Regression models for interval-censored semi-competing risks data with missing intermediate transition status

Jinheum Kim[1] & Jayoun Kim[2]

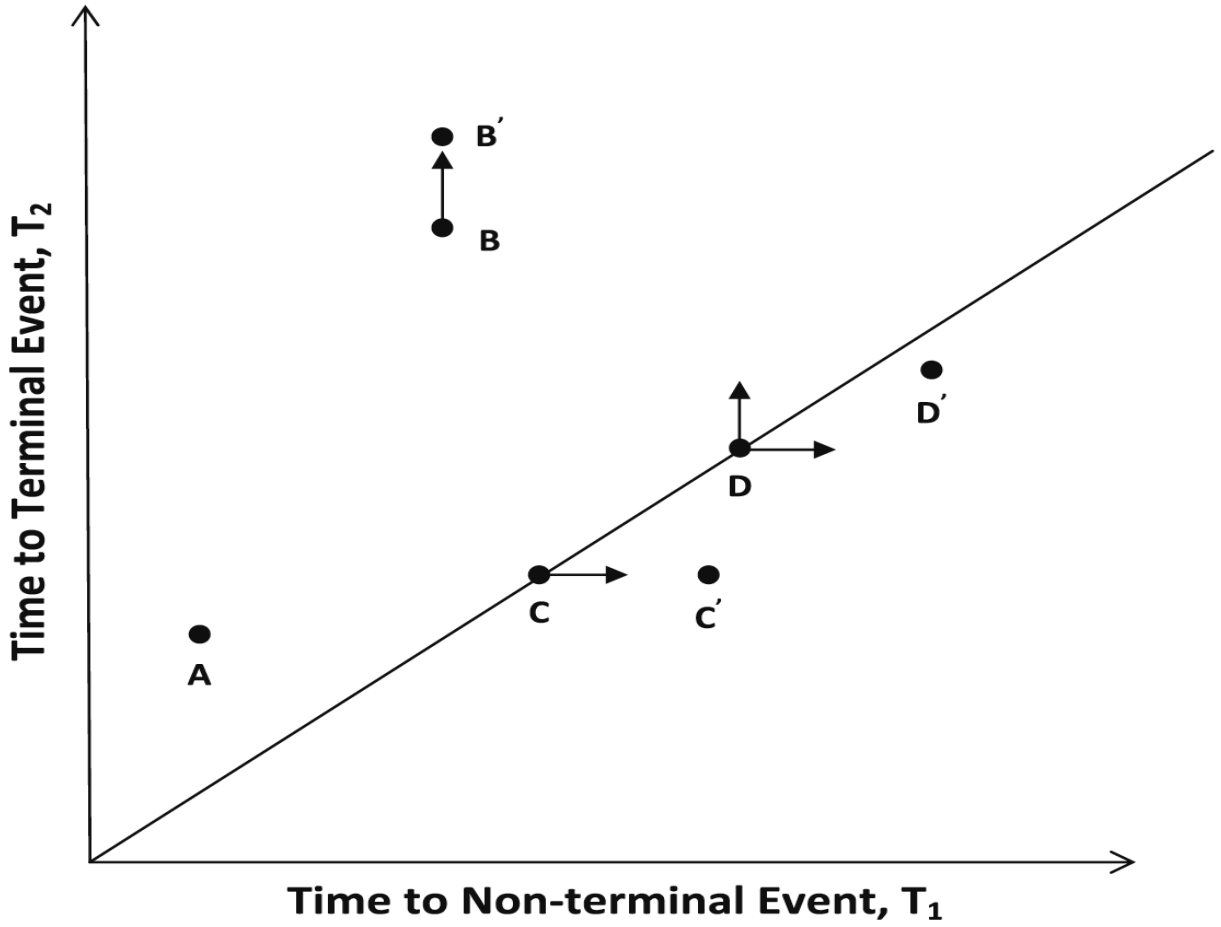
[1]Department of Applied Statistics, University of Suwon
[2]Research Coordinating Center, Konkuk University Medical Center

November 5, 2016

Outline

- Semi-competing risks data? Interval censoring?
- Model specification
- Likelihood construction
- Simulation studies
- Illustrative example: PAQUID data
- Concluding remarks

Representation on semi-competing risks data



PAQUID cohort: Interval censoring

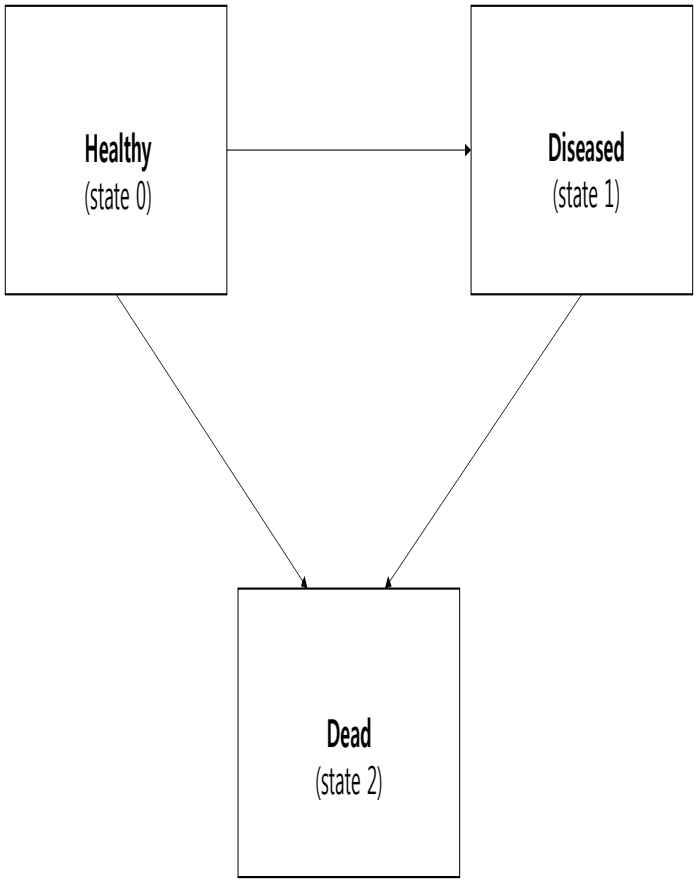
Table 1 Age in years at follow-up visits and diagnosis of dementia for four selected subjects of the PAQUID cohort, France, 1988–2010. Illustration of the possible configurations in the data with respect to the dementia and death status

			Follow-up visits											
	Entry		1 year		3 years		5 years		8 years		10 years		**Latest follow-up**	
Subject	Age	Dem	Age	Dem	Age	Dem	Age	Dem	Age	Dem	Age	Dem	Age	Dead[a]
Never diagnosed with dementia and alive at the latest follow-up ($n = 545$, i.e. 15% of subjects in the PAQUID study)														
A	66.5	0	N/A	N/A	70.5	0	N/A	N/A	74.4	0	76.8	0	88.2	0
Never diagnosed with dementia and died ($n = 2298$, i.e. 63% of subjects in the PAQUID study)														
B	70.4	0	71.5	0	73.6	0	N/A	N/A	N/A	N/A	N/A	N/A	85.2	1
Diagnosed with dementia and alive at the latest follow-up ($n = 193$, i.e. 5% of subjects in the PAQUID study)														
C	73.1	0	N/A	N/A	76.2	0	78.2	0	80.8	0	83.1	1	93.7	0
Diagnosed with dementia and died ($n = 639$, i.e. 17% of subjects in the PAQUID study)														
D	78.3	0	N/A	N/A	N/A	N/A	83.4	0	86.9	0	88.5	0	99.0	1

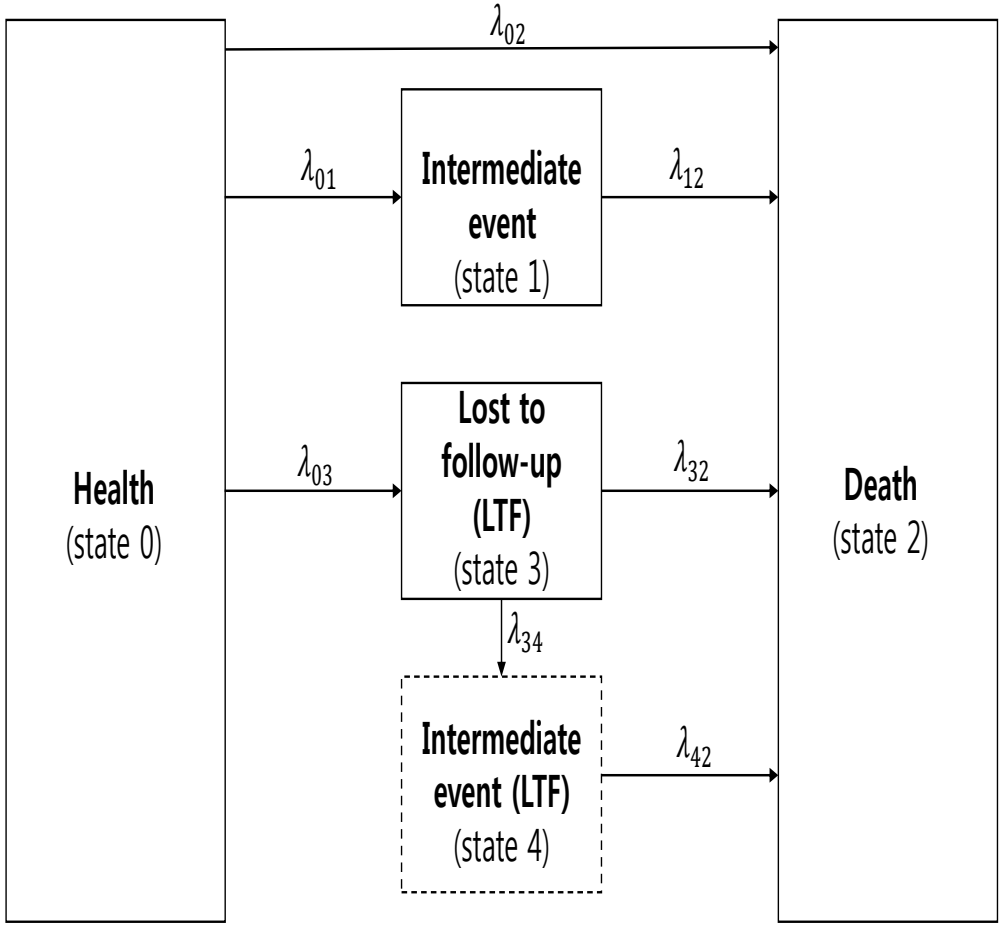
Dem, indicator of dementia (1 for yes, 0 for no); N/A, missing data because the subject missed the follow-up visit, or not applicable because the subject died earlier.

[a]Indicator of vital status at the latest news (1 for dead, 0 for alive).

Illness-death model *vs* multi-state model with LTF state



(a)



(b)

Previous works

- Leffondré et al. (2013): a semi-parametric illness-death model
- Frydman & Szarek (2009): nonparametric ML estimation
- Siannis et al. (2007) & Barrett et al. (2011): a multi-sate model with an unobserved state

Five-states model

- State 0: the health state
- State 1: a state to represent an intermediate event (IE)
- State 2: the terminal state (absorbing state)
- State 3: a state to represent loss to follow-up (LTF) for the intermediate process
- State 4: the unobservable state that represents an intermediate event experienced after the subject is LTF
- cf. Siannis et al. (2007)

Transition intensities

- $S = \{S(t),\ t \geq 0\}$: a multi-state process
  - $S(t) \in \{0, 1, 2, 3, 4\}$
- The intensity of a transition from state r to state s at time t is defined as

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

  - $\mathcal{A} = \{(r, s) : (r, s) = (0, 1), (0, 2), (0, 3), (1, 2), (3, 2), (3, 4), (4, 2)\}$
  - $\lambda_{rs}(t) = 0,\ (r, s) \notin \mathcal{A}$
  - $\lambda_{34}(t)$, $\lambda_{42}(t)$, and $\lambda_{32}(t)$: NOT identifiable $\Rightarrow$ need assumptions on both $\lambda_{34}(t)$ and $\lambda_{42}(t)$

Constraints on $\lambda_{34}(t)$ and $\lambda_{42}(t)$

- $\frac{\lambda_{02}(t)}{\lambda_{01}(t)} = r\frac{\lambda_{32}(t)}{\lambda_{34}(t)}, \ r > 0$
  - No information in the data concerning the value of r
- $\lambda_{42}(t) = \lambda_{12}(t)$
- cf. Siannis et al. (2007), Barret et al. (2011)

Model

- Given $\mathbf{x}$ and u, the transition intensity of $r \to s$ is assumed to be

$$\lambda_{kl}(t|\mathbf{x}, u) = \alpha_{kl}\gamma_{kl}t^{\gamma_{kl}-1}\exp(\boldsymbol{\beta}'_{kl}\mathbf{x} + u),\ (r,s) \in \mathcal{A}$$

  - $\alpha_{kl}(\gamma_{kl})$: the scale (shape) parameter of the weibull distribution
  - $\boldsymbol{\beta}_{kl}$: the vector of regression coefficients
  - u : an unobservable frailty, $u \sim N(0, \sigma^2)$
- $\boldsymbol{\zeta} = (\boldsymbol{\alpha}', \boldsymbol{\gamma}', \boldsymbol{\beta}'_{01}, \boldsymbol{\beta}'_{02}, \boldsymbol{\beta}'_{03}, \boldsymbol{\beta}'_{12}, \boldsymbol{\beta}'_{32}, \sigma^2)'$: the vector of parameters to be estimated
  - $\boldsymbol{\alpha} = (\alpha_{01}, \alpha_{02}, \alpha_{03}, \alpha_{12}, \alpha_{32})'$, $\boldsymbol{\gamma} = (\gamma_{01}, \gamma_{02}, \gamma_{03}, \gamma_{12}, \gamma_{32})'$
  - $\alpha_{34}, \alpha_{42}, \gamma_{34}, \gamma_{42}, \boldsymbol{\beta}_{34}, \boldsymbol{\beta}_{42}$: deterministic from the two constraints

Cumulative transition intensity functions

- The cumulative transition functions for leaving state 0, 1, 3, and 4 between t_1 and t_2 are given by, respectively,

$$
\begin{aligned}
H_0(t_1, t_2|\mathbf{x}, u) &= \int_{t_1}^{t_2} \{\lambda_{01}(s|\mathbf{x}, u) + \lambda_{02}(s|\mathbf{x}, u) + \lambda_{03}(s|\mathbf{x}, u)\} ds \\
&= \sum_{r=1}^{3} \alpha_{0r}(t_2^{\gamma_{0r}} - t_1^{\gamma_{0r}}) \exp(\boldsymbol{\beta}_{0r}'\mathbf{x} + u), \\
H_1(t_1, t_2|\mathbf{x}, u) &= \int_{t_1}^{t_2} \lambda_{12}(s|\mathbf{x}, u) ds = \alpha_{12}(t_2^{\gamma_{12}} - t_1^{\gamma_{12}}) \exp(\boldsymbol{\beta}_{12}'\mathbf{x} + u), \\
H_3(t_1, t_2|\mathbf{x}, u) &= \int_{t_1}^{t_2} \{\lambda_{32}(s|\mathbf{x}, u) + \lambda_{34}(s|\mathbf{x}, u)\} ds \\
&= \alpha_{32}(t_2^{\gamma_{32}} - t_1^{\gamma_{32}}) \exp(\boldsymbol{\beta}_{32}'\mathbf{x} + u) + \alpha_{34}(t_2^{\gamma_{34}} - t_1^{\gamma_{34}}) \exp(\boldsymbol{\beta}_{34}'\mathbf{x} + u), \\
H_4(t_1, t_2|\mathbf{x}, u) &= \int_{t_1}^{t_2} \lambda_{42}(s|\mathbf{x}, u) ds = \alpha_{42}(t_2^{\gamma_{42}} - t_1^{\gamma_{42}}) \exp(\boldsymbol{\beta}_{42}'\mathbf{x} + u)
\end{aligned}
$$

Notation

- R : time to an IE
- T : time to terminal event
- L : time to LTF
- C : censoring time
- $\mathcal{H}_0(s) = \{R \wedge L \wedge T > s\}$: the corresponding history to a subject who is in state 0 at time s
- $\mathcal{H}_{3,f}(s) = \{L = f, R \wedge T > s, f \leq s\}$: the corresponding history to a subject whose LTF have occurred at time f and who is in state 3 at time s

Six routes

- Route 1: 0→0; Route 2: 0→2; Route 3: 0→1; Route 4: 0→1→2; Route 5: 0→3; Route 6: 0→3→2
- I_{ij} $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, 6)$: an indicator function for subject i taking route j
- $\mathcal{B}_j = \{i : I_{ij} = 1\}$: a set of subjects taking route j

Observed data

- a_i : time at the visit before the diagnostic visit for subject i
- b_i : time at the diagnostic visit for subject i
- t_i : time at death or censoring for subject i
- Note that
  - $i \in \mathcal{B}_1 \cup \mathcal{B}_2 \Rightarrow a_i, b_i \geq t_i$
  - $i \in \mathcal{B}_3 \cup \mathcal{B}_4 \Rightarrow a_i < b_i \leq t_i$
  - $i \in \mathcal{B}_5 \cup \mathcal{B}_6 \Rightarrow a_i < t_i$, but $b_i < t_i$ or $b_i \geq t_i$
  - $i \in \mathcal{B}_1 \cup \mathcal{B}_3 \cup \mathcal{B}_5 \Rightarrow t_i$: time at censoring
  - $i \in \mathcal{B}_2 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \Rightarrow t_i$: time at death

Likelihood

- For route 1 ($i \in \mathcal{B}_1$),

$$\begin{aligned} Q_{i1}(t_i|\mathbf{x}_i, u_i) &= \Pr(R_i \wedge L_i \wedge T_i > t_i|\mathcal{H}_0(0), \mathbf{x}_i, u_i) \\ &= \exp\{-H_0(0, t_i|\mathbf{x}_i, u_i)\} \end{aligned}$$

- For route 2 ($i \in \mathcal{B}_2$),

$$\begin{aligned} Q_{i2}(t_i|\mathbf{x}_i, u_i) &= \Pr(T = t_i, R \wedge L > t_i|\mathcal{H}_0(0), \mathbf{x}_i, u_i) \\ &= Q_{i1}(t_i|\mathbf{x}_i, u_i) q_{02}(t_i|\mathbf{x}_i, u_i) \end{aligned}$$

Likelihood

- Using the algorithm proposed by Collett (2015), define a subset of the endpoints of $(a_i, b_i]$, $i \in \mathcal{B}_3 \cup \mathcal{B}_4$, as

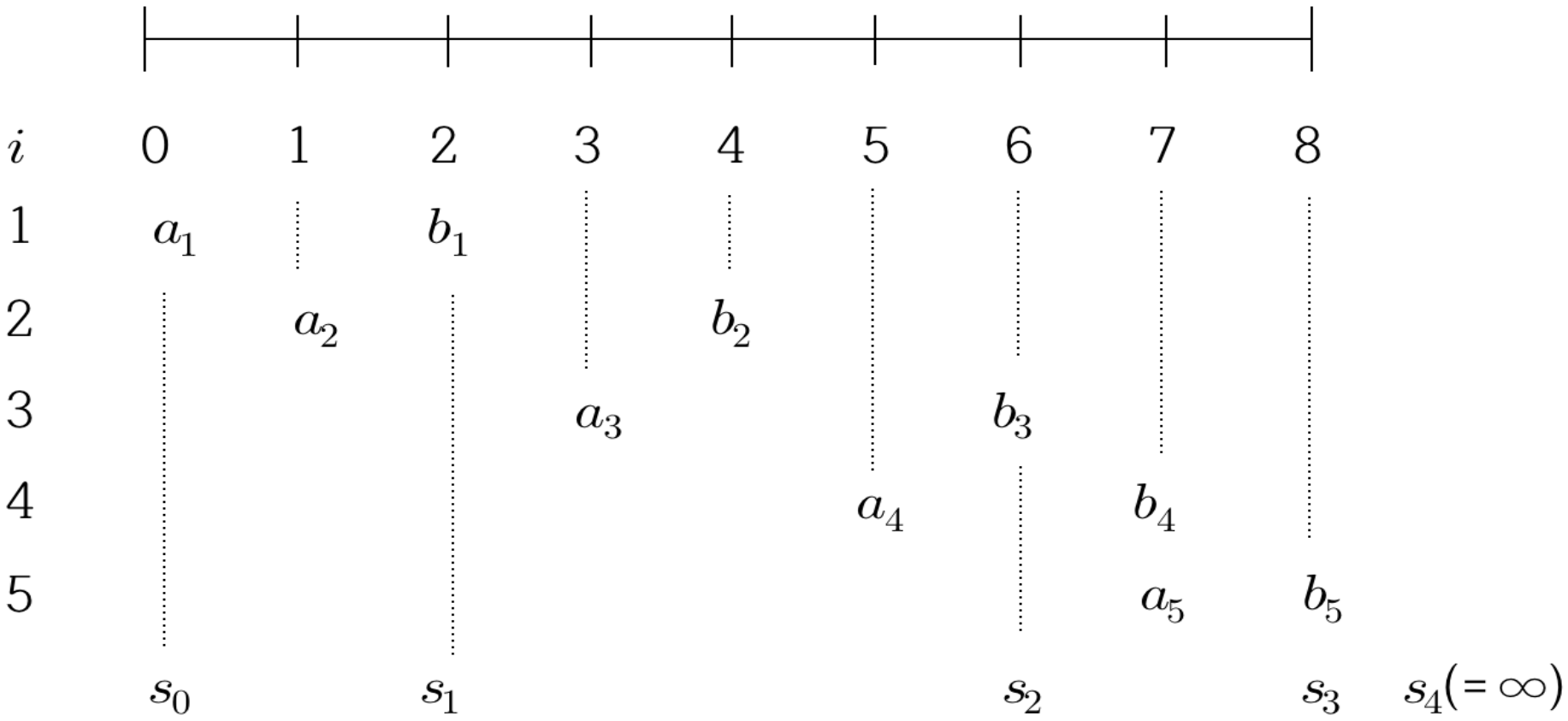
$$0 = s_0 < s_1 < s_2 < \cdots < s_l < s_{l+1} = \infty$$

  - s_1 : the smallest of the values of b_i
  - s_m $(m = 2, \ldots, l)$: the smallest of the values of b_i such that $a_i \geq s_{m-1}$
- Define the weight at time s_m for subject i as

$$w_{im} = \frac{d_{im} \exp\left\{-H_0(0, s_m|\mathbf{x}_i, u_i)\right\} \lambda_{01}(s_m|\mathbf{x}_i, u_i)}{\sum_{m'=1}^{l} d_{im'} \exp\left\{-H_0(0, s_{m'}|\mathbf{x}_i, u_i)\right\} \lambda_{01}(s'_m|\mathbf{x}_i, u_i)}$$

  - $d_{im} = I(s_m \in (a_i, b_i])$

Likelihood


i
0 1 2 3 4 5 6 7 8
1
2
3
4
5
a_1 b_1
a_2 b_2
a_3 b_3
a_4 b_4
a_5 b_5
s_0 s_1 s_2 s_3 $s_4(=\infty)$

Likelihood

- For route 3 ($i \in \mathcal{B}_3$),

$$
\begin{aligned}
Q_{i3}(a_i, b_i, t_i|\mathbf{x}_i, u_i) &= \Pr(R_i \in (a_i, b_i], L_i > t_i, T_i > t_i|\mathcal{H}_0(0), \mathbf{x}_i, u_i) \\
&= \exp\{-H_0(0, a_i|\mathbf{x}_i, u_i)\} \\
&\quad \times \sum_{m=1}^{l} \Big[d_{im} w_{im} \exp\{-H_0(a_i, s_m|\mathbf{x}_i, u_i)\} q_{01}(s_m|\mathbf{x}_i, u_i) \\
&\quad \times \exp\{-H_1(s_m, b_i|\mathbf{x}_i, u_i)\} \Big] \times \exp\{-H_1(b_i, t_i|\mathbf{x}_i, u_i)\} \\
&= \sum_{m=1}^{l} \Big[d_{im} w_{im} \exp\{-H_0(0, s_m|\mathbf{x}_i, u_i)\} q_{01}(s_m|\mathbf{x}_i, u_i) \\
&\quad \times \exp\{-H_1(s_m, t_i|\mathbf{x}_i, u_i)\} \Big]
\end{aligned}
$$

- For route 4 ($i \in \mathcal{B}_4$),

$$
\begin{aligned}
Q_{i4}(a_i, b_i, t_i|\mathbf{x}_i, u_i) &= \Pr(R_i \in (a_i, b_i], L_i > t_i, R_i < T_i = t_i|\mathcal{H}_0(0), \mathbf{x}_i, u_i) \\
&= Q_3(a_i, b_i, t_i|\mathbf{x}_i, u_i) q_{12}(t_i|\mathbf{x}_i, u_i)
\end{aligned}
$$

Likelihood

- For route 5 ($i \in \mathcal{B}_5$),

$$
\begin{aligned}
Q_{i5}(a_i, b_i, t_i|\mathbf{x}_i, u_i) &= \Pr(R_i \wedge T_i > t_i|\mathcal{H}_{3,a_i}(a_i), \mathbf{x}_i, u_i) \\
&\quad + \Pr(R_i \in (a_i, t_i], T_i > t_i|\mathcal{H}_{3,a_i}(a_i), \mathbf{x}_i, u_i) \\
&= \exp\{-H_0(0, a_i|\mathbf{x}_i, u_i)\} q_{03}(a_i|\mathbf{x}_i, u_i) \Big[\exp\{-H_3(a_i, t_i|\mathbf{x}_i, u_i)\} \\
&\quad + \int_{a_i}^{t_i} \exp\{-H_3(a_i, s|\mathbf{x}_i, u_i)\} q_{34}(s|\mathbf{x}_i, u_i) \exp\{-H_4(s, t_i|\mathbf{x}_i, u_i)\} ds \Big]
\end{aligned}
$$

- For route 6 ($i \in \mathcal{B}_6$),

$$
\begin{aligned}
Q_{i6}(a_i, b_i, t_i|\mathbf{x}_i, u_i) &= \Pr(R_i > T_i, T_i = t_i|\mathcal{H}_{3,a_i}(a_i), \mathbf{x}_i, u_i) \\
&\quad + \Pr(R_i \in (a_i, t_i], R_i < T_i = t_i|\mathcal{H}_{3,a_i}(a_i), \mathbf{x}_i, u_i) \\
&= \exp\{-H_1(0, a_i|\mathbf{x}_i, u_i)\} q_{03}(a_i|\mathbf{x}_i, u_i) \\
&\quad \Big[\exp\{-H_3(a_i, t_i|\mathbf{x}_i, u_i)\} q_{32}(t_i|\mathbf{x}_i, u_i) \\
&\quad + \Big\{ \int_{a_i}^{t_i} \exp\{-H_3(a_i, s|\mathbf{x}_i, u_i)\} q_{34}(s|\mathbf{x}_i, u_i) \exp\{-H_4(s, t_i|\mathbf{x}_i, u_i)\} ds \Big\} \\
&\quad \times q_{42}(t_i|\mathbf{x}_i, u_i) \Big]
\end{aligned}
$$

Likelihood

- Based on the complete data, the likelihood function is defined as

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

  - $\phi(\cdot)$: *pdf* of $N(0, \sigma^2)$

Parameter estimation

- Use the NLMIXED procedure (SAS Institute Inc., 2015)
- Using the adaptive Gaussian quadrature proposed by Pinheiro & Bates (1995), compute the marginal likelihood,

$$m(\zeta) = \int \cdots \int L(\zeta) du_1 \cdots du_n$$

  - $\hat{\zeta}$: a minimizder of $f(\zeta) = -\log m(\zeta)$ using the Newton-Raphson algorithm
  - $\mathrm{Var}(\hat{\zeta})$: the inverse of Hessian matrix evaluated at $\hat{\zeta}$

Simulation setup

- IE was assessed at
  - Scenario 1: 15, 31,. . ., 349, 365(days) with $\pm N(0, 3^2)$ days
  - Scenario 2: 7, 15, 22, 31,. . ., 334, 36(days) with $\pm N(0, 3^2)$ days
- Exponential baseline hazards with $\gamma_{rs} = 1$
  - Scale parameters(except α_{03}): $\alpha_{01} = \alpha_{02} = \alpha_{32} = 0.002,\ \alpha_{12} = 0.001$
  - For α_{03}, 0.001 (LTF%: low); 0.002 (LTF%: medium); 0.008(LTF%: high)
- x : 0-1 binary covariate with a success probability of 0.5
  - Regression parameters: $\beta_{01} = \beta_{02} = \beta_{03} = \beta_{12} = \beta_{32} = 1$
- u : $N(0, 1)$ frailty
- Fixed censoring at 365(days)
- r : 1 for the first constraint
- Compare the proposed method with both FDA (FDA, 2007; EMEA, 2006) and naive methods

Simulated data generation

- Step 1: generate T_{01}, T_{02}, and T_{03}
  - T_{01} : a solution to $\Lambda_{01}(t) + \ln(1 - U_{01}) = 0$ *wrt* t, $U_{01} \sim U[0, 1]$
  - T_{02} : a solution to $\Lambda_{02}(t) + \ln(1 - U_{02}) = 0$ *wrt* t, $U_{02} \sim U[0, 1]$
  - T_{03} : a solution to $\Lambda_{03}(t) + \ln(1 - U_{03}) = 0$ *wrt* t, $U_{03} \sim U[0, 1]$
  - If $C \leq T_{01} \wedge T_{02} \wedge T_{03}$, censored w/o being relapsed; stop $\Rightarrow$ Route 1
  - If $T_{01} = T_{01} \wedge T_{02} \wedge T_{03}$, relapsed; goto Step 2
  - If $T_{02} = T_{01} \wedge T_{02} \wedge T_{03}$, dead w/o being relapsed; stop $\Rightarrow$ Route 2
  - If $T_{03} = T_{01} \wedge T_{02} \wedge T_{03}$, LTF; goto Step 3
- Step 2: generate T_{12}
  - T_{12} : a solution to $\Lambda_{12}(t) + \ln(1 - U_{12}) = 0$ *wrt* t, $U_{12} \sim U[1 - \exp\{\Lambda_{01}(T_{01})\}, 1]$
  - If $C = T_{12} \wedge C$, censored w/ being relapsed; stop $\Rightarrow$ Route 3
  - If $T_{12} = T_{12} \wedge C$, dead w/ being relapsed; stop $\Rightarrow$ Route 4

Simulated data generation

- Step 3: genertae T_{32} and T_{34}
  - T_{32} : a solution to $\Lambda_{32}(t) + \ln(1 - U_{32}) = 0$ *wrt* t, $U_{32} \sim U[1 - \exp\{-\Lambda_{32}(T_{03})\}, 1]$
  - T_{34} : a solution to $\Lambda_{34}(t) + \ln(1 - U_{34}) = 0$ *wrt* t, $U_{34} \sim U[1 - \exp\{-\Lambda_{34}(T_{03})\}, 1]$
  - If $C \leq T_{32} \wedge T_{34}$, censored after LTF w/ being relapsed; stop $\Rightarrow$ Route 5
  - If $T_{32} = T_{32} \wedge T_{34}$, dead after LTF w/ being relapsed; stop $\Rightarrow$ Route 6
  - If $T_{34} = T_{32} \wedge T_{34}$, relapsed after LTF; goto Step 4
- Step 4: generate T_{42}
  - T_{42} : a solution to $\Lambda_{42}(t) + \ln(1 - U_{42}) = 0$ *wrt* t, $U_{42} \sim U[1 - \exp\{-\Lambda_{42}(T_{34})\}, 1]$
  - If $C = T_{42} \wedge C$, censored after LTF w/ being relapsed; stop $\Rightarrow$ Route 5
  - If $T_{42} = T_{42} \wedge C$, dead after LTF w/ being relapsed; stop $\Rightarrow$ Route 6

Simulation results: Unequal spacing visits

TABLE 1: The averages of bias(Bias) and standard error(SEM) and the coverage probability(CP) based on 500 replication with the sample size of 200 when the intermediate event is monitored with an unequal time interval

%		Proposed			FDA			Naive		
LTF	$\boldsymbol{\theta}$	Bias	SEM	CP	Bias	SEM	CP	Bias	SEM	CP
30	β_{01}	0.032	0.301	0.956	-0.107	0.332	0.922	-0.078	0.339	0.928
(13/17)	β_{02}	0.003	0.296	0.960	0.021	0.279	0.952	0.046	0.288	0.954
	β_{12}	0.038	0.443	0.958	0.048	0.548	0.958	0.097	0.561	0.968
	β_{03}	0.016	0.312	0.960						
	β_{32}	0.012	0.439	0.942						
	α_{01}	-1.17×10^{-4}	4.11×10^{-4}	0.912	-5.57×10^{-4}	3.45×10^{-4}	0.558	-7.41×10^{-4}	3.06×10^{-4}	0.328
	α_{02}	1.70×10^{-5}	4.31×10^{-4}	0.938	9.63×10^{-6}	4.08×10^{-4}	0.942	-8.64×10^{-6}	4.19×10^{-4}	0.944
	α_{12}	7.99×10^{-6}	3.58×10^{-4}	0.922	-2.24×10^{-4}	3.42×10^{-4}	0.756	-3.13×10^{-4}	3.08×10^{-4}	0.676
	α_{03}	2.22×10^{-5}	4.53×10^{-4}	0.968						
	α_{32}	1.36×10^{-4}	7.11×10^{-4}	0.950						
	σ^2	0.012	0.236	0.944	0.219	0.352	0.950	0.470	0.379	0.850

Sensitivity analysis: Choice of r

TABLE 2: The average of bias(Bias) and the coverage probability(CP) based on 500 replication with the sample size of 200 when the proportion of lost-to-follow-up is around 30%

	r							
	0.2		0.5		2		5	
$\boldsymbol{\theta}$	Bias	CP	Bias	CP	Bias	CP	Bias	CP
β_{01}	-0.038	0.962	-0.003	0.966	0.011	0.946	-0.028	0.926
β_{02}	-0.008	0.954	-0.015	0.954	0.036	0.946	-0.009	0.952
β_{12}	-0.015	0.958	0.009	0.958	0.021	0.952	-0.002	0.960
β_{03}	-0.050	0.938	-0.015	0.974	0.001	0.946	0.001	0.948
β_{32}	-0.019	0.960	-0.041	0.938	0.008	0.950	0.003	0.954
α_{01}	-7.40×10^{-5}	0.920	-9.36×10^{-5}	0.920	-1.08×10^{-4}	0.908	-5.35×10^{-5}	0.918
α_{02}	-7.31×10^{-6}	0.958	5.07×10^{-5}	0.936	-1.71×10^{-5}	0.928	1.85×10^{-5}	0.946
α_{12}	4.76×10^{-5}	0.924	5.53×10^{-5}	0.962	3.21×10^{-5}	0.928	5.28×10^{-5}	0.962
α_{03}	6.72×10^{-5}	0.944	4.79×10^{-5}	0.924	3.96×10^{-6}	0.926	-6.21×10^{-6}	0.938
α_{32}	1.33×10^{-4}	0.938	1.87×10^{-4}	0.942	1.75×10^{-4}	0.950	2.21×10^{-4}	0.962
σ^2	-0.071	0.916	-0.080	0.914	-0.057	0.900	-0.064	0.902

Sensitivity analysis: Misspecification of frailty distribution

TABLE 3: The average of bias(Bias) and the coverage probability(CP) based on 500 replication with the sample size of 200 when the proportion of lost-to-follow-up is around 30%

	$N(0,2)$			$t(4)$			$U(-2.45,2.45)$			$G(0.877,2.23)$		
θ	Bias	SEM	CP	Bias	SEM	CP	Bias	SEM	CP	Bias	SEM	CP
β_{01}	-0.059	0.335	0.958	-0.003	0.326	0.936	-0.043	0.343	0.952	-0.089	0.316	0.948
β_{02}	-0.007	0.328	0.950	0.007	0.320	0.946	-0.020	0.337	0.936	-0.053	0.310	0.944
β_{12}	-0.054	0.461	0.956	0.046	0.466	0.952	-0.039	0.463	0.958	0.015	0.429	0.958
β_{03}	-0.007	0.345	0.940	-0.009	0.336	0.942	0.010	0.354	0.948	-0.043	0.327	0.960
β_{32}	0.015	0.482	0.946	0.065	0.470	0.960	-0.072	0.496	0.934	-0.047	0.452	0.948
α_{01}	-3.81 $\times 10^{-5}$	4.72 $\times 10^{-4}$	0.928	-8.82 $\times 10^{-5}$	4.47 $\times 10^{-4}$	0.906	-1.15 $\times 10^{-4}$	4.65 $\times 10^{-4}$	0.898	9.95 $\times 10^{-5}$	4.75 $\times 10^{-4}$	0.958
α_{02}	8.14 $\times 10^{-5}$	4.94 $\times 10^{-4}$	0.930	2.66 $\times 10^{-5}$	4.68 $\times 10^{-4}$	0.952	1.25 $\times 10^{-5}$	4.89 $\times 10^{-4}$	0.920	1.90 $\times 10^{-4}$	4.90 $\times 10^{-4}$	0.966
α_{12}	8.81 $\times 10^{-5}$	3.88 $\times 10^{-4}$	0.944	-3.42 $\times 10^{-5}$	3.54 $\times 10^{-4}$	0.890	1.01 $\times 10^{-4}$	3.90 $\times 10^{-4}$	0.946	1.54 $\times 10^{-4}$	3.88 $\times 10^{-4}$	0.974
α_{03}	4.86 $\times 10^{-5}$	5.09 $\times 10^{-4}$	0.960	4.63 $\times 10^{-5}$	4.91 $\times 10^{-4}$	0.944	-4.72 $\times 10^{-5}$	4.98 $\times 10^{-4}$	0.920	2.01 $\times 10^{-4}$	5.16 $\times 10^{-4}$	0.964
α_{32}	1.95 $\times 10^{-4}$	7.98 $\times 10^{-4}$	0.942	1.20 $\times 10^{-4}$	7.50 $\times 10^{-4}$	0.954	2.81 $\times 10^{-4}$	8.53 $\times 10^{-4}$	0.954	5.83 $\times 10^{-4}$	8.75 $\times 10^{-4}$	0.970
σ^2	-0.147	0.355	0.878	-0.308	0.360	0.588	0.040	0.377	0.954	-0.621	0.280	0.402

Illustrative example: PAQUID data set

- Personnes Ages Quid (PAQUID): Helmer et al. (2001)
  - A prospective cohort study to investigate the impact of dementia on the risk of death
  - 3675 subjects aged 65 years or more, living in southwestern France, were recruited in 1988-90 and then screened for dementia every 2 or 3 years
- Instead analyze the 'Paq1000' data set included in the R package 'SmoothHazard'
- If time difference between the last visit and the latest follow-up of a subject is greater than 4 years, the subject is defined as being lost to follow-up since the last vist
  - 231 were LTF including 159 (68.8%) who died
  - Among 186 who were diagnosed with dementia, 127 (68.3%) died
  - Among 583 who were never diagnosed with dementia, 438 (75.1%) died
- Covariates: sex: 0=female, 1=male; primary school certificate: 0=w/ , 1=w/o certificate

Analysis of Paq1000 data set

TABLE 4: Regression parameter estimates(Est), their standard errors(SE), and p-values(P) for three estimation procedures

		Proposed			FDA			Naive		
Covariate	θ	Est	SE	P	Est	SE	P	Est	SE	P
Sex	β_{011}	-0.387	0.156	0.019	-0.376	0.158	0.023	-0.372	0.158	0.024
	β_{021}	0.371	0.097	< 0.001	0.224	0.083	0.011	0.224	0.083	0.011
	β_{121}	0.365	0.183	0.055	0.404	0.193	0.044	0.405	0.193	0.044
	β_{031}	-0.200	0.137	0.155						
	β_{321}	0.190	0.177	0.292						
Certificate	β_{012}	-0.366	0.194	0.068	-0.357	0.195	0.076	-0.357	0.195	0.077
	β_{022}	0.025	0.113	0.827	8.80×10^{-3}	0.097	0.928	9.36×10^{-3}	0.097	0.924
	β_{122}	-0.260	0.239	0.284	-0.298	0.256	0.253	-0.297	0.256	0.254
	β_{032}	-0.015	0.157	0.922						
	β_{322}	0.202	0.206	0.334						
	σ^2	4.55×10^{-3}	6.19×10^{-3}		6.10×10^{-3}	8.24×10^{-3}		6.59×10^{-3}	8.41×10^{-3}	

Concluding remarks

- Propose a Cox-type transition intensity model with a frailty to analyze the semi-competing risks data with a missing intermediate event
- Based on the simulation results, the proposed method satisfied the nominal level of CP
  - However, based on both FDA and naive methods, CPs of the regression parameter corresponding to the transition of $0 \rightarrow 1$ were less than 0.95 as the percentage of LTF increases
  - Those of the transition of $1 \rightarrow 2$ are greater than 0.95
- Moreover the proposed method was robust to the choice of r and a misspecification of the underlying frailty distribution

Thank you!