Quantifying the correlation of bivariate survival times by means of a novel self-consistency approach

Michael Schemper, Alexandra Kaider and Samo Wakounig

Section of Clinical Biometrics

Department of Medical Statistics and Informatics

Medical University of Vienna, Austria

Contents

- Practical relevance
- Current approaches
- Novel Iterative Multiple Imputation approach ('self-consistency')
- Three case studies
- Comparative empirical investigations

Practical relevance

- Correlations within pairs of possibly censored survival times are of interest in the analysis of :
  - Survival-type endpoints in studies of twins
  - Times till failure of paired organs
  - Surrogate survival-type endpoints
- Quantification by Kendall's τ or Spearman's r_s

Estimating correlations under censoring

- Fully non-parametric approaches based on the Kaplan-Meier estimate on the plane (i.e., for a bivariate survival function)

  $\longrightarrow$ 'finite region' estimates of correlation, if bivariate survival function does not drop to zero.

- Semiparametric approaches have a

  non-parametric component – transformation of survival times to survival probabilities, using Kaplan-Meier – and a

  parametric component – a parametric bivariate distribution is fitted over the whole range of transformed times

  $\longrightarrow$ 'infinite region' estimates of correlation (better interpretable, but extrapolation)

Copula approach (1)

Copula: bivariate parametric distribution with uniform margins U and V. Time $T_1 \to U$, time $T_2 \to V$.

A single parameter θ governs degree of correlation:

Bivariate distribution functions

Clayton $$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$$

Gumbel $$\exp\left\{-\left[\left(-\log u\right)^{\theta}+\left(-\log v\right)^{\theta}\right]^{1/\theta}\right\}$$

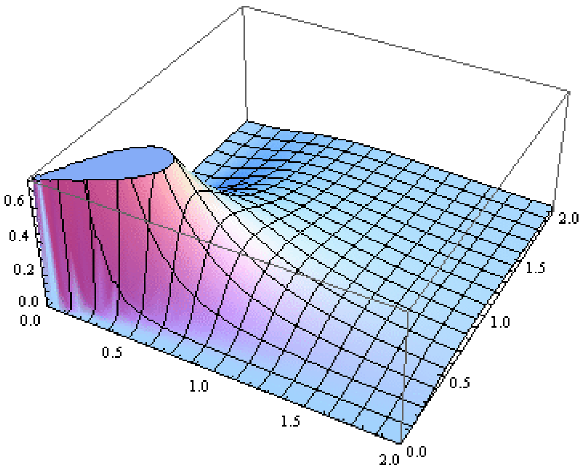
Frank $$-\frac{1}{\theta}\log\left\{1+\frac{\left[\exp\left(-\theta u\right)-1\right]\left[\exp\left(-\theta v\right)-1\right]}{\left[\exp\left(-\theta\right)-1\right]}\right\}$$

Normal $$\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^{2}}}\exp\left[-\frac{s_1^{2}-2\theta s_1 s_2+s_2^{2}}{2\left(1-\theta^{2}\right)}\right]ds_1\,ds_2$$

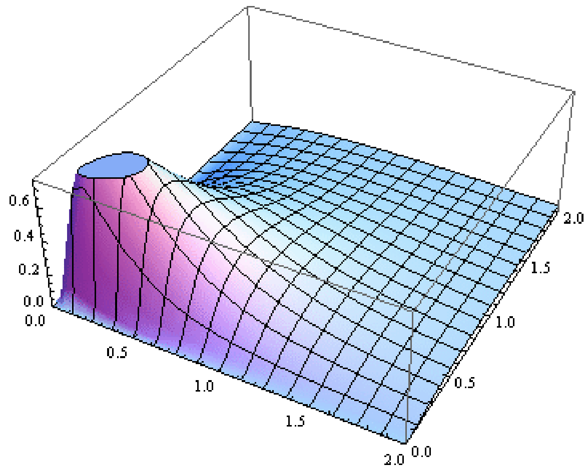
Copula approach (2)

Densities of copulas for r_s=0.6, lognormal marginal survival

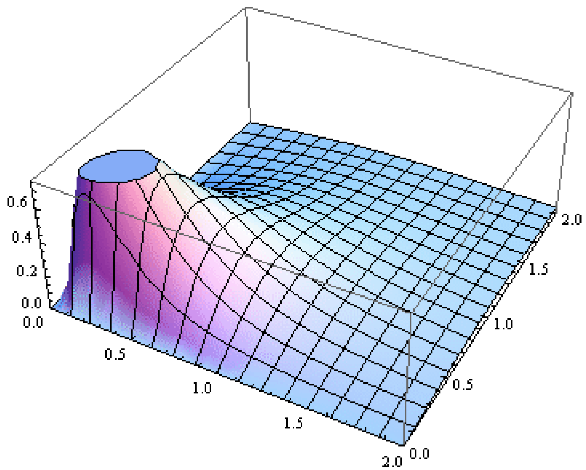
Clayton



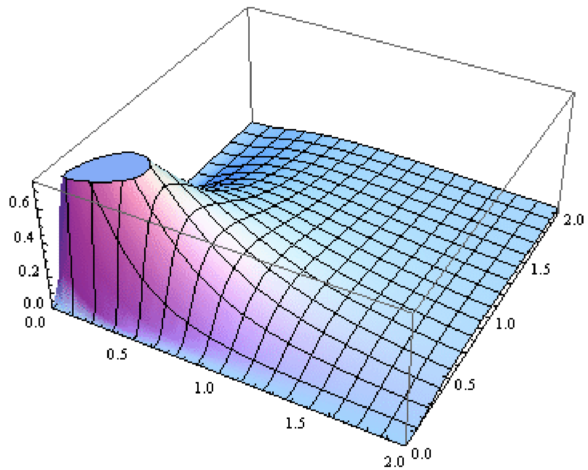
Gumbel



Frank



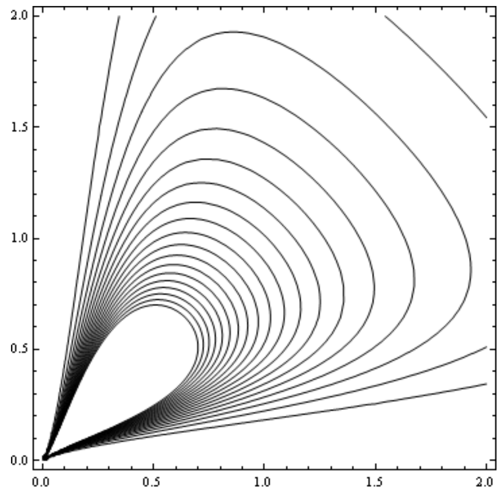
Normal



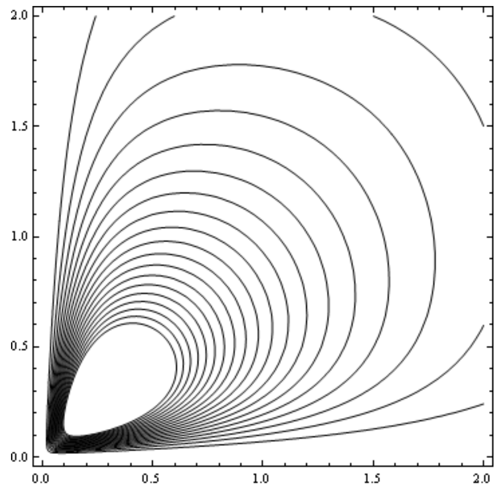
Copula approach (3)

Contour plots of copulas for $r_s = 0.6$

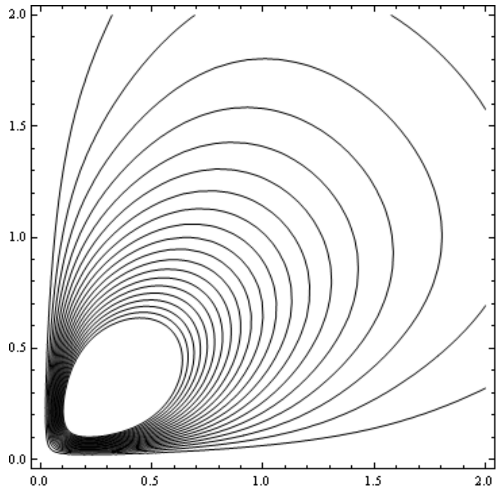
Clayton



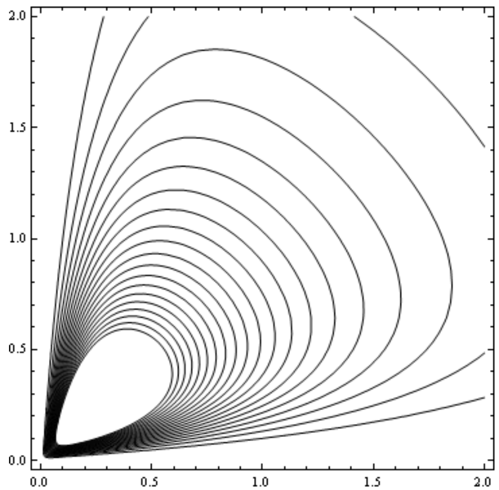
Gumbel



Frank



Normal



Iterative Multiple Imputation, IMI (Self-consistency approach)

Principle steps of algorithm:

1. Transform T_1 and T_2 to normal deviates N_1 and N_2

$$N_1 = \Phi^{-1}\left(S(T_1)\right) \qquad N_2 = \Phi^{-1}\left(S(T_2)\right)$$

normal copula idea

2. Obtain a starting value r_0 of the Pearson correlation from uncensored pairs (N_1, N_2)

normal scores Spearman

Iterative Multiple Imputation, IMI (2)

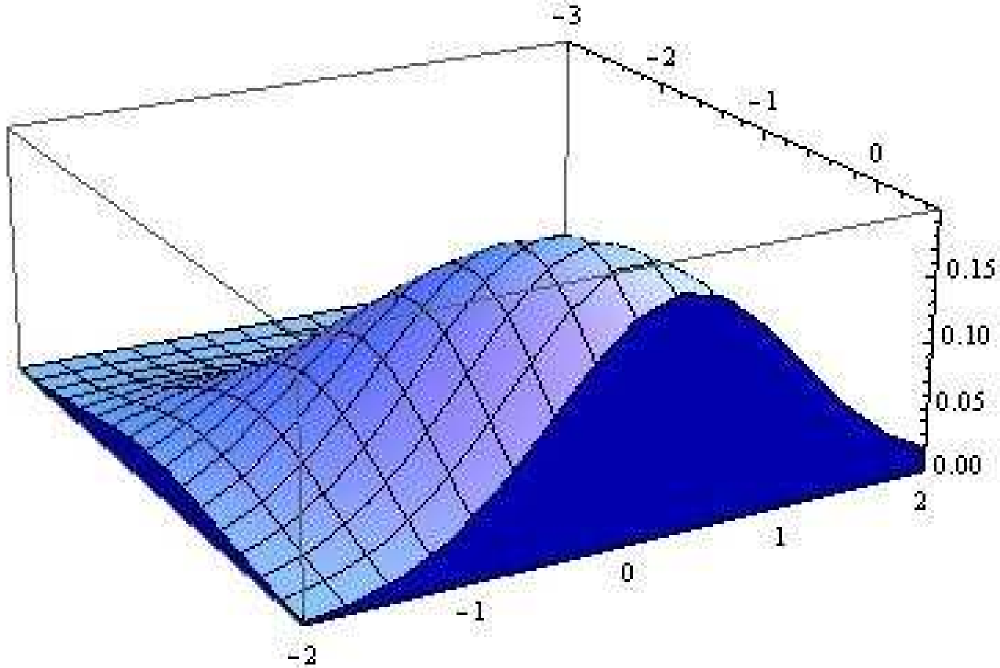
3. Produce M data sets with ‘uncensored‘ N_1 and N_2 :

- uncensored N_1 and N_2 are copied to each data set
- censored N_1 and/or N_2 are made ‘uncensored‘ by imputation, conditional on the censoring values N_1 and N_2 and using conditional normal distributions: e.g., $N_1 : N\left(r_0 N_2,\ (1-r_0^2)\right)$
- with both N_1 and N_2 censored a few MCMC steps are required.

Iterative Multiple Imputation, IMI (2a)

Conditional normal distributions:

$$N_1 : N\left(r_0 N_2,\ (1 - r_0^2)\right)$$



Iterative Multiple Imputation, IMI (3)

4. For each of the M data sets obtain r_1 from all, now 'uncensored' N_1 and N_2.
5. For each of the M data sets update the imputed N_1 and/or N_2 using r_1 .
6. The iterative process of **updating** N_1 and/or N_2 and then **updating** r is repeated until convergence of r within each data set.
7. The reported value of r is the average from the M data sets.

Iterative Multiple Imputation, IMI (4)

95% Confidence intervals for r :

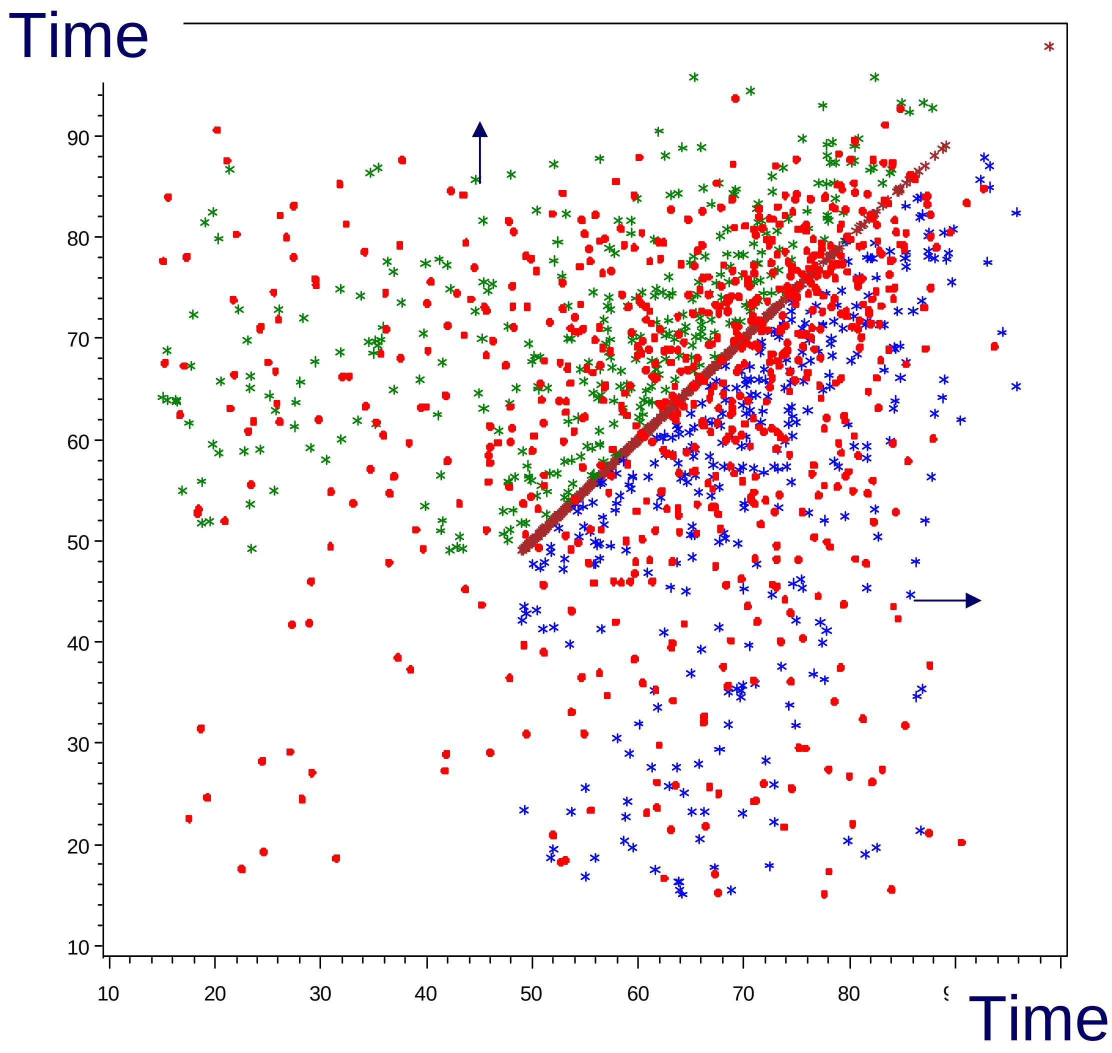
$$\left[\tanh\left(\varphi \pm 1.96\, SE(\varphi)\right)\right]$$

where $\varphi = \tanh^{-1}(r)$

and $SE(\varphi)$ is obtained from the ‘within‘ and ‘between‘ data sets variance of φ (Rubin, 1987)

Example 1

Danish Twin Survival Study (male, monozygotic)
(n = 1366, % censored times = 61)



Spearman correlations:

<u>Copulas:</u>

Clayton: $r_S = 0.10$

Gumbel: $r_S = 0.22$

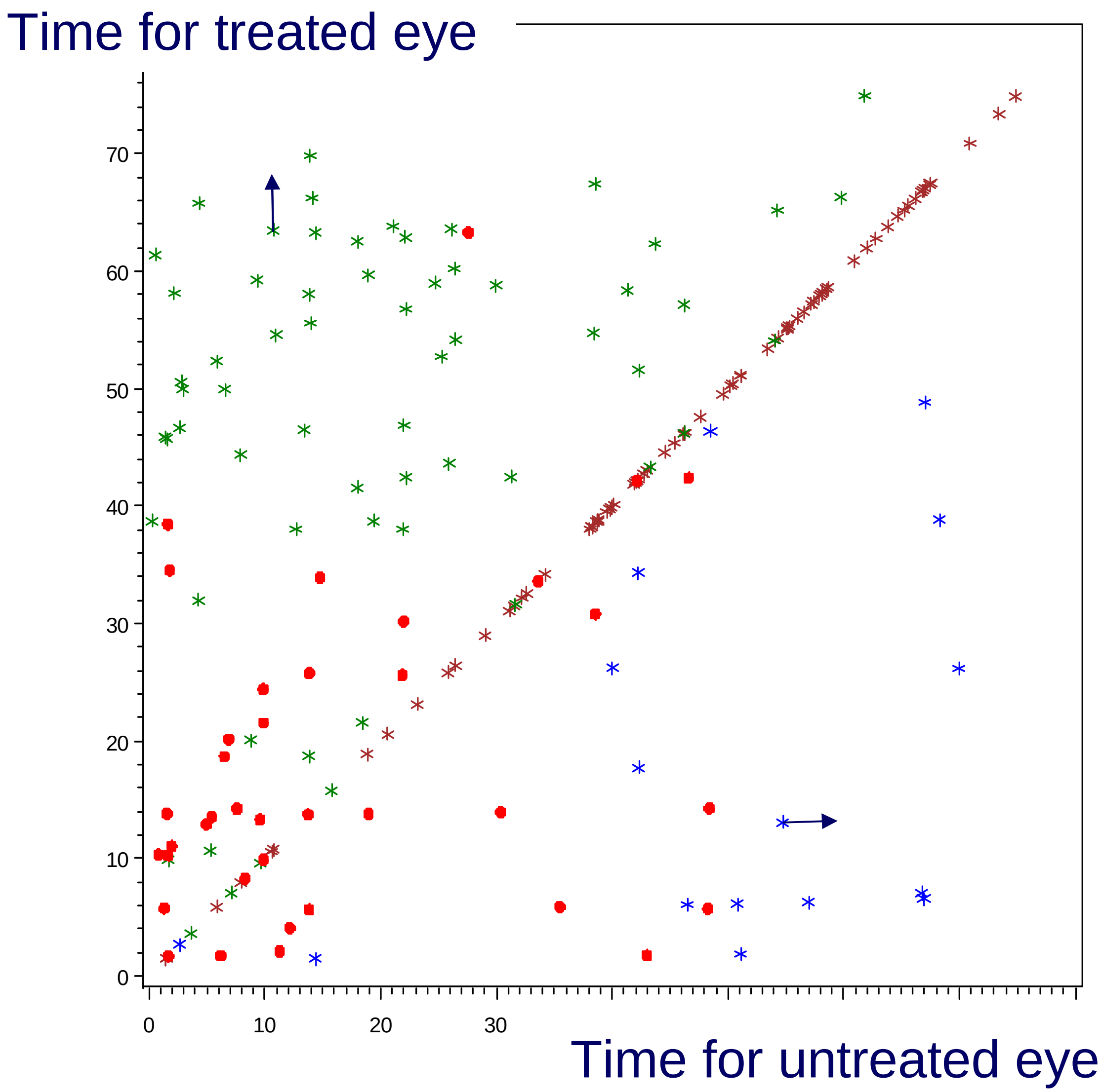
Frank: $r_S = 0.20$

Normal: $r_S = 0.18$

<u>IMI:</u> $r_S = 0.22$

Example 2

Diabetic Retinopathy Study
(n = 197, % censored times = 61)



Spearman correlations:

<u>Copulas:</u>

Clayton: $r_S = 0.22$

Gumbel: $r_S = 0.42$

Frank: $r_S = 0.37$

Normal: $r_S = 0.35$

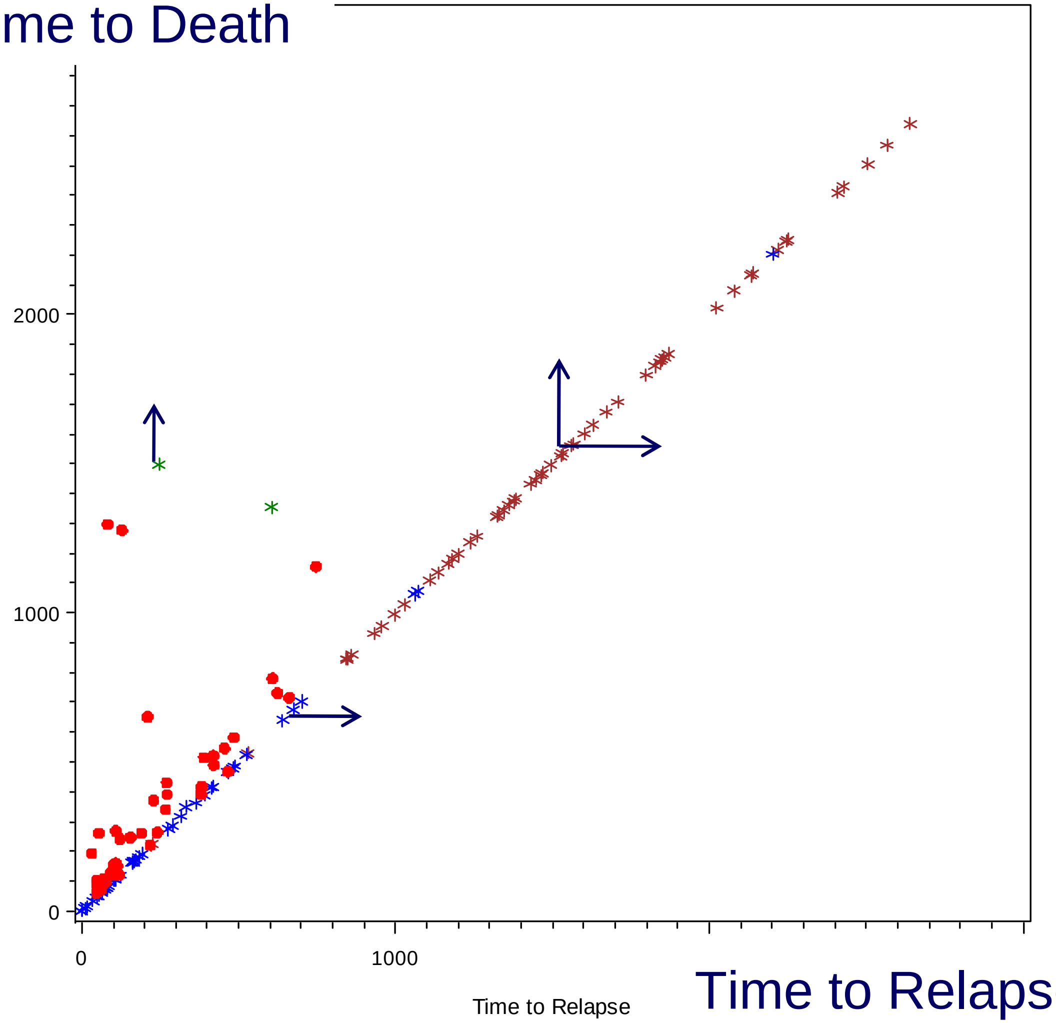
<u>IMI:</u> $r_S = 0.35$

Example 3

Bone Marrow Transplant Study

(n = 137, % censored: 41 (death) / 69 (relapse))

Time to Death



Time to Relapse

Spearman correlations:

<u>Copulas:</u>

Clayton: $r_S = 0.62$

Gumbel: $r_S = 0.83$

Frank: $r_S = 0.80$

Normal: $r_S = 0.79$

<u>IMI:</u> $r_S = 0.85$

Bivariate - exponential survival times
Entries: IMI / Normal Copula

r (population)	%cens = 30	%cens = 60	%cens = 90
0.0	0.00 / 0.00	0.00 / 0.00	0.00 / 0.02
0.3	0.30 / 0.28	0.30 / 0.28	0.30 / 0.29
0.6	0.60 / 0.58	0.60 / 0.58	0.60 / 0.56
0.9	0.90 / 0.89	0.90 / 0.90	0.90 / 0.89

N = 10.000

Gumbel survival times
Entries: IMI / Normal Copula

r (population)	%cens = 30	%cens = 60	%cens = 90
0.0	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00
0.3	0.30 / 0.29	0.27 / 0.26	0.22 / 0.23
0.6	0.60 / 0.58	0.56 / 0.54	0.50 / 0.48
0.9	0.89 / 0.88	0.88 / 0.86	0.84 / 0.83

N = 10.000

Clayton survival times
Entries: IMI / Normal Copula

r (population)	%cens = 30	%cens = 60	%cens = 90
0.0	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00
0.3	0.36 / 0.35	0.42 / 0.42	0.57 / 0.57
0.6	0.67 / 0.66	0.75 / 0.75	0.88 / 0.87
0.9	0.93 / 0.92	0.97 / 0.96	0.97 / 0.99

N = 10.000

Concluding remarks

- Normal copula has been considered as unsuitable for censored survival data and/or intractable – neither is the case as we have shown.
- In the examples IMI was always close to the copula chosen.
- We recommend either the semiparametric copula approach, choosing a certain copula by its max lik, or the IMI approach.
- The procedures discussed have been implemented in SAS-macros and R- functions and will be available for download from our website.