

# Semi-Competing Risks on A Trivariate Weibull Survival Model

Jenq-Daw Lee  
Graduate Institute of Political Economy  
National Cheng Kung University  
Tainan, Taiwan 70101  
ROC

Cheng K. Lee  
Loss Forecasting  
Home Loans & Insurance  
Bank of America  
Charlotte, NC 28244  
USA  
lee.abcde@gmail.com

## Abstract

A setting of a trivariate survival function using semi-competing risks concept is proposed, in which a terminal event can only occur after other events. The Stanford Heart Transplant data is reanalyzed using a trivariate Weibull distribution model with the proposed survival function.

**Keywords:** semi-competing risks; trivariate Weibull.

## 1. Introduction

Fine, Jiang and Chappell (2001) introduced the term “semi-competing risk” in which one event censors the other but not *vice versa*. In their article, the bivariate Clayton survival function was used to demonstrate the concept. Even before them, Li (1997) has worked on the same concept using the bivariate Weibull survival model by Lu *et al* (1990). In his dissertation, Li described the censoring event as “termination event” and, as did Fine *et al* (2001), the bivariate survival function was divided into two components with lower wedge and upper densities. Epstein *et al* (1996), Shen *et al* (1998), and Dignam *et al* (2007) worked on the likelihood function of a bivariate survival model with four types of censoring events. Different from those authors, in this article, a trivariate survival function for semi-competing risks with two-fatal and one non-fatal events is first constructed and then followed by the likelihood function.

In Section 2, the Stanford Heart Transplant Data is reconstructed for the analysis of semi-competing risks. In Section 3, the trivariate Weibull survival function is proposed followed by the likelihood function. Section 4 shows the results of the analysis of the data. Finally, the article is concluded with some discussion in Section 5.

## 2. Data Structure

The Stanford Heart Transplant Data has been analyzed by Aitkin *et al* (1983) using the Weibull, lognormal, and piecewise exponential models with the

consideration of pre-transplant and post transplant survival. They also surveyed the literature of analyzing the same data. In addition, the same data was studied by Miller *et al* (1982) using four regression techniques. The analyses by Noura (1990) and Loader (1991) are to find the change point of the hazard rate.

The Stanford Heart Transplant Data analyzed in this study is from the article by Crowley *et al* (1977) in which they denote  $T_1$  the date of acceptance to the study,  $T_2$  the date last seen, and  $T_3$  the date of transplantation.  $T_2$  is less than or equal to the last day of the data collection or the last day of the study which was April 1, 1974. An individual was to experience three events when the individual was accepted to the study. The 3 events are death before transplant ( $E_1$ ), transplant ( $E_2$ ), and death after transplant ( $E_3$ ). Therefore,  $E_1$  and  $E_3$  are terminal events or fatal events, and  $E_2$  is an intermediate event.

Let  $X_1$  be the time to  $E_1$ ,  $X_2$  be the time to  $E_2$ , and  $X_3$  be the time to  $E_3$ .  $X_1$ ,  $X_2$ , and  $X_3$  begin at  $T_1$  and are in days. The three events  $E_1$ ,  $E_2$ , and  $E_3$  are competing with each other for the occurrence to each individual. However, these three events must occur in some certain orders. When  $E_1$  occurs first, neither  $E_2$  nor  $E_3$  will occur because  $E_1$  is a terminal event. When  $E_2$  occurs,  $E_3$  may occur later but  $E_1$  will not occur because  $E_1$  and  $E_2$  are defined mutually exclusive. When  $E_3$  occurs,  $E_2$  must occur first because  $E_3$  is the event defined to occur after  $E_2$ . With these orders, an individual must fall into one and only one of the following four cases. First case, an individual experiences  $E_1$ , and, therefore, no possibility for the occurrence of  $E_2$  or  $E_3$ . The individual is said to be uncensored due to  $E_1$ . In this case,  $X_1 = T_2 - T_1$ , and  $X_2$  and  $X_3$  do not exist. Second case, an individual experiences  $E_2$  first, and then  $E_3$  with no occurrence of  $E_1$ . The individual is said to be uncensored due to  $E_2$  and  $E_3$ . In this case,  $X_2 = T_3 - T_1$ ,  $X_3 = T_2 - T_1$ , and  $X_1$  does not exist. Third case, an individual experiences only  $E_2$  before the end of the study. The individual is said to be uncensored due to  $E_2$ , and censored due to  $E_3$ . In this case,  $X_2 = T_3 - T_1$ ,  $X_3 = T_2 - T_1$ , and  $X_1$  does not exist. Fourth case, an individual does not experience  $E_1$ ,  $E_2$  or  $E_3$  before the end of the study. The individual is said to be censored due to  $E_1$ ,  $E_2$  and  $E_3$ . In this case,  $X_1 = T_2 - T_1$ ,  $X_2 = T_2 - T_1$ , and  $X_3 = T_2 - T_1$ . The original data of Crowley and Hu contains 103 observations. After deleting 3 observations with 0 in  $X_1$ ,  $X_2$  or  $X_3$ , and 4 observations of transplant with no mismatch score, there are total 96 observations included in this study.

### 3. The Model and the Likelihood Function

The Weibull distribution model is chosen for the marginal distribution of  $X_1$ ,  $X_2$ , and  $X_3$  as the model was adopted by Aitkin *et al* (1983) and Noura (1990). In order to study the relation among the three random variables, the following trivariate Weibull survival function is derived using Clayton copula (Bandeem-Roche & Liang, 1996).

$$S_{X_1, X_2, X_3} = \left( S_{X_1}^{1-\theta} + S_{X_2}^{1-\theta} + S_{X_3}^{1-\theta} - 2 \right)^{\frac{1}{1-\theta}} \quad (1)$$

where  $S_{X_1} = \text{Exp}\left(-\left(\frac{x_1}{\lambda_1}\right)^{\gamma_1}\right)$ ,  $S_{X_2} = \text{Exp}\left(-\left(\frac{x_2}{\lambda_2}\right)^{\gamma_2}\right)$ ,  $S_{X_3} = \text{Exp}\left(-\left(\frac{x_3}{\lambda_3}\right)^{\gamma_3}\right)$ ,  $0 < \lambda_1, \lambda_2, \lambda_3$

$< \infty$ ,  $0 < \gamma_1, \gamma_2, \gamma_3 < \infty$ , and  $1 < \infty$ .  $X_1, X_2$ , and  $X_3$  are independent when  $\alpha = 1$ . One of the features of the Clayton copula is that it allows positive and negative association between the random variables. To account for the effects of covariates, let  $\lambda_1$  be the exponential function of age at acceptance and previous surgery, and let both  $\lambda_2$  and  $\lambda_3$  be the exponential function of age at acceptance, previous surgery and mismatch score. Note that only individuals receiving transplant had mismatch score. The parameters in the proposed trivariate survival model are to be estimated by maximizing the likelihood function. When each of the three events can censor and be censored by other events, the proposed trivariate Weibull survival model is one of the components in the likelihood function. That is the component accounts for individuals of lost-to-follow-up or being censored at the end of the study. However, when semi-competing risks exist with the orders discussed in Section 2, the survival function for individuals censored at the end of the study or lost to follow-up becomes

$$S_{X_1}(t) + S_{X_2, X_3}(t, t) + \int_t^\infty \left[ \frac{\partial}{\partial x_3} S_{X_2, X_3}(x_2, x_3) \right]_{x_2=x_3} dx_3 \quad (2)$$

The detailed derivation is in the Appendix. As did Lawless (1982), the component in the likelihood for case 1 is the negative derivative of equation (2) with respect to  $x_1$ . The component for case 2 is the derivative of equation (2) with respect to  $x_2$  and  $x_3$ . The component for case 3 is the negative derivative of equation (2) with respect to  $x_2$ . And, the component for case 4 is equation (2) itself.

Therefore, the likelihood function for the proposed trivariate survival function with the four cases is

$$\begin{aligned} L(\theta) = & \prod_{i=1}^N f_{X_1}(t_{x_{1i}})^{p_i} \\ & \times f_{X_2, X_3}(t_{x_{2i}}, t_{x_{3i}})^{q_i} \\ & \times \left( \left[ -\frac{\partial}{\partial x_2} S_{X_2, X_3}(x_2, x_3) \right]_{x_2=t_{x_{2i}}, x_3=t_{x_{3i}}} \right)^{r_i} \\ & \times \left( S_{X_1}(t_i) + S_{X_2, X_3}(t_i, t_i) + \int_{t_i}^\infty \left( \frac{\partial}{\partial x_3} S_{X_2, X_3}(x_2, x_3) \right)_{x_2=x_3} dx_3 \right)^{1-p_i-q_i-r_i} \end{aligned} \quad (3)$$

where  $p, q$ , and  $r$  are event indices and  $t$  denotes the survival time.

#### 4. Results

The estimates and their corresponding 95% confidence intervals of the parameters are in the following table. The asymptotic covariance matrix is approximated by the inverse of the negative Hessian.

Parameter	Estimate	95% Confidence Interval	p-value
	1.677	( 1.147, 2.208)	< 0.001
age at acceptance ( $X_1$ )	0.087	( 0.060, 0.114)	< 0.001
previous surgery ( $X_1$ )	-1.316	(-5.527, 2.894)	0.540
$\beta_1$	0.342	( 0.258, 0.425)	< 0.001
age at acceptance ( $X_2$ )	0.076	( 0.061, 0.091)	< 0.001
previous surgery ( $X_2$ )	0.196	(-0.653, 1.045)	0.650
mismatch score ( $X_2$ )	-0.036	(-0.532, 0.460)	0.887
$\beta_2$	0.733	( 0.614, 0.852)	< 0.001
age at acceptance ( $X_3$ )	0.131	( 0.098, 0.165)	< 0.001
previous surgery ( $X_3$ )	1.993	(-0.242, 4.228)	0.080
mismatch score ( $X_3$ )	0.340	(-0.787, 1.467)	0.554
$\beta_3$	0.422	( 0.322, 0.523)	< 0.001

The results indicate only the age at acceptance is significantly different from zero at significance level of 0.05 for  $X_1$ ,  $X_2$ , and  $X_3$  while Crowley *et al* (1977) showed an insignificant result. The covariate previous surgery of  $X_1$ ,  $X_2$ , and  $X_3$ , and the covariate mismatch score of  $X_2$  and  $X_3$  are both insignificant that agrees to the findings of Crowley *et al* (1977), and Noura (1990). The overall association parameter is 1.677 that indicates  $X_1$ ,  $X_2$ , and  $X_3$  are not much correlated.

#### 5. Conclusions

In this article, the Clayton trivariate Weibull survival model with Weibull marginals is applied to the Stanford Heart Transplant data. Due to the order of the occurrences of the three events, a new formation of the likelihood function is proposed for the three random variables that are allowed for different sets of covariates. The work of this article can be expanded to accommodate more variables.

#### Acknowledge

The author thanks Mr. Daniel Warren Whitman for his proofreading this article.

#### Appendix

The fourth factor in the likelihood function is the probability that an individual is censored at  $T_2$ , the date last seen. After the censoring, although is unobservable,

the individual may experience  $E_1$  only, or  $E_2$  followed by  $E_3$ . Suppose the individual is censored at time  $t$ , then the probability for the occurrences of the three events is

$\Pr(t < X_1) + \Pr(t < X_2 < X_3)$ .  $\Pr(t < X_1)$  is simply equal to  $S_{X_1}(t)$ .

and,  $\Pr(t < X_2 < X_3) =$

$$\int_t^{\infty} \int_t^{x_3} f_{X_2, X_3}(x_2, x_3) dx_2 dx_3$$

$$= \int_t^{\infty} \left[ \int_t^{\infty} f_{X_2, X_3}(x_2, x_3) dx_2 - \int_{x_3}^{\infty} f_{X_2, X_3}(x_2, x_3) dx_2 \right] dx_3$$

$$= S_{X_2, X_3}(t, t) - \int_t^{\infty} \int_{x_3}^{\infty} f_{X_2, X_3}(x_2, x_3) dx_2 dx_3$$

Considering  $\int_{x_3}^{\infty} f_{X_2, X_3}(x_2, x_3) dx_2$ ,

$$\int_{x_3}^{\infty} f_{X_2, X_3}(x_2, x_3) dx_2$$

$$= \int_0^{\infty} f_{X_2, X_3}(x_2, x_3) dx_2 - \int_0^{x_3} f_{X_2, X_3}(x_2, x_3) dx_2$$

$$= f_{X_3}(x_3) - \int_0^{x_3} \left( \frac{\partial}{\partial x_2} \left( \frac{\partial}{\partial x_3} F_{X_2, X_3}(x_2, x_3) \right) \right) dx_2$$

$$= \frac{\partial}{\partial x_3} F_{X_3}(x_3) - \left[ \frac{\partial}{\partial x_3} F_{X_2, X_3}(x_2, x_3) \right]_{x_2=x_3} \quad (x_2 = x_3 \text{ denotes that } x_2 \text{ is replaced by } x_3)$$

$$= \left[ \frac{\partial}{\partial x_3} (F_{X_3}(x_3) - F_{X_2, X_3}(x_2, x_3)) \right]_{x_2=x_3}$$

$$= \left[ \frac{\partial}{\partial x_3} (-S_{X_2, X_3}(x_2, x_3)) \right]_{x_2=x_3}$$

Note that  $S_{X_2, X_3}(x_2, x_3) = 1 - F_{X_2}(x_2) - F_{X_3}(x_3) + F_{X_2, X_3}(x_2, x_3)$  (Li (1997)) where  $F_{X_2}$ ,  $F_{X_3}$  and  $F_{X_2, X_3}$  are, respectively, the cumulative density function of  $X_2$ ,  $X_3$ , and  $X_2$  and  $X_3$ .

Then,  $\Pr(t < X_2 < X_3) = S_{X_2, X_3}(t, t) + \int_t^{\infty} \left[ \frac{\partial}{\partial x_3} S_{X_2, X_3}(x_2, x_3) \right]_{x_2=x_3} dx_3$ .

Therefore, the fourth factor in the likelihood function is

$$\Pr(t < X_1) + \Pr(t < X_2 < X_3) = S_{X_1}(t) + S_{X_2, X_3}(t, t) + \int_t^{\infty} \left[ \frac{\partial}{\partial x_3} S_{X_2, X_3}(x_2, x_3) \right]_{x_2=x_3} dx_3$$

## References

1. Aitkin, M., Laird, N., Francis, B. (1983). A reanalysis of the Stanford heart transplant data. *Journal of the American Statistical Association* 78: 264-274.
2. Bandeen-Roche, K., Liang, K., (1996). Modeling Failure-Time Associations in Data with Multiple Levels of Clustering. *Biometrika* 83: 29-39.
3. Crowley, J., Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data. *Journal of the American Statistical Association* 72: 27-36.
4. Dignam, J., Wieand, K., Rathouz, P. (2007). A missing data approach to semi-competing risks problems. *Statistics in Medicine* 26: 837–856.
5. Epstein, D., Muñoz, A. (1996). A Bivariate Parametric Model for Survival and Intermediate Event Times. *Statistics in Medicine* 15: 1171-1185.
6. Fine J.P., Jiang, H., Chappell, R. (2001) On Semi-Competing Risks Data. *Bimoketrika* 88: 907-919.
7. Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons. New York, USA.
8. Li, C. L., (1997). A Model for Informative Censoring. Ph. D. Dissertation, The University of Alabama at Birmingham.
9. Loader, C. R. (1982). Inference for A Hazard Rate Change Point. *Biometrika* 78: 749-757.
10. Lu, J-C., Bhattacharyya, G. (1990). Some New Constructions of Bivariate Weibull Models. *Annals of The Institute of Statistical Mathematics* 42: 543-559.
11. Miller, R., Halpern, J. (1982). Regression with Censored Data. *Biometrika* 69: 521-531.
12. Noura, A. A. (1990). Proportional Hazards Chanepoint Models in Survival Analysis. *Applied Statistics* 39: 241-253.
13. Shen, Y., Thall, P. (1998). Parametric likelihoods for multiple non–fatal competing risks and death. *Statistics in Medicine* 17: 999–1015.