Presented at BayesComp 2025, Singapore, June 15–20

Bayesian analysis of clustered data within a semi-competing risks framework*

Jinheum Kim[1], Eunjeong Cha[2]

[1]University of Suwon, Department of Applied Statistics, South Korea; [2]University of Suwon, Department of Nursing Science, South Korea

Background

Traditional survival models typically focus on a single terminal event. However, clinical settings often involve more complex outcome structures. In *competing risks*, individuals face mutually exclusive events, while *semi-competing risks* arise when a non-terminal event (e.g., relapse) may be censored by a terminal event (e.g., death), which cannot be preceded by the non-terminal event—common in cancer studies.

Survival data are often clustered (e.g., by clinical center), requiring methods that account for intra-cluster correlation. Bayesian models provide a flexible framework to model both semi-competing risks and clustering. This study applies such an approach to multi-center breast cancer trial data.

Materials

- **Data source**: We used data from clinical trials involving 5,715 breast cancer patients across 36 centers in the US and Europe. The dataset, originally published by Haibe-Kains et al. (2012) and later reanalyzed by Peng et al. (2018), includes event times for relapse and death across six cohorts. It is publicly available as supplementary material from the original publication (link).
- **Outcome**: Relapse-free survival (RFS) and overall survival (OS) were analyzed under a semi-competing risks framework. RFS is defined as time from surgery to relapse (non-terminal), and OS as time from surgery to death (terminal). After excluding cases with missing data, 973 patients were included. Time was measured in years. The goal was to assess how clinical predictors influence both outcomes and their interdependence.
- **Predictors**: The model included two continuous predictors—age at diagnosis and tumor size—and five categorical predictors: estrogen receptor status, histological grade, nodal status, treatment, and clinical center.
- **Implementation**: Models were implemented in R using ***cmdstanr*** (Stan Development Team, 2022) with 4 chains, 1,000 warm-up iterations, and 2,000 sampling iterations per chain.

Methods

- **Observed data**: $\{(t_i, \delta_i, g_i, x_i), i = 1,2, \dots, n\}$
  - t_i: the observed survival or censoring time
  - $\delta_i \in \{0.1\}$: the event indicator
  - $x_i \in \mathcal{R}^p$: a covariate vector
  - $g_i \in \{1,2, \dots, K\}$: a dataset group to where subject i belongs (e.g, $K = 6$)
- **Survival models for relapse, death, and composite endpoints**
  - Model: $h_i(t|x_i, g_i) = h_0(t)\exp(x_i'\beta + \alpha_{g_i})$
    - β: a vector of regression coefficients
    - α_{g_i}: the fixed effect for the dataset group
    - $h_0(t)$: unspecified baseline hazard function
  - Log-partial likelihood:

$$\log \mathcal{L}(\beta, \alpha) = \sum_{i=1}^{n} \delta_i \left(x_i'\beta + \alpha_{g_i} - \log\left(\sum_{j:t_j \geq t_i} \exp\left(x_j'\beta + \alpha_{g_j} \right) \right) \right)$$

  - Priors: $\beta \sim N_p(0, I)$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)' \sim N_K(0, I)$
  - Posterior: $p(\beta, \alpha|\mathcal{D}) \propto \mathcal{L}(\beta, \alpha) \times \prod_{j=1}^{p} N(\beta_j|0,1) \times \prod_{k=1}^{K} N(\alpha_k|0,1)$
- **Semi-competing risks (SCR) model**
  - Model: $h_i^{(t)}(t|x_i, g_i) = h_0^{(t)}(t)\exp(x_i'\beta^{(t)} + \alpha_{g_i})$
  - $t \in \{1,2, \dots, T\}$: transition (e.g, $T = 3$)
  - $\beta^{(t)}$: a transition-specific vector of regression coefficients
  - Log-partial likelihood:

$$\log \mathcal{L}(\beta, \alpha) = \sum_{t=1}^{T} \sum_{i \in \varepsilon_t} \left(x_i'\beta^{(t)} + \alpha_{g_i} - \log\left(\sum_{j \in \mathcal{R}_i^{(t)}} \exp\left(x_j'\beta^{(t)} + \alpha_{g_j} \right) \right) \right)$$

    - ε_t: the set of individuals who experience an event of type t
    - $\mathcal{R}_i^{(t)}$: the risk set for subject i at the time of event for transition t
  - Priors: $\beta^{(t)} \sim N_p(0, I)$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)' \sim N_K(0, I)$
  - Posterior: $p(\beta, \alpha|\mathcal{D}) \propto \mathcal{L}(\beta, \alpha) \times \prod_{t=1}^{T} \prod_{j=1}^{p} N(\beta_{tj}|0,1) \times \prod_{k=1}^{K} N(\alpha_k|0,1)$

Results

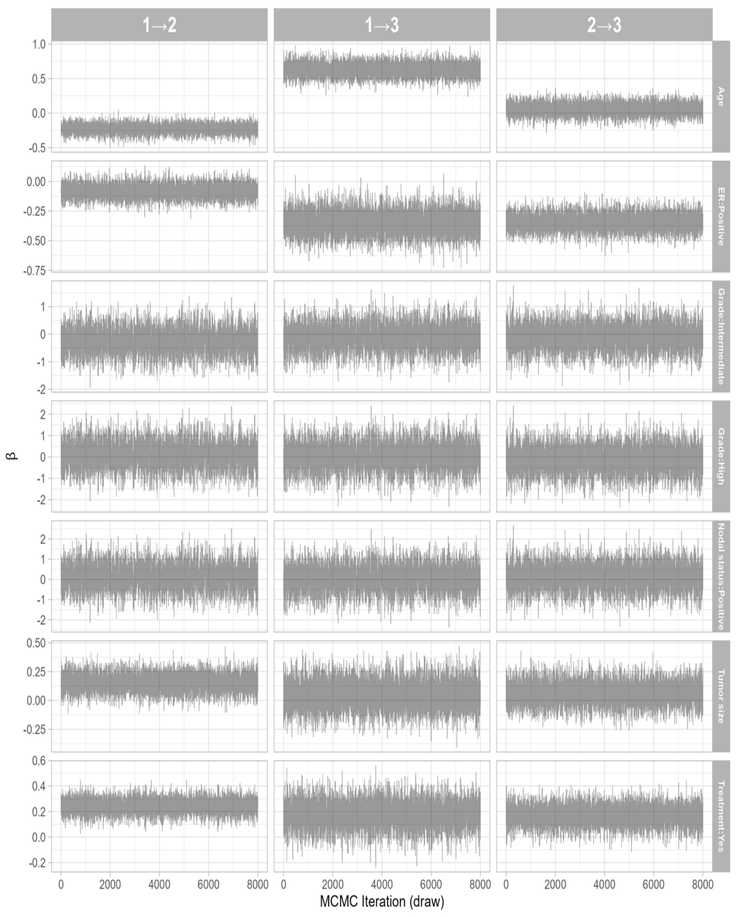
Pathway	Age	Estrogen receptor		Histological grade			Nodal status		Tumor size	Treatment		Dataset					
		Negative	Positive	Low	Intermediate	High	Negative	Positive		No	Yes	CAL[a]	NKI[b]	STNO2[c]	TRANSBIG[d]	UCSF[e]	UNC4[f]
Total	50.7 (±13.0)	287 (29.5%)	686 (70.5%)	153 (15.7%)	373 (38.3%)	447 (45.9%)	563 (57.9%)	410 (42.1%)	2.5 (±1.2)	401 (41.2%)	572 (58.8%)	106 (10.9%)	295 (30.3%)	86 (8.8%)	191 (19.6%)	113 (11.6%)	182 (18.7%)
Alive	51.0 (±12.4)	143 (25.5%)	418 (74.5%)	117 (20.9%)	207 (36.9%)	237 (42.2%)	342 (61.0%)	219 (39.0%)	2.4 (±1.2)	229 (40.8%)	332 (59.2%)	6 (1.1%)	189 (33.7%)	46 (8.2%)	107 (19.1%)	72 (12.8%)	141 (25.1%)
Realpsed	48.7 (±12.2)	24 (19.7%)	98 (80.3%)	12 (9.8%)	63 (51.6%)	47 (38.5%)	69 (56.6%)	53 (43.4%)	2.6 (±1.3)	59 (48.4%)	63 (51.6%)	27 (22.1%)	27 (22.1%)	11 (9.0%)	33 (27.0%)	9 (7.4%)	15 (12.3%)
Dead with relapse	58.8 (±15.8)	39 (38.2%)	63 (61.8%)	12 (11.8%)	43 (42.2%)	47 (46.1%)	43 (42.2%)	59 (57.8%)	2.7 (±1.2)	16 (15.7%)	86 (84.3%)	67 (65.7%)	5 (4.9%)	5 (4.9%)	0 (0.0%)	22 (21.6%)	3 (2.9%)
Dead w/o relapse	46.9 (±11.4)	81 (43.1%)	107 (56.9%)	12 (6.4%)	60 (31.9%)	116 (61.7%)	109 (58.0%)	79 (42.0%)	2.8 (±1.2)	97 (51.6%)	91 (48.4%)	6 (3.2%)	74 (39.4%)	24 (12.8%)	51 (27.1%)	10 (5.3%)	23 (12.2%)

[a]CAL: UCSF & Cal Pacific; [b]NKI: Netherlands Cancer Inst.; [c]STNO: Stanford/Norway; [d]TRANSBIG: TransBIG (EU); [e]UCSF: Univ. of California, SF; [f]UNC: Univ. of North Carolina.

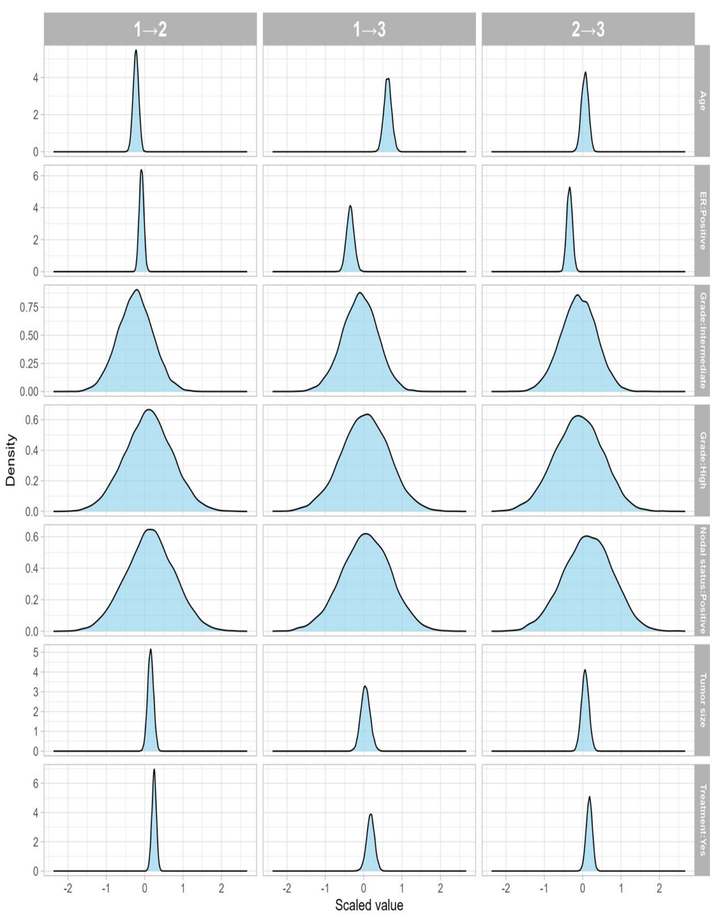
Hazard ratio and 95% HPD intervals for β

	Survival models						Semi-competing risks model					
	Relapse		Overall		Composite		1→2		1→3		2→3	
Covariate	HR	95% HPD	HR	95% HPD	HR	95% HPD	HR	95% HPD	HR	95% HPD	HR	95% HPD
Age	**0.78**	(0.69, 0.88)	**1.12**	(1.00, 1.25)	**0.79**	(0.70, 0.9)	**0.80**	(0.71, 0.90)	**1.87**	(1.59, 2.21)	1.06	(0.91, 1.23)
ER: Positive	0.92	(0.84, 1.02)	**0.76**	(0.69, 0.84)	0.91	(0.82, 1.00)	0.92	(0.83, 1.02)	**0.71**	(0.60, 0.83)	**0.71**	(0.63, 0.80)
Grade: Intermediate	**1.52**	(1.28, 1.83)	**1.31**	(1.09, 1.57)	**1.53**	(1.28, 1.84)	0.78	(0.37, 1.66)	0.91	(0.42, 1.91)	0.90	(0.42, 1.90)
Grade: High	**1.61**	(1.35, 1.95)	**1.51**	(1.27, 1.83)	**1.61**	(1.35, 1.95)	1.11	(0.40, 3.07)	1.04	(0.37, 2.84)	0.90	(0.32, 2.47)
Nodal status: Positive	**1.22**	(1.06, 1.39)	**1.14**	(1.01, 1.29)	**1.22**	(1.07, 1.40)	1.15	(0.40, 3.27)	1.06	(0.37, 3.02)	1.15	(0.39, 3.22)
Tumor size	**1.26**	(1.14, 1.39)	**1.25**	(1.13, 1.38)	**1.25**	(1.13, 1.38)	**1.17**	(1.03, 1.33)	1.05	(0.86, 1.28)	1.07	(0.91, 1.25)
Treatment: Yes	**0.81**	(0.70, 0.96)	0.89	(0.76, 1.05)	**0.81**	(0.69, 0.95)	**1.28**	(1.16, 1.40)	**1.20**	(1.00, 1.42)	**1.19**	(1.04, 1.34)

Trace plots for β by transition and covariate



Posterior density plots of β by transition and covariate



- Discussion: In the SCR framework, younger patients show higher relapse risk, while older patients face greater mortality, regardless of relapse. ER-positive patients have lower risks of both relapse and death compared to ER-negative patients. Histological and nodal status are not significant. Larger tumor size increases relapse risk, but its effect on death is not statistically significant. Treated patients exhibit higher risks of relapse and death than untreated patients.
- Limitation: The SCR model assumes a common cluster-specific fixed effect across all transitions. However, it can be extended to allow transition-specific effects, i.e., using $\alpha^{(t)}$ for $t = 1,2, \dots, T$, instead of a shared α.

*This research was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2023-00239958) (Jinheum Kim) and by the University of Suwon (Jinheum Kim & Eunjeong Cha).