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

30\(^{th}\) Annual Conference of the International Society of Clinical Biostatistics

Prague, 26 August 2009.
OUTLINE

1 Motivation

2 Interval censored semi-competing risks data

3 A model for \((T_1, T_2)\)

4 Estimation of \(\alpha\) and \(S_1(s)\)

5 Results

6 Conclusions
Interval-censored semi-competing risks data

**Motivation**

**Competing risks data**

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

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

- Event of interest: progression of the disease ($T_1$).
- Competing event: death ($T_2$).

**Competing Risks** to analyze ($T = \min(T_1, T_2), C = 1 \Rightarrow \text{CIF}$).
Often, **more information** is available: death can occur AFTER progression $\Rightarrow (T_1, T_2)$ can be estimated.

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

Since death is a terminating event, \(T_2\) censors \(T_1\), possibly dependently ⇒ **Semi-competing risks**

![Diagram of progression and death with events \(T_1\) and \(T_2\).]
Semi-competing risks data

In addition, $T_1$ is interval-censored in $D_1$.

Empirically,

- $S(s, t) = P(T_1 > s, T_2 > t)$ is estimable in $D_1$.
- $S_1(s) = P(T_1 > s)$ is not.

To recover $T_1$, we need to:

- Specify a valid model for $(T_1, T_2)$ in $D_1$.
- Derive the law of $T_1$ from the joint model.
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 $(T_1, T_2)$ in D1.
- Derive the law of $T_1$ from the joint model.
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....
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** \((L_i, R_i, Y_i, \delta_{1i}, \delta_{2i})\)

<table>
<thead>
<tr>
<th>(\delta_{2i})</th>
<th>(Y_i)</th>
<th>(T_{2i} \wedge C_i)</th>
<th>(\delta_{1i})</th>
<th>(T_{1i} \in)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>(T_2)</td>
<td>1</td>
<td>((L, R]) (T_2) exact, (T_1) interval-censored</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>(C)</td>
<td>1</td>
<td>((L, R]) (T_2) right-censored, (T_1) interval-censored</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>(T_2)</td>
<td>0</td>
<td>((L, \infty)) (T_2) exact, (T_1) right-censored</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>(C)</td>
<td>0</td>
<td>((L, \infty)) (T_2, T_1) right-censored</td>
</tr>
</tbody>
</table>

\(\delta_{2i} = \mathbb{1}\{T_{2i} \leq C_i\}\), \(\delta_{1i} = \mathbb{1}\{R_i < \infty\}\)
A model for the association: Clayton’s copula


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

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

- \( \alpha \) describes the association between \( T_1 \) and \( T_2 \).
- Given estimates for \( \alpha \), \( S_2(t) \) 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


The **joint survival function** in $D_1$ is modelled via the **Clayton copula**:

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

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

$$S_1(s) = \{S_T(s)^{1-\alpha} - S_2(s)^{1-\alpha} + 1\}^{\frac{1}{1-\alpha}}.$$
The estimation of $\alpha$ is based on the concordance indicator:

$$\Delta_{ij} = 1 \{ (T_{1i} - T_{1j})(T_{2i} - T_{2j}) > 0 \} .$$
When $T_{1i}$ and $T_{1j}$ are interval-censored, in general we cannot compute the concordance indicator $\Delta_{ij}$:
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:

$$\mathcal{H}_{ij} = \{(a_i, b_i, y_i, 1, 1), (a_j, \infty, y_j, 0, 1)\} \Rightarrow \left\{ \begin{array}{l}
T_{1i} \in (a_i, b_i), T_{2i} = y_i, \\
T_{1j} \in (a_j, \infty), T_{2j} = y_j
\end{array} \right\}$$
\[ Z_{ij} = \frac{1}{P[H_{ij}]} \left( \delta_{2i}\delta_{2j}P_1(i, j) + \delta_{2i}(1 - \delta_{2j})P_2(i, j) + (1 - \delta_{2i})\delta_{2j}P_2(j, i) \right), \]

with

\[ P_1(i, j) = P(\Delta_{ij} = 1, H_{ij}, \delta_{2i} = 1, \delta_{2j} = 1) = \int_{a_i}^{b_i} \int_{a_j}^{b_j} \mathbb{1}_{\{(x-u)(y_i-y_j) > 0\}} f(x, y_i) f(u, y_j) \, du \, dx \]

\[ P_2(i, j) = P(\Delta_{ij} = 1, H_{ij}, \delta_{2i} = 1, \delta_{2j} = 0) = \int_{\infty}^{\infty} \int_{y_j}^{b_i} \int_{a_j}^{b_j} \mathbb{1}_{\{(x-u)(y_i-v) > 0\}} f(x, y_i) f(u, v) \, du \, dx \, dv, \]

and \( f(s, t) = \partial S^2 / \partial s \partial t \implies Z_{ij} \text{ depends on } S_1, S_2 \text{ and } \alpha. \)
Estimating equations

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

- **Right-censoring:**

  \[
  U^R(\alpha) = \left( \begin{array}{c} n \\ 2 \end{array} \right)^{-1} \sum_{i<j} O^R_{ij} \left\{ \Delta_{ij} - \frac{\alpha}{\alpha + 1} \right\} = 0
  \]

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

- **Interval-censoring:**

  \[
  U_0(\alpha) = \left( \begin{array}{c} n \\ 2 \end{array} \right)^{-1} \sum_{i<j} O_{ij} \left\{ Z_{ij} - \frac{\alpha}{\alpha + 1} \right\} = 0,
  \]

  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 $O^R_{ij} = 1 \iff \Delta_{ij}$ 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 $O_{ij} = 1 \iff Z_{ij}$ is determined.
Fine et al. (2001) showed that $E[U^R(\alpha)] = 0$ and $\hat{\alpha}_R$ 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:

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

where $n_p$ is the proportion of individuals satisfying $O_{ij}^R = 1$ but $O_{ij} = 0$. 
Fine et al. (2001) showed that $E[U_R^R(\alpha)] = 0$ and $\hat{\alpha}_R$ is obtained as a root of $U_R^R(\alpha) = 0$.

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:

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

where $n_p$ is the proportion of individuals satisfying $O^R_{ij} = 1$ 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:

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

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

\[
U_1(\alpha) = \left(\frac{n}{2}\right)^{-1} \sum_{i<j} O_{ij} \left\{ Z_{ij} - \frac{\alpha}{\alpha + 1} \right\} - \hat{n}_p \frac{\alpha}{\alpha + 1} = 0.
\]
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:

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

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

\[
U_1(\alpha) = \binom{n}{2}^{-1} \sum_{i < j} O_{ij} \left\{ Z_{ij} - \frac{\alpha}{\alpha + 1} \right\} - \hat{n}_p \frac{\alpha}{\alpha + 1} = 0.
\]
The iterative estimation algorithm

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

**ITERATIVE PHASE**
Repeat until convergence:

1. Compute $Z_{ij}^{(k-1)} = Z_{ij}(\hat{\alpha}^{(k-1)}, \hat{S}_1(\cdot)^{(k-1)}, \hat{S}_2(\cdot))$.

2. Obtain $\hat{n}_p = n_p(\hat{\alpha}^{(k-1)}, \hat{S}_1(\cdot)^{(k-1)}, \hat{S}_2(\cdot))$.

3. Find $\hat{\alpha}^{(k)}$ as a solution of $U_1(\alpha; Z_{ij}^{(k-1)}, \hat{n}_p) = 0$.

4. Update $\hat{S}_1(s)^{(k)} = \{\hat{S}_T(s)^{1-\hat{\alpha}^{(k)}} - \hat{S}_2(s)^{1-\hat{\alpha}^{(k)}} + 1\}^{1-\hat{\alpha}^{(k)}}$. 
Simulated data set (n = 500): $T_1, T_2 \sim \text{Exp}$, $E[T_1] = 65$, $E[T_2] = 40$ observed in $[0, 100]$, and $\alpha = 3$. A 62% of dependent censoring results in the simulated data set.

### Estimation of $\alpha$:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Midpoint + SCR</td>
<td>4.15</td>
<td>0.38</td>
</tr>
<tr>
<td>ICSCR</td>
<td>3.44</td>
<td>0.57</td>
</tr>
</tbody>
</table>

![Graph showing probability over time for different estimation methods]
Simulation results:

Different scenarios considered varying $n$, $\alpha$, % dependent censoring and width of intervals.

**Figure: Bias**

\[\alpha=3, \text{narrow intervals, } n=200\]

- $\hat{\alpha}_m$
- $\hat{\alpha}_1$

**Figure: MSE**

\[\alpha=3, \text{narrow intervals, } n=200\]

- $\hat{\alpha}_m$
- $\hat{\alpha}_1$
Figure: Bias

$\alpha = 3$, wide intervals, $n=200$

% dependent censoring

Abs(Bias)

25% 50% 75%

0.00 0.10 0.20 0.30 0.40 0.50 0.60

$\hat{\alpha}_m$, $\hat{\alpha}_1$

Figure: MSE

$\alpha = 3$, wide intervals, $n=200$

MSE

25% 50% 75%

0.0 0.2 0.4 0.6 0.8 1.0 1.2

$\hat{\alpha}_m$, $\hat{\alpha}_1$
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 $\alpha$ and an iterative algorithm to jointly estimate $\alpha$ 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.