DOI: http://doi.org/10.4038/sljastats.v23i2.8071

Sri Lankan Journal of Applied Statistics, Vol (23-2)

Shared Frailty Model for Joint Survival Data - A Simulation Study

J.C. Liyanage*, and G.H.S. Karunarathna

Department of Statistics, Faculty of Science, University of Colombo, Colombo 03, Sri Lanka

*Correspnding author: jayliyanage1@gmail.com

Received: 26^{th} July 2022/ Revised: 15^{th} November 2022/ Published: 25^{th} November 2022

ABSTRACT

Performing a survival analysis, allows investigating factors that contribute to outcome over time. Though a considerable amount of literature has addressed this area, analysis of multiple survival responses has not received sufficient attention in past literature. The joint modeling fits well when there are multiple survival responses for the same study unit and it can provide improved results than fitting univariate models separately since the correlation between the responses can be captured through a joint model. Therefore, the aim of this study was to propose a joint modeling approach, in which the linkage between two survival responses(for simplicity, in this research bivariate lifetime data have been considered) was derived by sharing a common random effect under different random effect distributions through parametric forms of the baseline hazard function. In this study, Gamma and Normal random effect distributions and Exponential and Weibull parametric survival distributions were considered. The performance of the Shared Frailty model was compared with two Ordinary Proportional Hazard models through a simulation study, by fitting models for simulated data in three different sample sizes with 20% and 40% censoring proportions in different correlation structures. Bias, Coverage Probability (CP) and Mean Squared Error (MSE) were the performance measures used. Parameter estimates showing relatively low bias, high CP, and minimal MSE under the joint random effect model confirmed the suitability of the proposed model to capture joint survival data, surpassing the fit of two univariate models. Also, the results interpreted, Gamma distributed random effects to be more suitable with Exponential survival times while Normal random effects with Weibull survival times.

Keywords: survival, joint model, shared frailty, simulation

1 Introduction

Survival analysis was originally developed and used by medical researchers and data analysts to measure the lifetime of a certain population. Nevertheless, over the years, it has been applied to many other fields. As a branch of survival analysis, multivariate survival analysis has emerged; however, analyzing such multivariate survival data remains a challenging task due to the model specifications on the association between endpoints.

Multivariate survival data covers the field where independence between survival times cannot be assumed (Laux , 1985). Thus, in analyzing these data, the key element is an appropriate account for dependence between event times. Performing univariate analysis for each response separately may not be appropriate due to the correlation among the responses (Rondeau , 2019). Hence, it is important to fit a joint model that specifies independence among observed data items, conditional on a set of unobserved or latent variables. This is defined as Frailty models which are extensions of the proportional hazard model, where the random effect (the frailty) has a multiplicative effect on the baseline hazard function.

In common practice, shared frailty models are used in event times of related individuals, for similar organs, repeated measurements, etc., under the clustering scenarios by assuming to share the same frailty. Though clustering is not considered here, the shared frailty model concept has been used to capture the behavior of bivariate survival responses by forming one random effect as a multiplication of the other random effect, incorporating a shared constant.

With the recent surge of research studies on survival data, joint modeling has become popular over the past few years. Recently, Sunethra and Sooriyarachchi (2020) have introduced a novel method for joint modeling of survival data and count data for both simple randomized and cluster-randomized data, while Karunarathna and Sooriyarachchi (2019) have proposed a joint multilevel model for analyzing the length of stay of patients through competing endpoints in dengue epidemiology. Recently, Li et al. (2020) have introduced a multivariate joint frailty model for two types of recurrent events in the presence of a dependent terminal event with right-censored survival data. Joint modeling of multivariate longitudinal and multivariate survival data has been considered by Chi and Ibrahim (2006) in their research and Ratcliffe et al. (2004) have developed a joint model for the analysis of longitudinal and survival data in the presence of data clustering.

However, analyzing multiple survival responses with joint modeling approaches remains a gap in multivariate survival analysis. This work is novel since there is no such analysis was found in local literature where joint modeling of two survival responses was considered. Hence, this research is of great signifi-

cance, which will emphasize the efficiency of fitting a joint random effect model to the survival responses rather than fitting univariate models separately.

The main objective of this study is to propose a joint modeling approach (shared frailty model) for analyzing bivariate lifetime/survival data in the presence of censoring through different parametric forms of baseline hazard function and frailty term distributions. Apart from that, to emphasize the importance of modeling survival responses jointly, rather than fitting univariate models for each response separately.

The rest of the paper is organized into four sections, of which the first section is the introduction and some literature related to the research problem and the next succeeding sections follow as mentioned below.

Section 2: The theories and methodologies behind the simulation study have been discussed in this section.

Section 3: Implementation of the simulation study, the results obtained for the parameter estimates with the performance of the estimators are mentioned in this section.

Section 4: Summarizes major conclusions of the study and discusses the limitations and future work to be carried out in this area.

2 Theory and Methodology

This section discusses the theoretical background and the parameter estimation procedure of the Shared Frailty Model for non-clustered multivariate censored survival data. This mainly explores the distributional functions and models for survival responses, different frailty term distributions that capture the unobserved heterogeneity in models of survival data, and the key performance measures that are used in comparing the simulation study results.

2.1 Distributional functions in survival analysis

Survival data are summarized using the survival function and the hazard function. The survival function is defined as the probability of surviving beyond a specified time t. T denotes the time until the occurrence of a certain event (death) with Cumulative Distribution Function (CDF) F(t), on the interval [0,

∞), then its survival function is denoted by S.

$$S(t) = P(T > t) = \int_{t}^{\infty} f(u)du = 1 - F(t) \tag{1}$$

The hazard function, conventionally denoted h, is defined as the instantaneous probability of an event (death), which is the limit probability of an item or a person dying in the interval (t, t+δt) that has survived for time t.

$$h(t) = \lim_{\delta t \to 0} (\frac{t \leq T \leq t + \delta t | t \leq T}{\delta t}) = \frac{f(t)}{S(t)} \tag{2}$$

Approaches that have been widely used to model survival data are non-parametric and semi-parametric methods due to their less complexity with fewer assumptions than parametric methods and as there is no distributional assumption that needs to be imposed on the survival times. Despite these advantages, parametric models being an alternative to the Cox model, which is a semi-parametric model, have performed better with valid distributional assumptions, resulting in more efficient parameter estimates in the sense of having smaller standard errors as compared to those in the non-parametric models. Highlighting these reasons, this research study has been done under parametric modeling approach.

A parametric model is defined by a survival function or hazard function with a specified distributional form such as exponential, Weibull, Gompertz, gamma, log-normal and log-logistic (Lee and Go , 1997). Among these commonly used survival time distributions, only the exponential, Weibull, and Gompertz distributions share the assumption of proportional hazards with the cox model. In this research study, exponential and Weibull distributions were considered.

2.2 Frailty model

Multivariate survival analysis is considered a branch of survival analysis that deals with two or more events per subject. Since the various events occurred in the same individual, the survival times will not be independent in general. Thus, an appropriate account for the dependence between event times has been the key element when analyzing multivariate survival data.

The most commonly used and the very general approach is to specify the independence among the responses, conditional on a set of unobserved variables. This is defined under frailty models, where the heterogeneity of the model is captured by incorporating a random effect into the model (Ibrahim et al. ,

2001).

$$h(t_i) = h_0(t_i)exp(\beta_0 + \beta_1 X_i + V_i) \tag{3}$$

here i is the observation, t_i is the time, X_i is the covariate, β_0 and β_1 are the regression coefficients, V_i is the random effect, and $h_0(t)$ is the baseline hazard function.

Since the model is formulated through the hazard function, the appropriate survival times are not straightforward. For that, the inverse cumulative hazard method should be considered with respect to the specified distribution form of the survival times.

2.2.1 General considerations

The survival function of the random effect model (3) is given by,

$$S(t_i) = exp(-H_0(t_i) \times exp(\beta_0 + \beta_1 X_i + V_i)) \tag{4}$$

where $H_0(t) = \int_0^t h_0(u)du$ is formulated as the cumulative baseline hazard function.
Thus, the distribution function can be denoted as,

$$F(t_i) = 1 - exp(-H_0(t_i) \times exp(\beta_0 + \beta_1 X_i + V_i)) \tag{5}$$

If Y is taken as a random variable with distribution function F, then $U = F(Y)$ follows a uniform distribution on the interval $[0, 1]$, i.e. $U \sim Uni[0, 1]$ and if $U \sim Uni[0, 1]$, then also $(1 - U) \sim Uni[0, 1]$. More details can be referred from Mood et al. (1974).Then with reference to the equation (5),

$$U_i = exp(-H_0(t_i) \times exp(\beta_0 + \beta_1 X_i + V_i)) \sim Uni[0, 1] \tag{6}$$

The survival time T of the random effect model (3) can be expressed as follows, given that $h_0(t) > 0$ for all t, by modifying the equation stated by Bender et al. (2003),

$$T_i = H_0^{-1}[-log(U_i) \times exp(-(\beta_0 + \beta_1 X_i + V_i))] \tag{7}$$

This equation (7) has been used in the simulation study for the generation of survival times.

2.2.2 Exponential survival time distribution

The exponential distribution with scale parameter θ consists of a constant

baseline hazard function. Thus, the cumulative baseline hazard function is given by,

$$H_0(t) = \int_0^t \theta du = \theta t \tag{8}$$

Therefore,

$$H_0^{-1}(t) = \theta^{-1} t \tag{9}$$

By inserting (9) into equation (7), the expression for the survival time under exponential distribution can be obtained as follows.

$$\begin{aligned} T_i &= \theta^{-1}[-log(U_i) \times exp(-(\beta_0 + \beta_1 X_i + V_i))] \\ &= -\frac{log(U_i)}{\theta \times exp(\beta_0 + \beta_1 X_i + V_i)} \end{aligned} \tag{10}$$

2.2.3 Weibull survival time distribution

The Weibull distribution is characterized by two positive parameters, the parameter θ is known as the scale parameter and v is the shape parameter. According to Bender et al. (2005), when v = 1, the hazard function reduces to the exponential distribution and it decreases monotonically for 0 < v < 1 while it increases for v > 1. The cumulative baseline hazard function is given by,

$$H_0(t) = \int_0^t \theta \times v \times u^{v-1} du = \theta t^v \tag{11}$$

Therefore,

$$H_0^{-1}(t) = (\theta^{-1} t)^{\frac{1}{v}} \tag{12}$$

By inserting (12) into equation (7), the expression for the survival time under the Weibull distribution can be obtained as follows.

$$\begin{aligned} T_i &= (\theta^{-1}[-log(U_i) \times exp(-(\beta_0 + \beta_1 X_i + V_i))])^{\frac{1}{v}} \\ &= -(\frac{log(U_i)}{\theta \times exp(\beta_0 + \beta_1 X_i + V_i)})^{\frac{1}{v}} \end{aligned} \tag{13}$$

2.2.4 Distributional functions of random effects

The right choice of the frailty distribution has become an important aspect in arriving at a good description of the dependence structure between the correlated observations because the frailty distribution has got the ability to grab

these dependence changes. In this study, Gamma and the Normal distribution were used as the frailty term distributions as most of the software has limited the choice of the frailty distribution to these cases.

Models were fitted using the SAS NLMIXED procedure to obtain the maximum likelihood estimates from the marginal likelihood. Though Gamma and Normal random effect distributions were used, SAS NLMIXED only allows the random effects to be normally distributed. Thus, a numerical integration approximation (probability integral transformation) has been used in this study to transform a normal random effect into a non-normal random effect (Nelson et al. , 2004).

2.3 Estimation procedure

The parameters of the models can be estimated by the Maximum Likelihood Estimation (MLE) using the marginal likelihood, integrating over the random effects in SAS NLMIXED. The likelihood function for the survival function, in general, is denoted as,

$$L = \prod_{i=1}^{r} h(t_i) \prod_{i=1}^{n-r} S(t_i) \tag{14}$$

where h(t) and S(t) are defined as the hazard function and the survival function of the parametric distribution selected, and n and r are the total number of observations and the number of uncensored observations, respectively.

Since this study is focused on bivariate survival data, the estimation of the joint random effect model is considered. The maximum likelihood estimates are derived as the modes of the log-likelihood function corresponding to the joint distribution of the observed outcomes $(T_{1i}, \delta_{1i}, T_{2i}, \delta_{2i})$, under the assumption that the responses are independent relative to random effects. Thus, the log-likelihood contribution for the i^{th} observation can be formulated as follows (Rizopoulos, 2012).

$$\begin{aligned} log p(T_{1i}, \delta_{1i}, T_{2i}, \delta_{2i}; \theta) &= log \int p(T_{1i}, \delta_{1i}, T_{2i}, \delta_{2i}, V_i; \theta) dV_i \\ &= log \int p(T_{1i}, \delta_{1i} | V_i; \theta_{t1}, \beta) p(T_{2i}, \delta_{2i} | \lambda V_i; \theta_{t2}, \beta) f(V_i) dV_i \\ &= log \int p(T_{1i}, \delta_{1i} | V_i; \theta_{t1}, \beta) dV_i \\ &+ log \int p(T_{2i}, \delta_{2i} | \lambda V_i; \theta_{t2}, \beta) dV_i + log \int f(V_i) dV_i \end{aligned} \tag{15}$$

where, T_{1i} and T_{2i} are the two survival times, δ_{1i} and δ_{2i} are the censoring indicators, V_i is the random effect of i^{th} observation, corresponding to the first survival time T_{1i}, λ is the shared constant which formulated the random effect of T_{2i} and θ is the parameter set. If we take U_i as the second random effect, then $U_i = \lambda V_i$.

Here,

$$p(T_{1i}, \delta_{1i}|V_i; \theta_{t1}, \beta) = \prod_{i=1}^{r} h(t_{1i}, \delta_{1i)} \prod_{i=1}^{n-r} S(t_{1i}, \delta_{1i}) \tag{16}$$

$$p(T_{2i}, \delta_{2i}|U_i; \theta_{t2}, \beta) = \prod_{i=1}^{r} h(t_{2i}, \delta_{2i)} \prod_{i=1}^{n-r} S(t_{2i}, \delta_{2i}) \tag{17}$$

2.4 Performance measures

For comparing the performance of the joint random effect model with two ordinary proportional hazard models that can be fitted separately for the two survival responses, bias, Coverage Probability (CP) and Mean Squared Error (MSE) was used.

Biasedness:

The bias of an estimator is defined as,

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta \tag{18}$$

Where $\hat{\theta}$ is an estimator of θ, if $E(\hat{\theta}) = \theta$, then the estimator is unbiased. If $E(\hat{\theta}) \neq \theta$ then the estimator is said to have a positive or negative bias. That is, on average the estimator tends to over (or under) estimate the original parameter. Thus, the lower the bias the better the parameter estimates.

MSE:

MSE of an estimator is defined as,

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2 \tag{19}$$

It can be observed that the MSE is a one-to-one trade-off between the variance and the squared bias of the estimator. Smaller MSE values closer to zero interpret better parameter estimates.

CP:

Confidence intervals are considered an important and often underused area of statistical inference. The CP is defined as the proportion of the time that the interval contains the true value of interest. Thus, the higher the CP value, the better the parameter estimate.

Using bias alone as a performance measure to compare the parameter estimates of models is not a good practice because a biased estimator with a small variance may outperform an unbiased estimator with a large variance. Similarly, MSE or CP alone cannot evaluate the performance of models accurately. Hence, the MSE criterion along with bias and CP, have been used in this research study to evaluate the performance of the models.

2.5 Simulation study

The joint random effect model for the analysis of two non-clustered survival responses in the presence of censoring is considered in this simulation study. The shared frailty model under different parametric survival distributions by changing the frailty term distribution simultaneously is examined through a simulation study by considering various combinations of parameters with respect to the lower and higher correlation effects of the two survival responses. The theory behind the joint random effect model was presented in the previous section. This section explains the approach to carrying out the simulation study. Results and conclusions are given by examining the performance parameters of univariate random effect and joint random effect models.

2.5.1 Design outline of the simulation study

Data for the simulation study were generated using R software as follows,

1. A Single explanatory variable (X_i) was simulated with Bernoulli $\sim$ (0.5).

2. Shared random effects were generated as $V_i \sim$ N (0,σ_v^2), $V_i \sim$ Gamma (1,1) and $U_i = \lambda V_i$.

3. Two Survival times were generated by considering inverse cumulative method with reference to (Bender et al. , 2005), as $h_0^1(t) \sim Exp(1), h_0^2(t) \sim Exp(2)$ and $h_0^1(t) \sim Weibull(scale = 1, shape = 2)$, $h_0^2(t) \sim Weibull(scale = 3, shape = 4)$ with given parameters, selected using a trial and error method to reduce the complexity of the simulation study.

4. Exponential distribution was used to generate censoring time with 20% and 40% censoring proportions (Qian et al. , 2010).

Table 1: Parameters used in simulating data under Normal and Gamma random effect distributions

Survival Time Distribution	Correlation	β_{01}	β_{11}	β_{02}	β_{12}	λ	σ_v/k
Normal random effects							
Exponential	0.92	1	0.5	1	0.5	1.5	0.1
	0.33	1	-2	1	2	0.01	1
Weibull	0.97	1	0.5	1	0.5	1.5	0.1
Gamma random effects							
Exponential	0.95	1	1	1	1	0.8	1
	0.26	1	-2	1	2	0.01	1
Weibull	0.95	1	0.5	1	0.5	1.5	1

5. Censoring indicator variables δ_1 and δ_2 were obtained with regard to the two survival times indicating 1 if the event occurs during the study period and 0 if the survival time is censored by the end of the study period.

6. Finally, the new response variables Y_1 and Y_2 were created by taking the minimum time point out of survival time and censoring time.

The joint model consists of β_{01}, β_{02}, β_{11} and β_{12} as fixed effects parameters. Here, β_{01} and β_{02} are intercepts for the two survival responses and β_{11} and β_{12} are treatment effects respectively. λ is the shared constant used in obtaining random effects and σ_v^2 is the variance of the normally distributed random effect. Since there was no valid reference to refer for the parameters, various combinations of association between endpoints were taken into consideration as positive high correlation, positive low correlation, negative high correlation, and negative low correlation. Among the set of parameters for each correlation criterion, the set of parameters that gave the highest positive correlation and the lowest positive correlation was chosen. Obtaining a negative association between the two survival responses was a difficult task with reasonable values for shared constant and the random effect variance, which confirms the fact addressed in (Ibrahim et al. , 2001) as a limitation of using shared random effect in the model. Thus, only positive association was selected for the simulation study. The following table represents the selected parameters for the simulation study. Under Weibull survival time distribution, a positive low correlation could be obtained only with a high variation of random effects and a very low shared constant, which results in the non-convergence of datasets. Therefore, only positive high correlation was considered in this data simulation.

The sample size is an important feature of any empirical study of making inferences and it is also important in simulation studies. In this study, simulations were carried out corresponding to three different sample sizes small (n =

50), moderate (n = 500), and large (n = 1000). Under each sample size, 1000 datasets were generated by changing the parameters as stated in Table 1.

3 Results and Discussion

The above-explained procedure was followed for each of the three sample sizes and the models were fitted for 1000 data sets in each case separately. This section summarizes these results in order to evaluate the performance of the models.

Table 2: Simulation results under Exponential survival time distribution for sample size 50, censoring proportion 20%

Parameter	**Normal random effects (Corr(Y_1, Y_2) = 0.92)**					
	Bias		CP		MSE	
	Univariate	Joint	Univariate	Joint	Univariate	Joint
β_{01}	-0.436	-0.432	0.634	0.659	0.230	0.227
β_{11}	-0.209	-0.205	0.984	0.984	0.127	0.124
β_{02}	-0.463	-0.460	0.621	0.639	0.244	0.235
β_{12}	-0.298	-0.293	0.981	0.982	0.213	0.208
σ_v		-0.058		1		0.010
λ		-0.156		0.429		0.062
Gamma random effects (Corr(Y_1, Y_2) = 0.95)						
β_{01}	0.044	0.011	0.974	0.979	0.049	0.047
β_{11}	-0.202	-0.198	0.926	0.946	0.115	0.104
β_{02}	-0.356	-0.337	0.904	0.957	0.203	0.197
β_{12}	-0.288	-0.280	0.881	0.892	0.188	0.181
λ		0.007		1		0.005

When sample size increases, only a mild improvement in the results could be observed, and with the increase of the proportion of censored observations,

Table 3: Simulation results for positive low correlation under Exponential survival time distribution with Gamma distributed random effect for sample size 50, censoring proportion 20%

Parameter	**Corr(Y_1, Y_2) = 0.26**					
	Bias		CP		MSE	
	Univariate	Joint	Univariate	Joint	Univariate	Joint
β_{01}	0.054	0.031	0.774	0.745	0.051	0.046
β_{11}	-0.177	-0.117	0.741	0.752	0.254	0.232
β_{02}	-0.229	-0.185	0.652	0.741	0.234	0.211
β_{12}	-0.279	-0.220	0.611	0.612	0.248	0.199
λ		0.009		1		0.166

parameter estimates were more deviated from the true values irrespective of the distribution used. When dealing with real-life applications in survival analysis, analyzing a larger dataset is a rare situation. Thus, focusing on these aspects, only the results with the smallest sample size (50) and 20% censoring proportion are given in Table 2, Table 3, and Table 4.

When looking at the results shown in Table 2 and Table 3, it can be noticed that both univariate and joint models significantly performed well, showing relatively low bias, high CP, and minimal MSE under the joint model compared to the univariate models' results.

When focusing on the bias, both models of univariate, and joint random effect models acted similarly since there was not any significant variation in all models. It should be noted that all the bias values calculated for the parameters under normal random effect distribution are negative, interpreting an underestimation.

The simulation study was extended under normal random effect distribution to capture parameter estimates with positive low correlation ($\rho = 0.33$). But the simulation results that correspond to the initial parameters (β_{01} = 1, β_{11} = -2, β_{02} = 1, β_{12} = 2, σ_v = 1 and λ = 0.01) had converging issues. Therefore, those simulation results were not presented. With gamma random effect distribution, both positive high and positive low correlations could be obtained without any convergence issues and more satisfactory results were obtained compared to the simulation results with normal random effects.

With the insight of results, the joint random effect model performed comparatively better than univariate models under exponential survival time distribution. As per the prior results of the simulation under exponential survival time distribution, Table 4 indicates that the joint random effect model performed better than univariate models even with Weibull survival time distribution with respect to the performance indicators of relatively lower bias, high CP, and minimal MSE.

When comparing the results obtained under normal random effects and gamma random effects, it is visible that under Weibull survival time distribution, simulation results with normal random effects outperform the simulation results obtained with gamma random effects.

In summary, the simulation study indicates the suitability of the proposed joint random effect model over the univariate models in each and every combination presented in the study.

Table 4: Simulation results under Weibull survival time distribution for sample size 50, censoring proportion 20%

Parameter	Bias		CP		MSE	
	Univariate	Joint	Univariate	Joint	Univariate	Joint
Normal random effects (Corr(Y_1, Y_2) = 0.97)						
β_{01}	-0.134	-0.006	0.937	0.977	0.059	0.053
β_{11}	-0.080	-0.072	0.976	0.988	0.088	0.083
β_{02}	-0.073	0.046	0.950	0.968	0.060	0.049
β_{12}	-0.039	-0.031	0.972	0.983	0.089	0.082
σ_v		0.045		0.493		0.026
λ		-0.016		0.969		0.001
Gamma random effects (Corr(Y_1, Y_2) = 0.95)						
β_{01}	0.821	0.785	0.375	0.343	0.796	0.606
β_{11}	-0.140	-0.098	0.983	0.960	0.220	0.183
β_{02}	1.071	1.034	0.029	0.249	1.313	1.115
β_{12}	-0.062	-0.031	0.919	0.938	0.280	0.206
λ		-0.406		0.593		0.225

4 Conclusion

This section summarizes the final outcomes of the simulated data sets with limitations of the study and suggestions for possible further work.

4.1 Shared frailty model for analyzing multivariate survival data

Multivariate survival analysis which has become a popular area of study in the research field is the targeted topic in this paper. There has been much research on analyzing various forms of multivariate lifetime data, such as multilevel data, recurrent event data, and competing risks. For simplicity, the bivariate nature has been taken into consideration in most of the past studies as well as in this research. However, analysis of non-clustered multiple survival responses where each study unit experience more than one event has taken less attention. Thus, this research was able to fill that gap by introducing a joint random effect model, where the frailty term of one survival response is a multiplication of the other frailty term. Since all the subjects share the frailty terms with the same constant for the two survival responses, the name "shared frailty model" was given to the proposed joint random effect model. This model was fitted under Exponential and Weibull survival time distributions, also by changing the frailty term distribution to be Normal and Gamma.

With respect to the above requirements, thirty-six sets of data were simulated having 1000 datasets in each. The joint random effect model was compared with the univariate models that were fitted for the two survival responses sep-

arately to achieve the objectives of the study.

With regard to the simulated results under the changes of distributions, censoring proportions, and sample sizes, it was confirmed the suitability of the joint random effect model, in analyzing non-clustered correlated survival responses with the presence of censoring. Other than the demonstrated results, AIC values that were fitted for the models, also confirmed the suitability of the joint random effect model by giving a smaller AIC compared to the total of the two AIC values obtained with the univariate models.

Apart from that the results interpreted, Gamma distributed random effects to be more suitable with Exponential survival times and Normal random effects with Weibull survival times.

4.2 Limitations of the study

There were a few limitations when carrying out this study.

- Only positive association between the two survival responses could be captured with the use of shared random effects
- Was able to obtain negative associations with high variances and low shared constants (high negative values), which are considered to be extreme values, and thus, the convergence of datasets was difficult
- As PROC NLMIXED only allows normal random effects, to capture non-normal random effects, i.e. gamma random effects, probability integral transformation was applied. To follow the steps, scale and shape parameters were fixed to 1

4.3 Suggestions for further work

The following suggestions can be made for further work of this study.

- Extend the study by considering the Copula model
- Use correlated frailty terms instead of shared frailty terms
- Compare the model performances by changing different acceptable values for the parameters with more correlation structures
- Further, apply different frailty term distributions other than normal and gamma distributions

Acknowledgement

First and foremost, I would like to extend my sincere gratitude to my supervisor, Dr. Hasani Karunarathna, for giving me this great opportunity to do this research on her proposed topic. I am so much grateful for giving me her invaluable expertise, continuous guidance, clear and patient explanations with insightful feedback and for all the sacrifices done to make my research a success.

I am extremely grateful to my loving parents, husband, and my friends for the motivation and support given throughout this work and for always being with me in all ways. Also, I would like to offer my thanks to everyone whom I have not mentioned in this and who helped me even with a single word.

References

Bender, R., Augustin, T. and Blettner, M (2003) 'Bender, Augustin, Blettner: Generating Survival Times to Simulate Cox Proportional Hazards Models', *Survival*, 338.

Bender, R., Augustin, T. and Blettner, M. (2005) 'Generating survival times to simulate Cox proportional hazards models', (July 2004), pp. 1713–1723. DOI: 10.1002/sim.2059.

Chi, Y. Y. and Ibrahim, J. G. (2006) 'Joint models for multivariate longitudinal and multivariate survival data', *Biometrics*, DOI: 10.1111/j.1541-0420.2005.00448.x.

Ibrahim, J. G., Chen, M.-H. and Sinha, D. (2001) 'Frailty Models', *Biometrics*, pp. 100–154. DOI: 10.1007/978-1-4757-3447-8_4.

Karunarathna, G. H. S. and Sooriyarachchi, M. R. (2019) 'Joint Multilevel Model for Analyzing Length of Stay through Competing Endpoints in Dengue Epidemiology',*Sri Lankan Journal of Applied Statistics*, 19(1), p. 45. DOI: 10.4038/sljastats.v19i1.7997.

Laux, S. E. (1985) Statistics for Biology and Health Series, *IEEE Transactions on Electron Devices*

Lee, E. T. and Go, O. T. (1997) 'SURVIVAL ANALYSIS IN PUBLIC HEALTH RESEARCH'.

Li, G., Lesperance, M. and Wu, Z. (2020) 'Joint Modeling of Multivariate Survival Data With an Application to Retirement', *Sociological Methods and Research*. DOI: 10.1177/0049124120914928.

Mood, A., Graybill, F., and Boes, D. (1974) Introduction to the Theory of Statistics. 3rd Edition, McGraw-Hill, New York, 122-123.

Nelson, K. P. et al. (2004) 'Using the probability integral transformation for non-normal random effects in non-linear mixed models By', (March 2004).

Qian, J., Li, B. and Chen, P. (2010) 'Generating survival data in the simulation studies of Cox model'. DOI: 10.1109/ICIC.2010.294.

Ratcliffe, S. J. et al. (2004) 'Joint Modeling of Longitudinal and Survival Data via a Common Frailty', (December), pp. 892–899.

Rondeau, V. (2019) 'Survival Analysis with Correlated Endpoints : Joint Frailty-Copula Models Main Title : Survival Analysis with Correlated Endpoints Subtitle : Joint Frailty-Copula Models', (March).

Sunethra, A. A. and Sooriyarachchi, M. R. (2020) 'A novel method for joint modeling of survival data and count data for both simple randomized and cluster-randomized data', *Communications in Statistics - Theory and Methods*, 0(0), pp. 1–23. DOI: 10.1080/03610926.2020.1713366.