at a 7:3 ratio.
- We fit the model using the training set to obtain the parameter estimates.
- For the testing set, we predict the RMST given the disease history up to each examination.
- For comparison, we redo the above steps using the msm package.

Prediction Error

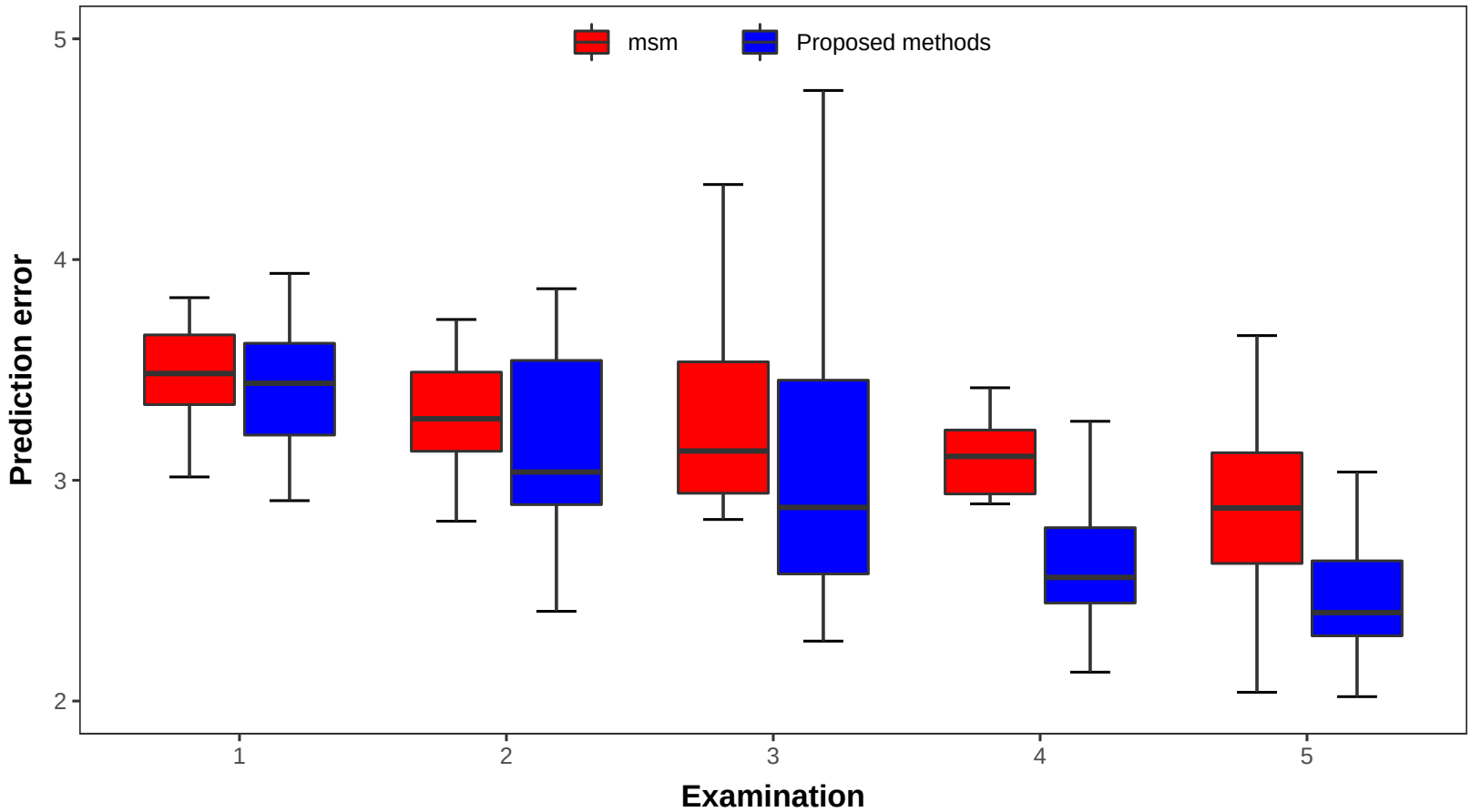


Figure 2: Boxplots of prediction error over 20 replicates at each examination.

Extension

- Our models can be extended to the competing risks set-up with more than one absorbing states, e.g., different causes of death.
- Joint modelling (with shared random effects) can be used when time-dependent covariates are measured only at a finite number of time points, or when there are more than one disease processes.
- Model diagnostics techniques, e.g., goodness-of-fit test.

Thank you!