Interval-censored semi-competing risks data

N. Porta[1] M.L. Calle[2] G. Gómez[1]

[1]Universitat Politècnica de Catalunya

[2]Universitat de Vic

30th Anual Conference of the
International Society of Clinical Biostatistics

Prague, 26 August 2009.

grup de recerca en anàlisi estadística de la supervivència
grass

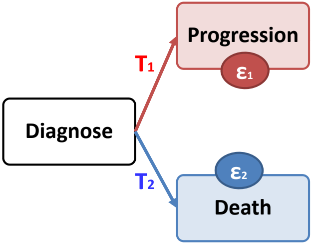
Departament d'Estadística i Investigació Operativa
UNIVERSITAT POLITÈCNICA DE CATALUNYA

OUTLINE

Competing risks data

- Event of interest: **progression of the disease** ($\mathbf{T_1}$).
- Competing event: **death** ($\mathbf{T_2}$).

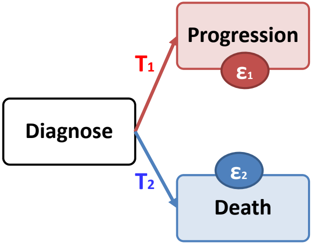
Competing Risks to analyze ($T = \min(T_1, T_2), C = 1$) $\Rightarrow$ CIF.

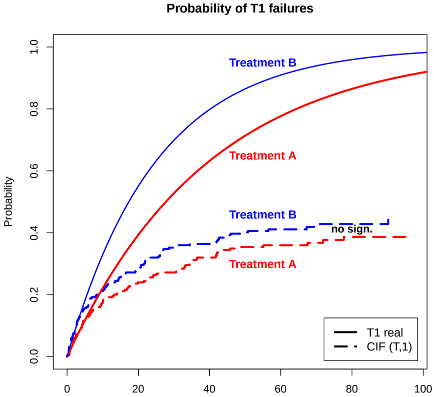


Competing risks data

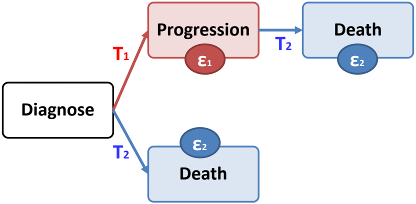
- Event of interest: **progression of the disease** ($\mathbf{T_1}$).
- Competing event: **death** ($\mathbf{T_2}$).

Competing Risks to analyze $(T = \min(T_1, T_2), C = 1) \Rightarrow$ CIF.

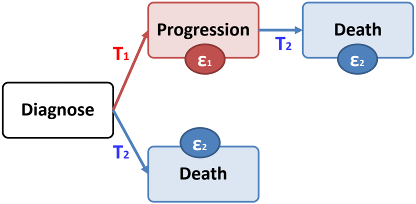


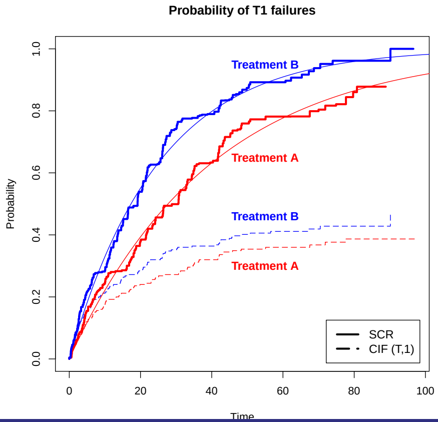


- Often, **more information** is available: death can occur AFTER progression $\Longrightarrow$ (T_1, T_2) can be estimated.
- Since death is a terminating event, T_2 censors T_1, possibly dependently $\Longrightarrow$ **Semi-competing risks**

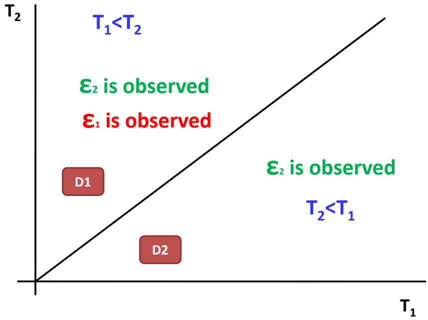


- Often, **more information** is available: death can occur AFTER progression $\Longrightarrow$ (T_1, T_2) can be estimated.
- Since death is a terminating event, T_2 censors T_1, possibly dependently $\Longrightarrow$ **Semi-competing risks**





Semi-competing risks data



Empirically,

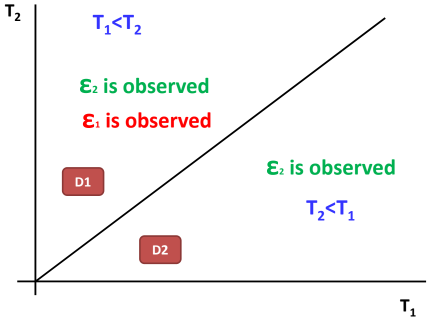
- $S(s,t) = P(T_1 > s, T_2 > t)$ is estimable in D1.
- $S_1(s) = P(T_1 > s)$ is not.

To recover T_1, we need to:

- Specify a valid model for (T_1, T_2) in D1.
- Derive the law of T_1 from the joint model.

In addition, T_1 is interval-censored in D1....

Semi-competing risks data



In addition, T_1 is interval-censored in D1....

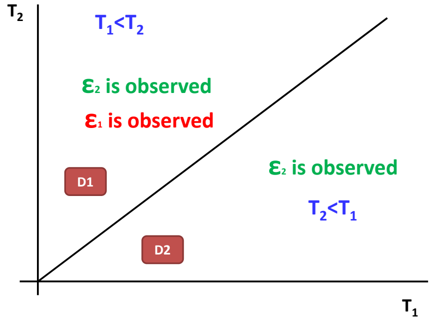
Empirically,

- $S(s,t) = P(T_1 > s, T_2 > t)$ is estimable in D1.
- $S_1(s) = P(T_1 > s)$ is not.

To recover T_1, we need to:

- **Specify a valid model for $(\mathbf{T}_1, \mathbf{T}_2)$ in D1.**
- **Derive the law of T_1 from the joint model.**

Semi-competing risks data



In addition, T_1 is interval-censored in D1....

Empirically,

- $S(s,t) = P(T_1 > s, T_2 > t)$ is estimable in D1.
- $S_1(s) = P(T_1 > s)$ is not.

To recover T_1, we need to:

- **Specify a valid model for $(\mathbf{T_1}, \mathbf{T_2})$ in D1.**
- **Derive the law of T_1 from the joint model.**

Interval-censored semi-competing risks data

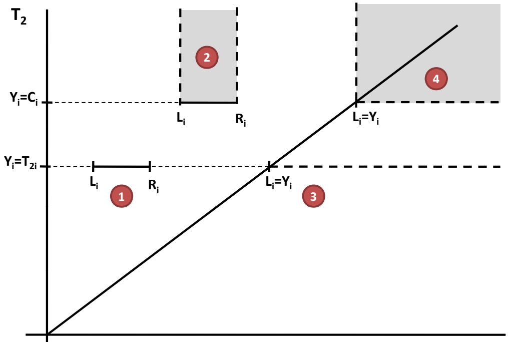
Consider a **semi-competing risks data** situation for (T_1, T_2) where T_1 is **interval-censored** :

- In $D1$, there exists L and $R < T_2$ such that $T_1 \in (L, R]$.
- T_2 is exactly observed or right-censored by independent C.
- Assume (L, R, C) censors non-informatively (T_1, T_2).

Observed data $(L_i, R_i, Y_i, \delta_{1i}, \delta_{2i})$

	δ_{2i}	$Y_i = T_{2i} \wedge C_i$	δ_{1i}	$T_{1i} \in$	
1	1	T_2	1	$(L, R]$	T_2 exact, T_1 interval-censored
2	0	C	1	$(L, R]$	T_2 right-censored, T_1 interval-censored
3	1	T_2	0	(L, ∞)	T_2 exact, T_1 right-censored
4	0	C	0	(L, ∞)	T_2, T_1 right-censored

* $\delta_{2i} = \mathbb{1}_{\{T_{2i} \leq C_i\}}$, $\delta_{1i} = \mathbb{1}_{\{R_i < \infty\}}$



A model for the association: Clayton's copula

Goal: To extend the methodolgy for right-censored SCR data proposed by Fine, J. P., Jiang, H., and Chappell, R. (2001). **On semi-competing risks data.** *Biometrika*, 88(4): 907-919.

The **joint survival function** in $D1$ is modelled via the **Clayton copula**:

$$S(s,t) = P(T_1 > s, T_2 > t) = \{S_1(s)^{1-\alpha} + S_2(t)^{1-\alpha} - 1\}^{\frac{1}{1-\alpha}} \qquad \alpha > 1$$

- α describes the association between T_1 and T_2.
- Given estimates for α, $S_2(t)$ and $S_T(t) = S(t,t)$, we can estimate

$$S_1(s) = \{S_T(s)^{1-\alpha} - S_2(s)^{1-\alpha} + 1\}^{\frac{1}{1-\alpha}}.$$

A model for the association: Clayton's copula

Goal: To extend the methodolgy for right-censored SCR data proposed by Fine, J. P., Jiang, H., and Chappell, R. (2001). **On semi-competing risks data.** *Biometrika*, 88(4): 907-919.

The **joint survival function** in $D1$ is modelled via the **Clayton copula**:

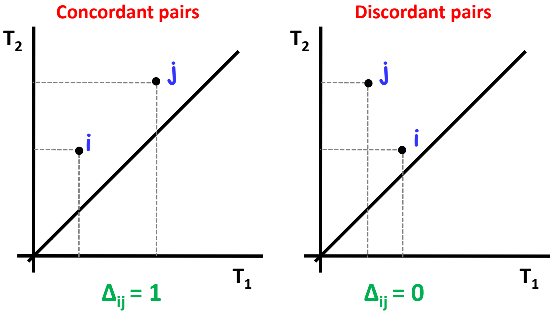
$$S(s,t) = P(T_1 > s, T_2 > t) = \{S_1(s)^{1-\alpha} + S_2(t)^{1-\alpha} - 1\}^{\frac{1}{1-\alpha}} \qquad \alpha > 1$$

- α describes the association between T_1 and T_2.
- Given estimates for α, $S_2(t)$ and $S_T(t) = S(t,t)$, we can estimate

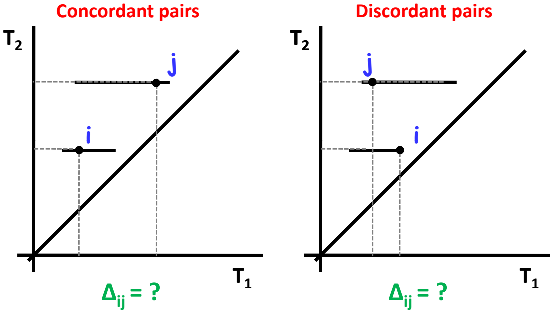
$$S_1(s) = \{S_T(s)^{1-\alpha} - S_2(s)^{1-\alpha} + 1\}^{\frac{1}{1-\alpha}}.$$

Estimation of α

The estimation of α is based on the **concordance indicator:**
$\Delta_{ij} = \mathbb{1}_{\{(T_{1i}-T_{1j})(T_{2i}-T_{2j})>0\}}$.



When T_{1i} and T_{1j} are interval-censored, in general we cannot compute the concordance indicator Δ_{ij}:



The expected concordance Z_{ij}

Definition (Expected concordance)

$$Z_{ij} = E[\Delta_{ij}|\mathcal{H}_{ij}] = P[\Delta_{ij} = 1|\mathcal{H}_{ij}]$$

where

$$\mathcal{H}_{ij} = \{(a_i, b_i, y_i, \delta_{1i}, \delta_{2i}), (a_j, b_j, y_j, \delta_{1j}, \delta_{2j})\}$$

is the observed data for the pair (i, j).

Example of observed data:

$$\mathcal{H}_{ij} = \{(a_i, b_i, y_i, 1, 1), (a_j, \infty, y_j, 0, 1)\} \Rightarrow \begin{cases} T_{1i} \in (a_i, b_i], \; T_{2i} = y_i, \\ T_{1j} \in (a_j, \infty), \; T_{2j} = y_j \end{cases}$$

$$Z_{ij} = \frac{1}{P[\mathcal{H}_{ij}]} \left(\delta_{2i}\delta_{2j}P_1(i,j) + \delta_{2i}(1-\delta_{2j})P_2(i,j) + (1-\delta_{2i})\delta_{2j}P_2(j,i) \right),$$

with

$$\begin{aligned} P_1(i,j) &= P(\Delta_{ij} = 1, \mathcal{H}_{ij}, \delta_{2i} = 1, \delta_{2j} = 1) \\ &= \int_{a_i}^{b_i} \int_{a_j}^{b_j} \mathbb{1}_{\{(x-u)(y_i-y_j)>0\}} f(x,y_i) f(u,y_j) du dx \\ P_2(i,j) &= P(\Delta_{ij} = 1, \mathcal{H}_{ij}, \delta_{2i} = 1, \delta_{2j} = 0) \\ &= \int_{y_j}^{\infty} \int_{a_i}^{b_i} \int_{a_j}^{b_j} \mathbb{1}_{\{(x-u)(y_i-v)>0\}} f(x,y_i) f(u,v) du dx dv, \end{aligned}$$

and $f(s,t) = \partial S^2 / \partial s \partial t \Longrightarrow$ Z_{ij} **depends on** S_1**,** S_2 **and** α.

Estimating equations

Since $E[\Delta_{ij}] = E[Z_{ij}] = \frac{\alpha}{\alpha+1}$ under Clayton's copula model,

- **Right-censoring:**

$$U^R(\alpha) = \binom{n}{2}^{-1} \sum_{i<j} O_{ij}^R \left\{ \Delta_{ij} - \frac{\alpha}{\alpha+1} \right\} = 0$$

where $O_{ij}^R = 1 \Leftrightarrow \Delta_{ij}$ is determined.

- **Interval-censoring:**

$$U_0(\alpha) = \binom{n}{2}^{-1} \sum_{i<j} O_{ij} \left\{ Z_{ij} - \frac{\alpha}{\alpha+1} \right\} = 0,$$

where $O_{ij} = 1 \Leftrightarrow Z_{ij}$ is determined.

Estimating equations

Since $E[\Delta_{ij}] = E[Z_{ij}] = \frac{\alpha}{\alpha+1}$ under Clayton's copula model,

- **Right-censoring:**

$$U^R(\alpha) = \binom{n}{2}^{-1} \sum_{i<j} O^R_{ij} \left\{ \Delta_{ij} - \frac{\alpha}{\alpha+1} \right\} = 0$$

where $O^R_{ij} = 1 \Leftrightarrow \Delta_{ij}$ is determined.

- **Interval-censoring:**

$$U_0(\alpha) = \binom{n}{2}^{-1} \sum_{i<j} O_{ij} \left\{ Z_{ij} - \frac{\alpha}{\alpha+1} \right\} = 0,$$

where $O_{ij} = 1 \Leftrightarrow Z_{ij}$ is determined.

- Fine et al.(2001) showed that $E[U^R(\alpha)] = 0$ and $\widehat{\alpha}_R$ is obtained as a root of $U^R(\alpha) = 0$.
- For known $S_1(\cdot)$ and $S_2(\cdot)$, equation $U_0(\alpha) = 0$ is biased, because the comparable pairs are not selected at random. In fact:

$$E\left[U_0(\alpha)\right] = \underbrace{E\left[U^R(\alpha)\right]}_{0} + n_p \frac{\alpha}{\alpha + 1}$$

where n_p **is the proportion of individuals satisfying** $O_{ij}^R = 1$ **but** $O_{ij} = 0$**.**

- Fine et al.(2001) showed that $E[U^R(\alpha)] = 0$ and $\widehat{\alpha}_R$ is obtained as a root of $U^R(\alpha) = 0$.
- For known $S_1(\cdot)$ and $S_2(\cdot)$, equation $U_0(\alpha) = 0$ is biased, because the comparable pairs are not selected at random. In fact:

$$E\left[U_0(\alpha)\right] = \underbrace{E\left[U^R(\alpha)\right]}_{0} + n_p \frac{\alpha}{\alpha + 1}$$

where n_p **is the proportion of individuals satisfying** $O_{ij}^R = 1$ **but** $O_{ij} = 0$**.**

For ICSCR, n_p is never observed, but can be estimated from a subsample of the non-comparable pairs ($O_{ij} = 0$) from expressions like:

$$\widehat{n}_p = \frac{1}{\binom{n}{2}} \sum_{(i,j)} P(T_{1i} \in (a_i, a_j], T_{2i} = y_i \big| T_{1i} \in (a_i, b_i], T_{2i} = y_i, y_i > y_j, a_i < a_j).$$

Then, given $S_1(\cdot)$ and $S_2(\cdot)$ known, an unbiased estimating equation is obtained:

$$U_1(\alpha) = \binom{n}{2}^{-1} \sum_{i<j} O_{ij} \left\{ Z_{ij} - \frac{\alpha}{\alpha+1} \right\} - \widehat{n}_p \frac{\alpha}{\alpha+1} = 0.$$

For ICSCR, n_p is never observed, but can be estimated from a subsample of the non-comparable pairs ($O_{ij} = 0$) from expressions like:

$$\widehat{n}_p = \frac{1}{\binom{n}{2}} \sum_{(i,j)} P(T_{1i} \in (a_i, a_j], T_{2i} = y_i \big| T_{1i} \in (a_i, b_i], T_{2i} = y_i, y_i > y_j, a_i < a_j).$$

Then, given $S_1(\cdot)$ and $S_2(\cdot)$ known, an unbiased estimating equation is obtained:

$$U_1(\alpha) = \binom{n}{2}^{-1} \sum_{i<j} O_{ij} \left\{ Z_{ij} - \frac{\alpha}{\alpha + 1} \right\} - \widehat{n}_p \frac{\alpha}{\alpha + 1} = 0.$$

The iterative estimation algorithm

INITIAL PHASE Obtain $\hat{S}_2(\cdot)$, $\hat{S}_T(\cdot)$, $\widehat{\alpha}^{(0)}$, $\hat{S}_1(\cdot)^{(0)}$ and O_{ij} for all pairs $i < j$.

ITERATIVE PHASE

Repeat until convergence:

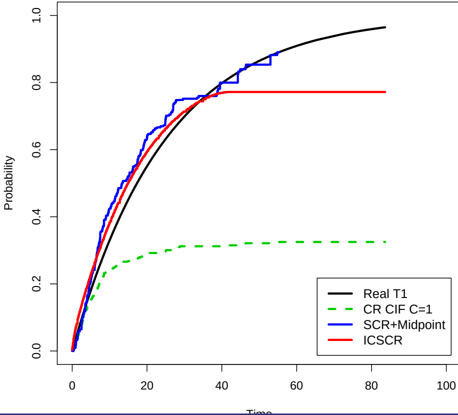
1. Compute $Z_{ij}^{(k-1)} = Z_{ij}\big(\widehat{\alpha}^{(k-1)}, \hat{S}_1(\cdot)^{(k-1)}, \hat{S}_2(\cdot)\big)$.
2. Obtain $\hat{n}_p = n_p\big(\widehat{\alpha}^{(k-1)}, \hat{S}_1(\cdot)^{(k-1)}, \hat{S}_2(\cdot)\big)$.
3. Find $\widehat{\alpha}^{(k)}$ as a solution of $U_1(\alpha; Z_{ij}^{(k-1)}, \hat{n}_p) = 0$.
4. Update $\hat{S}_1(s)^{(k)} = \{\hat{S}_T(s)^{1-\widehat{\alpha}^{(k)}} - \hat{S}_2(s)^{1-\widehat{\alpha}^{(k)}} + 1\}^{\frac{1}{1-\widehat{\alpha}^{(k)}}}$.

Illustration

Simulated data set ($n = 500$): $T_1, T_2 \sim$Exp, $E[T_1] = 65$, $E[T_2] = 40$ observed in $[0, 100]$, and $\alpha = 3$. A **62%** of dependent censoring results in the simulated data set.

Estimation of α:

	Estimate	SE
Real	3	
Midpoint+SCR	4.15	0.38
ICSCR	3.44	0.57



Simulation results:

Different scenarios considered varying *n*, α, % **dependent censoring** and **width of intervals**.

Figure: Bias

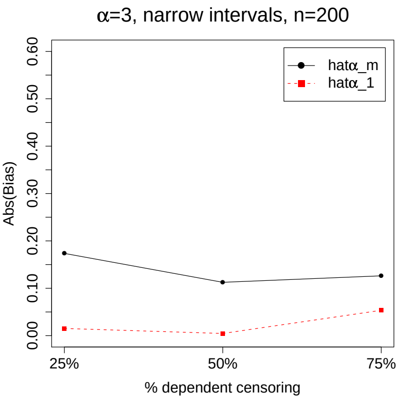


Figure: MSE

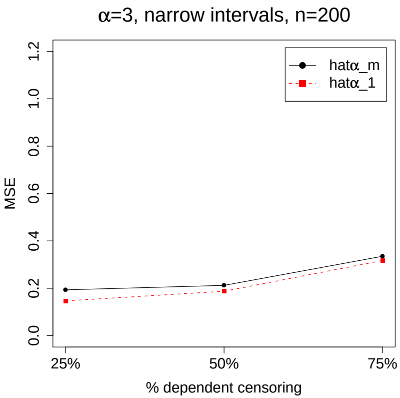


Figure: Bias

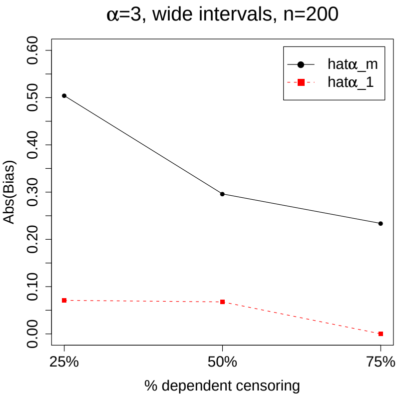
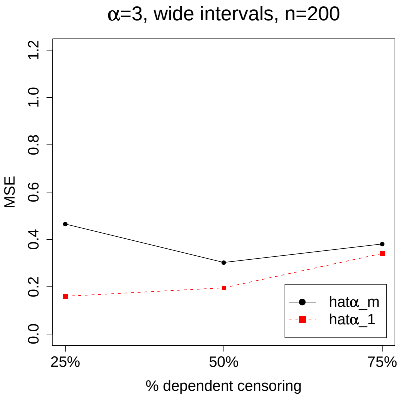


Figure: MSE



Conclusions

- Goals on a semi-competing risks data analysis:
  - association between T_1 and T_2, and
  - the marginal distribution of T_1.
- Under Clayton's copula model, we have proposed a method when T_1 is interval-censored, by considering the **expected concordance** Z_{ij}, new estimating equations for α and an iterative algorithm to jointly estimate α and $S_1(s)$.
- Our method ICSCR performs better than **midpoint imputation** which reduces the problem to right-censored data.
- **On-going work**: asymptotic properties.

Thanks for your attention!!!!

nuria.porta-bleda@upc.edu
http://www−eio.upc.es/research/grass/

This work is partially supported by grant MTM2008–06747–C02-00 from the Ministerio de Ciencia y Tecnología, Spain. Núria Porta is a recipient of a research fellowship from the Commission for Universities and Research of the Ministry of Innovation, Universities and Enterprise of the Government of Catalunya, and the European Social Funds.