Copula-based modelling and analysis of semi-competing risk data, with application to renal transplant

Lexy Sorrell[1], **Yinghui Wei**[1], **Malgorzata Wojtys**[1], **Peter Rowe**[2]

[1] *Centre for Mathematical Sciences, School of Engineering, Computing and Mathematics, University of Plymouth, UK*

[2] *South West Transplant Centre, University Hospitals Plymouth NHS Trust, UK*

Introduction

Copulas are multivariate distribution functions that allow for dependence between random variables to be modelled separately from their marginal distributions. Copula functions are defined as follows [1]. For a p-dimensional random vector $\boldsymbol{U} = (U_1, ..., U_p)$ uniformly distributed on the unit cube $[0,1]^p$ with observed values $u_1, ...u_p$, a copula $C : \mathbf{I}^p \to \mathbf{I}$ is defined by,

$$C(u_1, ...u_p) = P(U_1 \leq u_1, ..., U_p \leq u_p),$$

where $\mathbf{I} = [0,1]$ denotes the unit interval. Sklar's Theorem [1] states that a copula can be derived from any joint distribution function and also any copula with marginal distributions is a multivariate distribution function. Bivariate copulas can be used to join two marginal distributions with an association parameter, which can be transformed into commonly used correlation coefficients such as Spearman's rank, ρ [2]:

$$\rho = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3. \qquad (1)$$

We consider the Normal, Clayton, Frank and Gumbel copulas to model the association between non-terminal and terminal events. The definitions of each are given below, respectively.

$$C_N(u,v) = \frac{1}{2\pi\sqrt{1-\rho_r^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left(\frac{-(s^2 - 2\rho_r st + t^2)}{2(1-\rho_r^2)}\right) dsdt$$

$$C_C(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$$

$$C_F(u,v) = -\frac{1}{\theta}\log\left(\frac{(1-e^{-\theta}) - (1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$$

$$C_G(u,v) = \exp(-((-\log(u))^\theta + (-\log(v))^\theta)^{\frac{1}{\theta}})$$

In Figure 1, we show the contour plots of each copula function defined above with correlation $\rho = 0.75$.

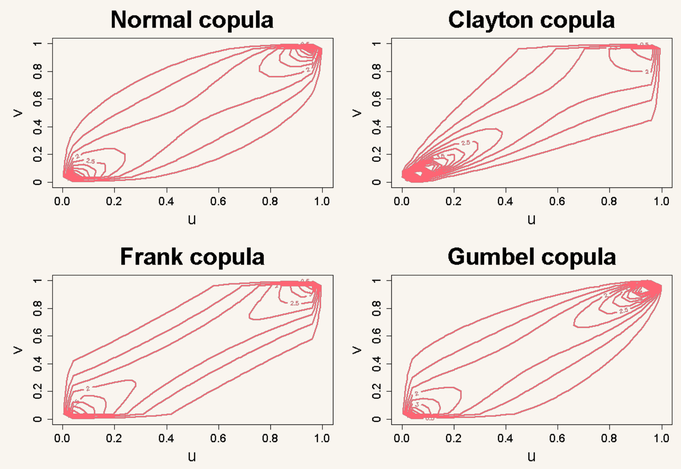

Figure 1: **Contour plots of the Normal, Clayton, Frank and Gumbel copulas.** Contour plots of each copula with marginal distribution functions of u and v, with Spearman's correlation coefficient, $\rho = 0.75$ for each. We see the Normal and Frank copulas have roughly constant association, whereas the Clayton copula has a stronger lower tail density and the Gumbel copula has a stronger upper tail density.

Semi-competing risk events are bivariate survival endpoints where one event can censor another, but not vice-versa. For instance death can censor graft failure, but graft failure cannot censor death.

We let T_1 and T_2 be true times to the non-terminal and terminal events, respectively. We aim to estimate the correlation between T_1 and T_2 as if there was no censoring. However, these events are subject to censoring, T_1, the non-terminal event can be censored by the terminal event and an independent censoring time, C, however T_2 can only be censored by C. We define the following data set for each individual:

- $X = \min(T_1, T_2, C)$
- $Y = \min(T_2, C)$
- $d_1 = 1$ if $T_1 \leq \min(T_2, C)$, $d_1 = 0$ otherwise
- $d_2 = 1$ if $T_2 \leq C$, $d_2 = 0$ otherwise

Methods

We define the joint survival function to be

$$S_B(t_1, t_2) = C(S_{U1}(x), S_{U2}(y))$$

for any copula function C, with marginal survival functions $S_{U1}(x)$ for the non-terminal event and $S_{U2}(y)$ for the terminal event. The copula function, C, has density function

$$c(S_{U1}(x), S_{U2}(y)) = \frac{\partial^2 C(S_{U1}(x), S_{U2}(y))}{\partial S_{U1}(x) \partial S_{U2}(y)}.$$

We propose a likelihood function [3], that accounts for each of the four possibilities of survival outcomes for each individual, $i = 1...n$:

$$L(\theta) = \prod_{i=1}^{n} \left(c(S_{U1}(x_i), S_{U2}(y_i)) f_1(x_i) f_2(y_i)\right)^{d_{i1}d_{i2}}$$
$$\times \left(\frac{\partial C(S_{U1}(x_i), S_{U2}(y_i))}{\partial S_{U1}(x_i)} f_1(x_i)\right)^{d_{i1}(1-d_{i2})}$$
$$\times \left(\frac{\partial C(S_{U2}(x_i), S_{U2}(y_i))}{\partial S_{U2}(y_i)} f_2(y_i)\right)^{(1-d_{i1})d_{i2}}$$
$$\times \left(C(S_{U1}(x_i), S_{U2}(y_i))\right)^{(1-d_{i1})(1-d_{i2})}.$$

Where d_1 and d_2 are the indicators of the non-terminal and terminal events occurring, respectively. We assume the survival distributions for the non-terminal and terminal events follow the Exponential distribution, $S_{U1}(x) = \exp(-\lambda_1 x)$ and $S_{U2}(y) = \exp(-\lambda_2 y)$, and $f_1(x)$ and $f_2(y)$ are the respective probability density functions.

Sklar's Theorem has been extended to account for the conditioning of the bivariate variables of interest on covariates [4]. We allow both the non-terminal and terminal hazard rates, along with the association parameter θ, to vary with covariates of interest.

Application to renal transplant

The NHS Blood and Transplant renal transplant data set consists of 40,348 individuals who have undergone renal transplant with the events of interest graft failure and death. We include covariates donor, sex and age group in the analysis. We use our likelihood function to estimate the hazard rates of graft failure and death, along with Spearman's rank correlation coefficient. The Frank copula is chosen by the AIC.

Covariate	Graft failure	Death
Age group (ref: ≤ 50)	1.36 (1.31, 1.41)	3.85 (3.68, 4.03)
Sex (ref: males)	1.00 (0.96, 1.04)	0.93 (0.89, 0.97)
Donor (ref: deceased)	0.59 (0.56, 0.62)	0.54 (0.51, 0.57)

Table 1: Hazard ratios for graft failure and death estimated by the Frank copula model. The reference groups are given in brackets.

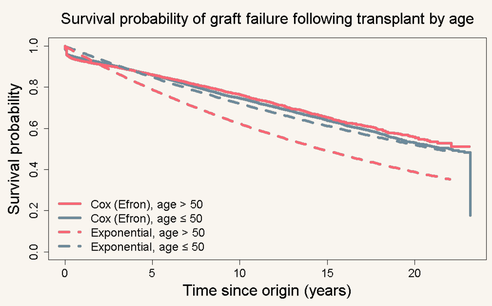

Figure 2: **Survival probabilities of graft failure.** The survival curves estimated by the Cox proportional hazards model show a higher survival probability in the older age group, however the copula-based model with the Frank copula and Exponential marginal distributions shows the older age group having a lower survival probability.

We estimate the correlation coefficient to be stronger with the older age group and the living donor recipients compared to the deceased donor recipients. There is no difference in the association between the two events between males and females.

Simulation study

We simulate semi-competing risk data, mimicking the NHSBT renal transplant data set. We design and implement a simulation study to compare the copula-based methods to the Cox model, and investigate whether the inclusion of covariates in the association parameter improves the estimation of the hazard ratios of both events.

Table 2 displays the results from the simulation study, comparing the estimation of the hazard ratios of the non-terminal and terminal events between 3 models. Model 1 is the Cox proportional hazards model. Model 2 includes covariates in the estimation of the marginal event parameters but not the association parameter. Model 3 includes covariates in the estimation of marginal and association parameters.

Model	Covariate	Graft failure			Death		
		Bias	CP	MSE	Bias	CP	MSE
1	Age group	0.509	0.0	0.517	0.234	94.2	0.600
	Sex	0.065	94.1	0.284	0.053	95.5	0.261
	Donor	0.048	91.8	0.236	0.036	95.1	0.215
2	Age group	0.149	59.3	0.318	0.260	93.3	0.606
	Sex	0.055	94.4	0.266	0.048	96.2	0.247
	Donor	0.037	95.3	0.216	0.035	95.4	0.210
3	Age group	0.073	95.7	0.307	0.214	95.8	0.568
	Sex	0.054	94.6	0.263	0.049	94.8	0.251
	Donor	0.037	96.0	0.217	0.033	94.4	0.205

Table 2: Hazard ratios for graft failure and death estimated by the Frank copula model (the true distribution). Bias represents the mean of bias, CP represents the coverage probability and MSE represents the mean squared error. The reference categories for each covariate are ≤ 50, males and deceased donor recipients, with the hazard ratios being given for > 50, females and living donor recipients, respectively.

Conclusion

We propose a copula-based approach to estimate the correlation coefficient between semi-competing risk endpoints and the hazard ratios of covarites of interest. These models improve the estimation of the non-terminal hazard ratio compared to the Cox proportional hazards model when the variables of interest affect the association between the events. We find that the inclusion of covariates in the estimation of the association parameter improves the estimation of the hazard ratios of the non-terminal event.

Future Work

The survival distributions may be extended to include the Weibull and Gompertz distributions. This may also be extended to include continuous covariates.

References

[1] Sklar A (1959) *Distribution functions of n dimensions and margins* (Public Institute Statistical University 8:229–231)

[2] Nelsen RB (2007) *An introduction to copulas* (Springer Science & Business Media)

[3] Fu H *et al.* (2013) *Joint modeling of progression-free survival and overall survival by a Bayesian normal induction copula estimation model* (Stat Med 32(2):240-254)

[4] Patton AJ (2006) *Modelling asymmetric exchange rate dependence* (International Economic Review 47(2):527-556)

UNIVERSITY OF PLYMOUTH