Additive-multiplicative Hazards Regression Models for Interval-censored Semi-competing Risks Data with Missing Intermediate Events

Jinheum Kim[1], Jayoun Kim[2]
[1]University of Suwon, Department of Applied Statistics, South Korea,
[2]Medical Research Collaborating Center, Seoul National University Hospital, South Korea

ABSTRACT

We propose a multi-state model for analyzing semi-competing risks data with interval-censored or missing intermediate events. This model is an extension of the illness-death model (IDM), which includes three states: healthy, diseased, and dead. The diseased state can be regarded as the intermediate event. Two more states are added to the IDM to account for missing events, which are caused by a loss of follow-up before the end of the study; one is a state called the lost-to-follow-up (LTF) and the other is an unobservable state that represents an intermediate event experienced after the LTF occurred. We employ the additive and multiplicative hazards model with a log-normal frailty and construct the conditional likelihood to estimate the transition intensities among states in the multi-state model. Marginalization of the full likelihood is accomplished using adaptive importance sampling, and the optimal solution of the regression parameters is achieved through the iterative quasi-Newton algorithm. Simulation is performed to investigate the finite-sample performance of the proposed estimation method in terms of the relative bias and coverage probability of the regression parameters. Our proposed method is also illustrated using a dataset initially analyzed by Helmer et al. (2001).

CONTACT

Jinheum Kim
Email: jkimdt65@gmail.com

INTRODUCTION

In clinical trials, the occurrence of a non-fatal event can be detected in conjunction with possibly incessant monitoring during periodic follow-up. For illustration purposes of our methodologies, a dataset named PAQUID is analyzed to conduct a longitudinal study to investigate the meaningful prognostic factors associated with dementia. These data were initially analyzed by Helmer et al. (2001) using the conventional Cox (Cox, 1972) model. In this paper, we employ a semi-competing risks model where death may occur after dementia has occurred, but death censors the disease. As shown in the PAQUID data, dementia can be censored informatively by death. Furthermore, participants may be excluded from the study due to LTF. This makes it important to consider both cases; dementia is censored or not.

Most of non-fatal event times are not observed exactly but lies on an interval of the form $(L, R]$. We could emulate Barrett et al. (2011) and assume that a non-fatal event of a subject occurs uniformly on the interval $(L, R]$. However, using the methods proposed by Lindsey and Ryan (1998), we instead partition the interval $(L, R]$ into a few sub-intervals, in which a non-fatal event can occur.

In addition, we propose an additive-multiplicative model by combining the Cox (Cox, 1972) proportional hazards model with the additive risk model of Lin and Ying (1994), in accordance with a multi-state model.

MODELS AND PARAMETER ESTIMATIONS

As depicted in Figure 1, the proposed model consists of five states. Let t be the time from study entry. S_t is defined as the state that each subject can take at t.

$$\mathcal{A} = \{(0,1), (0,2), (0,3), (1,2), (3,2), (3,4), (4,2)\}.$$

Define $\lambda_{rs}(t)$ to be the transition intensity from states r to s at t,

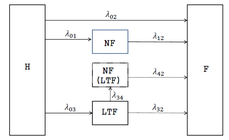
$$\lambda_{rs}(t) = \lim_{dt \to 0} \frac{Pr\,(S_{t+dt} = s | S_t = r)}{dt}, (r,s) \in \mathcal{A},$$

and $\lambda_{rs}(t) = 0,\ (r,s) \notin \mathcal{A}$. The data corresponding to transitions $3 \to 4$ and $4 \to 2$ are not observable, requiring the following assumptions for $\lambda_{34}(t)$ and $\lambda_{42}(t)$:

$$\lambda_{34}(t) = \lambda_{01}(t),$$
$$\lambda_{42}(t) = \lambda_{12}(t).$$

Given covariates $\boldsymbol{z}$ and $\boldsymbol{w}$, along with frailty u, we consider additive and multiplicative models defined as

Figure 1 : A multi-state model



$$\lambda_{rs}(t|\boldsymbol{z}, \boldsymbol{w}, u) = \eta(\boldsymbol{\beta}_{rs}{}'\boldsymbol{z} + \exp(\boldsymbol{\alpha}_{rs}{}'\boldsymbol{w})\theta_{rs}\gamma_{rs}t^{\gamma_{rs}-1}),$$

where $\eta = \exp(u)$ is a log-normal frailty and u follows a normal distribution $N(0, \sigma^2)$. The parameter vector estimated is $\boldsymbol{\zeta} = (\boldsymbol{\theta}^*, \boldsymbol{\gamma}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \sigma^2)'$.

A total of six routes can be experienced by a subject: route 1 ($0 \to 0$), route 2 ($0 \to 2$), route 3 ($0 \to 1$), route 4 ($0 \to 1 \to 2$), route 5 ($0 \to 3$), and route 6 ($0 \to 3 \to 2$). Therefore, likelihood functions Q_1 and Q_2 can be constructed for routes 1 and 2, respectively.

$$Q_{i1}(t_i) = \exp\{-H_0(0, t_i)\}.$$
$$Q_{i2}(t_i) = Q_{i1}(t_i)\lambda_{02}(t_i).$$

We can have a refined set of time points:

$$0 = s_0 < s_1 < s_2 < \cdots < s_l < s_{l+1} = \infty.$$

We can define the weight $w_{i'm}$ at s_m $(m = 1,2,\ldots,l)$ for subject i':

$$w_{i'm} = \frac{d_{i'm}\exp\{-H_0(0, s_m)\}\lambda_{01}(s_m)}{\sum_{m'=1}^{l} d_{i'm'}\exp\{-H_0(0, s_{m'})\}\lambda_{01}(s_{m'})}.$$

Likelihood functions can also be constructed for routes 3 and 4:

$$Q_{i3}(a_i, b_i, t_i) = \sum_{m=1}^{l} d_{im}\, w_{im}\exp\{-H_0(0, s_m)\}\lambda_{01}(s_m).$$
$$Q_{i4}(a_i, b_i, t_i) = Q_3(a_i, b_i, t_i)\lambda_{12}(t_i).$$

Finally, likelihood functions for routes 5 and 6 are given by

$$Q_{i5}(a_i, b_i, t_i) =$$
$$\exp\{-H_0(0, a_i)\}\lambda_{03}(a_i)[\exp\{-H_3(a_i, t_i)\}$$
$$+ \int_{a_i}^{t_i} \exp\{-H_3(a_i, s)\}\lambda_{34}(s)\exp\{-H_4(s, t_i)\}ds].$$

$$Q_{i6}(a_i, b_i, t_i) =$$
$$\exp\{-H_0(0, a_i)\}\lambda_{03}(a_i)\ [\exp\{-H_3(a_i, t_i)\}\lambda_{32}(t_i)$$
$$+ \left\{\int_{a_i}^{t_i} \exp\{-H_3(a_i,)\}\lambda_{34}(s)\exp\{-H_4(s, t_i)\}ds\right\}\lambda_{42}(t_i)].$$

Therefore, the likelihood function for $\boldsymbol{\zeta}$ is

$$L(\boldsymbol{\zeta}) = \prod_{i=1}^{n}\left\{\prod_{j=1}^{6} Q_{ij}^{I_{ij}}\right\}\phi(0, \sigma^2; u_i).$$

In our analysis, we use the NLMIXED procedure of the SAS software to estimate $\boldsymbol{\zeta}$. We define the marginal likelihood as

$$m(\boldsymbol{\zeta}) = \int\cdots\int L(\boldsymbol{\zeta})du_1\cdots du_n.$$

Then, we find the value of $\boldsymbol{\zeta}$ that minimizes $f(\boldsymbol{\zeta}) = -\log m(\boldsymbol{\zeta})$, which is referred to as $\hat{\boldsymbol{\zeta}}$. Consequently, the inverse of the Hessian matrix evaluated at $\hat{\boldsymbol{\zeta}}$ is defined as the estimated variance-covariance matrix of $\hat{\boldsymbol{\zeta}}$.

RESULTS

Table 1 : Empirical results for the averages of the relative bias (r.Bias) and standard errors (SEM), standard deviation (SD), and coverage probability (CP) for the regression parameters and variance parameter of the log-normal frailty based on the 'imputed-by-the-right-endpoint' and 'proposed' methods under three types of LTF proportions ('low', 'moderate', and 'high')

		low (LTF(%)=22.6)				moderate (LTF(%)=34.2)				high (LTF(%)=47.3)			
	true	r.Bias	SD	SEM	CP	r.Bias	SD	SEM	CP	r.Bias	SD	SEM	CP
parameter	value	(%)	($\times 10^5$)	($\times 10^5$)	(%)	(%)	($\times 10^5$)	($\times 10^5$)	(%)	(%)	($\times 10^5$)	($\times 10^5$)	(%)
imputed-by-the-right-endpoint method													
α_{01}	0.01	81.7	15026	15104	96.8	73.3	15105	14975	95.2	20.6	14450	15186	96.2
α_{02}	0.01	176.2	22297	22961	95.4	224.5	24408	24066	94.8	28.8	27643	27426	95.4
α_{03}	0.01	133.2	27006	26429	95.4	86.1	16223	17086	96.2	75.3	13291	13250	95.2
α_{12}	0.01	-145.7	28481	25849	95.2	-113.4	29996	26279	93.6	74.1	27309	25620	96.2
α_{32}	0.01	109.7	65499	55615	93.2	77.1	36087	33747	96.8	50.8	26680	25219	94.8
β_{01}	0.004	-10.1	81	87	93.8	-9.4	80	88	91.6	-8.6	87	88	92.6
β_{02}	0.004	1.1	92	92	94.6	-1.4	97	96	94.4	-0.4	103	101	94.0
β_{03}	0.004	-7.2	87	86	91.2	-6.4	105	105	93.4	-9.3	136	133	91.6
β_{12}	0.004	3.5	115	115	97.0	5.7	112	116	95.6	3.4	116	115	95.8
β_{32}	0.004	-0.3	217	199	96.8	7.6	174	173	96.4	5.3	148	152	97.2
σ^2	0.01	904.5	8164	8782	93.8	815.5	7848	8084	92.4	771.6	7626	7076	90.4
proposed method													
α_{01}	0.01	32.2	15236	15136	96.2	64.8	15344	15084	95.0	20.7	14496	15279	96.2
α_{02}	0.01	70.2	22692	22951	95.0	286.1	24010	24133	95.2	100.4	28030	27569	95.2
α_{03}	0.01	289.4	27525	26349	95.0	121.3	16740	17161	95.0	86.3	13465	13391	95.0
α_{12}	0.01	-24.6	26718	25661	95.8	-83.4	29355	26230	94.0	-37.0	27102	25736	96.2
α_{32}	0.01	123.2	67188	56122	93.6	93.4	36291	33577	95.6	54.0	26591	25204	94.2
β_{01}	0.004	-5.5	89	90	95.0	-5.0	85	91	93.8	-5.3	88	91	94.4
β_{02}	0.004	5.9	95	97	96.2	4.0	99	100	94.2	4.8	107	106	94.8
β_{03}	0.004	-1.7	92	91	93.2	-1.7	107	110	95.6	-3.3	141	139	93.4
β_{12}	0.004	-1.9	109	110	96.4	0.9	108	112	95.2	0.1	110	111	96.0
β_{32}	0.004	1.0	214	199	96.6	8.4	179	175	96.2	4.4	150	153	97.0
σ^2	0.01	1016.2	9069	9074	92.2	897.8	7743	8198	91.0	874.1	7616	7423	87.2

Table 3 : P-values of the test used to check the proportional hazard assumption for each transition model

	transition models				
covariate	$0 \to 1$	$0 \to 2$	$1 \to 2$	$0 \to 3$	$3 \to 2$
gender	0.063	< 0.001	0.093	0.354	0.062
cerificate	0.963	< 0.001	0.147	0.754	0.148

Table 2 : Sensitivity analysis of the 'proposed' method depending on the underlying frailty distribution in terms of the averages of the relative bias (r.Bias) as well as the standard errors (SEM) and coverage probability (CP) when the LTF proportion is 'moderate'

		$U(-0.173, 0.173)$				$DE(0.007)$				$G(100.5, 0.01)$			
		(LTF(%)=34.4)				(LTF(%)=34.5)				(LTF(%)=34.5)			
	true	r.Bias	SEM	CP		r.Bias	SEM	CP		r.Bias	SEM	CP	
parameter	value	(%)	($\times 10^5$)	(%)		(%)	($\times 10^5$)	(%)		(%)	($\times 10^5$)	(%)	
α_{01}	0.01	40.9	15169	94.2		36.7	15215	94.8		-28.7	15089	96.2	
α_{02}	0.01	-83.8	24639	97.0		89.0	24401	94.4		45.0	24623	96.6	
α_{03}	0.01	58.9	17266	93.8		-110.5	16925	93.8		119.9	17198	93.6	
α_{12}	0.01	-27.9	26141	94.8		27.1	26318	93.2		159.3	25682	95.2	
α_{32}	0.01	-80.1	33158	94.0		-22.5	33619	94.0		-82.4	33595	91.8	
β_{01}	0.004	-5.8	90	94.0		-3.3	91	94.0		-6.7	90	92.2	
β_{02}	0.004	4.9	100	94.8		5.4	101	97.0		4.4	99	95.4	
β_{03}	0.004	-2.9	109	91.0		-2.8	110	94.4		-2.9	108	91.8	
β_{12}	0.004	2.3	111	95.0		0.6	110	94.8		-0.3	111	95.6	
β_{32}	0.004	0.8	169	96.0		2.9	171	96.8		5.2	170	96.4	
σ^2	0.01	974.2	8325	89.4		926.1	8284	89.6		928.7	8462	89.2	

Table 4 : Regression parameter estimates (Est) with the accompanying standard errors (SE) and p-values (P)

covariate	parameter	Est	SE	P
gender	β_{01}	-7.70×10^{-4}	3.32×10^{-4}	0.027
	β_{02}	1.93×10^{-3}	5.35×10^{-4}	0.001
	β_{12}	0.0626	0.0339	0.074
	β_{03}	-5.50×10^{-4}	3.69×10^{-4}	0.144
	β_{32}	6.66×10^{-3}	0.0116	0.570
certificate	α_{01}	-0.264	0.163	0.115
	α_{02}	0.0251	0.131	0.850
	α_{12}	-0.300	0.296	0.319
	α_{03}	-0.0245	0.144	0.866
	α_{32}	0.101	0.201	0.620
	σ^2	0.00352	0.00598	0.560

CONCLUSIONS

The proposed model was illustrated using PAQUID data and yielded several promising results. The risk of transition from a healthy state to dementia is higher for women; however, the risk of death after being diagnosed with dementia is higher for men. Similar trends are found for non-diagnosed groups. Meanwhile, the risk of transition from a healthy state to dementia is higher for the educated group; the risk of death after being diagnosed with dementia is also higher for this group. There is, however, a reversed result for non-diagnosed targets.

Furthermore, we conducted simulations with finite-sample sizes to investigate the efficiency of the proposed estimators. In particular, we considered three different types of LTF proportions. In general, the coverage probabilities of the regression parameters are close to a nominal level of 0.95 in most cases. The proposed estimators turned out to be robust in terms of the misspecification of frailty distributions.

REFERENCES

Barrett J K, Siannis F, Farewell VT. (2011). A semi-competing risks model for data with interval-censoring and informative observation: An application to the MRC cognitive function and ageing study. Statistics in Medicine, **30**, 1-10.