Pukyong National University

2018 / 7 / 6

A joint frailty-copula model for clustered semi-competing risks data

Takeshi Emura

Graduate Institute of Statistics

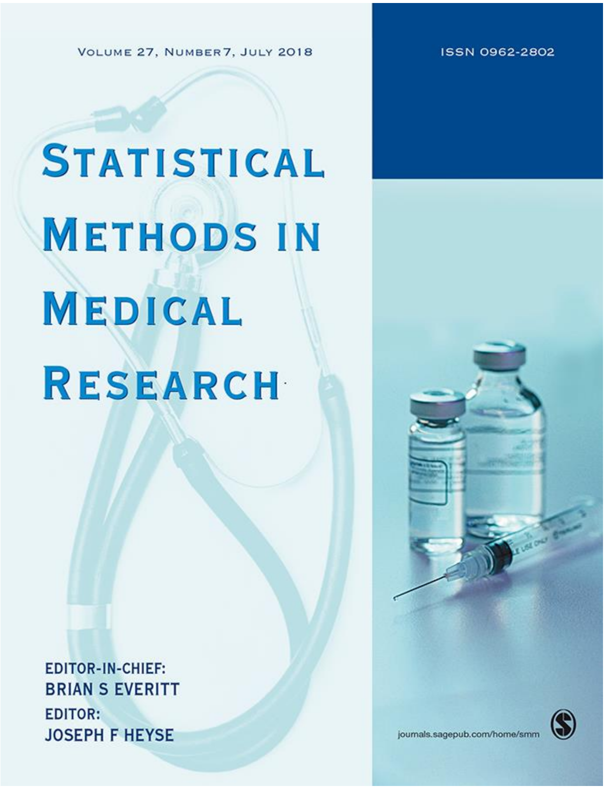
National Central University (Taiwan)

Part I (25 min):
Model & Estimation

Stat Methods Med Res **(2017) 26(6): 2649-66**

Part II (20 min):
Prediction & high-dimensional covariates

Stat Methods Med Res **(2018-) doi:10.1177/0962280216688032**



VOLUME 27, NUMBER7, JULY 2018

ISSN 0962-2802

STATISTICAL METHODS IN MEDICAL RESEARCH

EDITOR-IN-CHIEF:
BRIAN S EVERITT
EDITOR:
JOSEPH F HEYSE

journals.sagepub.com/home/smm

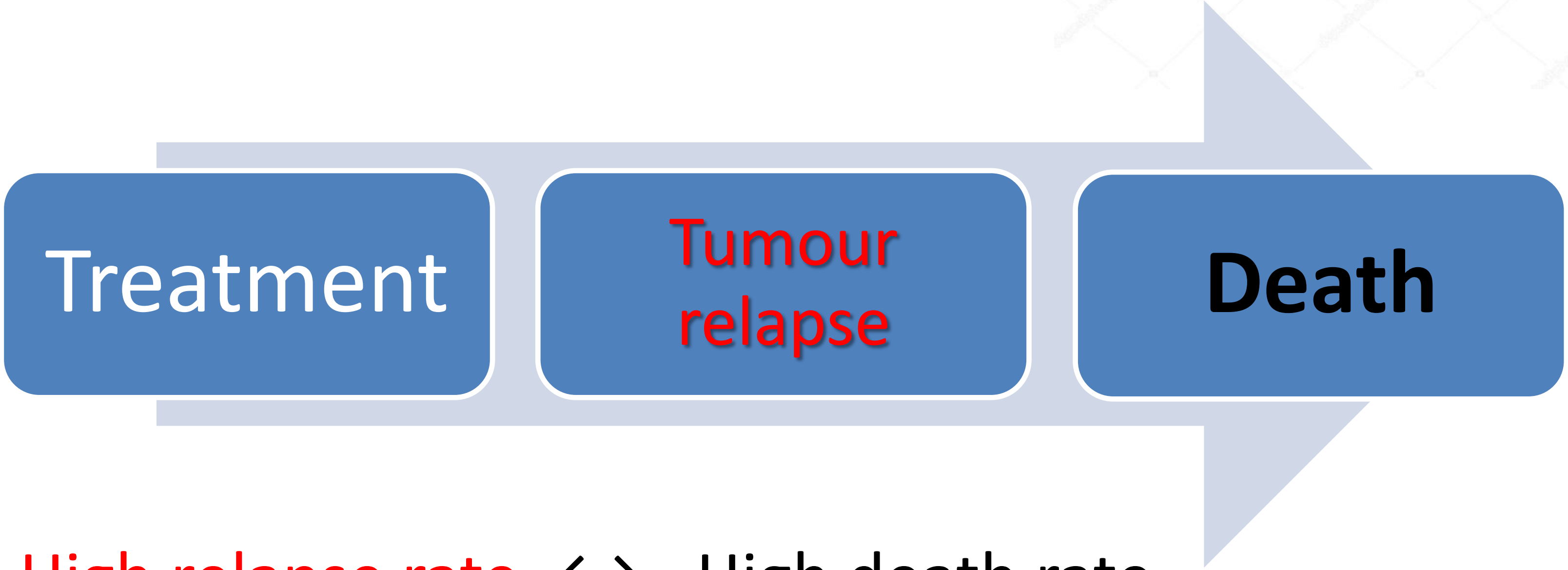
Endpoints for cancer patients

- Time-to-progression (TTP)

  (e.g., relapse)

- Overall survival (OS)

  (=time-to-death)



Treatment

Tumour relapse

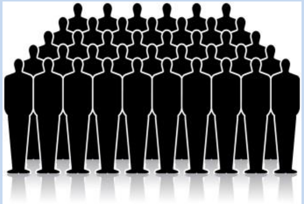
Death

High relapse rate ↔ High death rate

Short time of relapse ↔ Short time of death

Patients collected from 4 studies

Cluster 1 (Medium risk)

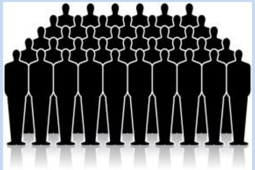


N_1=110

Cluster 3 (Medium risk)

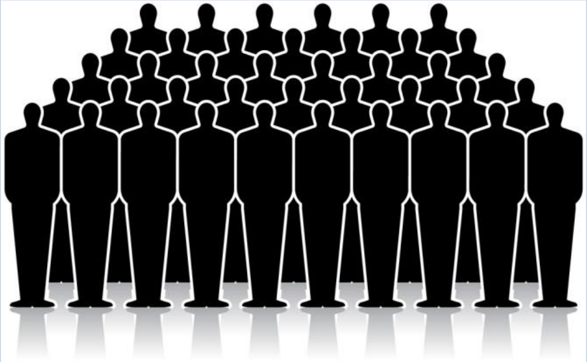
N_3=278

Cluster 2 (High risk)



N_2=58

Cluster 4 (Low risk)



N_4=557

Combined sample size :

$$\sum_{i=1}^{4} N_i = 110 + 278 + 58 + 557$$
$$= 1003$$

Motivation: Ovarian cancer (Ganzfried et al. 2013)

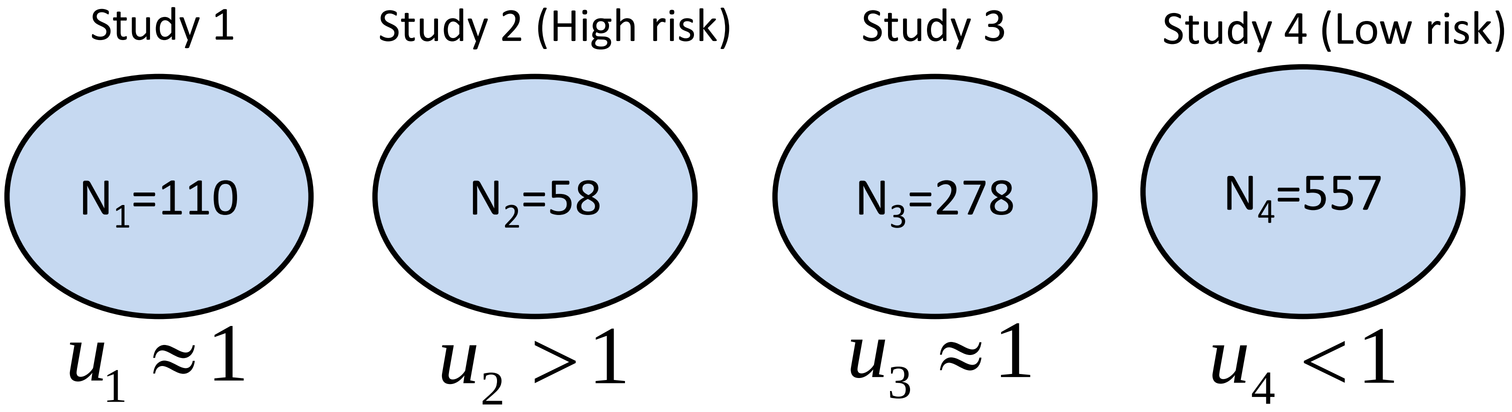
A meta-analytic data of ovarian cancer patients.

	Dataset[a]	Sample size	The number of observed events (event rates %)		
			Relapse	Death	Censoring
			($\delta_{ij}=1$)	($\delta^*_{ij}=1$)	($\delta^*_{ij}=0$)
	GSE17260	$N_1=110$	76 (69%)	46 (42%)	64 (58%)
High risk →	GSE30161	$N_2=58$	48 (83%)	36 (62%)	22 (38%)
	GSE9891	$N_3=278$	185 (67%)	113 (41%)	165 (59%)
Low risk →	TCGA	$N_4=557$	266 (48%)	290 (52%)	267 (48%)
	Total	$\sum_{i=1}^{4} N_i=1003$	575 (57%)	485 (48%)	518 (52%)

Risks of relapse are heterogeneous

- We shall account the heterogeneity by a random effect, called ***frailty***.

- Random effect (called frailty)



Study 1 | Study 2 (High risk) | Study 3 | Study 4 (Low risk)

N_1=110 | N_2=58 | N_3=278 | N_4=557

$u_1 \approx 1$ | $u_2 > 1$ | $u_3 \approx 1$ | $u_4 < 1$

Gamma frailty :

$$u_i \sim f_\eta(u) = \frac{1}{\Gamma(1/\eta)\eta^{1/\eta}} u^{\frac{1}{\eta}-1} \exp\left(-\frac{u}{\eta}\right),$$

$$\begin{cases} E[u_i] = 1 \\ Var[u_i] = \eta \end{cases}$$

Ref:(Burzykowski et al. 2001; Duchateau and Janssen 2007 Rondeau et al. 2011; Ha et al. 2018)

$$X_{ij} = \text{TTP} \ (\text{Time to progression due to relapse})$$

$$D_{ij} = \text{OS} \ (\text{Overall survival} = \text{time to death})$$

Two marginal hazard functions

$$\begin{cases} r_{ij}(t \mid u_i) = \Pr(t \le X_{ij} < t + dt \mid X_{ij} \ge t, u_i) & (\text{hazard for } X_{ij}) \\ \lambda_{ij}(t \mid u_i) = \Pr(t \le D_{ij} < t + dt \mid X_{ij} \ge t, u_i) & (\text{hazard for } D_{ij}) \end{cases}$$

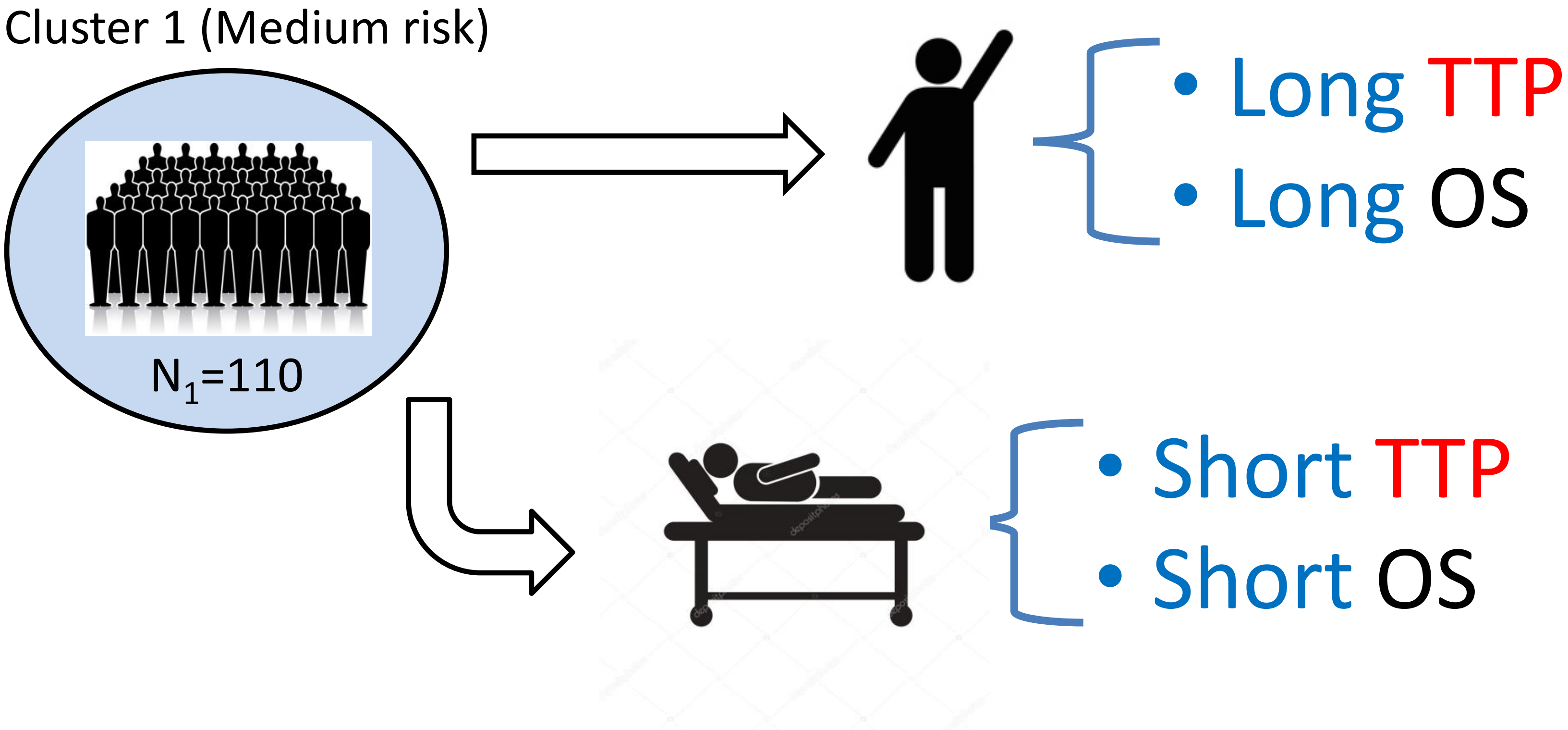
Joint frailty model (Rondeau et al., 2011)

$$\begin{cases} r_{ij}(t \mid u_i) = u_i r_0(t) \exp(\boldsymbol{\beta}_1' \mathbf{Z}_{1ij}) & (\text{hazard for } X_{ij}) \\ \lambda_{ij}(t \mid u_i) = u_i^{\alpha} \lambda_0(t) \exp(\boldsymbol{\beta}_2' \mathbf{Z}_{2ij}) & (\text{hazard for } D_{ij}) \end{cases}$$

$$u_i = \text{frailty} \ (u_i < 1\text{: Low risk;} \quad u_i > 1\text{: High risk})$$

$$\begin{cases} \mathbf{Z}_{1ij} = \text{covariates for } X_{ij} \\ \mathbf{Z}_{2ij} = \text{covariates for } D_{ij} \end{cases}$$

Patient level dependence between TTP and OS



Cluster 1 (Medium risk)

N_1=110

- Long TTP
- Long OS

- Short TTP
- Short OS

➔There exist **patient level dependence** between TTP and OS !

Joint frailty-copula model (Proposed)

$$\begin{cases} r_{ij}(\,t\,|\,u_i\,) = u_i r_0(t)\exp(\,\boldsymbol{\beta}_1'\mathbf{Z}_{1,ij}\,) & \text{for } X_{ij} \\ \lambda_{ij}(\,t\,|\,u_i\,) = u_i^{\alpha}\lambda_0(t)\exp(\,\boldsymbol{\beta}_2'\mathbf{Z}_{2,ij}\,) & \text{for } D_{ij} \\ \Pr(\,X_{ij} > x\,,\,D_{ij} > y\,|\,u_i\,) = C_\theta[\,S_{Xij}(x\,|\,u_i), S_{Dij}(y\,|\,u_i)\,] & \end{cases}$$

$C_\theta[\,v, w\,]$: copula function,

θ: dependence parameter

← Patient-level dependence

$$S_{Xij}(t\,|\,u_i) = \exp\left\{-\int_0^t r_{ij}(t\,|\,u_i)\right\}:$$ marginal survival for X_{ij},

$$S_{Dij}(t\,|\,u_i) = \exp\left\{-\int_0^t \lambda_{ij}(t\,|\,u_i)\right\}:$$ marginal survival for D_{ij},

Copulas

The independence copula:

$$C(u,v) = uv\,,$$

The Clayton copula (Clayton 1978):

$$C_\theta(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \qquad \theta > 0\,,$$

The Gumbel copula (Gumbel 1960), also known as the Hougaard copula:

$$C_\theta(u,v) = \exp\left[-\{ (-\log u)^{\theta+1} + (-\log v)^{\theta+1} \}^{\frac{1}{\theta+1}} \right], \qquad \theta \geq 0\,,$$

The Frank copula (Frank 1979):

$$C_\theta(u,v) = -\frac{1}{\theta}\log\left\{ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right\}, \qquad \theta \neq 0\,.$$

Example: Clayton copula

$$C_\theta(v, w) = (v^{-\theta} + w^{-\theta} - 1)^{-1/\theta}$$

$$\theta + 1 = \frac{\Pr(X = x, D = y)\Pr(X > x, D > y)}{\Pr(X = x, D > y)\Pr(X > x, D = y)}$$

$$= \text{Odds ratio} \Rightarrow$$

	Relapse	Relapse-free
Death	X=x,D=y	X>x, D=y
Alive	X=x,D>y	X>x, D>y

$\theta > 0$: Positive dependence: (relapse) ↔ (death)

$-1 < \theta < 0$: Negative dependence: (relapse) ↔ (death)

- Kendall's tau $= \dfrac{\theta}{\theta + 2}$

Baseline hazard functions

- Cubic M-splines

$$r_0(t) = \sum_{\ell=1}^{5} g_\ell M_\ell(t) = \mathbf{g}'\mathbf{M}(t)$$

$$\lambda_0(t) = \sum_{\ell=1}^{5} h_\ell M_\ell(t) = \mathbf{h}'\mathbf{M}(t)$$

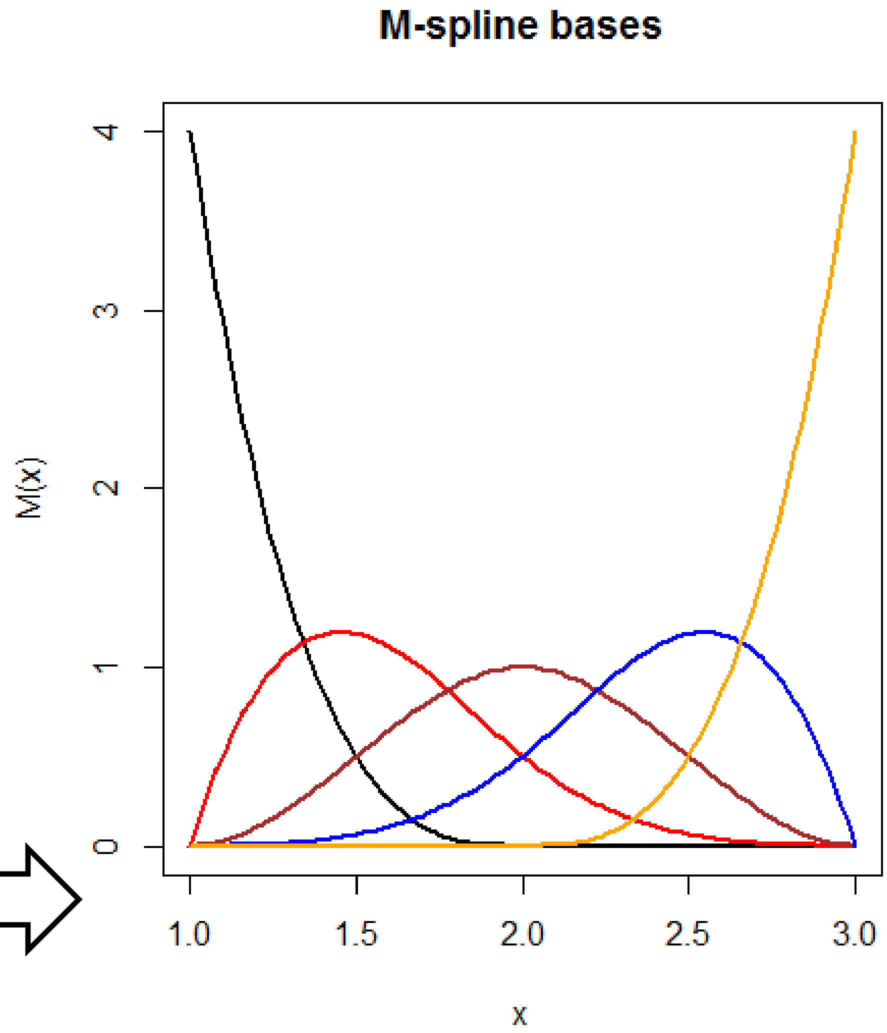
- Unknown parameters

$$\mathbf{g}' = (\, g_1, \;\ldots, g_5 \,)$$

$$\mathbf{h}' = (\, h_1, \;\ldots, h_5 \,)$$

- 5 Basis functions

$$\mathbf{M}(t) = (\, M_1(t), \;\ldots, M_5(t)\,)'$$



Joint frailty-copula model (Proposed)

$u_i \sim Gamma(1/\eta, \eta)$

Frailty

Splines

$$r_0(t) = \sum_{\ell=1}^{5} g_\ell M_\ell(t) = \mathbf{g}'\mathbf{M}(t)$$

$$\lambda_0(t) = \sum_{\ell=1}^{5} h_\ell M_\ell(t) = \mathbf{h}'\mathbf{M}(t)$$

$$\begin{cases} r_{ij}(\,t \mid u_i\,) = u_i r_0(t) \exp(\,\boldsymbol{\beta}_1' \mathbf{Z}_{1,ij}\,) & \text{for } X_{ij} \\ \lambda_{ij}(\,t \mid u_i\,) = u_i^{\alpha} \lambda_0(t) \exp(\,\boldsymbol{\beta}_2' \mathbf{Z}_{2,ij}\,) & \text{for } D_{ij} \\ \Pr(\,X_{ij} > x\,,\, D_{ij} > y \mid u_i\,) = C_\theta[\,S_{Xij}(x \mid u_i),\, S_{Dij}(y \mid u_i)\,] & \end{cases}$$

Clayton copula

$$C_\theta(\,v, w\,) = (\,v^{-\theta} + w^{-\theta} - 1\,)^{-1/\theta}$$

- Unknown parameters: $(\,\alpha, \eta, \theta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, r_0, \lambda_0\,)$

Ovarian cancer data (Ganzfried et al., 2013)

n ca

		Sample size	The number of observed events (event rates)		
			Relapse	Death	Censoring
Japanese	Study 1	$N_1 = 84$	59 (70%)	38 (45%)	46 (55%)
American	Study 2	$N_2 = 58$	48 (83%)	36 (62%)	22 (38%)
Australian	Study 3	$N_3 = 260$	185 (71%)	113 (43%)	147 (57%)
American	Study 4	$N_4 = 510$	252 (49%)	278 (55%)	232 (45%)
	Total	$\sum_{i=1}^{4} N_i = 912$	544 (60%)	465 (51%)	447 (49%)

Notes: The data are extracted from R Bioconductor *curatedOvarianData*

Between-study heterogeneity (via gamma frailty)

Patient-level dependence (via Clayton copula)

Data structure

$X_{ij} = \text{TTP (Time to progression due to recurrence, Relapse, etc.)}$

$D_{ij} = \text{OS (Overall survival = time to death from any cause)}$

$C_{ij} = \text{Administrative censoring time (e.g., study end)}$

<u>Observations :</u>

$$(T_{ij} , T_{ij}^{*} , \delta_{ij} , \delta_{ij}^{*} , \mathbf{Z}_{1ij} , \mathbf{Z}_{2ij}), \quad i = 1, 2, ..., G, \quad j = 1, 2, ..., N_i$$

$*$ First occuring event time

$$T_{ij} = \min(X_{ij} , D_{ij} , C_{ij}), \qquad \delta_{ij} = \mathbf{I}(T_{ij} = X_{ij})$$

Indicator of progression

$*$ Terminal event time

$$T_{ij}^{*} = \min(D_{ij} , C_{ij}), \qquad \delta_{ij}^{*} = \mathbf{I}(T_{ij}^{*} = D_{ij})$$

Indicator of death

Data structure

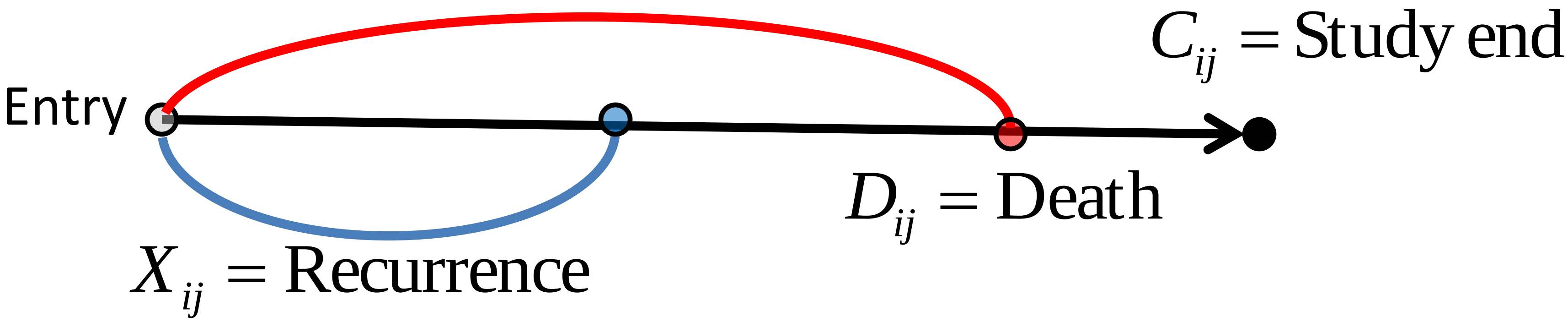


Fig. Case of $\delta_{ij} = 1, \quad \delta^{*}_{ij} = 1$

$*$ First occuring event time :

$$T_{ij} = \min(X_{ij}, D_{ij}, C_{ij}) = X_{ij}$$

$$\delta_{ij} = \mathbf{I}(T_{ij} = X_{ij}) = 1$$

$*$ Terminal event time :

$$T^{*}_{ij} = \min(D_{ij}, C_{ij}) = D_{ij}$$

$$\delta^{*}_{ij} = \mathbf{I}(T^{*}_{ij} = D_{ij}) = 1$$

4 patterns

	Likelihood contribution
Relapse → Death T_{ij} T_{ij}^*	$\Pr(X_{ij} = T_{ij}, D_{ij} = T_{ij}^* \mid u_i)$
Relapse → Censoring T_{ij} T_{ij}^*	$\Pr(X_{ij} = T_{ij}, D_{ij} > T_{ij}^* \mid u_i)$
Death (without relapse) $T_{ij} = T_{ij}^*$	$\Pr(X_{ij} > T_{ij}, D_{ij} = T_{ij}^* \mid u_i)$
Censoring (neither relapse nor death) $T_{ij} = T_{ij}^*$	$\Pr(X_{ij} > T_{ij}, D_{ij} > T_{ij}^* \mid u_i)$

Log-likelihood (proposed)

$$
\begin{aligned}
&\ell(\alpha, \eta, \theta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, r_0, \lambda_0) \\
&= \sum_{i=1}^{G} \left[\sum_{j=1}^{N_i} \{ \delta_{ij} \log r_{ij}(T_{ij}) + \delta_{ij}^{*} \log \lambda_{ij}(T_{ij}^{*}) \} \right. \\
&\qquad + \log \int_0^\infty \left\{ u_i^{m_i + \alpha m_i^{*}} \prod_{j=1}^{N_i} \psi_\theta[\, u_i R_{ij}(T_{ij}), u^{\alpha} \Lambda_{ij}(T_{ij}^{*}) \,]^{\delta_{ij}} \, \psi_\theta^{*}[\, u_i R_{ij}(T_{ij}), u_i^{\alpha} \Lambda_{ij}(T_{ij}^{*}) \,]^{\delta_{ij}^{*}} \right. \\
&\qquad\qquad \left. \left. \times \Theta_\theta[\, u_i R_{ij}(T_{ij}), u_i^{\alpha} \Lambda_{ij}(T_{ij}^{*}) \,]^{\delta_{ij}\delta_{ij}^{*}} D_\theta[\, u_i R_{ij}(T_{ij}), u_i^{\alpha} \Lambda_{ij}(T_{ij}^{*}) \,] \right\} f_\eta(u_i) du_i \right],
\end{aligned}
$$

where $r_{ij}(t) = r_0(t)\exp(\boldsymbol{\beta}_1' \mathbf{Z}_{ij})$, $\lambda_{ij}(t) = \lambda_0(t)\exp(\boldsymbol{\beta}_2' \mathbf{Z}_{ij})$,

$D_\theta[s, t] = C_\theta[\exp(-s), \exp(-t)]$, $\psi_\theta = D_\theta^{[1,0]} / D_\theta$, $D_\theta^{[1,0]} = -\partial D_\theta / \partial s$, $\psi_\theta^{*} = D_\theta^{[0,1]} / D_\theta$,

$D_\theta^{[0,1]} = -\partial D_\theta / \partial t$, $\Theta_\theta = D_\theta^{[1,1]} D_\theta / D_\theta^{[1,0]} D_\theta^{[0,1]}$ and $D_\theta^{[1,1]} = \partial^2 D_\theta / \partial s \partial t$.

Derivatives of copula

- Penalized likelihood with cubic M-spline

  ➔ Directly follow Rondeau et al. (2011)

$$\ell(\alpha, \eta, \theta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, r_0, \lambda_0) - \kappa_1 \int \ddot{\gamma}_0(t)^2 dt - \kappa_2 \int \ddot{\lambda}_0(t)^2 dt$$

$$\int \ddot{r}_0(t)^2 dt = \sum_{k=1}^{L_r} \sum_{\ell=1}^{L_r} g_k g_\ell \int \ddot{M}_k(t) \ddot{M}_\ell(t) dt, \quad \int \ddot{\lambda}_0(t)^2 dt = \sum_{k=1}^{L_\lambda} \sum_{\ell=1}^{L_\lambda} h_k h_\ell \int \ddot{M}_k(t) \ddot{M}_\ell(t) dt$$

- κ_1 = Smoothing parameter for the hazard of TTP
- κ_2 = Smoothing parameter for the hazard of OS

The values ($\hat{\kappa}_1, \hat{\kappa}_2$) chosen by LCV (Joly, et al. 1998)

$$(\hat{\alpha}, \hat{\eta}, \hat{\theta}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{r}_0, \hat{\lambda}_0) = \underset{(\alpha, \eta, \theta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, r_0, \lambda_0)}{\arg\max} \left[\ell(\alpha, \eta, \theta, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, r_0, \lambda_0) - \hat{\kappa}_1 \int \ddot{\gamma}_0(t)^2 dt - \hat{\kappa}_2 \int \ddot{\lambda}_0(t)^2 dt \right]$$

•Implementation: R *joint.Cox* package

Simulation setting: G=5, N_i=100 or 200

- Frailty: $u_i \sim$ Gamma $(1/\eta, \eta)$ where $\eta = 0.5$
- Covariate: $Z_{ij} \sim \text{Unif}(0, 1)$
- Proportional hazard model with frailty

$$R_{ij}(x \mid u_i) = u_i r_0 x \exp(\beta_1 Z_{ij}), \quad \Lambda_{ij}(y \mid u_i) = u_i \lambda_0 y \exp(\beta_2 Z_{ij})$$

where r_0 =1 and λ_0 =1 (Exponential distribution)

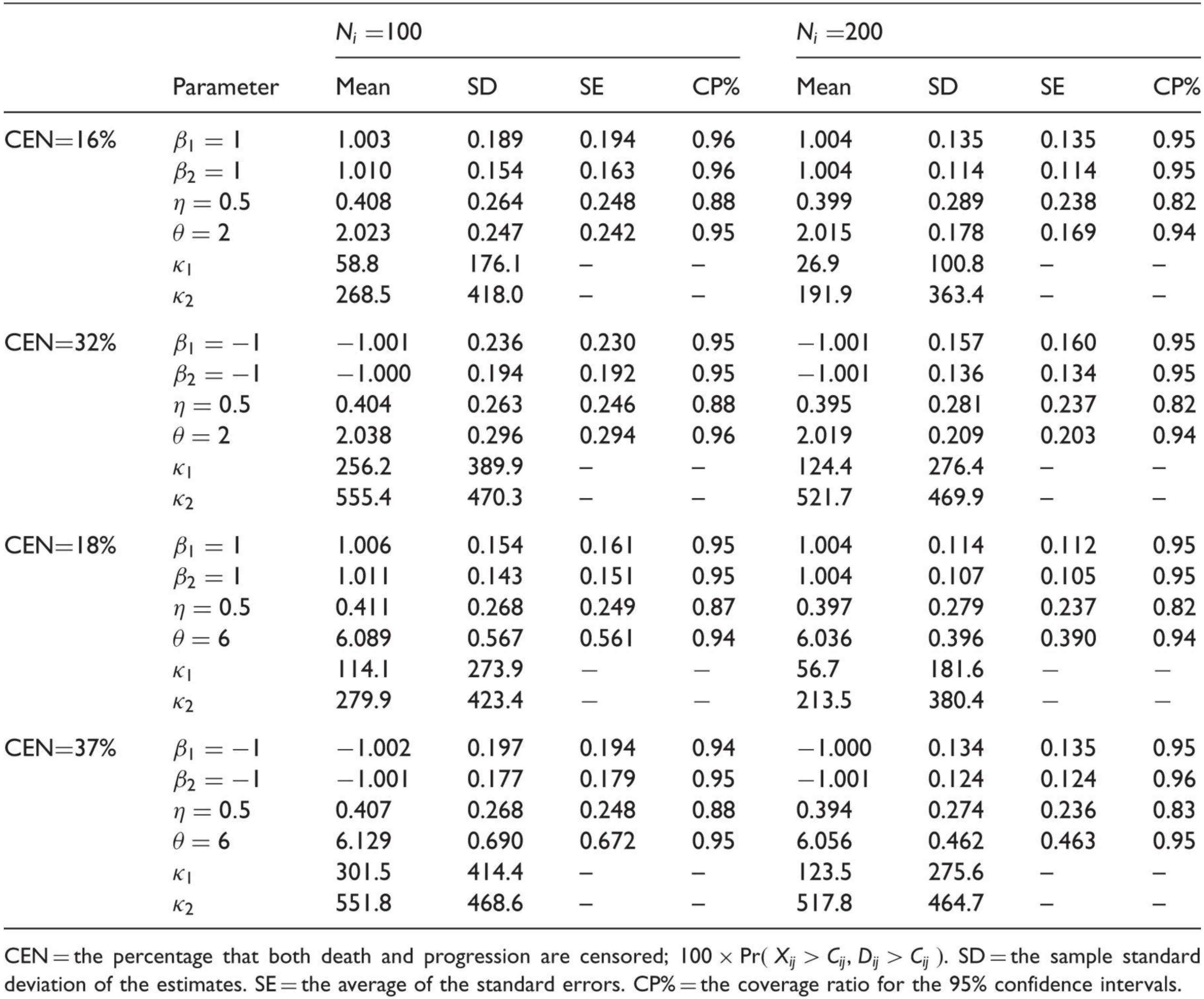
- Joint frailty-copula model

$$\Pr(X_{ij} > x , D_{ij} > y \mid u_i) = [\exp\{ \theta R_{ij}(x \mid u_i) \} + \exp\{ \theta \Lambda_{ij}(y \mid u_i) \} - 1]^{-1/\theta} ,$$

at $\theta = 2$ ➔ $\tau(X_{ij}, D_{ij} \mid u_i) = 0.5$.

- Censoring, $C_{ij} \sim \text{Unif}(0, 5)$ ➔ 16~37% censored subjects

Simulation results for the proposed method (G = 5 studies) based on 500 replications.

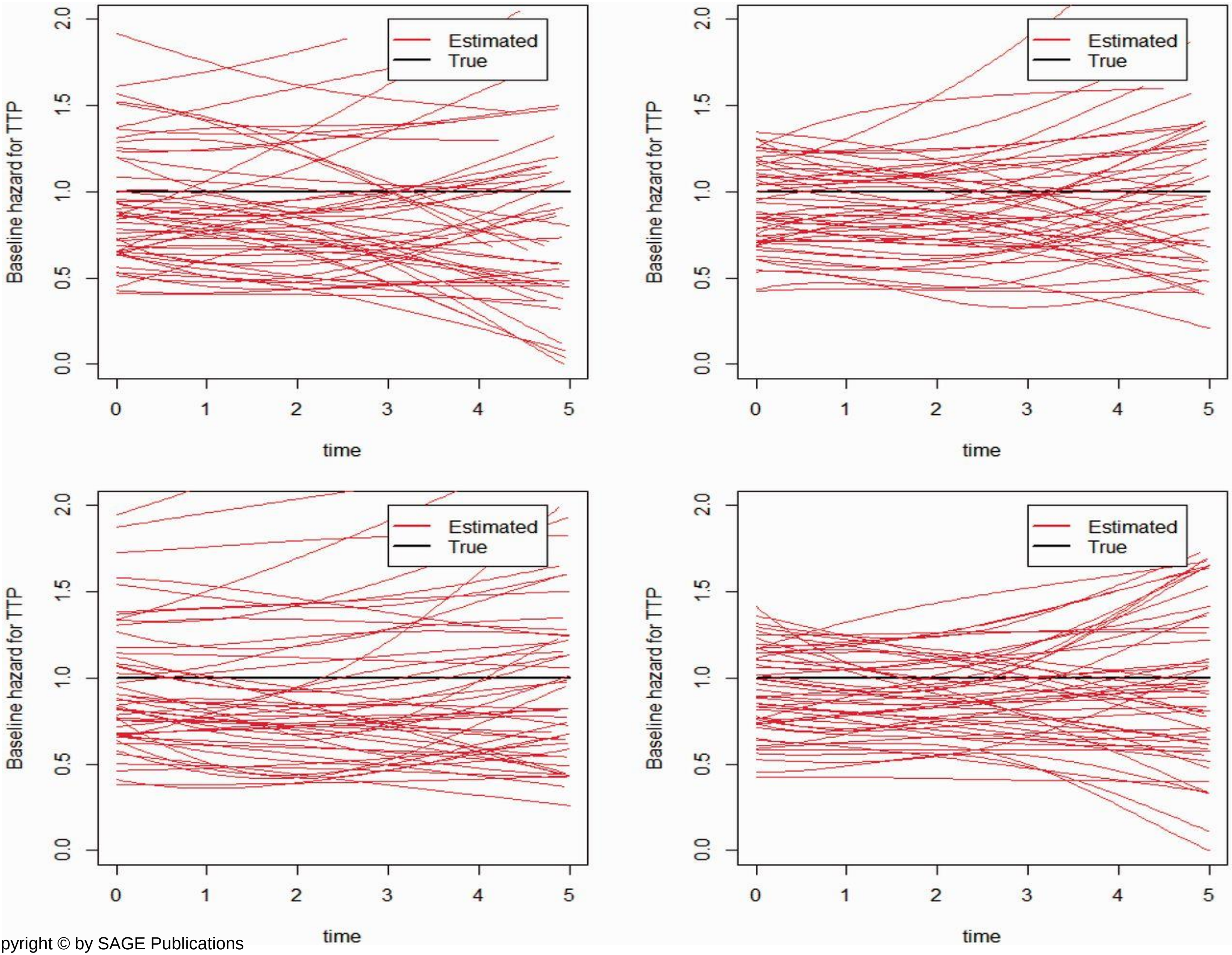


		$N_i = 100$				$N_i = 200$			
	Parameter	Mean	SD	SE	CP%	Mean	SD	SE	CP%
CEN=16%	$\beta_1 = 1$	1.003	0.189	0.194	0.96	1.004	0.135	0.135	0.95
	$\beta_2 = 1$	1.010	0.154	0.163	0.96	1.004	0.114	0.114	0.95
	$\eta = 0.5$	0.408	0.264	0.248	0.88	0.399	0.289	0.238	0.82
	$\theta = 2$	2.023	0.247	0.242	0.95	2.015	0.178	0.169	0.94
	κ_1	58.8	176.1	–	–	26.9	100.8	–	–
	κ_2	268.5	418.0	–	–	191.9	363.4	–	–
CEN=32%	$\beta_1 = -1$	−1.001	0.236	0.230	0.95	−1.001	0.157	0.160	0.95
	$\beta_2 = -1$	−1.000	0.194	0.192	0.95	−1.001	0.136	0.134	0.95
	$\eta = 0.5$	0.404	0.263	0.246	0.88	0.395	0.281	0.237	0.82
	$\theta = 2$	2.038	0.296	0.294	0.96	2.019	0.209	0.203	0.94
	κ_1	256.2	389.9	–	–	124.4	276.4	–	–
	κ_2	555.4	470.3	–	–	521.7	469.9	–	–
CEN=18%	$\beta_1 = 1$	1.006	0.154	0.161	0.95	1.004	0.114	0.112	0.95
	$\beta_2 = 1$	1.011	0.143	0.151	0.95	1.004	0.107	0.105	0.95
	$\eta = 0.5$	0.411	0.268	0.249	0.87	0.397	0.279	0.237	0.82
	$\theta = 6$	6.089	0.567	0.561	0.94	6.036	0.396	0.390	0.94
	κ_1	114.1	273.9	–	–	56.7	181.6	–	–
	κ_2	279.9	423.4	–	–	213.5	380.4	–	–
CEN=37%	$\beta_1 = -1$	−1.002	0.197	0.194	0.94	−1.000	0.134	0.135	0.95
	$\beta_2 = -1$	−1.001	0.177	0.179	0.95	−1.001	0.124	0.124	0.96
	$\eta = 0.5$	0.407	0.268	0.248	0.88	0.394	0.274	0.236	0.83
	$\theta = 6$	6.129	0.690	0.672	0.95	6.056	0.462	0.463	0.95
	κ_1	301.5	414.4	–	–	123.5	275.6	–	–
	κ_2	551.8	468.6	–	–	517.8	464.7	–	–

CEN = the percentage that both death and progression are censored; $100 \times \Pr(X_{ij} > C_{ij}, D_{ij} > C_{ij})$. SD = the sample standard deviation of the estimates. SE = the average of the standard errors. CP% = the coverage ratio for the 95% confidence intervals.

Takeshi Emura et al. Stat Methods Med Res ,2015;0962280215604510

Simulation results for estimating the baseline hazard based on 50 replications.



Ovarian cancer meta-analysis

(Ganzfried et al. 2013)

Dataset[a]	Sample size	The number of observed events (event rates %)		
		Relapse $(\delta_{ij}=1)$	Death $(\delta^*_{ij}=1)$	Censoring $(\delta^*_{ij}=0)$
GSE17260	$N_1=110$	76 (69%)	46 (42%)	64 (58%)
GSE30161	$N_2=58$	48 (83%)	36 (62%)	22 (38%)
GSE9891	$N_3=278$	185 (67%)	113 (41%)	165 (59%)
TCGA	$N_4=557$	266 (48%)	290 (52%)	267 (48%)
Total	$\sum_{i=1}^{4} N_i=1003$	575 (57%)	485 (48%)	518 (52%)

- Goal 1: Marginal analysis of relapse (TTP) and death (OS)

$$\begin{cases} r_{ij}(t\,|\,u_i)=u_i r_0(t)\exp(\;\beta_1\times\mathrm{CXCL12}\;) & (\text{ hazard for TTP })\\ \lambda_{ij}(t\,|\,u_i)=u_i^{\alpha}\lambda_0(t)\exp(\;\beta_2\times\mathrm{CXCL12}\;) & (\text{ hazard for OS }) \end{cases}$$

- Goal 2: Association analysis of TTP and death

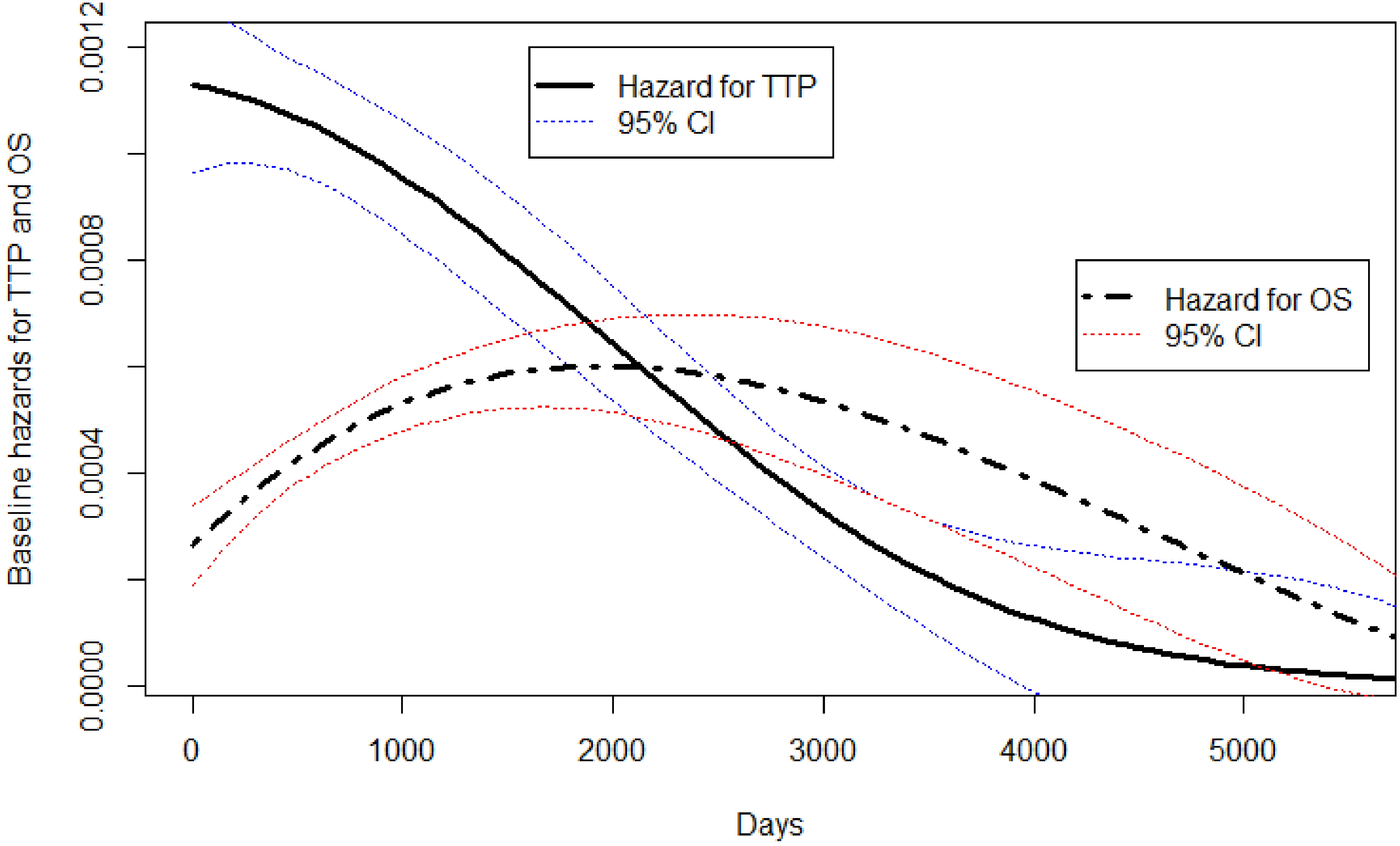
$\underline{\text{Copula model}}$:

$$\Pr(\,X_{ij}>x\,,D_{ij}>y\,|\,u_i\,)=C_{\theta}[\,\exp\{-R_{ij}(x\,|\,u_i)\},\exp\{-\Lambda_{ij}(y\,|\,u_i)\}\,]$$

Table 5. The joint analysis of recurrence (TTP) and death (OS) for the meta-anal (four studies, 1003 patients) for ovarian cancer patients of Ganzfried et al.[19].

	Proposed method: Estimate (95% CI)
RR[a] for relapse (TTP) : $\exp(\beta_1)$	1.22 (1.13-1.32)
RR[a] for death (OS) : $\exp(\beta_2)$	1.18 (1.08-1.29)
Heterogeneity: $\eta = Var_\eta(u_i)$	0.033 (0.006-0.186)
Copula parameter: θ	2.35 (1.90-2.90)
RR for death after relapse: $\theta + 1$	3.35 (2.90-3.90)
Kendall's tau: $\tau = \theta / (\theta + 2)$	0.54 (0.49-0.59)
Maximum penalized log-likelihood	-8604.093

Notes: [a]RR (Relative Risk) of *CXCL12* expression on the hazards are examined.



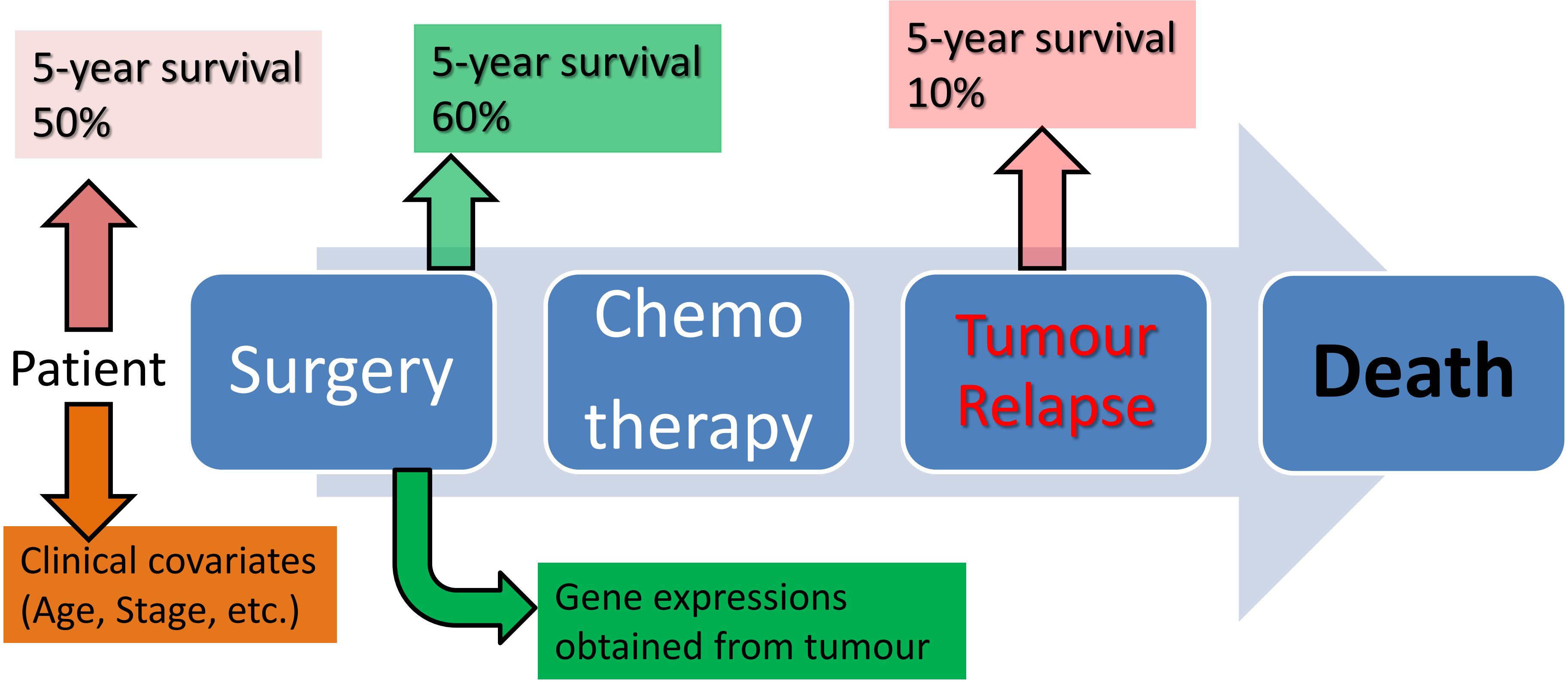
Estimated baseline hazards:

$$\hat{\lambda}_0(x) = 0.211 M_1(x) + 1.084 M_2(x) + 1.001 M_3(x) + 0.180 M_4(x)$$

$$\hat{r}_0(x) = 0.907 M_1(x) + 1.711 M_2(x) + 0.040 M_4(x)$$

Part II:
Dynamic prediction & high-dimensional covariates

Follow-up for a cancer patient



5-year survival 50%

5-year survival 60%

5-year survival 10%

Patient

Surgery

Chemo therapy

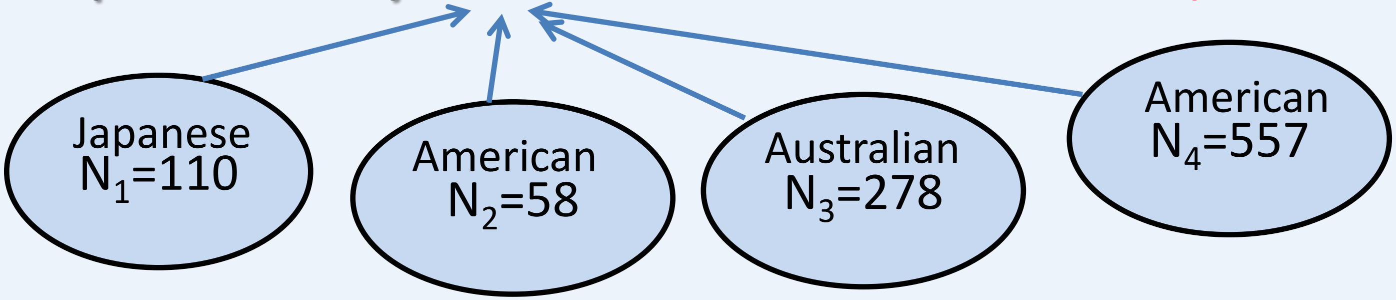
Tumour Relapse

Death

Clinical covariates (Age, Stage, etc.)

Gene expressions obtained from tumour

Death probability = $\hat{F}$(Clinical, Gene, Relapse, Timing)



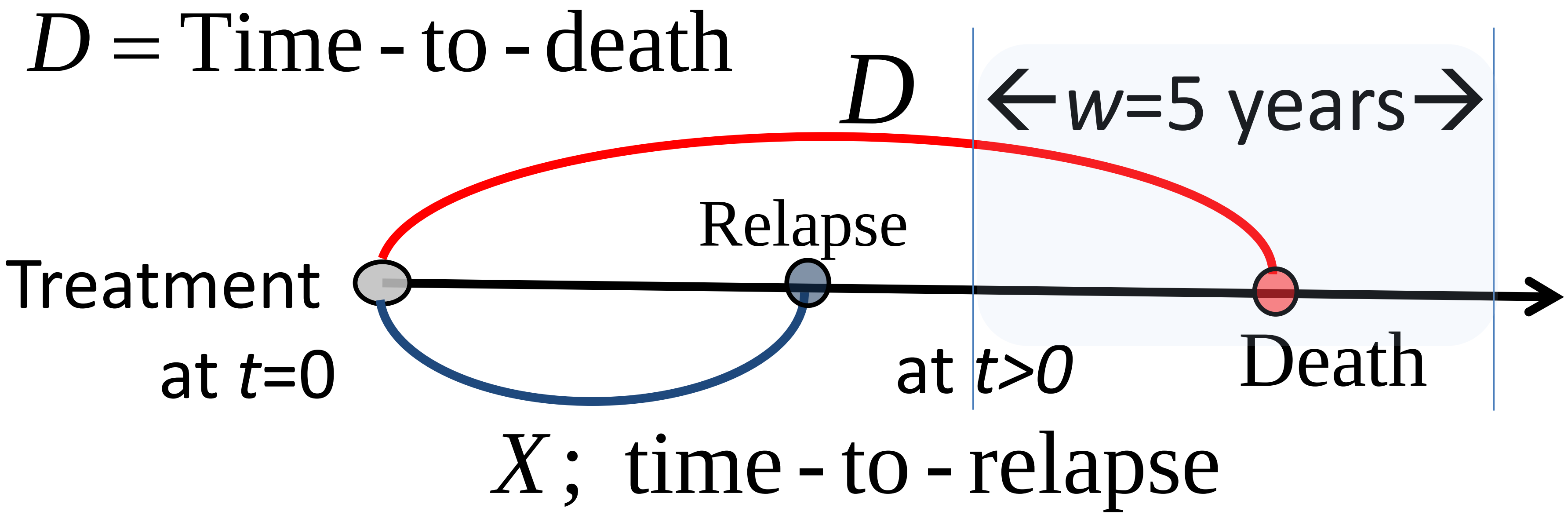
Japanese N_1=110

American N_2=58

Australian N_3=278

American N_4=557

Dynamic Prediction



$D = \text{Time-to-death}$

D

←w=5 years→

Relapse

Treatment

at t=0

at t>0

Death

X; time-to-relapse

$$F(t, t+w \mid X, \mathbf{Z}) = \Pr(D \le t+w \mid D > t, X, \mathbf{Z})$$

↑Conditional failure function (van Houwelingen and Putter 2013)

How to construct the prediction formula?

1) Landmark model (Conditional Cox models fitted at different time points)
2) Time-dependent covariate ? (Cox model is only for exogenous TDC)
3) Joint model (use a copula on (X, D))

Copula model

$$\Pr(X > x, D > y) = C_\theta[\Pr(X > x), \Pr(D > y)]$$

Clayton copula: $C_\theta(v, w) = (v^{-\theta} + w^{-\theta} - 1)^{-1/\theta}$

$$\theta + 1 = \frac{\Pr(X = x, D = y)\Pr(X > x, D > y)}{\Pr(X = x, D > y)\Pr(X > x, D = y)} = \text{Odds ratio in } 2\times 2\text{table}$$

$$\begin{cases} \theta > 0: & \text{Positive dependence} \\ -1 < \theta < 0: & \text{Negative dependence} \end{cases}$$

- $\text{Kendall's tau} = \dfrac{\theta}{\theta + 2}$

	Relapse	Relapse-free
Death	$X=x,D=y$	$X>x, D=y$
Alive	$X=x,D>y$	$X>x, D>y$

Ovarian cancer data (Ganzfried et al., 2013)

	Sample size	The number of observed events (event rates)			The number
		Relapse	Death	Censoring	of genes
Japanese	$N_1 = 84$	59 (70%)	38 (45%)	46 (55%)	18,548
American	$N_2 = 58$	48 (83%)	36 (62%)	22 (38%)	18,524
Australian	$N_3 = 260$	185 (71%)	113 (43%)	147 (57%)	18,524
American	$N_4 = 510$	252 (49%)	278 (55%)	232 (45%)	12,211
Total	$\sum_{i=1}^{4} N_i = 912$	544 (60%)	465 (51%)	447 (49%)	Common=11,756

Notes: The data are extracted from R Bioconductor *curatedOvarianData* package

Heterogeneity (random effects)

Dependence (Clayton copula)

High-dimensional covariates

Methods for high-dimensional covariates

- Lasso (Cox-regression with L_1 penalty)

  Tibshirani (1997 Stat Med), Gui & Li (2005 Bioinformatics)

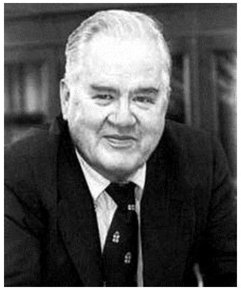
- Ridge regression (Cox-regression with L_2 penalty)

  Verveij & van Howelingen(1994 Stat. Med.), Zhao et al. (2011 PONE)

- Univariate selection (forward selection via univariate Cox –

  regression Jenssen et al. (2002 Nature Med), Chen et al. (2007 NEJM)

- Compound covariate (adopted for this research)



John Tukey

Tukey (1993 Controlled Clinical Trial), Matsui (2006, BMC Bioinfomatics),

Simon et al (2011 Boinfo), Matsui et al (2012 Clin Can Res)

Emura et al (2012 PONE), Emura et al. (2018- CMPB)

Univariate Feature Selection & Compound covariate

Step1: Univariate Cox model for each gene

$$\lambda(t \mid V_k) = \lambda_0(t)\exp(\beta_k V_k),$$

$V_k = k$ -th gene expression ($k = 1, \ldots, p$)

Step2: **Wald test via:** $z_j = \hat{\beta}_j / SE(\hat{\beta}_j)$

$$H_o : \beta_k = 0 \quad \text{vs.} \quad H_1 : \beta_k \neq 0$$

Step3 : **Select genes** with P-value < 0.001

$$\Omega = \{\, k : P_k < 0.0001 \,\} \qquad P_j = \Pr(|Z| > |z_j|)$$

Step4 : Compound covariates based on selected genes

$$CC = \hat{\beta}_1 V_1 + \cdots + \hat{\beta}_q V_q$$

Proposed method (1/3)

- **Step 1: Selected genes**

$$\begin{cases} \mathbf{V}_{ij} = (V_{ij,1}, \ldots, V_{ij,q_1}) & \text{: associated with relapse } X_{ij} \\ \mathbf{W}_{ij} = (W_{ij,1}, \ldots, W_{ij,q_2}) & \text{: associated with death } D_{ij} \end{cases}$$

$$\begin{cases} r_{ij}(t) = r_0(t)\exp(b_k V_{ij,k}), & q_1 \text{ : the number of genes with } P < 0.001 \\ \lambda_{ij}(t) = \lambda_0(t)\exp(c_k W_{ij,k}), & q_2 \text{ : the number of genes with } P < 0.001 \end{cases}$$

for *k*-th gene

- **Step 2: compound covariate (CC) predictors**

$$\begin{cases} \mathrm{CC}_{1,ij} = \hat{b}_1 V_{ij,1} + \cdots + \hat{b}_{q_1} V_{ij,q_1} & \text{: associated with relapse } X_{ij} \\ \mathrm{CC}_{2,ij} = \hat{c}_1 W_{ij,1} + \cdots + \hat{c}_{q_2} W_{ij,q_2} & \text{: associated with death } D_{ij} \end{cases}$$

Proposed method (2/3)

- **Step 3**: Fit the joint frailty-copula model (Emura et al. 2015 *SMMR*)

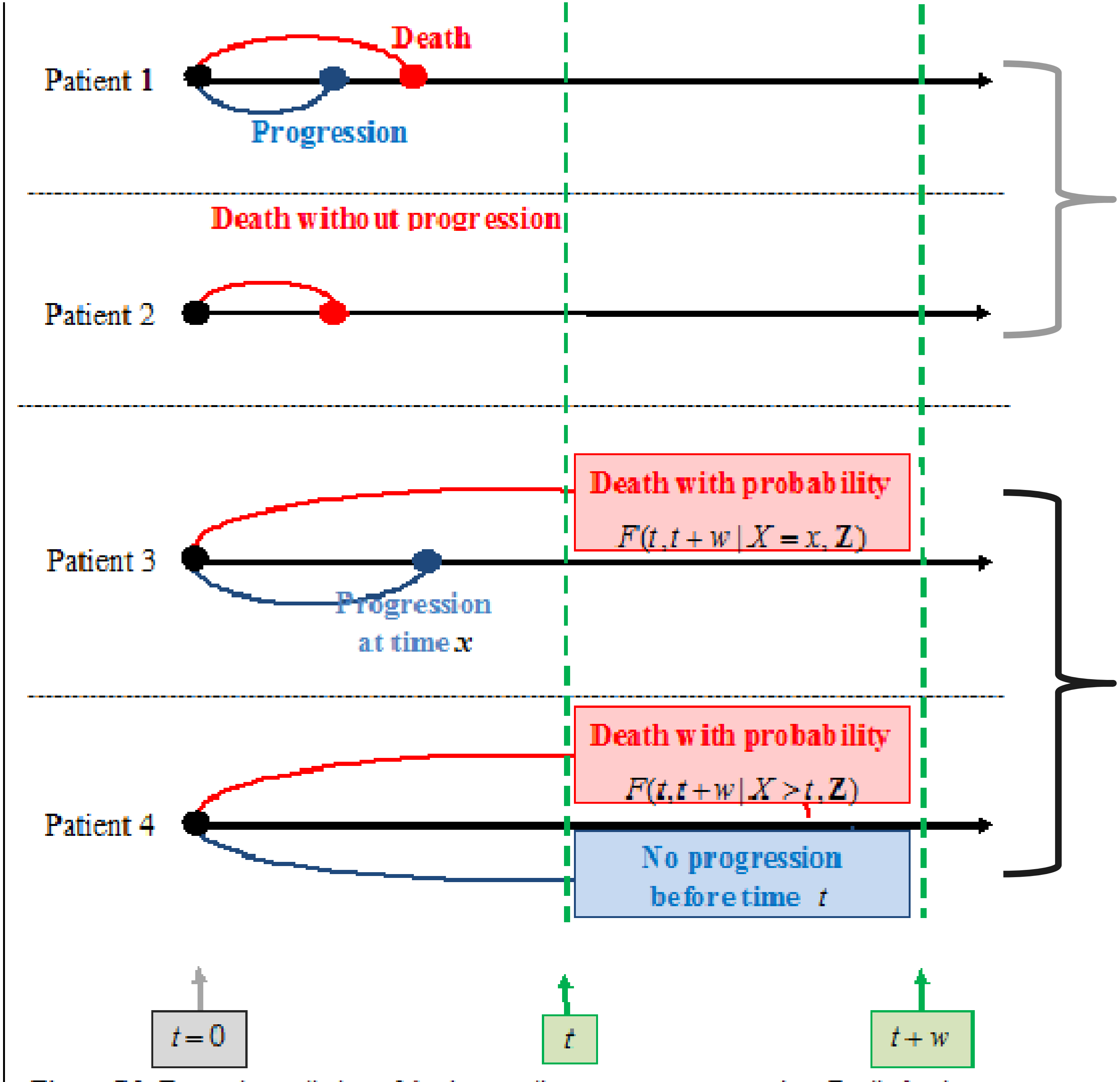
$$\begin{cases} r_{ij}(\, t \mid u_i \,) = u_i r_0(t) \exp(\, \boldsymbol{\beta}_1' \mathbf{Z}_{1,\,ij} + \gamma_1 \mathrm{CC}_{1,ij} \,) & \text{for} \quad X_{ij} \\ \lambda_{ij}(\, t \mid u_i \,) = u_i^{\alpha} \lambda_0(t) \exp(\, \boldsymbol{\beta}_2' \mathbf{Z}_{2,\,ij} + \gamma_2 \mathrm{CC}_{2,ij} \,) & \text{for} \quad D_{ij} \\ \Pr(\, X_{ij} > x \,,\, D_{ij} > y \mid u_i \,) = C_\theta[\, S_X(x \mid u_i), S_D(y \mid u_i) \,] & \end{cases}$$

for the ***i***-th study and ***j***-th patient

The Clayton copula: $C_\theta(\, v, w \,) = (\, v^{-\theta} + w^{-\theta} - 1 \,)^{-1/\theta}, \quad \theta \geq 0$

Estimator $(\, \hat{\theta}, \hat{\eta}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\gamma}_1, \hat{\gamma}_2, \hat{r}_0, \hat{\lambda}_0 \,)$

→ R package *joint.Cox* (Emura, 2017 on CRAN)



Out of interest

Two different formulas

Proposed method (3/3)

- If the patient does not experience tumour progression before *t*,

$$\hat{F}(t,t+w\mid X>t,\mathbf{Z})=\Pr(D\le t+w\mid D>t,\,X>t,\,\mathbf{Z})$$

$$=\frac{\int_0^\infty\left(C_\theta[\,S_X(t\mid u),\,S_D(t\mid u)\,]-C_\theta[\,S_X(t\mid u),\,S_D(t+w\mid u)\,]\right)f_\eta(u)du}{\int_0^\infty C_\theta[\,S_X(t\mid u),\,S_D(t\mid u)\,]f_\eta(u)du}$$

$$(\,\hat{\theta},\hat{\eta},\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2,\hat{\gamma}_1,\hat{\gamma}_2,\hat{r}_0,\hat{\lambda}_0\,)$$

- If the patient experiences tumour progression before *t*,

$$\hat{F}(t,t+w\mid X=x,\,\mathbf{Z})=\Pr(D\le t+w\mid D>t,\,X=x,\,\mathbf{Z})$$

$$=\frac{\int_0^\infty\left(C_\theta^{[1,0]}[\,S_X(x\mid u),\,S_D(t\mid u)\,]-C_\theta^{[1,0]}[\,S_X(x\mid u),\,S_D(t+w\mid u)\,]\right)uS_X(x\mid u)f_\eta(u)du}{\int_0^\infty C_\theta^{[1,0]}[\,S_X(x\mid u),\,S_D(t\mid u)\,]uS_X(x\mid u)f_\eta(u)du}$$

Data analysis (Ganzfried et al., 2013)

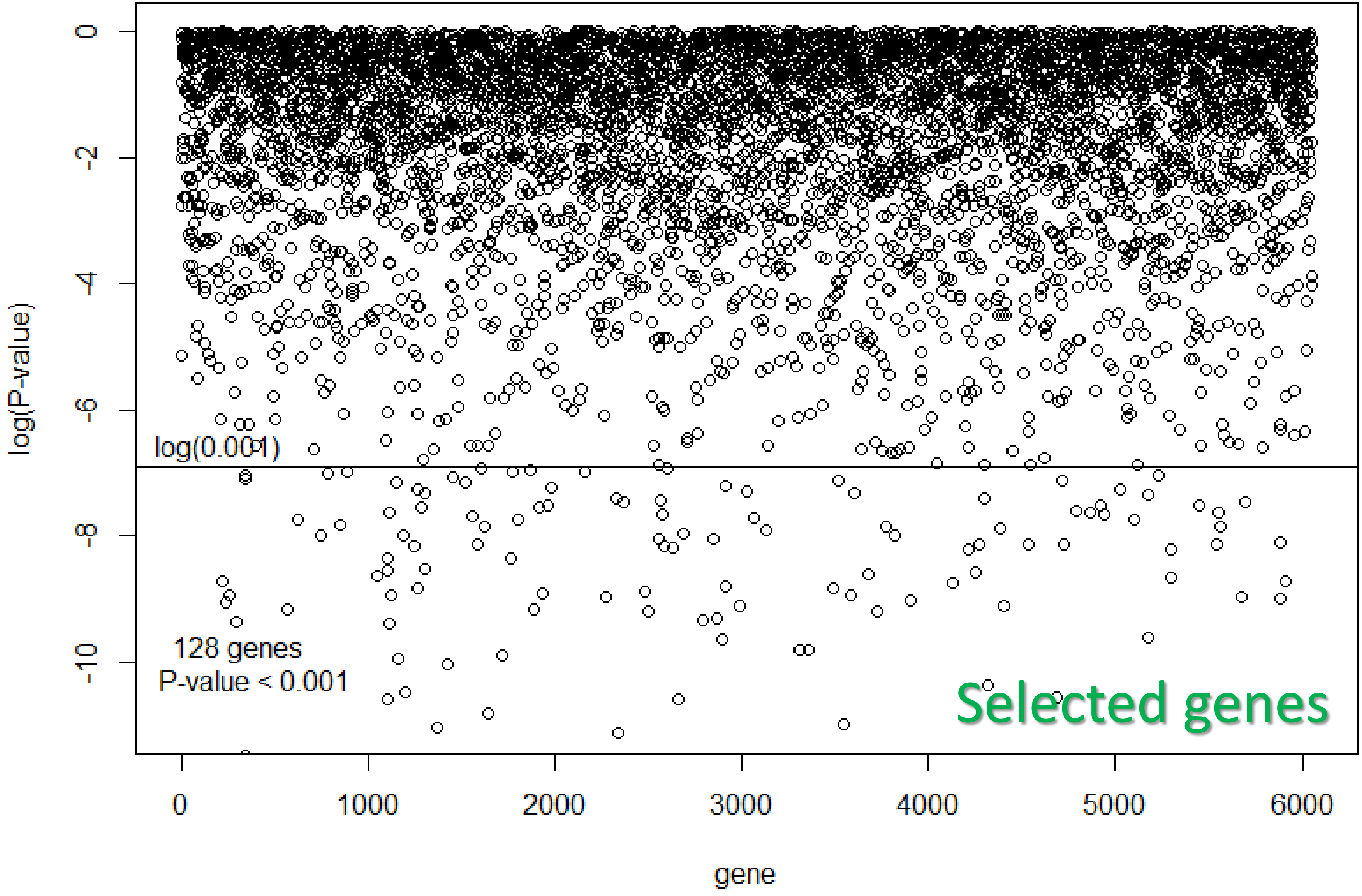
A meta-analytic data combining the four independent studies of ovarian cancer patients

	Sample size	The number of observed events (event rates)			The number of genes
		Relapse	Death	Censoring	
Study 1	$N_1 = 84$	59 (70%)	38 (45%)	46 (55%)	18,548
Study 2	$N_2 = 58$	48 (83%)	36 (62%)	22 (38%)	18,524
Study 3	$N_3 = 260$	185 (71%)	113 (43%)	147 (57%)	18,524
Study 4	$N_4 = 510$	252 (49%)	278 (55%)	232 (45%)	12,211
Total	$\sum_{i=1}^{4} N_i = 912$	544 (60%)	465 (51%)	447 (49%)	Common=11,756

Notes: The data are extracted from R Bioconductor *curatedOvarianData* package

Select genes with P-value =0.001

Univariate association between gene and time-to-death



Selected genes

Data Analysis: model fitting

Joint frailty-copula model

$$\begin{cases} r_{ij}(t \mid u_i) = u_i r_0(t) \exp(\gamma_1 \mathrm{CC}_{1,ij}) & \text{(for time to relapse } X_{ij}\text{)} \\ \lambda_{ij}(t \mid u_i) = \lambda_0(t) \exp(\beta_2 Z_{2,ij} + \gamma_2 \mathrm{CC}_{2,ij}) & \text{(for time to death } D_{ij}\text{)} \end{cases}$$

Clinical covariate:

$\mathbf{Z}_{2,ij}$=the residual tumour size at surgery (<1cm vs. ≥ 1cm)

Compound covariate (CC):

- $\mathrm{CC}_{1,ij}$ =(0.249**CXCL12*)+(0.235**TIMP2*)+(0.222**PDPN*)+⋯+(-0.152**MMP12*),

  involving 158 genes (P-value < 0.001 for time-to-relapse)

- $\mathrm{CC}_{2,ij}$ =(0.237**NCOA3*)+(0.223**TEAD1*)+(0.263**YWHAB*)+⋯+(-0.157**KCNH4*),

  invloving 128 genes (P-value < 0.001 for time-to-death).

Data Analysis: model fitting

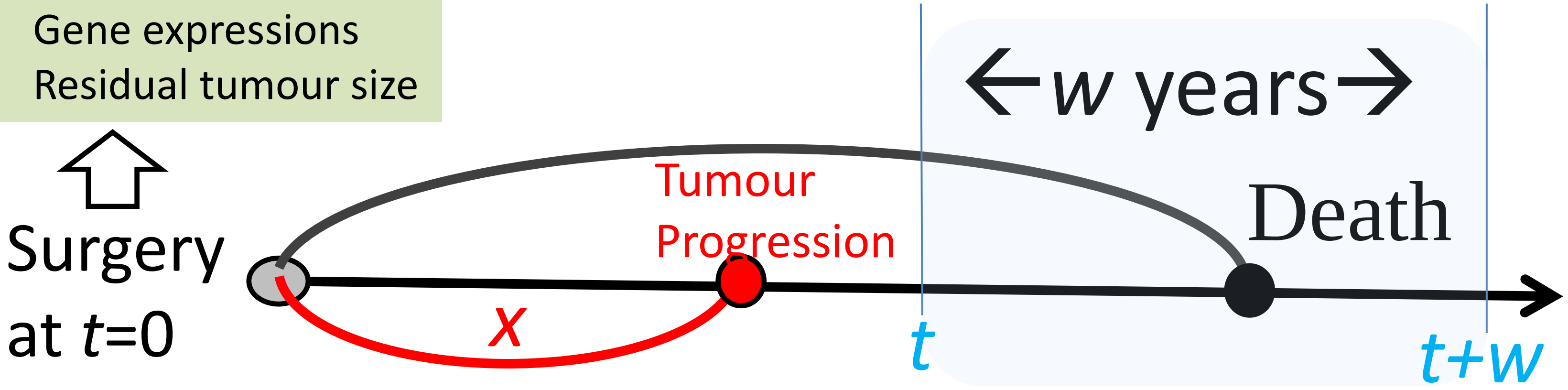
$$\begin{cases} r_{ij}(t \mid u_i) = u_i r_0(t)\exp(\gamma_1 \mathrm{CC}_{1,ij}) & \text{(for time to relapse } X_{ij}\text{)} \\ \lambda_{ij}(t \mid u_i) = \lambda_0(t)\exp(\beta_2 Z_{2,ij} + \gamma_2 \mathrm{CC}_{2,ij}) & \text{(for time to death } D_{ij}\text{)} \end{cases}$$

$$\Pr(X_{ij} > x, D_{ij} > y \mid u_i) = C_\theta[S_X(x \mid u_i), S_D(y \mid u_i)]$$

	Parameter	Estimate	95% CI
Relapse	$\exp(\gamma_1)$	1.48	1.37-1.59
Death	$\exp(\beta_2)$	1.18	1.03-1.35
	$\exp(\gamma_2)$	1.56	1.44-1.70
Copula	θ	1.90	1.49-2.42
	$\tau = \theta/(\theta+2)$	0.49	0.32-0.65

Estimated prediction formula

- Gene expressions
- Residual tumour size



Estimated conditional failure function

$$\hat{F}(t, t+w \mid X = x, \mathbf{Z}) = \hat{\Pr}(D \le t+w \mid D > t,\ X = x,\ \mathbf{Z})$$

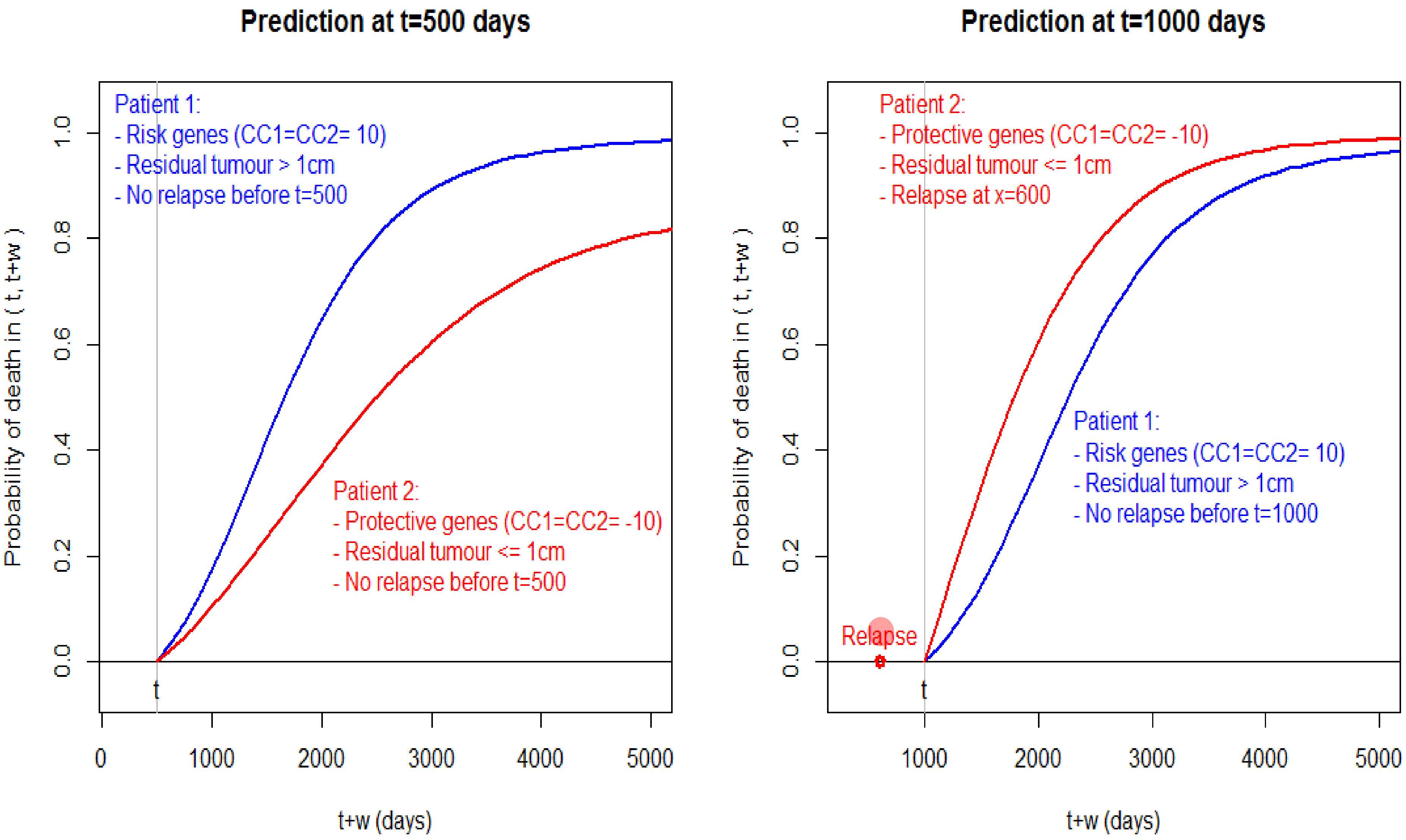
$$= \frac{\int_0^\infty \left(C_{\hat\theta}^{[1,0]}[\,\hat{S}_X(x \mid u), \hat{S}_D(t \mid u)\,] - C_{\hat\theta}^{[1,0]}[\,\hat{S}_X(x \mid u), \hat{S}_D(t+w \mid u)\,] \right) u \hat{S}_X(x \mid u) f_{\hat\eta}(u)\,du}{\int_0^\infty C_{\hat\theta}^{[1,0]}[\,\hat{S}_X(x \mid u), \hat{S}_D(t \mid u)\,] u \hat{S}_X(x \mid u) f_{\hat\eta}(u)\,du},$$

$$\hat{S}_X(\,t \mid u\,) = \exp\left\{-u\hat{R}_0(t)\exp(\,\hat\gamma_1 \mathrm{CC}_1\,)\right\},$$

$$\hat{S}_D(\,t \mid u_i\,) = \exp\left\{-u^{\hat\alpha}\hat\Lambda_0(t)\exp(\,\beta_2 Z_2 + \hat\gamma_2 \mathrm{CC}_2\,)\right\},$$

$$\left.\begin{aligned} \mathrm{CC}_{1,ij} &= (0.249^*\mathit{CXCL12}) + (0.235^*\mathit{TIMP2}) + (0.222^*\mathit{PDPN}) + \cdots + (-0.152^*\mathit{MMP12}) \\ \mathrm{CC}_{2,ij} &= (0.237^*\mathit{NCOA3}) + (0.223^*\mathit{TEAD1}) + (0.263^*\mathit{YWHAB}) + \cdots + (-0.157^*\mathit{KCNH4}) \end{aligned}\right\} \text{Compound covariate}$$

$$F(t, t+w \mid X = x, \mathbf{Z}) = \Pr(D \le t+w \mid D > t, X = x, \mathbf{Z})$$



Future works

- Gamma frailty

  ➔ Log-normal frailty (bivariate)

  Ongoing research with Dr. Rondeau
- Splines ➔ Weibull (conjugate for gamma frailty)

  Ongoing research with Wu BH (master student)
- Relationship between the joint frailty-copula model and sub-distribution hazard model (Ha et al. 2016)

  discussing with Prof. Ha
- Welcome to propose a new future work!