FRAILTY MODELS WITH APPLICATIONS IN MEDICAL RESEARCH : OBSERVED AND SIMULATED DATA

By:

Gachau W. Susan

I56/64582/2011

Supervisor:

Prof. JAM Ottieno

A project submitted in partial fulfillment for the Master of Science in Mathematical Statistics in the University of Nairobi.

June, 2014

DECLARATION

This project is my original work and has never been presented in any learning institution for any academic award.

GACHAU SUSAN WANJA (I56/64582/2011)

Signature..... Date.....

This thesis has been submitted for examination with my approval as university supervisor.

PROF. JAM OTTIENO.

Signature..... Date.....

ABSTRACT

Multivariate or cluster failure time data are common in survival analysis and finding an appropriate method to model the correlation among the observations is a very important issue for valid and reliable statistical inference. The primary objective of this project was to review various models for clustered survival data with focus on frailty models and their properties. Semi parametric Cox marginal and frailty models were used to analyze observed right censored data from a multicenter clinical trial. A simulation study was conducted to assess the impact of frailty distribution mis-specification on parameters estimates. Different settings in terms of the number of centers and true heterogeneity parameter were considered.

From the observed data, the estimated heterogeneity parameters were small yielding insignificant center effect. From the simulation study, the regression coefficient was less affected by mis-specification of the frailty distribution and initial simulation settings compared to the heterogeneity parameter. In conclusion, in the absence of center effect, event times were homogenous between and within the centers. From simulation study, gamma frailty model would be a practical choice in real data analysis with time to event endpoint when the regression parameters are of primary interest and when the choice of frailty distribution is not straightforward.

Key words: *Frailty, Heterogeneity parameter, Regression parameter.*

ACKNOWLEDGEMENTS

First I thank the almighty God for giving me life and strength to successfully complete this masters project. My sincere appreciation goes to my supervisor, Prof. JAM Ottieno for all his guidance, time, constructive advice and suggestions that led to the successful completion of this project.

I would like to acknowledge my lecturers at School of Mathematics for imparting part of their knowledge to me. In particular, I would like to thank Professor Weke and Professor Manene for all the assistance in the processing of scholarship grant from Eastern African Universities Mathematics Programme (EAUMP) in conjuction with International Science Programme (ISP) Uppsala University, Sweden which enabled me to pursue this valuable Masters program.

My appreciation goes to my family and relatives for the love and support in any way during my student years. To my classmates and friends, I appreciate your love, friendship, advice and all the encouragement.

Susan Gachau.

$27^{th} June, 2014.$

DEDICATION

To my family and friends.

Contents

List of Abbreviations

CI : Confidence Interval

EM : Expectation- Maximization

HR : Hazard Ratio

K-M : Kaplan - Meier estimator

NCIC-CTG : National Cancer Institute of Canada- Clinical Trial Group

OS : Overall Survival

PFS : Progression -Free Survival

PPL : Penalized Partial Likelihood

RB : Relative Bias

REML :Restricted Maximum Likelihood Estimate

SD : Standard Deviation

SE : Standard Error

STS : Soft-Tissue Sarcomas

List of Figures

List of Tables

ï»¿

Chapter 1

General Introduction

1.1 Background

Survival analysis also referred to as "time to event analysis" is a statistical method for data analysis where the outcome variable of interest is the time to the occurrence of an event. It is applied in a number of fields such as public health, social science and engineering. In medical science, time to event can be time until tumor recurrence in a cancer study, time to death or time to infection. In the social sciences, interest can lie in analyzing time to events such as job changes and so forth. In the engineering sciences, survival analysis is called failure-time analysis since the main focus is in modeling the lifetimes of machines or electronic components. Although different disciplines may emphasize slightly different approaches and techniques, survival analysis is the name that is most widely used and recognized (Xin, 2009).

One important survival analysis technique is the non-parametric methods developed by Kaplan and Meier (1958). These methods work well for homogeneous samples, but do not determine whether or not certain variables are related to the survival times. This shortcoming led to the development of methods such as univariate Cox proportional hazards (PH) model for analyzing survival data in the presence of covariates or prognostic factors.

Multivariate or cluster failure time data are also common in survival analysis and finding an appropriate method to model the correlation among the observations is a very important issue for valid and reliable statistical inference. For instance, in randomized controlled trials, subjects are recruited at multiple study centers with an aim to provide adequate sample sizes to enhance generalizability of study results. However, factors that vary by center, patient characteristics and medical practice patterns potentially lead to clustering or dependence between outcomes at each center (Demissie, 2009). Therefore, data analysts must choose when and how to incorporate center effects into the analysis to avoid misleading study outcomes.

In the presence of clustered survival data, a natural framework for estimating the unexplained variability is through a frailty (random effects) model. The introduction of random effects in survival data modelling dates back to Beard (1959), who, in modelling mortality, introduced it in a univariate setting and called it longevity factor. Vaupel *et al.* (1979) on the other hand introduced frailty models as a generalization of the Cox's proportional hazards model allow-

ing for random effects as a result of unobserved heterogeneity of each individual or a group of people. In this model, the unobserved frailty shared by individual members in a cluster acts multiplicatively as a factor on the hazard function and is normally modelled parametrically (Li *et al.*, 2007; Legrand *et al.*, 2006; Govindarajulu *et al.*, 2009; Ha and Gilbert, 2010).

1.2 Problem statement and consequence of ignoring the frailty

Although more heterogeneous trials lead to more general conclusions as they are based on a wider patient population, Duchateau *et al.* (2002) noted that heterogeneity decreases the power to detect clinically important treatment differences. Ha *et al,*. (2012) further noted that such heterogeneity may alter the reporting and interpretation of the treatment effect. It is therefore important to find out what factors cause this heterogeneity as it might help to improve the quality of patient care (Legrand *et al.*, 2006). These factors include patient-specific factors and center-specific factors.

The objective of the Cox proportional hazard model is to assess the effects of the covariates by estimating their coefficients. However, the covariates do not always fully account for the true differences in risk especially in clustered survival data. Therefore, including the unobserved frailty term in the model enhances correct measure of covariates effect avoiding underestimation or overestimation of the parameters.

Many frailty distributions amongst them the lognormal and the power variance function family comprising of gamma, Inverse Gaussian, positive stable and compound Poisson distributions have been studied by different authors. However, some of these distributions are not used in practice due to software limitations. Moreover, there is lack of sound estimation procedures for more complex frailty models. Additionally, due to the latent nature of the frailty term, it can be dif-

ficult to determine an appropriate frailty distribution for a particular data set. Thus, mis-specification of this unobserved covariate can occur, leading to biased estimates, reduced efficiency of the model estimates hence misleading conclusions (Li et al., 2007; Moreno, 2008).

1.3 Study objectives

With the problem and consequence associated with ignoring the frailty in clustered survival data, it is important in practice to examine to what extent mis-specification of the frailty distribution affects the validity of the regression coefficients and heterogeneity parameter estimates. In this regard, the primary objective of this project is to review various survival models for clustered data with focus on frailty models and their properties. To achieve this, the problem is broken down into specific sub-sections as follows:

1. Apply semi-parametric frailty models and marginal models to observed clinical trial data and compare parameter estimates and assess the estimated heterogeneity parameters.

2. Investigate the impact of frailty distribution mis-specification on the parameters of interest i.e. treatment log hazard effect and heterogeneity parameter as well as assess the sensitivity of these parameter estimates in terms of bias with respect to varying baseline hazard distributions and simulation settings.

1.4 Report outline

This project report is organized as follows: In chapter 2 literature review on previous studies on frailty model applications is presented. Chapter 3 reviews statistical methods for survival data. Chapter 4 discusses the properties of various frailty distributions as well as some baseline hazard distributions. In Chapter 5, estimation methods and their properties for parametric and semi-parametric frailty models are presented while descriptions of a case study and corresponding results are found in chapter 6. In chapter 7, a simulation scheme and simulations results for the parameters of interest are presented while chapter 8 provides the discussion. In chapter 9, the conclusion, limitations of the study and recommendations for further research are provided. The last sections present the references and appendix respectively.

Chapter 2

Critical Literature Review on Studies on Previous Multi-center Clinical Trials Studies

2.1 Introduction

Several studies on multi-center clinical trials have been conducted and this chapter presents a critical review on a few studies aimed at identifying areas where a particular study performed well, its limitations and gaps for possible improvements based on: study objectives, statistical methods and simulations schemes, results and conclusions.

2.2 Previous case studies

The first study considered was by Duchateau and Janssen, (2008) involving peri-operative breast cancer multi-center clinical trial study data. In this study, a semi-parametric marginal model and semi-parametric frailty models were fitted and parameters of interest compared . Additionally, a simulation study was conducted to investigate how the bias and the spread of the estimated heterogeneity parameter θ around its true value was influenced by

(i) the size of the multi-centre trial (which is determined by the number of clusters and the number of patients per cluster $(n_i = n)$)

(ii) the event rate $h_0(t)$ (assumed to be constant over time: $h_0(t) = h_0$)

(iii) the size of the true heterogeneity parameter θ

(iv) the size of the true treatment effect β (expressed in terms of the hazard ratio $HR = exp(\beta)$).

In their simulation scheme, the number of centres varied between 15 and 30 centers with 20, 40 or 60 patients per center. They studied two types of breast cancer trials: the early breast cancer clinical trial with a low yearly constant hazard rate set at $h_0 = 0.07$ and metastatic breast cancer clinical trial with a high yearly constant hazard rate set at $h_0 = 0.22$. They assumed an accrual period of 5 years (with constant accrual rate) and a further follow-up period of 3 years. Time at risk for a particular patient consisted of the time at risk before the end of the accrual period (ranging from 0 to 5 years) plus the follow-up time. This resulted in approximately 30% and 70% of the patients having the event in

the early breast cancer and metastatic breast cancer clinical trial, respectively, at the end of the study with the remaining patients censored. As true values for the heterogeneity parameter, 0, 0.1, and 0.2 were used as this is the most likely range of values to be observed in breast cancer clinical trials. For each parameter setting 6500 data sets were generated.

To investigate the robustness of the gamma frailty distribution assumption with respect to the lognormal distribution (model mis-specification), the gamma frailty model was used to fit clustered data generated from a lognormal frailty model. The results revealed that the downward bias of the variance estimator was more pronounced in the mis-specified model, for both the mean and the median of the variance estimates. Increasing the magnitude of θ from 0.1 to 0.2 lead to further discrepancy. They conclude that for small values of θ working with a mis-specified model still lead to acceptable estimates for the heterogeneity parameter but robustness was an issue for large values of θ.

One limitation of this study was the assumption of a constant accrual rate i.e. patients were enrolled into the study at a fixed rate. This is in contrast to real life clinical trials where patients within the same center and across different centers are normally enrolled into the study in a random manner at some point during the accrual period. In this regard, it would have been more appropriate to accrue patients following a random uniform distribution.

From this study, it is further noted that for both observed and simulated data, the heterogeneity parameter θ from the frailty models was estimated using the

Penalised Partial Likelihood (PPL) technique and the p-value of the loglikelihood ratio test for $H_0 : \theta = 0$ versus the alternative hypothesis $H_0 : \theta > 0$, assessed based on a χ^2_1 distribution. The use of the χ^2_1 distribution was inappropriate in this case since the value for θ in the null hypothesis is at the boundary of the parameter space. Although they acknowledged the problem of using χ^2_1 distribution, they did not provide an alternative test method. In such a problem a likelihood ratio test based on $\chi^2_{0:1}$ distribution is more appropriate because the test involves comparing a model with and without frailties.

Another study was conducted by Tundo, (2009) on frailty models for the between center variation in survival following rectum cancer diagnosis. The objectives of this study were to evaluate the performance of different regression outcome approaches for right-censored survival data in the presence of small centers. Specifically, the survival models considered in this study were the fully parametric exponential model, the semi-parametric Cox model with a dummy per center (fixed effects model), the semi-parametric frailty models with correct lognormal and mis-specified gamma distributed frailties. A simulation study was also conducted where the survival times of the patients were randomly simulated out of center specific exponential distributions with each center having its appointed log hazard rate. Censoring times for the patients were generated from a uniform distribution with zero as the lower limit and 5 as upper limit assuming that everybody is censored or died within 5 years. To obtain the observation time or censored failure time the minimum of the survival and censoring time

per patient was taken. The PPL estimation method was employed in this study. The root mean squared error (RMSE) was used as a measure of evaluating the performance of the different models.

Based on observed and simulated data, the overall performance of the frailty models was far much better than that of the fixed effects model or the fully parametric exponential model especially in handling centers with no events. In this study, convergence problems were experienced in fitting fixed effects models with many small centers. The strength of this study is that unlike most simulation studies on multi-center clinical trials in literature where the centers are taken to be of equal size, centers of varying sizes were used reflecting the situation in real life multi-center clinical studies.

Some limitations of this study were ambiguities and inconsistences in different sections of the report making it hard for the reader to understand or replicate the work. For instance, it was not stated in the study objectives what the parameter of interest were i.e. either the log hazard or the heterogeneity parameter or both. Furthermore, the choice of models and the initial parameters for example the true heterogeneity parameter for the two frailty densities considered in the simulation study was not clearly motivate. In this project, a mis-specified gamma frailty model was fitted and concluded that the models with mis-specified gamma and correct lognormal distributed frailties did not differ substantially. However, it was not indicated in the report how the mis-specification aspect came about. Additionally, there was no standardization of heterogeneity parameters used in

the simulation of frailties. For this reason, it was not possible to compare the results of correctly and mis-specified models.

Glidden and Vittinghoff (2004) conducted a study on Modelling clustered survival data from multi-center clinical trials. They surveyed approaches to multi-centrer clinical trials for censored time to event data. A simulation study was undertaken to compare the performance of the three centre-specific models i.e. stratified, fixed effects Cox models and the frailty model with respect to bias, root mean squared error (RMSE) and empirical coverage of 95 per cent confidence intervals. Simulation settings varied with respect to total sample size (N =100, 400), number of subjects per centre (n=2, 10, 20), magnitude of intra-center dependence and the frailty distribution. The frailty distribution considered in this study were generated from the gamma, inverse Gaussian and positive stable densities. For each of the frailty distributions, the choice of parameters values was motivated by values that gave Kendall's τ of 0.50. A Weibull baseline hazard function with shape parameter 1.3 and scale parameter 5.0 respectively. They also examined the performance of the marginal Cox model in the setting of no centre effects.

From a broad range of simulation settings, they found that the frailty model approach compared favorably with competing methods (fixed effects and stratified approaches). With a small number of centres, the frailty model was only slightly less efficient than the population averaged (marginal) model and gave confidence intervals with considerably better coverage properties. No results are

presented for the fixed effects model with N =400 and n=2 because the models did not reliably converge. A problem also encountered by Tundo, (2009). In addition, their simulation results suggested that regression coefficient estimates were minimally affected by frailty distribution mis-specification (gamma frailty model fitted to inverse Gaussian generated frailties). The shortcoming of their simulation study is that they did not standardize the mean and variance of the generated frailties thus making comparability impossible.

2.3 Summary

From the above reviewed literature and others not included, there should be consistency between the study objectives and methods used. In case of simulation studies, it is important to standardize the parameters to allow for comparison and reduce chances of biased results. It is also important that the choice of initial parameters are clearly motivated and proper formal tests used depending on the problem and the methods used. In summary, we can examine the studies in terms of:

(i)Determinants of bias and the spread of estimated parameters.

(ii) Estimation techniques.

(iii)Performance measures in terms of bias.

Chapter 3

Statistical Methods for Survival Data

3.1 Introduction

The analysis of survival data requires special techniques because the data are almost always incomplete due to censoring and familiar parametric assumptions might be unjustifiable. For instance in biomedical research, the investigators follow patients until they reach a pre-specified endpoint for example, death or disease progression. However, some patients withdraw from the study or the study comes to an end before the endpoint is reached. In these cases, the survival times are *censored* i.e. subjects survived to a certain time beyond which their status is unknown. The uncensored survival times are often referred to as *event times*. There are at least three types of possible censoring schemes. Right

censoring is the most common type of censoring. For right censored data, all that is known for some individuals is a time beyond which the subject is still alive. In the left censoring, a failure time is only known to be before a certain time while interval censoring data reflects uncertainty as to the exact time the units failed within an interval (Demissie, 2009).

3.2 Relationships between $S(t)$, $f(t)$ and $h(t)$

Let T be a random variable denoting the survival time. The distribution of survival times is characterized by any of three functions: the survival function ($S(t)$), the probability density ($f(t)$) or the hazard function ($h(t)$). The survival function is defined as the probability that the survival time is greater or equal to t and is defined for both discrete and continuous T. Similarly, the probability density and hazard functions are easily specified for discrete and continuous T.

T discrete

For a discrete random variable T taking well-ordered values $0 \leq t_1 < t_2 < ...$, let the probability mass function be given by $P(T = t_j) = f(t_j), j = 1, 2, ...$, then the survival function is

$$s(t) = \sum_{j|t_j} f(t_j)$$

$$= \sum f(t_j) I_{(t_j > t)} \tag{3.1}$$

where the indicator function

$$I_{(t_j>t)} = \begin{cases} 0, & \text{if } t_j < t. \\ 1, & \text{if } t_j \geq t. \end{cases}$$

$$h_j = h(t_j) = P(T = t_j | T \geq t_j) = \frac{f(t_j)}{S(t_j)} = \frac{S(t_j) - S(t_{j+1})}{S(t_j)} = 1 - \frac{S(t_{j+1})}{S(t_j)}$$

Thus

$$1 - h(t_j) = \frac{S(t_{j+1})}{S(t_j)}$$

and taking the product on both sides, we get

$$\prod_{j|t_j<t} (1 - h(t_j)) = \frac{S(t_2)}{S(t_1)} \times \frac{S(t_3)}{S(t_2)} \times ... \times \frac{S(t_{j+1})}{S(t_j)} = S(t) \tag{3.2}$$

Since $S(t_1) = 1$ and $S(t) = S(t_{j+1})$.

$$\prod_{j|t_j<t} (1 - h(t_j)) = S(t)$$

Moreover,

$$f(t_j) = h(t_j)S(t)$$

Therefore substituting for $S(t)$ in the equation above, we have

$$f(t_j) = h(t_j) \prod_{i=1}^{j-1} (1 - h(t_i) \tag{3.3}$$

T continuous

For an absolutely continuous variable T, the hazard function gives the instantaneous failure rate at t given that the individual has survived up to time t. i.e.

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T \leq t + \Delta t | T > t)}{\Delta t} \tag{3.4}$$

T is nonnegative and represents the future lifetime of an individual. Thus

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T \leq t + \Delta t | T > t)}{Pr(T > t) * \Delta t}$$

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T \leq t + \Delta t)/\Delta t}{Pr(T > t)}$$

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{3.5}$$

The hazard rate, $h(t)$, is obtained from the conditional probability that an event occurs in the interval $[t, t + \Delta t]$ given that the event did not occur yet before time t. By definition;

$$S(t) = 1 - F(t) \tag{3.6}$$

Therefore

$$h(t) = \frac{f(t)}{S(t)}$$

and

$$f(t) = F^{'}(t) = -S^{'}(t) \tag{3.7}$$

Therefore

$$h(t) = \frac{-S^{'}(t)}{S(t)} \tag{3.8}$$

$$h(t) = \frac{-d}{dt} ln(S(t))$$

$$-\int h(t)dt = ln(S(t)) \tag{3.9}$$

After integrating and exponentiating we have

$$S(t) = exp(-\int h(t)dt)$$

$$S(t) = exp(-H(t))$$

$$H(t) = -ln[S(t)] \tag{3.10}$$

These three functions give mathematically equivalent specification of the distributions of the survival time T. If one of them is known, the other two are easily determined. One of these functions can be chosen as the basis of statistical analysis according to the particular situation. The survival function is most useful for comparing the survival progress of two or more groups while the hazard function gives a more useful description of the risk of failure at any time point (Qi, 2009).

3.3 Non-parametric methods

In general, survival data can be conveniently summarized through estimates of the survival function and hazard function. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time. These methods are said to be non-parametric methods since they require

no assumptions about the distribution of survival time. To compare the survival distribution of two or more groups, the log-rank tests can be used (Collet, 1994).

3.3.1 The Kaplan-Meier estimate of the survival function

In clinical trial studies, individual's data is normally available on time to death or time to last seen alive. The Kaplan-Meier (K-M) estimator (1958) for the survival curves which is a non-parametric method is usually used to explore such data. For example, Suppose that r individuals have failures in a group of individuals, let $0 \leq t_{(1)} < ... < t_{(r)} < \infty$ be the observed ordered death times. Let r_j be the size of the risk set at $t_{(j)}$, where risk set denotes the collection of individuals alive and uncensored just before $t_{(j)}$. Let d_j be the number of observed events at $t_{(j)}, j = 1, ..., r$. Then the K-M estimator of S(t) is defined by

$$\widehat{S(t)} = \prod_{j:t_{(j)}<t} (1 - \frac{d_j}{r_j}) \tag{3.11}$$

This estimator is a step function that changes values only at the time of each event. For further illustration purposes, the maximum likelihood for a K-M estimator for discrete case will be shown next. Suppose that the distribution is discrete, with τ_j at finitely many specified points $0 \leq \tau_1 < \tau_2 < ... < \tau_j$. As described in Section 3.2, the survival function $S(t)$ may be expressed in terms of the discrete hazard function h_j as

$$S(t) = \prod_{j|\tau_j<t} (1 - h_j) \tag{3.12}$$

To derive the full likelihood from a sample of n observations, we collect all the terms corresponding to the τ_j . Let $b_i = j$ if the i^{th} individual dies at τ_j : Using equation (3.3), the contribution to the total log likelihood is

$$logh_{b_i} + \sum_{k<b_i} log(1-h_k)$$

Let $e_i = j$ if the i^{th} individual is censored at τ_j ; using the equation above, the log likelihood contribution to the total likelihood is

$$\sum_{k<e_i} log(1-h_k)$$

Then the total log likelihood is given by

$$l = \sum_{death_i} logh_{b_i} + \sum_{death_i} [\sum log(1-h_k)] + \sum_{censor_i} [\sum_{k\leq e_i} log(1-h_k)]$$

$$= \sum_j d_j logh_j + \sum_k [\sum_{j>k} d_j] log(1-h_k) + \sum_k [\sum_{j\geq k} c_j] log(1-h_k)$$

$$= \sum_j d_j logh_j + (r_j - d_j) log(1-h_j) \tag{3.13}$$

where d_j is the number of observed death at τ_j, c_j is the number censored at $[\tau_j, \tau_{j+1})$ and r_j is the number of living and uncensored at τ_j. h_j is the solution of

$$\frac{\partial l}{\partial h_j} = \frac{d_j}{h_j} - \frac{r_j - d_j}{1-h_j} = 0$$

By solving the above equation, the maximum likelihood estimate of h_j is given by

$$\widehat{h_j} = d_j / r_j \tag{3.14}$$

This maximizes the likelihood since the total log likelihood function is concave down (Qi, 2009). Substitituing for h_j in equation (3.12), the K-M estimator of the survival function is

$$\widehat{S(t)} = \prod_{j|\tau_j<t} (1 - \frac{d_j}{r_j})$$

Therefore, the K-M estimator is the maximum likelihood estimator. The K-M estimator gives a discrete distribution. If the observations are known to come from unknown continuous distribution, then the maximum likelihood estimator does not exist (Johansen, 1978). One shortcoming of the K-M method is that it does not control for covariates and it requires categorical predictors.

3.4 Conditional and Marginal Cox models

In the presence of dependence induced by cluster effects, two distinct approaches are available i.e. the conditional (or cluster-specific) and the marginal (or population-averaged) models. These two approaches differ in estimation methods as well as interpretation, (Glidden and Vittinghoff, 2004).

3.4.1 Marginal Cox Proportion Hazard model

The marginal model (Cox proportional hazards mode) is a popular model in survival data and was proposed by Cox (1972). This model does not take the clustering into account and acts as if the event times are independent of each other, even if they belong to the same cluster,(Duchateau and Janssen, 2008). The hazard rate is expressed as

$$h(t|x) = h_0 exp(x_j^t \beta) \tag{3.15}$$

where h_0 is the baseline hazard function at time t, x is the vector of explanatory variables and β is a vector of unknown regression coefficients. When the baseline hazard h_0 is specified, the models is commonly referred to as parametric Cox PH model. On the other hand, this model is referred to as a semi-parametric Cox PH model when no parametric form is imposed on h_0, Collet (1994). In this case, parameters can be estimated by partial likelihood method presented by Cox (1972). Although the estimates are less efficient compared to the maximum likelihood estimates (for parametric baseline hazard), unspecified baseline hazard serves as a remedial virtue against mis-specification (Keele, 2007). The corresponding survival functions are related as follows:

$$S(t|x) = S_0(t)^{exp(\sum_{i=1}^{n} \beta_i x_i)} \tag{3.16}$$

As earlier mentioned, the marginal Cox PH model leave the structure of the intra-cluster association unspecified but adjust for it in the inference. The regression coefficients are assumed to be the same for all individuals hence interpreted at population averaged level as the log-hazard ratio; the hazard ratio is the measure of effect. The hazard ratio of two individuals with different covariates x and x^* is

$$(\widehat{HR}) = \frac{h_0(t)exp(\widehat{\beta}'x)}{h_0(t)exp(\widehat{\beta}'x^*)}$$

$$= exp(\sum \widehat{\beta}'(x - x^*)) \tag{3.17}$$

This hazard ratio is time-independent, the reason why this model is called the proportional hazards model.

3.4.2 Conditional models for survival data

The Fixed effects Cox model

This model assumes that the cluster effect is modeled by a fixed effect and is formulated as follows

$$h_{ij}(t) = h_0(t)exp(x_{ij}^t\beta + c_i) \tag{3.18}$$

where c_i is the fixed effect for the i^{th} cluster. This model is often over-parameterized and therefore the first cluster is set as the reference cluster i.e. $c_1 = 0$. An advantage of this method is that it does not put constraints on the distribution of the center-effects hence no chances of mis-specification (Tundo, 2009). According to Glidden and Vittinghoff (2004), the fixed effects model is most appropriate when the cluster effects are of essential interest. A major drawback of this approach is the large number of parameters that have to be estimated especially in the presence of many small clusters (Tundo, 2009).

The stratified Cox model

Stratified models are conditional models formulated as follows

$$h_{ij}(t) = h_{i0}(t)exp(x_{ij}^t\beta) \tag{3.19}$$

where h_{i0} is the baseline hazard for the i^{th} cluster. This model assumes that the baseline hazards are completely unrelated nuisance functions and could have dif-

ferent shapes with some or the entire hazard functions unequal. It also assumes that regression coefficients are the same in each stratum. The ease of computation and the applicability across a wide variety of settings make the stratified Cox model an appealing tool, especially if clustering is of no essential interest or if frailties act non-proportionally on the baseline risk. However, a major drawback of this approach is that it results to discarding a considerable amount of information from the sample. Glidden and Vittinghoff, (2004) noted that for a fixed sample size, the loss of information increases with the number of clusters.

Frailty models (Random effects models)

In frailty models, the variability of survival times can be divided into two parts. One part is the observed risk factors, known as covariates and the other part is unobserved risk factors, known as frailty. Including the frailty term in the model allows to correctly measure the covariate effects avoiding underestimation or overestimation of the parameters (Li *et al.*, 2007). The advantage of frailty models over other conditional models is that they use a single parameter to index the degree of dependence; in contrast to the fixed effects model, where the number of parameters to describe cluster effects grow with the number of clusters. Frailty models are used to make adjustments for overdispersion/underdispersion. When unobserved or unmeasured effects are ignored, the estimates of survival may be misleading. Therefore, corrections for this overdispersion/underdispersion is needed in order to allow for adjustments for those important frailties.

3.4.3 Univariate frailty models

The univariate frailty model presents the population as a mixture in which baseline hazard is common to all individuals but each individual has his/her own frailty. Suppose we have a sample of j observations in a study. Some of these observations fail earlier than others due to unobserved heterogeneity. The proportional hazards model assumes that conditional on the frailty, the hazard function for an individual at time $t > 0$ is

$$h_j(t) = h_0(t)exp(x_j^t\beta + W_j\psi), j = 1, ..., n; \tag{3.20}$$

where W_j is a frailty term from a probability distribution. If W_j could be measured and included in the model, then ψ would go to 0 and we would obtain the marginal Cox PH model. The hazard function conditional on both covariates and frailty can be rewritten as

$$h_j(t) = h_0(t)u_jexp(x_j^t\beta), j = 1, ..., n \tag{3.21}$$

where $u_j = exp(W_j)$. This shows that the hazard of an individual also depends on an unobservable random variable, u_j , which acts multiplicatively on the hazard rate. If frailty is not taken into account, then $u_j = 1$.

3.4.4 Shared frailty models

This frailty model allows the individuals in the same cluster to share the same frailty value (Ulviya, 2013). In this regard, a random effect is introduced for each cluster so that subjects from one cluster are more alike than subjects from

different clusters. The random effect describes the unobserved influences common to all subjects of that particular cluster (Legrand *et al.*, 2006) and the variance of these random effects is a measure of the heterogeneity in the outcome between clusters. The conditional hazard function at time t for the j^{th} subject in the i^{th} cluster is given by

$$h_{ij}(t) = h_0(t)exp(x_{ij}^t\beta + w_i), j = 1, ..., n_i \tag{3.22}$$

where h_0 is the baseline hazard at time t (can either be specified parametrically with a distribution or left unspecified). x_{ij}^t is the vector of subject specific covariates and β is the corresponding vector of regression coefficients (unknown parameters). w_i is the random effect for center i. Though the random effects $w_i^{'}s$,i=1,...,G are unobserved, it is assumed that they are independent and identically distributed from a density $f_W(\bullet)$. The corresponding frailty model can be re-written as follows

$$h_{ij}(t) = h_0(t)exp(w_i)exp(x_{ij}^t\beta)$$

$$h_{ij}(t) = h_0(t)u_iexp(x_{ij}^t\beta) \tag{3.23}$$

where $u_i = exp(w_i)$ is known as the frailty and acts multiplicatively on the hazard rate for the j^{th} patient in the i^{th} center (Nguti, 2003). Model (3.22) is called the shared frailty model because subjects in the same cluster all share the same frailty factor (Duchateau and Janssen, 2008). For this model, regression coefficients are interpreted conditional on the center random effect. In this project, main focus will be on shared frailty models with the assumption that the center random effect operates at a group (center) level.

Heterogeneity parameter

In frailty models, θ is estimated to get an idea on heterogeneity in the outcome between clusters. When θ is large and differs significantly from zero; it reflects heterogeneity between clusters and a strong association among individuals in the same cluster. On the other hand, when θ is equal to zero, the frailties are identically equal to one which implies that the cluster effects are not present and events are independent within and across centers (Glidden and Vittinghoff, 2004). The likelihood ratio test comparing the models with and without frailties is normally used for testing the null hypothesis $\theta = 0$ versus the alternative hypothesis $\theta > 0$. Since the null hypothesis is at the boundary of the parameter space, a mixture of chi-square distribution with 0 and 1 degree of freedom was used as suggested by Duchateau and Janssen (2008).

Kendall's τ measures of dependence

Most dependence measures have been developed for bivariate data, we describe the measures for such data. For two randomly chosen clusters i and k of size two, the event times are (T_{i1}, T_{i2}) and (T_{k1}, T_{k2}). The assumption is that the covariate information is the same in each cluster.

Kendall's τ (Kendall, 1938) is a global measure of dependence and is defined as

$$\tau = E[sign((T_{i1} - T_{k1})(T_{i2} - T_{k2}))] \tag{3.24}$$

where $sign(x) = -1$, 0,1 for $x < 0$, $x = 0$, $x > 0$. An alternative formulation for

continuous distributions (Genest and MacKay, 1986) is given by

$$\tau = P((T_{i1} - T_{k1})(T_{i2} - T_{k2}) > 0) - P((T_{i1} - T_{k1})(T_{i2} - T_{k2}) < 0$$

$$= 2P((T_{i1} - T_{k1})(T_{i2} - T_{k2}) > 0) - 1$$

$$2p - 1$$

3.5 Handling ties in survival data

In survival analysis, it is common for the data sets to contain ties in events that is, two or more individuals share the same time. Usually, time is considered to be a continuous variable, in which case the probability of a tie is zero. In practice, the accuracy of measurement is often more limited and two observed times can have the same value. For instance when the measurement unit is in years, two people that died in the same year will have the same event time recorded, even though it is very unlikely that they died at the same moment of time. One assumption of the Cox proportional hazards model is that there are no tied data, however in real applications, tied event times are commonly observed and a modification of Cox's partial likelihood function needed to handle these ties. The following hypothetical data set was used to demonstrates on how to handle ties between event times:

Table 3.1: *Demonstration of Ties between Event Times*

id.	**Fail**	**Trans**	**Event**
1	12	10	1
2	12	8	1
3	16	NA	1
4	9	NA	1
5	20	7	0
6	9	5	1
7	11	NA	0

In Table 3.1, *id* represents the identification number of the patients. The $Fail$ column is the lifetime or censoring time for each individual; the *transpl* column represents the time that the patients get a heart transplant where NA indicates that the particular patient never got a transplant before the end of study. The last column *Event* is an indicator variable in which 1 means the patient died at the $fail$ time and 0 means the patient did not die before the end of the study and therefore was censored. In this data set, patients 1 and 2, as well as 4 and 6 have the same failure time. Therefore there are ties between patients 1 and 2 and between patients 4 and 6. Several methods have been developed to perform survival analysis with tied data. The exact method (Allison, 2010), Breslow approximation (Breslow, 1974), Efron approximation (Efron, 1977) and the discreet method are the most commonly used methods. The following section

presents details and an illustration of how to handle ties for each of the above named methods using the hypothetical data set in Table 3.1.

3.5.1 Exact method

This method assumes that the ties occur as a result of imprecise measurement of continuous time, hence there exists an underlying ordering for the tied events. As a result, when calculating the partial likelihood for the fitted model, all the possible orderings need to be taken into consideration (Allison, 2010). In the hypothetical data set, there exists a tie between patient 1 and patient 2. The assumption of the exact method is that because of the limit of fineness of the measurement, patient 1's event time can either be before or after patient 2's event time, which gives us two possibilities. The partial likelihood will include both possibilities and therefore includes the sum of all the possibilities of all possible orderings (Xin, 2011). From the data set, at time 12

$$L_{12} = \frac{1}{2}\left(\frac{e^{\beta x_1}}{e^{\beta x_1} + e^{\beta x_2} + e^{\beta x_3} + e^{\beta x_5}}\right)\left(\frac{e^{\beta x_2}}{e^{\beta x_2} + e^{\beta x_3} + e^{\beta x_5}}\right)$$

$$+\frac{1}{2}\left(\frac{e^{\beta x_2}}{e^{\beta x_1} + e^{\beta x_2} + e^{\beta x_3} + e^{\beta x_5}}\right)\left(\frac{e^{\beta x_1}}{e^{\beta x_1} + e^{\beta x_3} + e^{\beta x_5}}\right) \tag{3.25}$$

The partial likelihood contribution shown above consists of the sum of two products. If we compare each product with Cox's partial likelihood in Section 5.1, the first product is for the possibility that patient 1 fails before patient 2 and the second is for the possibility that patient 2 fails before patient 1. The exact method is a very precise method; however, since it is based on permutations,

this method can become computationally infeasible when there is a multitude of time points that contain a large number of ties (Xin, 2011; Allison, 2010).

3.5.2 Breslow approximation

This approximation method assumes that event times are continuous and the hazard of event is constant in the interval (t_i, t_{i+1}) (Breslow, 1974). Furthermore, an individual whose censoring time falls in the interval (t_i, t_{i+1}) is assumed to have been censored at the start of the interval, that is at time t_i. Letting x_j be the vector of covariates for the j^{th} individual, set D_i consist of d_i individuals who failed at time t_i. Also, letting R_i be the risk set at time t_i, such that R_i contains all the individuals that are alive or at risk at time t_i, the partial likelihood for the Breslow approximation (Hertz-Picciotto and Rockhill, 1997) is:

$$L(\beta) = \prod_{i=1}^{k} \frac{exp[(\sum_{j \in D_i} x_j)\beta]}{[\sum_{j \in R_i} exp(x_j \beta)]^{d_i}} \tag{3.26}$$

Consequently, the contribution to the partial likelihood for time 12 can be approximated by:

$$L(\beta) = \frac{e^{\beta x_1} e^{\beta x_2}}{[e^{\beta x_1} + e^{\beta x_2} + e^{\beta x_3} + e^{\beta x_5}]^2} \tag{3.27}$$

The Breslow approximation is obtained by setting the denominators of each ratio in L_{12} to $e^{\beta x_1}+e^{\beta x_2}+e^{\beta x_3}+e^{\beta x_5}$ which includes all the patients in the risk set at the first event time. The calculation is much simpler than the partial likelihood for the exact method. However, the Breslow approximation becomes more complex as the number of ties at a particular time point becomes large relative to the number of patients in the risk set (Kalbeisch and Prentice, 2002).

3.5.3 Efron approximation

The Efron approximation is more accurate than the other approximations (Allison, 2010; Kalbeisch and Prentice, 2002). However as the percentage of ties increases, the performance of all approximations becomes worse since their partial likelihoods will become more different from the exact partial likelihood. The partial likelihood function for the Efron's approximation (Hertz-Picciottoand Rockhill, 1997) is:

$$L(\beta) = \prod_{i=1}^{k} \frac{exp[(j \in D_i x_j)\beta]}{\prod_{l=1}^{d_i}[\sum_{j \in R_i} exp(x_j\beta) - \frac{l-1}{d_i}\sum_{j \in D_i} exp(x_j\beta)]} \tag{3.28}$$

As a result, the contribution to the partial likelihood for time 12 can be approximated by:

$$\frac{e^{\beta x_1} e^{\beta x_1}}{[e^{\beta x_1} + e^{\beta x_2} + e^{\beta x_3} + e^{\beta x_5}][(e^{\beta x_1} + e^{\beta x_2} + e^{\beta x_3} + e^{\beta x_5}) - \frac{1}{2}(e^{\beta x_1} + e^{\beta x_2})]}$$

Compared to the Breslow approximation, the partial likelihood of the Efron approximation is closer to the exact method since the denominators are not simply treated as the same term $e^{\beta x_1} + e^{\beta x_2} + e^{\beta x_3} + e^{\beta x_5}$ as in the Breslow approximation. At the same time, the calculation for the Efron approximation still remains simpler than the exact method.

3.5.4 Discrete method

The method is also an exact method but based on a different model. With the assumption that time is really discrete, there is no underlying ordering as in the exact method as assumed in the original Cox proportional hazards model

(Allison,2010; Cox, 1972). In the hypothetical data set when there is a tie at time 12 between patient 1 and patient 2, the discrete method treats these events as if they did happen at the same time point. Cox proposed a similar model for discrete-time data, which is sometimes called the proportional odds model (Cox, 1972; Allison, 2010). However, since it is still computationally time-consuming, another approximation was developed by Kalbeisch and Prentice (1973). Let d_i be the number of individuals who fail at time t_i. Let q be one subset of d_i and Q_i be a set that includes all possible q, i.e., Q_i is the set of all subsets of size d_i from the risk set at time t_i. Let $s_q = \sum_j^d z_{qj}$, which is the sum of the covariate values for a specific subset q. Then the partial likelihood for the discrete method proposed by Kalbeisch and Prentice (1973) is:

$$L(\beta) = \prod_{i=1}^{k} \frac{exp[(\sum \sum j \in D_i x_j)\beta]}{\sum_{q \in Q_i} exp(s_q^* \beta)}$$

Similar to the Breslow and Efron approximations, the calculation of the discrete method is simpler than the exact method. However, when the percentage of events is large in the study and the measurement unit is imprecise so there are many ties, the estimated parameters from this approximation are more likely to be biased (Xin,2011; Kalbeisch and Prentice, 1973). In this project, the Breslow approximation method was used to handle ties in the observed and simulated data sets.

Chapter 4

Statistical Frailty Distributions and Parametric Baseline Distributions

4.1 Introduction

There are various frailty models that have been developed and suggested in the literature and any distribution with a positive random variable can be used to model frailty (Ulviya, 2013). Several authors have noted that unlike standard random effects models, inferential methods have been less developed in frailty models because of censoring and truncation. The frailty distributions most often applied are the gamma distribution (Clayton, 1978; Vaupel *et al.*, 1979; Oakes, 1982; Hougaard, 2000; Wienke *et al.*, 2002; Wienke *et al.*, 2003a; Hanagal and Sharma, 2012), the positive stable distribution (Hougaard, 1986b), the power Variance function (PVF) distribution (Hougaard, 1986a), the inverse Gaussian

distribution (Hougaard, 1984), the compound Poisson distribution (Aalen, 1988) and the log-normal Distribution (McGilchrist and Aisbett, 1991). This chapter, presents details and properties of various frailty distributions.

4.2 Gamma distribution

Gamma frailty model belongs to the power variance function family (Hougaard, 1986b) and can be expressed in terms of its Laplace transform from which properties such as mean and variance are easily derived (Duchateau and Janssen, 2008). Assuming a two-parameter gamma density with $\delta > 0$ and $\gamma > 0$ as shape and scale parameters respectively, the density function is given by

$$f_U(u) = \frac{\gamma^\delta u^{\delta-1} exp(-\gamma u)}{\Gamma(\delta)} \tag{4.1}$$

with $\delta > 0$ and $\gamma > 0$. The corresponding Laplace transform is given by

$$L(s) = \int_0^\infty exp(-us) f_U(u) du \tag{4.2}$$

$$= \int_0^\infty exp(-us) \frac{\gamma^\delta u^{\delta-1} exp(-\gamma u)}{\Gamma(\delta)} du$$

$$= \frac{\gamma^\delta}{\Gamma(\delta)} \int_0^\infty exp - u(s+\gamma) u^{\delta-1} du$$

Letting $y = u(s+\gamma)$; $u = \frac{y}{(s+\gamma)}$ and $du = \frac{dy}{(s+\gamma)}$

Substituting for u and du we get

$$= \frac{\gamma^\delta}{\Gamma(\delta)} \int_0^\infty exp - y(\frac{y}{(s+\gamma)})^{\delta-1} \frac{dy}{(s+\gamma)}$$

$$= \frac{\gamma^{\delta}}{(s+\gamma)^{\delta}\Gamma(\delta)} \int_0^{\infty} exp - y(y)^{\delta-1} dy$$

Where $\Gamma\delta = \int_0^{\infty} exp - y(y)^{\delta-1} dy$

Hence

$$L(s) = (\frac{\gamma^{\delta}}{(s+\gamma)^{\delta}\Gamma(\delta)}) * \Gamma\delta$$
$$= \frac{\gamma^{\delta}}{(s+\gamma)^{\delta}}$$
$$= \gamma^{\delta}(s+\gamma)^{-\delta} \tag{4.3}$$

L(s) exists in the neighbourhood of zero and the mean and variance can be obtained by using the first and second derivatives of the Laplace transform.

$$L^{(1)}(s) = -\delta\gamma^{\delta}(s+\gamma)^{-\delta-1} \tag{4.4}$$

$$L^{(2)}(s) = \delta(\delta+1)\gamma^{\delta}(s+\gamma)^{-\delta-2} \tag{4.5}$$

Evaluating these derivatives at s=0, the expected mean and variance are;

$$E(U) = (-1)L^{(1)}(0) = \delta/\gamma \tag{4.6}$$

$$Var(U) = L^{(2)}(0) - (-L^{(1)}(0))^2$$

$$= \frac{\delta(\delta+1}{\gamma^2} - (\frac{\delta}{\gamma})^2$$

$$= \frac{\delta^2 + \delta - \delta^2}{\gamma^2}$$

$$= \frac{\delta}{\gamma} \tag{4.7}$$

In gamma frailty models, restriction $\delta = \gamma$ is used, which results in expectation of 1. The variance of the frailty variable is then $\frac{1}{\gamma}$. Assuming that the frailty term u is a gamma with $E(U) = 1$ and $Var(U) = \theta$, then $\delta = \gamma = 1/\theta$ (Ulviya, 2013). The distribution function of the frailty term u is therefore a one-parameter gamma distribution given by

$$f_U(u) = \frac{u^{1/\theta} exp(-u/\theta)}{\theta^{1/\theta}\Gamma\theta} \tag{4.8}$$

where $\theta > 0$ and $u > 1$ indicates that individuals in group i are frail, whereas $u < 1$ indicates that individuals are strong and have lower risk. The corresponding Laplace transform is given by;

$$L(s) = (1 + \theta s)^{1/\theta} \tag{4.9}$$

Once the frailty is integrated out, accounting for unobserved heterogeneity is reduced to estimating the variance of the frailty term. The variance θ of the frailty term represents the heterogeneity among clusters while the mean is constrained to 1 in order to make the average hazard identifiable (Duchateau *et al.*, 2002; Nguti, 2003; Glidden and Vittinghoff, 2004; Duchateau and Janssen, 2008). The ease of interpretation coupled with the analytic simplicity and variety of forms as the parameter varies has popularized the use of the gamma frailty model in the correlated failure time analysis (Li et al., 2007). However, there are no known biological reasons which make the gamma distribution preferable than other distributions (Hougaard, 1995).

4.3 Inverse Gaussian distribution

The inverse Gaussian density function is given by

$$f_U(u) = (\alpha/2\pi)^{1/2}u^{-3/2}exp(\frac{-\alpha}{2u\mu^2}(u-\mu)^2) \tag{4.10}$$

with $\mu > 0$ and $\alpha > 0$. The corresponding Laplace transform is

$$L(s) = \int_0^\infty exp(-su)(\frac{\alpha}{2\pi})^{1/2}u^{-3/2}exp(\frac{-\alpha}{u\mu^2}(u-\mu)^2)du$$

$$= (\frac{\alpha}{2\pi})^{1/2}exp(\frac{\alpha}{\mu})\int_0^\infty u^{-3/2}exp(-(\frac{\alpha}{2\mu^2}+s)u)exp(-\frac{\alpha}{2u})du$$

$$\begin{aligned} &= exp(\frac{\alpha}{\mu})exp(-2(\frac{\alpha}{2})^{1/2}(\frac{\alpha}{2\mu^2})^{1/2}) \\ &= exp(\frac{\alpha}{\mu} - (\frac{\alpha^2}{\mu^2}+2\alpha s)^{1/2}) \end{aligned} \tag{4.11}$$

The first and second derivatives of the Laplace transforms are given by

$$L^{(1)}(s) = -\alpha exp(\frac{\alpha}{\mu})exp(-(\frac{\alpha^2}{\mu^2}+2\alpha s)^{1/2})(\frac{\alpha^2}{\mu^2}+2\alpha s)^{-1/2} \tag{4.12}$$

$$\begin{aligned} L^{(2)}(s) = \alpha^2 exp(\frac{\alpha}{\mu})exp(-(\frac{\alpha^2}{\mu^2}+2\alpha)^{1/2})(\frac{\alpha^2}{\mu^2}+2\alpha s)^{-1} + \\ \alpha^2 exp(\frac{\alpha}{\mu})exp(-(\frac{\alpha^2}{\mu^2}+2\alpha s)^{1/2})(\frac{\alpha^2}{\mu^2}+2\alpha s)^{-3/2} \end{aligned} \tag{4.13}$$

Evaluating the derivatives at $s = 0$ we have

$$E(U) = -L^1(0) = \mu \tag{4.14}$$

$$Var(U) = L^{(2)}(0) - (-L^1(0))^2 = \mu^3/\alpha \tag{4.15}$$

Substituting $\mu = 1$ in equation (4.11) we obtain the following simplified Laplace transform

$$L(s) = exp[\alpha(1 - (1 + 2\alpha^{-1}s)^{1/2})]$$

For $\mu = 1$ we have that $\theta = Var(U) = 1/\alpha$ (so $\alpha = \infty$ corresponds with no heterogeneity($\theta = 0$)).

4.4 Positive stable distribution

In general, the stable distributions have the property that, with $Y_1, ..., Y_n$ independent and identically distributed (iid) random variables, for each n there exists a normalising constant $c(n)$ such that $D(\sum_{i=1}^{n} Y_i) = D(c(n)Y_i)$ where$D(Y)$ means the distribution (law) of Y. The constant $c(n)$ takes form $n^{1/\theta}$ with $\theta \in (0, 2]$ θ being called the characteristic exponent (Duchateau and Janssen, 2008). The standard normal density function is a stable density. For iid standard normal distributed random variables $Y_1, ..., Y_n$ we have $D(\sum_{i=1}^{n} Y_i) = D(n^{1/2}Y_i)$ i.e. $\theta = 2$. The stable distributions on the positive half line have $\theta \in (0, 1]$ ($\theta = 1$) corresponds to the degenerate distribution. To link this with frailty distributions let $U = Y_1$ and the density function is then given by

$$f_U(u) = -\frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{\Gamma(k\theta + 1)}{k!} (-u^{-\theta})^k sin(\theta k \pi) \tag{4.16}$$

with $0 < \theta < 1$. This density function has infinite mean and the variance is therefore also undetermined. Although an infinite mean is more difficult to work with, it is actually one of the main reasons why this density function was pro-

posed (Duchateau and Janssen, 2008). Only density functions with infinite mean have the property that the heterogeneity parameter is independent from the covariate information (Hougaard ,1986b). Furthermore, it is often stated as an attractive property of the positive stable frailty distribution that the proportionality property for the conditional hazard is inherited by the population hazard (Duchateau and Janssen, 2008).

The Laplace transform has the simple form

$$L(s) = exp(-s^{\theta}) \tag{4.17}$$

Since $L(s)$ does not exist in the neighbourhood of zero, the mean does not exist. This is shown by taking the right limit of $L(s)$ for $s \to 0$ we get that the mean is infinite:

$$\lim_{s\to 0} L^{(1)}(s) = -\theta \lim_{s\to 0} \frac{exp(-s^{\theta})}{s^{1-\theta}} = -\infty \tag{4.18}$$

4.5 The power variance function distribution

The family of the power variance function distributions was introduced as an extension of the positive stable distribution by Hougaard (1986b). It contains the gamma, inverse Gaussian and positive stable distributions; they are obtained for choices of the parameters at the boundary of the parameter space (Duchateau and Janssen, 2008). The density function is given by

$$f_U(u) = exp(-\frac{\nu}{\theta}(\frac{u}{\mu} + \frac{1}{\nu - 1}))$$
$$\times \frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{(\nu/\theta)^{k\nu}(u/\nu)^{k(\nu-1)}\Gamma(1 - k(\nu - 1))sin(\pi k(\nu - 1))}{k!(\nu - 1)^k} \tag{4.19}$$

with $\mu > 0$, $\theta > 0$ and $0 < \nu \leq 1$. The corresponding Laplace transform given by Aalen (1992) is

$$L(s) = exp[\frac{\nu}{\theta(1-\nu)}(1 - (1 + \frac{\theta\mu s}{\nu})^{1-\nu})] \tag{4.20}$$

with the first and second derivatives given by

$$L^{(1)}(s) = -L(s)\mu(1 + \frac{\theta\mu s}{\nu})^{-\nu}$$

$$L^{(2)}(s) = L(s)\mu^2((1 + \frac{\theta\mu s}{\nu})^{-2\nu} + \theta(1 + \frac{\theta\mu s}{\nu})^{-\nu-1})$$

Assuming that the derivatives above exist in the neighbourhood of zero (not the case for the positive stable distribution), we can evaluate them at $s = 0$ to find

$$E(U) = (-1)L^{(1)}(0) = \mu \tag{4.21}$$

$$Var(U) = L^{(2)}(0) - (-L^{(1)}(0))^2 = \theta\mu^2 \tag{4.22}$$

4.6 The compound Poisson distribution

In some application, a proportion of the subjects is not susceptible for the event under consideration and to model this, a frailty term U for which $P(U = 0)$ is positive is used. We therefore consider a distribution of the frailty term that has two parts: the positive probability at zero and a continuous subdensity on the positive real line. With $\mu > 0$, $\theta > 0$, and $\nu > 1$ we have

$$P(U = 0) = exp(\frac{-\nu}{\theta(\nu - 1)}) \tag{4.23}$$

and

$$f_U(u) = exp(-\frac{\nu}{\theta}(\frac{u}{\mu} + \frac{1}{\nu - 1}))$$

$$\times \frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{(\nu/\theta)^{k\nu}(u/\nu)^{k(\nu-1)}\Gamma(1 - k(\nu - 1))sin(\pi k(\nu - 1))}{k!(\nu - 1)^k} \tag{4.24}$$

The corresponding Laplace transform is

$$L(s) = exp[\frac{\nu}{\theta(1 - \nu)}(1 - (1 + \frac{\theta\mu s}{\nu})^{1-\nu})] \tag{4.25}$$

4.7 Lognormal distribution

The use of lognormal distribution in frailty models originates from the link with generalized mixed models with a standard assumption that the random effects w_i follow a zero-mean normal distribution with variance σ^2 (Duchateau and Janssen, 2008). The corresponding lognormal frailty distribution is given by

$$f_U(u) = \frac{1}{u\sqrt{2\pi\sigma^2}} exp(-\frac{logu^2}{2\sigma^2}) \tag{4.26}$$

The mean and variance are expressed as

$$E(U) = exp(\sigma^2/2) \tag{4.27}$$

$$Var(U) = exp(\sigma^2)(exp(\sigma) - 1) \tag{4.28}$$

It is noted that the mean and variance of the lognormal frailty density are both functions of the parameter σ^2. Although lognormal frailty distribution has no explicit evaluation of the Laplace transform, it allows a relatively simple extension to the multivariate case with general variance-covariance matrix which is far

more complex to pursue with other distributions (Nguti, 2003).

In this project, application focused on frailty models with gamma and lognormal distributions and θ was used to denote the heterogeneity parameter for both distrutions.

4.8 Baseline hazard distributions for parametric frailty models

The risk of an event occurring can be constant over time or with more complicated hazard rates that increase and decrease over time or that increase or decrease at faster or slower rates. Exactly how the hazard rate varies with time is generally referred to as time dependency. The logic of parametric duration models is that they assume a particular shape for the hazard rate. Below are some of the commonly used baseline hazards distributions:

4.8.1 Exponential

For the exponential model, the hazard rate is characterized by:

$$h(t) = \lambda \tag{4.29}$$

This implies that the conditional probability of an event is constant over time (and that events occur according to a Poisson process). That is, the risk of an event occurring is flat with respect to time (Jenkins, 2008). The corresponding

cumulative hazard is given by;

$$H(t) = \lambda t \tag{4.30}$$

Recall from the section (3.2) that $H(t) = -ln[S(t)]$. As a result, we have

$$S(t) = e^{-H(t)} = e^{-\lambda t} \tag{4.31}$$

This means that the density is

$$f(t) = S(t)h(t) = \lambda e^{-\lambda t} \tag{4.32}$$

Having defined $h(t)$, $f(t)$, and $S(t)$ and their relationships in section (3.2), it is easy to construct the sample likelihood for the exponential model as shown below:

$$L = \prod_{i=1}^{N} \{f(t)\}^{d_i} \{S(t)\}^{1-d_i}$$

$$L = \prod_{i=1}^{N} \{\lambda e^{-\lambda t}\}^{d_i} \{e^{-\lambda t}\}^{1-d_i} \tag{4.33}$$

4.8.2 Weibull

The baseline hazard, $h(t)$ can be chosen to follow a Weibull(λ, ρ) distribution which is more general and flexible than the exponential distribution. The Weibull baseline hazard allows for hazard rates that are non-constant but monotonic (Jenkins, 2008). The probability density function is given by

$$f(t) = \lambda \rho t^{\rho-1} exp(-\lambda t^{\rho}) \tag{4.34}$$

Where $\lambda > 0$ and $\rho > 0$ are shape and scale parameters respectively. The corresponding survival function is given by;

$$S(t) = Pr(T > t) = \int_t^{\infty} \lambda\rho x^{\rho-1} exp(-\lambda x^{\rho}) dx \tag{4.35}$$

Using integration by substitution, we let $u = \lambda x^{\rho}$ thus, $\lambda\rho x^{\rho-1}dx = du$ and substituting in equation (4.35) above

$$\begin{aligned} S(t) &= \int_{\lambda t^{\rho}}^{\infty} exp(-u) du \\ &= -\exp(-u)|_{\lambda t^{\rho}}^{\infty} \\ &= \exp(-\lambda t^{\rho}) \end{aligned} \tag{4.36}$$

The corresponding hazard is given by;

$$\begin{aligned} h(t) = \frac{f(t)}{S(t)} &= \frac{\lambda\rho t^{\rho-1} exp(-\lambda t^{\rho})}{\exp(-\lambda t^{\rho})} \\ &= \lambda\rho t^{\rho-1} \end{aligned} \tag{4.37}$$

and the cumulative hazard is

$$\begin{aligned} H(t) &= \int_0^t h(x) dx \\ &= \int_0^x \lambda\rho x^{\rho-1} dx \\ &= \lambda t^{\rho} \end{aligned} \tag{4.38}$$

The hazard rises if $\rho > 1$, constant if $\rho = 1$ and decreases if $\rho < 1$. Exponential distribution is a special case of Weibull distribution when the shape parameter

λ is 1. The likelihood function for the Weibull model is constructed as follows:

$$L = \prod_{i=1}^{N} \{\lambda\rho t^{\rho-1} exp(-\lambda t^{\rho})\}^{d_i} \{\exp(-\lambda t^{\rho})\}^{1-d_i} \tag{4.39}$$

4.8.3 Gompertz

Gompertz (1825) idea of exponential aging, postulated that $h(t)$ satisfies the simple differential equation

$$\frac{dh(t)}{dt} = \rho h(t) \tag{4.40}$$

Solving this

$$\frac{dh(t)}{h(t)} = \rho dt$$

$$\int \frac{dh(t)}{h(t)} = \int \rho dt$$

The Gompertz distribution is characterized by the fact that the log of the hazard is linear in t, so

$$lnh(t) = \rho t + c$$

$$h(t) = \lambda e^{\rho t} \tag{4.41}$$

where $\lambda = e^{X\beta}$ and ρ is the shape parameter. The corresponding survival function is

$$S(t) = e^{-\lambda\rho^{-1}(e^{\rho t}-1)} \tag{4.42}$$

The Gompertz model is useful for monotone hazard rates that either increase or decrease exponentially with time. The shape parameter satisfies the following conditions:

•$If\ \rho < 1$, then the hazard is monotonically decreasing with time.

•$If \rho > 1$, then the hazard is monotonically increasing with time.

•$If \rho = 1$, then the hazard is flat and we have the exponential model.

In other words, this implies that a person's probability of dying increases at a constant exponential rate as age increases. This distribution provides a remarkably close fit to adult mortality in contemporary developed countries (Rodriguez, 2010). The corresponding cumulative hazard is given by;

$$\begin{aligned} H(t) &= \int_0^t h(x)dx \\ &= \int_0^t \lambda e^{\rho x} dx \\ &= \frac{\lambda}{\rho}(e^{\rho} t - 1) \end{aligned} \tag{4.43}$$

4.8.4 Log-logistic

In the log-logistic model, the hazard rate is characterized by:

$$h(t) = \frac{\lambda^{\frac{1}{\rho}} t[(\frac{1}{\rho}) - 1]}{\rho[1 + (\lambda t)^{\frac{1}{\rho}}]} \tag{4.44}$$

Similar to the Weibull model, the log-logistic model has two parameters, λ, the location parameter and ρ, the shape parameter. The log-logistic allows for non-monotonic unimodal hazards - in this case inverted U-shapes (Jenkins, 2008). The shape parameter satisfies the following conditions:

•$If \rho < 1$, then the conditional hazard first rises, then falls.

•$If \rho \geq 1$, then the hazard is declining.

For the log-logistic model, the hazard can never be monotonically rising and the

corresponding survival function is:

$$S(t) = \frac{1}{1 + (\lambda t)^{\frac{1}{\rho}}} \tag{4.45}$$

with a density function :

$$f(t) = h(t) * S(t) = \frac{\lambda^{\frac{1}{\rho}} t[(\frac{1}{\rho}) - 1]}{\{\rho[1 + (\lambda t)^{\frac{1}{\rho}}]^2\}} \tag{4.46}$$

the corresponding cummulative hazard function is given by:

$$H(t) = 1 + (\lambda t)^{\frac{1}{\rho}} \tag{4.47}$$

Having defined $h(t)$, $f(t)$ and $S(t)$, it is easy to construct the likelihood for the log-logistic model as shown below

$$L = \prod_{i=1}^{N} \{\frac{\lambda^{\frac{1}{\rho}} t[(\frac{1}{\rho}) - 1]}{\{\rho[1 + (\lambda t)^{\frac{1}{\rho}}]^2\}} \times \frac{1}{1 + (\lambda t)^{\frac{1}{\rho}}}\}^{d_i} \{\frac{1}{1 + (\lambda t)^{\frac{1}{\rho}}}\}^{1-d_i} \tag{4.48}$$

The expected duration for the log-logistic has a closed form solution when $\rho < 1$ (Klein and Moeschberger, 2003).

4.8.5 Lognormal

The hazard function of the log-normal distribution increases from 0 to reach a maximum and then decreases monotonically, approaching 0 as $t \to \infty$. Using the lognormal distribution with parameters μ and σ; $w = ln(x)$ $N(\mu, sigma)$

$$S(t) = 1 - \phi(\frac{lnt - \mu}{\sigma}) \tag{4.49}$$

where ϕ is the standard Normal cdf and $\mu = X\beta$. The density function is:

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} e^{\frac{(lnt-\mu)^2}{2\sigma^2}} \tag{4.50}$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{1}{\sigma t\sqrt{2\pi}} e^{\frac{(lnt-\mu)^2}{2\sigma^2}}}{1 - \phi(\frac{lnx-\mu}{\sigma})} \tag{4.51}$$

The hazard rate is similar to that for the log-logistic for the case where $\rho < 1$, i.e. it first rises and then falls.

4.8.6 Exponential power

Using the exponential power density with survival function;

$$S(t) = exp(1 - e^{\lambda t^\alpha}) \tag{4.52}$$

$$h(t) = \frac{-d}{dt} lnS(t)$$

$$h(t) = \frac{-d}{dt}(1 - e^{\lambda t^\alpha})$$

$$h(t) = \alpha(exp(1 - e^{\lambda t^\alpha})) \tag{4.53}$$

From the relationships shown in chapter 2, the corresponding cumulative hazard is given by

$$H(t) = -ln(S(t)))$$

$$H(t) = -ln(exp(1 - e^{\lambda t^\alpha}))$$

$$H(t) = e^{\lambda t^\alpha} - 1 \tag{4.54}$$

4.8.7 Generalized Gamma Model

The generalized gamma model has a quite complicated specification involving two shape parameters (Jenkins, 2008). The density of the generalized gamma distribution is:

$$f(t) = \frac{\lambda\rho(\lambda t)^{\rho\kappa-1}e^{-(\lambda t)^{\rho}}}{\Gamma\kappa} \tag{4.55}$$

where

$$\lambda_i = e^{-(x_i\beta)}$$

and ρ andκ are the two shape parameters. The two shape parameters allow for quite a flexible hazard rate including a U-shape. An attractive characteristic of the generalized gamma model is that it nests several of the other parametric models as special cases: Weibull, exponential, log-normal, and the standard gamma (Balakrishnan and Peng, 2006). Thus, this model is good for adjudicating between (some) competing parametric models. The shape parameters work in the following way:

•$If\kappa = 1$, then the Weibull distribution is implied.

•$If\kappa = \rho = 1$, the exponential is implied.

•$If\kappa = 0$, the log-normal is implied.

•$If\rho = 1$, the gamma distribution is implied.

Among these parametric distributions, only the exponential, the Weibull and the Gompertz model share the assumption of proportional hazards with the Cox regression model (Bender et al., 2005). The characteristics of these distributions are summarized in Table 4.1.

Table 4.1: *Summary of characteristics of the exponential, the Weibull and the Gompertz distribution*

Characteristic	**Exponential**	**Weibull**	**Gompertz**
Parameter	scale parameter $\lambda > 0$	scale parameter$\lambda > 0$ shape parameter $\rho > 0$	scale parameter$\lambda > 0$ shape parameter $\alpha \in (-\infty, \infty)$
Range	$[0, \infty)$	$[0, \infty)$	$[0, \infty)$
Hazard function	$h_0(t) = \lambda$	$h_0(t) = \lambda\rho x^{\rho-1}$	$h_0(t) = exp(\alpha x)$
Cummulative hazard	$H_0(t) = \lambda t$	$H_0(t) = \lambda t^{\rho}$	$H_0^{-1}(t) = \frac{1}{\alpha}(log(\frac{\alpha}{\lambda}t + 1)$
Inverse cumm hazard	$H_0^{-1}(t) = \lambda^{-1}t$	$H_0^{-1}(t) = (\lambda^{-1}t)^{1/\rho}$	$H_0(t) = \frac{1}{\alpha}(exp(\alpha t) - 1)$
Density function	$f_0(t) = \lambda exp(-\lambda t)$	$f_0(t) = \lambda\rho t^{\rho-1}exp(-\lambda t^{\rho})$	$f_0(t) = \lambda exp(\alpha t)exp(\frac{\lambda}{\alpha}(1 - exp(\alpha t)))$
Survival function	$S_0(t) = exp(-\lambda x)$	$S_0(t) = exp(-\lambda t^{\rho})$	$S_0(x) = exp(\frac{\lambda}{\alpha}(1 - exp(\alpha t)))$
Mean	$E(T) = \frac{1}{\lambda}$	$E(T) = \frac{1}{\sqrt[\rho]{\lambda^2}}\Gamma(\frac{1}{\rho} + 1)$	$E(T) = \frac{1}{\lambda}G(\frac{\lambda}{\alpha})$ where $G(x) = \int_x^{\infty}\frac{1}{y}exp(-y)dy$

According to Bender *et al.* (2005), the effect of covariates in the Cox model have to be translated from the hazards to the survival times. This is because the standard software packages for Cox models require the individual survival time data, not the hazard function. Table 4.2 presents the formulas for the survival time and the hazard function of Cox models using the exponential, the Weibull and the Gompertz distribution.

Table 4.2: *Formulae for the survival time and the hazard function of Cox models using the exponential, the Weibull and the Gompertz distribution.*

Characteristic	**Cox-Exponential**	**Cox-Weibull**	**Cox-Gompertz**
Survival time	$T = -\frac{log(U)}{\lambda exp(\beta' x)}$	$T = (-\frac{log(U)}{\lambda exp(\beta' x)})^{1/\rho}$	$T = \frac{1}{\alpha} log(1 - \alpha \frac{log(U)}{\lambda exp(\beta' x)})$
Hazard function	$h(t\|x) = \lambda exp(\beta' x)$	$h(t\|x) = \lambda \rho x^{\rho-1} exp(\beta' x)$	$h(t\|x) = \lambda exp(\alpha t) exp(\beta' x)$

Chapter 5

Estimation Methods

5.1 Introduction

In survival analysis, estimation methods vary depending on the model of interest and the amount of information available i.e. parametric or semi-parametric model, conditional or marginal Cox proportional hazard model. This chapter presents the various estimation methods for survival models.

5.2 Estimation in semi-parametric Cox PH model

By fitting the Cox proportional hazards model, we wish to estimate the vector of regression coefficients, β. A popular estimation approach was proposed by Cox (1972) in which a partial likelihood function that does not depend on $h_0(t)$ is obtained for β. Partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameters ($h_0(t)$) in

the Cox PH model. In this section, we will construct the partial likelihood function based on the proportional hazards model. Let $t_1, t_2, ..., t_n$ be the observed survival time for n individuals. Let the ordered death time of r individuals be $t_{(1)} < t_{(2)} < ... < t_{(r)}$ and let $R(t_{(j)})$ be the risk set just before $t_{(j)}$ i.e. the group of individuals who are alive and uncensored at a time just prior to $t_{(j)}$. The conditional probability that the i^{th} individual dies at $t_{(j)}$ given that one individual from the risk set on $R(t_{(j)})$ dies at $t_{(j)}$ is;

P(individual i dies at $t_{(j)}$|one death from the risk set $R(t_{(j)})$ at $t_{(j)}$)

$$\text{P(individual } i \text{ dies at } t_{(j)})/\text{P(one death at } t_{(j)})$$

$$\simeq \frac{\lim \Delta t \downarrow 0 P\{individualidiesatt_{(j)}, t_{(j)} + \Delta t)\}/\Delta t}{\lim \Delta t \downarrow 0 \sum_{k \in R(t_{(j)})} P\{individualkdiesat(t_{(j)}, t_{(j)+\Delta t})\}/\Delta t \Delta}$$

$$= \frac{h_i(t_j)}{\sum_{k \in R(t_{(j)})} h_k(t_{(j)})}$$

$$= \frac{h_0(t_{(j)}) exp(\beta' x_i(t_{(j)}))}{\sum_{k \in R(t_{(j)})} h_0(t_{(j)}) exp(\beta' x_k(t_{(j)}))}$$

$$= \frac{exp(\beta' x_i(t_{(j)}))}{\sum_{k \in R(t_{(j)})} exp(\beta' x_k(t_{(j)}))}$$

Then, the partial likelihood function for the Cox PH model is given by

$$L(\beta) = \prod_{j=1}^{r} \frac{exp(\beta' x_i(t_{(j)}))}{\sum_{k \in R(t_{(j)})} exp(\beta' x_k(t_{(j)}))} \tag{5.1}$$

in which $x_i(t_{(j)})$ is the vector of covariate values for individual $i$who dies at $t_{(j)}$. Note that this likelihood function is only for the uncensored individuals. Let

$t_1, t_2, ..., t_n$ be the observed survival time for n individuals and δ_i be the event indicator, which is *zero* if the i^{th} survival time is censored, and *one* otherwise. The likelihood function in equation (5.1) can be expressed by;

$$L(\beta) = \prod_{i=1}^{n} [\frac{exp(\beta' x_i(t_{(i)}))}{\sum_{k \in R(t_{(i)})} exp(\beta' x_k(t_{(i)}))}]^{\delta_i} \tag{5.2}$$

where $R(t_i)$ is the risk set at time t_i: The partial likelihood is valid when there are no ties in the dataset i.e. there are no two subjects who have the same event time.

5.3 Estimation in semi-parametric frailty models

5.3.1 The Expectation-Maximization (EM) Algorithm

In a semi parametric approach, the baseline hazard is unspecified and the frailties (u_i) are unobserved. For these reasons, it is difficult to maximize the likelihood to estimate the parameters (Nguti, 2003). One solution to this kind of problem is the Expectation-Maximization (EM) algorithm which is typically used in the presence of unobserved (latent) information. The EM algorithm iterates between the expectation and maximization step.

Expectation step

In the expectation step, the expected values of the unobserved frailties conditional on the observed information and the current parameter estimates are obtained.

Maximization step

In the maximization step, the expected values obtained in the E-step are considered to be the true information and new estimates of the parameters of interest are obtained by maximization of the likelihood, given the expected values. The applicability of the EM algorithm for a particular problem depends on two conditions. First, it should be easy to obtain expected values for the unobserved information. Second, the maximization of the likelihood, conditional on the expected values of the unobserved information, should be straightforward as the EM algorithm is based on performing these two steps iteratively. The execution of the EM algorithm is computer intensive and slow.

5.3.2 The Penalized Partial Likelihood (PPL)

An alternative estimation method is the Penalized Partial Likelihood (PPL) presented by Therneau and Grambsch (2000) where the random effect is treated as a penalty term. The PPL approach is preferred over EM algorithm since it is faster and is implemented in most standard software.

The PPL for normal random effects

The use of PPL method for the lognormal frailty is motivated by the Laplace approximation to the full likelihood similar to the arguments used in the context of generalized linear mixed models (McGilchrist, 1993). The full likelihood is presented as follows;

$$l_{full}(h_0(\cdot), \theta, \beta) = logf(z, u|h_0(\cdot), \theta, \beta)$$

$$= logf(z, |h_0(\cdot), \beta, u) + logf(u|\theta)$$

$$= l_{full,1}(h_0(\cdot), \beta) + l_{full,2}(h_0(\theta) \tag{5.3}$$

In PPL approach, $logf(u|\theta)$ part of the likelihood is considered to be a penalty term such that if the actual value of the random effect is far away from its 0 (zero) mean, the absolute value of the logarithm of the density function evaluated at this value will be large and the penalty term has a large negative contribution to the full data loglikelihood.

Taking the random effects $(w_i's)$ as another set of parameters in the first part of the likelihood, this likelihood part can be transformed into a partial likelihood expression as follows;

$$l_{PPL}(\theta, \beta, w) = l_{part}(\beta, w) - l_{pen}(\theta, w) \tag{5.4}$$

The first part $l_{part}(\beta, w)$ represents the conditional likelihood of the data given the frailties, the second part $l_{pen}(\theta, w)$ stands for the distribution of the frailties. The frailties are thus in both parts of the penalized partial likelihood. The second term penalizes random effects that are far away from the mean value zero by reducing the penalized partial likelihood. This corresponds to shrinking the random effects towards the zero-mean.

if $\eta_{ij} = x_{ij}^t\beta + w_i$ and $\eta = (\eta_{11}, ..., \eta_{cn_s})$;

$$l_{part}(\beta, w) = \sum_{i=1}^{G}\sum_{j=1}^{n_i} \delta_{ij}[\eta_{ij} - log(\sum_{qw \in R(y_{ij})} exp(\eta_{qw}))] \tag{5.5}$$

$$l_{pen}(\theta, w) = -\sum_{i=1}^{G} logf_W(w_i) \tag{5.6}$$

so for random effects $w_i, i = 1, ..., G$ with mean 0 normal density and variance θ

we get

$$l_{pen}(\theta, w) = \frac{1}{2}\sum_{i=1}^{G}(\frac{w_i^2}{\theta} + log(2\pi\theta))$$

Maximization in PPL approach is a double iterative process that alternates between an inner and an outer loop until convergence. In the inner loop, the Newton-Raphson procedure is used to maximize, for a provisional value of θ, β and w, (best linear unbiased predictors, BLUPs) (Duchateau *et al.*, 2002). For both gamma and lognormal frailty distributions, this step is identical.

In the outer loop of a lognormal distribution, the restricted maximum likelihood estimator (REML) for θ is obtained using the best linear unbiased predictors, BLUPs. Details are as follows. Let l denote the outer loop index and k the inner loop index. Let $\theta^{(l)}$ be the estimate for θ at the l^{th} iteration in the outer loop. Given $\theta^{(l)}$, $\beta^{(l,k)}$and $w^{(l,k)}$ are the estimates and predictions for β and w at the k^{th} iterative step in the inner loop. Starting from initial values $\beta^{(1,0)}$ and $w^{(1,0)}$,$\theta^{(0)}$ and $\theta^{(1)}$, the k^{th} iterative step for Newton-Raphson, given $\theta^{(l)}$, is given by

$$\begin{bmatrix} \beta^{(l,k)} \\ w^{(l,k)} \end{bmatrix} = \begin{bmatrix} \beta^{(l,k-1)} \\ w^{(l,k-1)} \end{bmatrix} - V \begin{bmatrix} 0 \\ (\theta^{(l)}) - w^{(l,k-1)} \end{bmatrix} + V \begin{bmatrix} X & Z \end{bmatrix} \frac{dl_{part(\beta,w)}}{d\eta}$$

where

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

is the inverse of the square $(p+s)$ -dimensional matrix A with A given by

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} X^t \\ Z^t \end{bmatrix} \left(\frac{-d^2 l_{part}(\beta, w)}{d\eta d\eta^t}\right) \begin{bmatrix} X & Z \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & (\theta(l))^{-1} I_G \end{bmatrix}$$

Once the Newton-Raphson procedure has converged for the current value of $\theta^{(l)}$, a REML estimate for θ is given by

$$\theta^{(l+1)} = \frac{\sum_{i=1}^{G} (w_i^{(l,k)})^2}{G - r}$$

where $r = trace(V_{22})/\theta^{(l)}$. This outer loop is iterated until the absolute difference between two sequential values for θ, $|\theta^{(l)} - \theta^{(l-1)}|$ is sufficiently small.

The penalized partial likelihood for the gamma frailty

The penalised partial likelihood can be written in the same way as for the normal random effects equation (5.25) but with penalty function given by

$$l_{pen}(\theta, w) = \frac{1}{\theta} \sum_{i=1}^{G} (w_i - exp(w_i)) \tag{5.7}$$

Since a REML estimate is not available, the outer loop of a gamma frailty distribution is based on the maximization of a profiled version of marginal likelihood (Duchateau *et al.*, 2002; Duchateau and Janssen, 2008). For gamma frailty model, PPL and EM algorithm lead to the same estimates.

5.4 Estimation in parametric frailty model

When a parametric baseline hazard is assumed, maximum likelihood estimates can be obtained by maximizing the likelihood function. This not only makes

estimation easier, but also describe explicitly the effect of the frailty on hazard ratios over time. Survival data consists of event times and censored observations and the likelihood function under random right censoring is given by;

$$L = \prod_{j=1}^{n} [(1 - H_j(t)) f_j(t)]^{\delta_j} [(1 - F_j(t)) h_j(t)]^{1-\delta_j} \tag{5.8}$$

where δ_j is the censoring indicator, h and H are the density function and the cumulative distribution function of the censoring time respectively. f and F are the density function and the cumulative distribution function of the event time respectively. The distribution of censoring times in the likelihood function can be ignored because it does not depend on the parameters of interest related to the survival function (Ulviya, 2013). Therefore, the likelihood function for the j^{th} subject assuming right censoring is of the form;

$$L = \prod_{j=1}^{n} (f_j(t)^{\delta_j} (S_j(t))^{1-\delta_j}$$

Following the idea above, the likelihood function for the j^{th} subject in the i^{th} cluster is given by;

$$L_i = \prod_{j=1}^{n} (f_{ij}(t)^{\delta_{ij}} (S_{ij}(t))^{1-\delta_{ij}} \tag{5.9}$$

From the relationships given in section (3.2), we can rewrite the conditional likelihood function in equation(5.4) as;

$$L_i = \prod_{j=1}^{n} (h_{ij}(t)^{\delta_{ij}} (S_{ij}(t)) \tag{5.10}$$

From these relationships, we can derive the forms of the conditional and marginal likelihood functions of the frailty models. From section (3.4.4), the Cox PH model

with frailties is given by;

$$h_{ij}(t) = h_0(t)u_i exp(x_{ij}^t \beta) \tag{5.11}$$

Equation (5.6) can be rewritten as

$$\frac{f_{ij}(t)}{S_{ij}(t)} = h_0(t)u_i exp(x_{ij}^t \beta) \tag{5.12}$$

Integrating both sides of the equation (5.7), we get an expression for the survival function.

$$\int_0^\infty \frac{f_{ij}(t)}{S_{ij}(t)} dt = \int_0^\infty h_0(t)u_i exp(x_{ij}^t \beta) dt$$

$$-ln(S_{ij}(t)) = H_0(t)u_i exp(x_{ij}^t \beta)$$

Therefore,

$$S_{ij}(t) = exp(-H_0(t)u_i exp(x_{ij}^t \beta)) \tag{5.13}$$

The conditional likelihood function for the i^{th} subgroup is then given by

$$L_i(\psi, \beta | u_i) = \prod_{j=1}^{n_i} (h_0(t_{ij})u_i exp(x_{ij}^t \beta))^{\delta_{ij}} exp(-H_0(t_{ij})u_i exp(x_{ij}^t \beta)) \tag{5.14}$$

where ψ is a vector of parameters of the baseline hazard. It follows that the marginal likelihood function for the i^{th} cluster is

$$L_i(\psi, \theta, \beta) = \prod_{j=1}^{n_i} \int_0^\infty (h_0(t_{ij})uexp(x_{ij}^t \beta))^{\delta_{ij}} exp(-H_0(t_{ij})uexp(x_{ij}^t \beta)) f(u) du \tag{5.15}$$

where *f(u)* is the probability distribution function of frailties $u_1, ..., u_G$. The following section illustrates derivation of the marginal loglikelihood for the gamma frailty model.

5.4.1 Parametric Cox proportion hazard model with gamma frailty

To obtain the marginal loglikelihood for the gamma frailty model, first we integrated out the gamma frailties in the conditional survival likelihood. This leads to explicit and simple marginal likelihood function which only contains the parameters of interest. The marginal likelihood function for the i^{th} cluster is given by

$$L_i(\psi,\theta,\beta)=\prod_{j=1}^{n_i}\int_0^{\infty}(h_0(t_{ij})ue^{(x_{ij}^t\beta)})^{\delta_{ij}}e^{(-H_0(t_{ij})uexp(x_{ij}^t\beta))}\times\frac{u^{1/\theta-1}e^{u/\theta}}{\Gamma(1/\theta)\theta^{1/\theta}}du \quad (5.16)$$

where ψ contains the baseline hazard parameters. For the exponential baseline hazard ψ =(λ) ,ψ =(λ and ρ) for the Weibull baseline hazard and ψ =(λ and α) for the Gompertz baseline hazard. Rearranging the terms in equation (5.11), we obtain the following expression

$$L_i(\psi,\theta,\beta)=\prod_{j=1}^{n_i}h_0(t_{ij})^{\delta_{ij}}exp(x_{ij}^t\beta)^{\delta_{ij}}\int_0^{\infty}\frac{u^{1/\theta+d_i-1}e^{u/\theta}exp(-\sum_{j=1}^{n_i}H_0(t_{ij})uexp(x_{ij}^t\beta))}{\Gamma(1/\theta)\theta^{1/\theta}}du$$

$$=\prod_{j=1}^{n_i}h_0(t_{ij})^{\delta_{ij}}exp(x_{ij}^t\beta)^{\delta_{ij}}\frac{\Gamma(1/\theta+d_i)\theta^{(1/\theta+d_i)}}{\Gamma(1/\theta)\theta^{1/\theta}}$$

$$\times\int_0^{\infty}\frac{u^{1/\theta+d_i-1}exp-u(1/\theta+\sum_{j=1}^{n_i}H_0(t_{ij})u_iexp(x_{ij}^T\beta))}{\Gamma(1/\theta+d_i)\theta^{(1/\theta+d_i)}}du \quad (5.17)$$

where $d_i=\sum_{j=1}^{n_i}\delta_{ij}$.

We integrate out the frailty term u, so as to make the problem more tractable. The term under the integral is the moment generating function (mgf) of a gamma

distribution with a pdf $\Gamma(1/\theta + di, 1/\theta)$. Using this fact, we can derive the expression for marginal likelihood function as;

$$L_i(\psi,\theta,\beta) = \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} exp(x_{ij}^t\beta)^{\delta_{ij}} \frac{\Gamma(1/\theta + d_i)\theta^{(1/\theta+d_i)}}{\Gamma(1/\theta)\theta^{1/\theta}\theta^{(1/\theta+d_i)}(1/\theta + \sum_{j=1}^{n_i} H_0(t_{ij})exp(x_{ij}^t\beta))^{(1/\theta+d_i)}}$$

$$\times \int_0^\infty \frac{u^{1/\theta+d_i-1}e^{-u(1/\theta+\sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^t\beta})}[1/\theta + \sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^t\beta}]^{(1/\theta+d_i)}}{\Gamma(1/\theta + d_i)}du$$

$$= \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} exp(x_{ij}^t\beta)^{\delta_{ij}} \frac{\Gamma(1/\theta + d_i)}{\Gamma(1/\theta)\theta^{1/\theta}(1/\theta + \sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^t\beta})^{(1/\theta+d_i)}}$$

$$\times \int_0^\infty \frac{u^{1/\theta+d_i-1}e^{-u(1/\theta+\sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^t\beta})}[1/\theta + \sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^t\beta}]^{(1/\theta+d_i)}}{\Gamma(1/\theta + d_i)}du \tag{5.18}$$

It is observed that the term under the integral is the pdf of $\Gamma(1/\theta + d_i, 1/\theta + \sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^t\beta})$, which integrates to 1. Thus the obtained marginal likelihood function is

$$L_i(\psi,\theta,\beta) = \frac{\Gamma(1/\theta + d_i)\prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} e^{x_{ij}{}^T\beta\delta_{ij}}}{(1/\theta + \sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^t\beta})^{(1/\theta+d_i)}\Gamma(1/\theta)\theta^{1/\theta}} \tag{5.19}$$

Taking the logarithm of this expression and summing over the G clusters. We obtain the marginal loglikelihood function, $l(\psi,\theta,\beta)$.

$$l(\psi,\theta,\beta) = \sum_{i=1}^{G}[d_i log(\theta) - log(\Gamma(1/\theta)) + log(\Gamma(1/\theta) + d_i) - (1/\theta + d_i)log(1+$$

$$\theta\sum_{j=1}^{n_i} H_0(t_{ij}))exp(x_{ij}^t\beta)) + \sum_{j=1}^{ni} \delta_{ij}(x_{ij}^t\beta + log(h_0(t_{ij})))]$$

By maximizing this loglikelihood function, we can obtain maximum likelihood estimatesfor ψ, θ, β. We consider parametric forms of baseline hazards so that the marginal likelihood is also parametric and we can use classical maximum likelihood techniques to estimate the parameters of interest. As an illustration, we work out the Hessian matrix for the frailty model with Weibull baseline hazard with respect to the parameters $\zeta = (\lambda, \rho, \theta, \beta)$ and one covariate. First we show the contribution to the first derivative of the marginal loglikelihood of a particular cluster i which can then be summed over all clustered to obtain the first derivative of the marginal loglikelihood. As earlier shown, the hazard and cumulative hazard functions for the Weibull distribution are given by

$$h_0(t) = \lambda \rho t^{\rho - 1}$$

and

$$H_0(t) = \lambda t^{\rho}$$

respectively. The corresponding marginal loglikelihood function for gamma frailty with Weibull baseline hazard rate is

$$l(\lambda, \rho, \theta, \beta) = \sum_{i=1}^{G} [d_i log(\theta) - log(\Gamma(1/\theta)) + log(\Gamma(1/\theta) + d_i) - (1/\theta + d_i) log(1+$$

$$\theta \sum_{j=1}^{n_i} \lambda t^{\rho} exp(x_{ij}^t \beta)) + \sum_{j=1}^{ni} \delta_{ij} (x_{ij}^t \beta + log(\lambda \rho t^{\rho - 1}))] \tag{5.20}$$

The maximum likelihood estimates can be obtained by setting each of the first-order derivatives to 0 and solving for the parameter of interest.

The first partial derivative of the scale parameter λ is given by;

$$\frac{\partial l_{marg,i} l(\zeta)}{\partial \lambda} = \frac{-(d_i + 1/\theta)\theta \sum_{j=1}^{n_i} t_{ij}^{\rho} exp(x_{ij}\beta)}{1 + \theta \sum_{j=1}^{n_i} \lambda t^{\rho} exp(x_{ij}\beta)} + d_i \lambda^{-1}$$

$$= \frac{-(d_i\theta + 1) \sum_{j=1}^{n_i} t_{ij}^{\rho} exp(x_{ij}^t \beta)}{1 + \theta \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta)} + d_i \lambda^{-1} \tag{5.21}$$

The first partial derivative of the shape parameter ρ is given by;

$$\frac{\partial l_{marg,i} l(\zeta)}{\partial \rho} = \frac{-(d_i\theta + 1) \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta) logt_{ij}}{1 + \theta \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta)} + \sum_{j=1}^{n_i} \delta_{ij} (log\lambda + log\rho + (\rho - 1) logt_{ij})$$

$$= \frac{-(d_i\theta + 1) \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta) logt_{ij}}{1 + \theta \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta)} + \sum_{j=1}^{n_i} \delta_{ij} (1/\rho + logt_{ij}) \tag{5.22}$$

The first partial derivative of the variance parameter estimate θ is given by;

$$\frac{\partial l_{marg,i} l(\zeta)}{\partial \theta} = \frac{-(d_i + \theta^{-1}) \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta)}{1 + \theta \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta)} + \theta^{-2} log(1 + \theta \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta))$$

$$-I(d_i > 0) \sum_{l=0}^{d_i - 1} (\theta + l\theta^2)^{-1} + d_i \theta^{-1} \tag{5.23}$$

The partial first derivative of the regression parameter estimate β is given by;

$$\frac{\partial l_{marg,i} l(\zeta)}{\partial \beta} = \frac{-(d_i\theta + 1) \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta) x_{ij}}{1 + \theta \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta)} + \sum_{j=1}^{n_i} \delta_{ij} x_{ij} \tag{5.24}$$

The corresponding second derivatives for for the parameters of interest are given by;

The second partial derivative of the scale parameter λ is given by;

$$\frac{\partial^2 l_{marg,i} l(\zeta)}{\partial \lambda^2} = \frac{(d_i + 1/\theta)(\sum_{j=1}^{n_i} t_{ij}^{\rho} exp(x_{ij}\beta))^2}{(1 + \theta \sum_{j=1}^{n_i} \lambda t^{\rho} exp(x_{ij}\beta))^2} + d_i \lambda^{-2}$$

$$= \frac{(d_i + 1/\theta)}{(1/\theta + \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta))^2} (\sum_{j=1}^{n_i} t_{ij}^{\rho} exp(x_{ij}^t \beta))^2 - d_i \lambda^{-2} \tag{5.25}$$

The second partial derivative of the shape parameter ρ is given by

$$\frac{\partial^2 l_{marg,i} l(\zeta)}{\partial \rho^2} = \frac{(d_i + 1/\theta)}{(1/\theta + \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta))^2} + [(\sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta) log t_{ij})^2$$

$$-(1/\theta + \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta)) \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta)(log t_{ij})^2] - d_i \rho^{-2} \tag{5.26}$$

The second partial derivative of the regression parameter β is given by

$$\frac{\partial^2 l_{marg,i} l(\zeta)}{\partial \beta^2} = \frac{(d_i + \theta^{-1})}{(\theta^{-1} + \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta))^2} [(\sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta) x_{ij})^2$$

$$- \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta) x_{ij}^2 (\theta^{-1} + \sum_{j=1}^{n_i} \lambda t_{ij}^{\rho} exp(x_{ij}^t \beta))] \tag{5.27}$$

The marginal loglikelihood function for gamma frailty with exponential baseline hazard rate is;

$$l(\psi, \theta, \beta) = \sum_{i=1}^{G} [d_i log(\theta) - log(\Gamma(1/\theta)) + log(\Gamma(1/\theta) + d_i) - (1/\theta + d_i) log(1+$$

$$\theta \sum_{j=1}^{n_i} \lambda t exp(x_{ij}^t \beta)) + \sum_{j=1}^{ni} \delta_{ij} (x_{ij}^t \beta + log(\lambda))]$$

The parameters of interest can be obtained by solving the first partial derivatives in a similar manner as above.

Chapter 6

Soft-Tissue Sarcoma Clinical-Trial Data Analysis

6.1 Introduction

Soft-Tissue Sarcomas (STS) are a rare and heterogeneous group of tumors of mesenchymal origin. STS occur mainly in support and connective tissues of the body such as fat cells, muscle, tendons, nerves, blood vessels or lymph vessels (Cancer.Net, 2013). STS can start in any part of the body with about 60% beginning in arms or legs, 30% start in the torso or abdomen while 10% occur in the head or neck. STS accounts for about 1% of all adult cancers and about 15% of all cancers in children (Cancer.Net, 2013). There are over 50 different subtypes of STS which exhibit great differences in terms of genetic alterations pathogenesis, histopathological features and clinical behaviours. Unlike most other types

of cancer which are usually named for the part of the body where the cancer began, the specific types of sarcoma are named according to the normal tissue cells they most closely resemble (Garcia et al., 2004). However, for the purpose of treatment, all the subtypes are grouped under the heading STS.

In this chapter, after presenting the the study design of a randomized multicenter Soft-Tissue Sarcoma clinical trial we will apply exploratory data analysis techniques (non-parametric methods) and statistical analysis methods(semi-parametric survival models), present the corresponding results and compare the main methods: semi-parametric marginal Cox model and frailty models(gamma and lognormal frailty models).

6.2 Study design

The data analyzed in this chapter came from a randomized phase III, open label, multicenter study conducted by the National Cancer Institute of Canada-Clinical Trial Group (NCIC-CTG) between April, 2003 and July, 2012. The study enrolled 450 patients from 36 centers. Patients enrolled were between 18 and 63 years of age and had histological evidence of high grade Soft-Tissue Sarcoma with advanced unresectable or metastatic disease (2=Intermediate and 3=High). Histological types considered were 1=Leiomyosarcoma, 2=Synovial sarcoma, 3=Liposarcoma, 4=Others. Eligible patients were randomized to re-

ceive either a single agent treatment or a combination of two treatment agents using minimization technique. Treatment was administered until progression of the disease, unacceptable levels of toxicity or patient's refusal, up to a maximum of 6 cycles of chemotherapy. Overall Survival (OS), the primary endpoint of interest was computed from the date of randomization to the date of death, whatever the cause. The secondary endpoint was Progression-Free Survival (PFS) computed from the date of randomization to the first documented date of progression or death. Patients that were alive and progression-free at the time of the analysis were censored at the date of last follow-up. Randomization was stratified by center, performance status, age group and presence of liver metastases.

6.3 Data description

The variables in the data and their coding are presented in Table 6.1. The data are right censored and all the variables considered were categorical except age which was continuous.

Table 6.1: *Summary of variables in the data set*

Variables	**Description**	**Codes/values**
HOSPNO	Hospital identifier	
PATID	Patient identification number	
AGE	Age of patients	Years
CenPFS	Progression status	1=Censored, 2= Event
Censur	Survival status	1=Alive, 2=Dead
Timepro	Progression free survival	Days
Timesur	Overall survival	Days
Trt1	Treatment arms	1= A, 2= B
Grad-rand (Tumor grade)	Tumor grade	2=Intermediate, 3=High
Qval114 (Perform status)	Performance status	0=Able to carry out normal activities 1=Restricted in some or all activities
Qval132 (liver meta)	Presence of liver metastases at baseline	0=No , 1= Yes
Hisloc	Histological type	1=Leiomyosarcoma,2=Synovial sarcoma, 3=Liposarcoma,4=Others

6.4 Exploratory data analysis (EDA)

A total of 450 patients were enrolled across 36 centers. However, 12 centers with 2 or less patients were dropped from the analysis to avoid estimation-related problems in subsequent statistical analysis. For this reason, the total number of patient reduced to 427 with 49.6% of the patients randomized to treatment A and 50.4% randomized to treatment B. As observed in Table A1 in the Appendix, the remaining 24 centers accrued between 5 and 38 patients with mean and median of 15 and 18 patients respectively. In this study, the mean age was 45.2 years with a standard deviation of 10.6 years. In the analysis age was categorized into two groups whereby 57% of the patients were younger than 50 years and 43% were 50 years old or more. Four hundred (94%) patients had an event in PFS while 350 (82%) patients had an event in both PFS and OS. There was no missing data for either of the endpoints or covariates of interest.

6.4.1 Kaplan-Meier survival curves

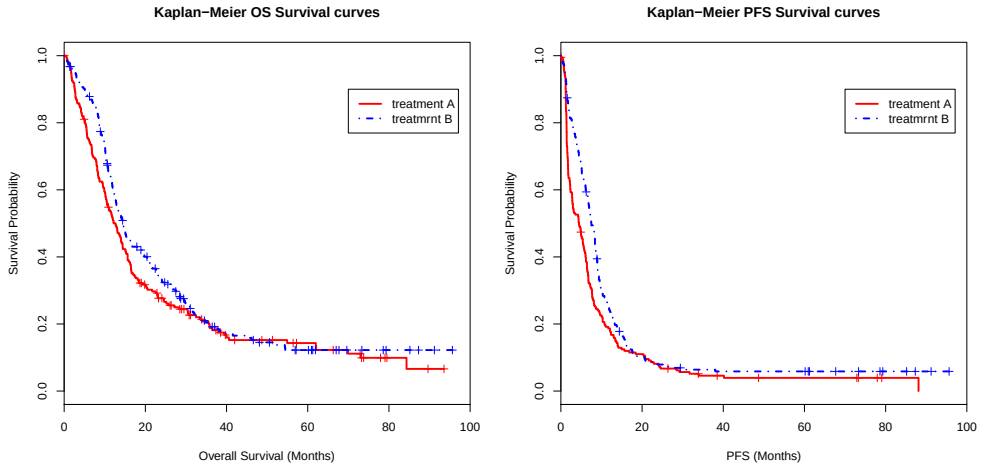

Figure 6.1: *Kaplan Meier OS curves and PFS curves by treatment*

Figure 6.1 shows Kaplan-Meier OS and PFS survival curves by treatment. The estimated median OS time was 12.7 months with 95% confidence interval (CI) [10.4, 14.4] in treatment A and 14.3 months in treatment B with a 95% CI [12.7, 16.8]. Similarly, the median PFS was 4.5 months with 95% CI [2.8, 5.6] and 7.5 months with 95% CI [6.8, 8.4] for arms A and B respectively. The survival curves were crossing suggesting violation of proportional hazard assumptions.

Figure 6.2 presents Kaplan-Meier OS and PFS curves stratified by centers. From the plot, there seems to be variability in the outcome between centers for the OS. Similar observations were made from PFS curves stratified by centers. Based on a classical log rank test, the P-values were 0.711 and 0.344 for OS and PFS endpoints respectively.

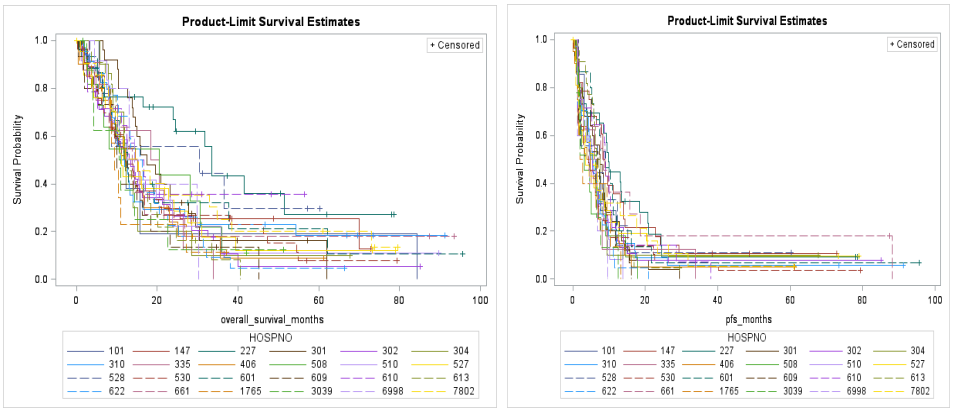

Figure 6.2: *Kaplan Meier OS curves and PFS curves stratified by centers*

6.5 Semi-parametric marginal and frailty models

6.5.1 Comparison of parameter estimates for OS

Table 6.2 presents the marginal (population averaged) and center-specific model (gamma and lognormal frailty models) results for OS endpoint. It is observed that for all the covariates, the Hazard Ratio (HR) with the corresponding 95% confidence interval (CI) were close for the three models but slightly higher for the marginal model. However, it is important to bear in mind that parameter interpretation for marginal and frailty models differs and examining their magnitude alone is of no relevant consequence. For example, in the case of marginal model, on average, the risk of an individual in arm B dying was 0.786 times lower compared to an individual in arm A. On the other hand, for either of the frailty models, for a given center, the risk of an individual in arm B dying was 0.785 times lower compared to an individual in arm A (evaluated at reference levels of other covariates). The corresponding 95% CIs did not contain the value 1;

Table 6.2: *Overall survival Hazard Ratio (95% CI) from frailty and marginal Cox models*

	Gamma frailty		**Lognormal frailty**		**Marginal Cox model**	
parameter	**HR**	**(95% CI)**	**HR**	**(95% CI)**	**HR**	**(95% CI)**
Treatment: B	0.785	(0.634, 0.973)	0.785	(0.633, 0.972)	0.786	(0.635, 0.973)
Hisloc: 1	0.841	(0.642, 1.103)	0.840	(0.640, 1.103)	0.842	(0.644, 1.101)
Hisloc: 2	0.919	(0.667, 1.266)	0.916	(0.664, 1.263)	0.923	(0.671, 1.270)
Hisloc: 3	0.579	(0.401, 0.836)	0.577	(0.400, 0.834)	0.583	(0.404, 0.840)
Tumor grade: 2	0.764	(0.616, 0.949)	0.764	(0.615, 0.949)	0.765	(0.617, 0.949)
Liver meta: 0	0.716	(0.535, 0.960)	0.715	(0.534, 0.959)	0.717	(0.536, 0.960)
Perform status: 0	0.565	(0.456, 0.699)	0.563	(0.455, 0.698)	0.566	(0.457, 0.701)
Age $\geq$ 50	1.157	(0.925, 1.447)	1.156	(0.924, 1.448)	1.158	(0.927, 1.447)

therefore, there is a significant difference between the treatment arms. All other parameter estimates can be interpreted in a similar manner.

6.5.2 Assessing the heterogeneity parameter for OS

The similarity between the marginal and shared frailty models could be further attributed to the fact that for the frailty models, the heterogeneity parameters were very small i.e. 0.005 and 0.008 for gamma and lognormal frailty model respectively. Furthermore, the random effects estimates for all the centers were not significantly different from 0. A formal test for the need of center effect was conducted by comparing the partial log-likelihood for the models with and without the frailty term. For the lognormal frailty, the change in the partial log-likelihood was -2 (-1836.376 +1836.3) = 0.152 which was compared to a mixture of chi-square with zero and one degree of freedom ($\chi^2_{0:1}$). Based on the resulting P-value, 0.348, there was no sufficient evidence to reject the null hypothesis of homogeneity between the centers. Similarly, for the gamma frailty model, the change in partial log-likelihood with inclusion of the frailty was 0.305 and compared to ($\chi^2_{0:1}$) , the resulting P-value was 0.291. From these results, there was no sufficient evidence to reject the null hypothesis; suggesting that events were independent within and across centers.

6.5.3 Comparison of parameter estimates and assessing the heterogeneity parameter estimate for PFS

Table 6.3: *PFS Hazard Ratio (95% CI) from frailty and marginal Cox models*

	Gamma frailty		**Lognormal frailty**		**Marginal Cox model**	
parameter	**HR**	**(95% CI)**	**HR**	**(95% CI)**	**HR**	**(95% CI)**
Treatment: B	0.703	(0.551, 0.826)	0.700	(0.550, 0.823)	0.699	(0.554, 0.848)
Hisloc: 1	0.969	(0.751, 1.251)	0.967	(0.749, 1.249)	0.930	(0.715, 1.210)
Hisloc: 2	0.922	(0.679, 1.252)	0.919	(0.676, 1.249)	0.956	(0.695, 1.315)
Hisloc: 3	0.700	(0.505, 0.970)	0.697	(0.503, 0.967)	0.601	(0.417, 0.865)
Tumor grade: 2	0.816	(0.665, 1.000)	0.816	(0.665, 1.001)	0.727	(0.587, 0.902)
Liver meta: 0	0.766	(0.581, 1.010)	0.766	(0.581, 1.010)	0.709	(0.533, 0.943)
Perform status: 0	0.709	(0.580, 0.867)	0.708	(0.579, 0.866)	0.674	(0.545, 0.834)
Age $\geq$ 50	0.901	(0.728, 1.114)	0.900	(0.728, 1.114)	0.966	(0.773, 1.208)

Table 6.3 presents the results for the PFS endpoint. The HR (95% CI) for most covariates obtained under marginal model were relatively lower (narrower) compared to frailty models (gamma and lognormal). Furthermore, the estimated heterogeneity parameters for gamma and lognormal frailty models were 0.023 and 0.029 respectively. Although the estimated heterogeneity parameters were larger for PFS compared OS, all the center random effects estimated were not significantly different from 0 (Table A6 and A7 in the Appendix). Additionally, a

formal test for the need of center random effect was conducted by comparing the partial log-likelihood for the models with and without the frailty term. Based on a mixture of chi-square with zero and one degree of freedom the resulting P-values were 0.157 and 0.145 for gamma and lognormal frailty models respectively. Therefore, we failed to reject the null hypothesis of homogeneity between centers. For both OS and PFS, it was observed (Table A2 and A3 in the Appendix) that the standard error for the estimated heterogeneity parameter was available for lognormal frailty model and missing for gamma frailty model and a comparison of the parameter estimates is not straightforward because these two frailty densities have different means. For instance, considering PFS endpoint, the estimated frailty mean was $exp(0.029/2) = 1.015$ for lognormal frailty model and 1 for the gamma frailty model.

Chapter 7

Simulation Study

7.1 Introduction

This chapter presents a simulation study that was undertaken to evaluate the performance and robustness of parametric frailty models with respect to bias of the treatment log hazard ($\widehat{\beta}_1$) and the heterogeneity parameter ($\widehat{\theta}$) estimates around the trues initial values. The simulated data was designed to reflect some aspects of the observed clinical trial data analysed in chapter 6 with respect to PFS endpoint. Details of the simulation study and results are presented in the following sections.

7.2 Simulation scheme

Three baseline hazard distributions were considered i.e. exponential, Weibull and the Gompertz distributions. First, assuming a fixed constant event rate λ,

time to event (survival time) for each patient was randomly generated from an exponential distribution expressed by

$$T_{ij} = \frac{log(U)}{\lambda u_i exp(\beta_1 * treatment)} \tag{7.1}$$

For a Weibull baseline hazard, the survival time for subject j in center i corresponded to;

$$T_{ij} = (\frac{log(U)}{\lambda u_i exp(\beta_1 * treatment)})^{1/\rho} \tag{7.2}$$

For the Gompertz baseline hazard, the survival time for each subject was generated as follows

$$T_{ij} = \frac{1}{\alpha} log(1 - \frac{\alpha log(U)}{\lambda u_i exp(\beta_1 * treatment)}) \tag{7.3}$$

where U is a random variable following a uniform distribution in the interval [0,1] (Bender et al., 2005). The true treatment log hazard β_1, was as estimated from the observed data in chapter 6. Patients were assumed to have been accrued into the study at some point during an 84 month period with their entry time generated from a uniform distribution between time zero and 84 months; an approach suggested and applied by Morden et al. (2011). A follow-up period of 24 months was considered. Time at risk for a particular subject was calculated as the time at risk before the end of accrual period plus the follow-up time. A patient j in center i with time to event longer than time at risk was censored with time to censoring equal to time at risk such that $x_{ij} = Min(T_{ij}, C_{ij})$ where C_{ij} is the censoring time independent of T_{ij} and $\delta_{ij} = I(C_{ij} > T_{ij})$ is the censoring indicator as described by Moreno, (2008). The frailties were generated from three distributions. First, a one-parameter gamma with mean 1 and variance θ

was considered. The second frailty distribution considered was a transformation of lognormal distribution which according to Duchateau and Janssen (2008) is expressed as

$$f_U(u) = \frac{1}{u\sqrt{2\pi\sigma^2}} exp(-\frac{(logu - \mu)^2}{2\sigma^2}) \tag{7.4}$$

The mean and variance are expressed as;

$$E(U) = exp(\mu + \sigma^2/2) = 1$$

$$Var(U) = exp(2\mu + \sigma^2)(exp(\sigma) - 1) = \theta$$

Where $\mu = -log(\theta + 1)/2$ and $\sigma^2 = (\theta + 1)$ were used to ensure a lognormal distribution with mean 1 and variance θ. The third distribution considered was a discrete distribution. For this distribution, the frailty was sampled from two values i.e. x_1 and x_2 with probabilities 0.2 and 0.8 respectively. For each initial value of θ, x_1 and x_2 were obtained by solving the following set of constrained mean and variance equations.

$$E(X) = \sum_{1}^{2} prob(x_i)x_i = 1 \tag{7.5}$$

$$Var(X) = \sum_{1}^{2} prob(x_i)(x_i - E(X))^2 = \theta \tag{7.6}$$

Though not a frailty distribution, the discrete distribution was considered so as to study the impact of extreme mis-specification of the frailty distribution on the parameters of interest. The mean and variance were fixed to 1 and θ respectively for the three distributions to allow for comparability. In this simulation study, three values of θ were considered as true values of the heterogeneity parameter

i.e. 0.02, 0.2 and 2 to study the impact of increasing or decreasing the variance of the frailty distributions. 1000 datasets were generated for each parameter setting where N is the fixed sample size and c is the number of centers under consideration. The multiplicative frailty model in section (3.4.4) was fitted to the simulated data. The mean, median, percent relative bias (RB %), standard deviation (SD) and the mean of the standard error (SE) were determined to describe the spread and the bias of around. The RB % and SD for are respectively defined as;

$$RB\% = |\frac{\overline{\beta_1} - \beta_1}{\beta_1}| * 100$$

and

$$SD = \{\sum_i (\widehat{\beta_1}^{(i)} - \overline{\beta}_1)^2/999\}^{1/2}$$

where $\overline{\beta_1} = \sum_i \widehat{\beta_1}^{(i)}/1000$ is the mean of $\widehat{\beta_1}^{(i)}s$ estimated in the i^{th} simulation. Similarly, the mean, median, per cent relative bias (RB %) and standard deviation (SD) for the heterogeneity parameter were obtained as above replacing accordingly.

7.3 Statistical software

Statistical analysis was conducted using SAS version 9.3 and R version 3.0.1. Specifically, *lifetest* procedure in SAS was used to obtain the Kaplan-Meier survival curves and log-rank test. The semi parametric Cox marginal and frailty models were fitted using *phreg* procedure using PPL estimation method. All

simulations were conducted in R and *Survreg* function in Survival Package used to fit parametric frailty models. Ties in the observed and simulated data were handled using the Breslow method, the default method in both SAS and R. Statistical tests were conducted at 5% level of significance and 95% confidence intervals computed where necessary.

7.4 Simulation Results

7.4.1 Comparison of gamma and lognormal generated frailties

The simulated data consisted of a fixed sample size of 450 patients. For simplicity, two settings that varied with respect to number of centers (c) were considered, i.e. 10 centers each having 45 patients and 25 centers each having 18 patients. The true treatment log hazard $(\beta) = -0.353$ was estimated from a proportional hazard model with treatment as the only covariate. Randomization for each patient to receive either the treatment or control was generated from a binomial distribution with success probability 0.5. Three baseline hazard distributions considered were: Exponential, Weibul and the Gompertz distribution. For the exponential baseline hazard a constant event rate, $\lambda = 0.180$ was chosen. For the Weibull baseline hazard a scale parameter $\lambda = 2$ and shape parameter $\rho = 0.8$ were chosen. On the other hand, a scale pamameter $\lambda = 1.5$ and shape parameter $\rho = 5$ were chosen for the Gompertz distribution. All these parameters were

chosen such that approximately 8% of all patients were censored. Figure 7.1

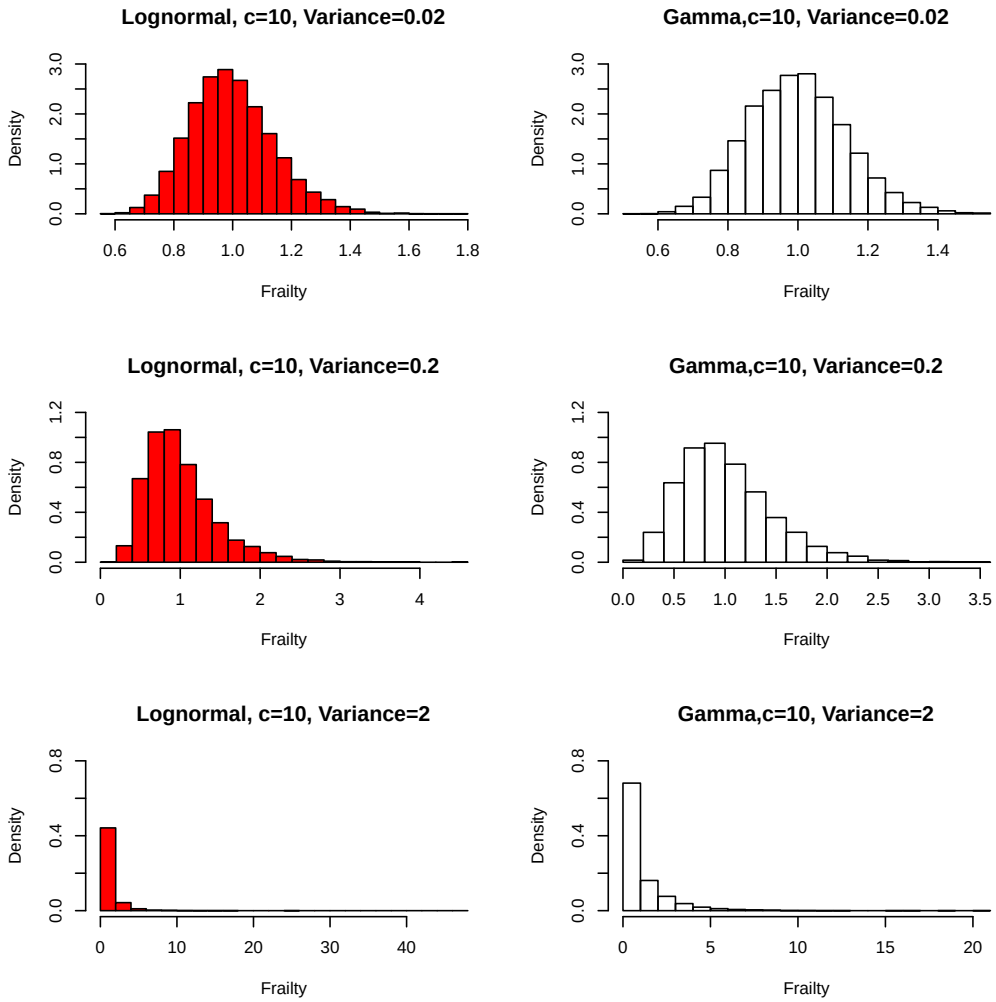

Figure 7.1: *Lognormal and gamma distributed frailties in 10 centers over 1000 iterations*

above presents the histograms of generated frailties under gamma and lognormal frailty densities with mean 1 and variances (θ): 0.02. 0.2 and 2 over 1000 iterations. It is observed that for a given value of θ, these two distributions have approximately similar shapes but deviate from each other with increasing size of θ. Additionally, these densities become more left skewed for larger variances. For

a particular variance (θ), the range of generated lognormal frailties was wider compared to that of gamma distributed frailties. Specifically, for $\theta = 2$, the range of lognormal was twice the range of gamma frailties. When the number of centers was increased to 25 (Figure B.1 in the Appendix), a similar trend was observed but the range reduced accordingly for each variance θ.

Simulation study results of correctly specified lognormal frailty model (lognormal frailty model fitted to clustered data generated from a lognormal distribution) were not evaluated. This is because by default, a lognormal frailty model fitted in *Survreg* function has a mean $E(\widehat{U}) \neq 1$ while the mean of generated lognormal frailties was constrained to 1. Moreover, as noted earlier, the mean and variance of lognormal distribution are linked which led to inflated bias.

7.4.2 Simulation results: Regression coefficient

Table 7.1: *Simulation results for estimated β_1 from correctly specified gamma (True $\beta_1 = -0.353$)*

	Centers =10			Centers = 25		
	θ=0.02	θ=0.2	θ=2	θ=0.02	θ=0.2	θ=2
True frailty distribution: gamma						
Baseline hazard: Exponential						
Mean	-0.356	-0.355	-0.354	-0.353	-0.354	-0.354
Median	-0.355	-0.351	-0.356	-0.351	-0.355	-0.350
RB%	0.979	0.554	0.253	0.088	0.653	0.476
SD	0.096	0.101	0.105	0.099	0.101	0.103
SE	0.097	0.098	0.102	0.097	0.100	0.105
Baseline hazard: Weibull						
Mean	-0.356	-0.354	-0.354	-0.353	-0.356	-0.354
Median	-0.355	-0.352	-0.356	-0.351	-0.354	-0.349
RB%	0.978	0.556	0.253	0.088	0.553	0.382
SD	0.098	0.101	0.105	0.099	0.101	0.104
SE	0.097	0.098	0.102	0.097	0.100	0.105
Baseline hazard: Gompertz						
Mean	-0.355	-0.354	-0.355	-0.353	-0.355	-0.354
Median	-0.353	-0.352	-0.353	-0.351	-0.355	-0.352
RB%	0.977	0.559	0.511	0.089	0.453	0.368
SD	0.094	0.101	0.102	0.099	0.101	0.099
SE	0.096	0.096	0.099	0.097	0.100	0.101

Table 7.1 presents the simulation results for log hazard estimated from correctly specified gamma frailty model (gamma frailty model fitted to clustered data generated from a gamma distribution) with three baseline distributions. Generally, the mean and median of the estimated β were close to true β with a 0.08 % to 0.9 % bias range. Additionally, the differences in terms of the RB% and SD for the three baseline distributions were very negligible suggesting that the baseline hazard distribution did not affect the estimation of the regression parameter. It was further noted that for a particular true θ, the RB% slightly increased

when the number of centers increased from 10 to 25 except for true $\theta = 0.02$ where a decrease was observed. Considering a 10 center scenario, the RB % decreased with increasing magnitude of true θ. However, for a 25 center scenario, no particular trend was observed. The standard error (SE) estimates over 1000 simulations were very close to the SD and both were increasing with an increase in size of true.

7.4.3 Impact of frailty mis-specification on regression coefficient

To assess the impact and sensitivity to mis-specification of the frailty distribution on regression coefficient, a gamma frailty model was first fitted to clustered data generated from a lognormal distribution. From the results presented in Table 7.2, it was observed that the mean and median of the estimated β were very similar. In general, the RB% ranged between 0.7% and 1.58%. A cross the two center settings i.e. 10 and 25, the RB% corresponding to $\theta = 0.02$ and $\theta = 0.2$ decreased when the centers increased from 10 to 25. On the other hand, a slight decrease was observed for true $\theta = 2$. Furthermore, the bias from this mis-specified model and the correctly specified model did not vary substantially suggesting robustness of the gamma frailty with respect to lognormal distribution.

Table 7.2: *Simulation results for estimated β_1 from Misspecified gamma frailty model (True $\beta_1 = -0.353$)*

	Centers =10			**Centers = 25**		
	θ=0.02	θ=0.2	θ=2	θ=0.02	θ=0.2	θ=2
True frailty distribution: Lognormal						
Baseline hazard: Exponential						
Mean	-0.358	-0.359	-0.358	-0.356	-0.357	-0.358
Median	-0.350	-0.357	-0.357	-0.358	-0.355	-0.358
RB%	1.427	1.639	1.289	0.751	1.188	1.038
SD	0.099	0.100	0.096	0.099	0.101	0.101
SE	0.097	0.098	0.098	0.097	0.100	0.101
Baseline hazard: Weibulll						
Mean	-0.357	-0.358	-0.356	-0.356	-0.357	-0.357
Median	-0.354	-0.358	-0.356	-0.358	-0.355	-0.348
RB%	1.383	1.587	1.060	0.751	1.187	1.316
SD	0.099	0.100	0.090	0.099	0.101	0.101
SE	0.097	0.098	0.102	0.097	0.099	0.104
True frailty distribution: discrete						
Baseline hazard: Exponential						
Mean	-0.348	-0.333	-0.291	-0.349	-0.334	-0.292
Median	-0.347	-0.334	-0.291	-0.347	-0.334	-0.292
RB%	1.489	5.677	17.46	1.24	5.465	17.31
SD	0.094	0.092	0.099	0.095	0.092	0.095
SE	0.096	0.096	0.099	0.096	0.096	0.099
Baseline hazard: Weibull						
Mean	-0.347	-0.333	-0.297	-0.348	-0.334	-0.297
Median	-0.347	-0.334	-0.296	-0.350	-0.332	-0.299
RB%	1.435	5.589	15.79	1.322	5.431	15.66
SD	0.094	0.091	0.090	0.096	0.093	0.093
SE	0.097	0.096	0.096	0.098	0.098	0.098

When the gamma frailty model was fitted to the data generated from a discrete distribution, the estimated regression parameter was somewhat sensitive to mis-specification as observed in Table 7.2 with a bias range of 1.24 % to 17.46%. Within a particular center scenario, i.e. either 10 or 25, the RB% substantially increased with an increase in size of the true θ. On the other hand for a particular θ, the RB% decreased by a small margin when the centers increased from 10 to 25. This implied that the regression coefficient was not greatly affected by center size. It is also noted that under the mis-specified models, the SE and SD tended to be smaller compared to correctly specified model.

For the mis-specified models, only the exponential and Weibull baseline distributions were considered. Similar to the correctly specified model, the RB did not vary significantly for the two mis-specified models across the two baseline hazard distributions suggesting that the baseline hazard did not affect the estimation of the regression parameter.

7.4.4 Simulation results: Heterogeneity parameter

Table 7.3: *Simulation results for estimated θ from correctly specified gamma frailty model.*

	Centers =10			Centers = 25		
	θ=0.02	θ=0.2	θ=2	θ=0.02	θ=0.2	θ=2
True frailty distribution: gamma						
Baseline hazard: Exponential						
Mean	0.017	0.179	2.005	0.019	0.201	1.976
Median	0.013	0.160	1.984	0.009	0.186	1.942
RB%	13.26	6.513	0.266	4.410	0.711	1.165
SD	0.019	0.068	0.308	0.021	0.091	0.498
Baseline hazard: Weibull						
Mean	0.017	0.128	1.892	0.019	0.201	1.978
Median	0.013	0.113	1.815	0.009	0.186	1.943
RB%	13.89	6.95	5.504	4.415	0.713	1.097
SD	0.018	0.018	0.713	0.022	0.091	0.499
Baseline hazard: Gompertz						
Mean	0.016	0.113	1.871	0.019	0.201	1.949
Median	0.013	0.114	1.804	0.009	0.186	1.929
RB%	13.90	6.23	6.419	4.414	0.713	2.561
SD	0.018	0.192	0.088	0.021	0.091	0.469

From Table 7.3, simulation results for estimated θ obtained under correctly specified gamma frailty model are presented. It was observed that the RB% range was between 0.27% and 13.9%. For a 10 centers scenario, the RB% decreased with increasing size of true θ. However, this trend was not observed in 25 centers scenario. Furthermore, for a particular true θ, the RB% decreased when the number of centers increased from 10 to 25. The standard deviation (SD) was increasing with an increase in magnitude of true θ and number of centers. For a particular true θ in either 10 or 25 center setting, the RB% for the exponential baseline hazard was slightly lower compared to the other baseline hazards. On the other hand, the relative bias remained within the same range with negligible

differences for the Weibull and Gompertz baseline hazards models.

7.4.5 Impact of mis-specification on heterogeneity parameter

Table 7.4: *Simulation results for estimated θ from mis-specified frailty models.*

	Centers =10			**Centers = 25**		
	θ=0.02	θ=0.2	θ=2	θ=0.02	θ=0.2	θ=2
True frailty distribution: Lognormal						
Baseline hazard: Exponential						
Mean	0.017	0.267	0.970	0.018	0.163	0.918
Median	0.013	0.243	0.949	0.009	0.154	0.905
RB%	14.11	33.29	51.50	10.46	18.74	54.09
SD	0.018	0.160	0.275	0.019	0.067	0.214
Baseline hazard: Weibull						
Mean	0.017	0.266	969	0.018	0.163	0.915
Median	0.013	0.243	0.949	0.009	0.152	0.905
RB%	14.11	33.27	51.50	10.48	18.73	54.10
SD	0.018	0.168	0.274	0.194	0.069	0.213
True frailty distribution: Discrete						
Baseline hazard: Exponential						
Mean	0.002	0.003	0.003	0.004	0.004	0.005
Median	0.000	0.000	0.000	0.000	0.000	0.000
RB%	89.47	98.69	99.84	77.96	97.88	99.77
SD	0.006	0.006	0.007	0.010	0.009	0.010
Baseline hazard: Weibull						
Mean	0.002	0.002	0.002	0.004	0.004	0.005
Median	0.000	0.000	0.000	0.000	0.000	0.000
RB%	87.16	98.66	99.86	77.16	97.56	99.74
SD	0.006	0.006	0.005	0.010	0.009	0.010

Similar to the regression coefficient, sensitivity to mis-specification of the frailty distribution was assessed with respect to the estimated heterogeneity parameter θ. From Table 7.4, moderate to high RB% was observed for each of the assumed true θ. The RB% was much higher for the two mis-specified models compared

to the correctly specified frailty model. Specifically, for a gamma frailty model fitted to discrete generated frailties, serious downward bias was observed with a RB% ranging between 77.16% and 99.8%. Additionally, the RB% increased with an increase in size of true θ.

For gamma frailty model fitted to lognormal generated frailties, the RB% ranged between 10.46% and 54.10%. For $\theta = 0.02$, the bias was close to that of correctly specified gamma frailty model. These observations were consistent with results of generated frailties whereby for $\theta = 0.02$ and $\theta = 0.2$, the range of frailties for the two distributions were close whereas for $\theta = 2$, the range was much wider for lognormal frailties compared to gamma distributed frailties. A slight decrease in RB% was also observed when the number of centers increased from 10 to 25. These results show that the mis-specified gamma frailty model was not successful in estimating the underlying true heterogeneity parameter. Similar to correctly specified gamma frailty model, the standard deviation (SD) generally increased with an increase in size of initial θ.

For a particular value of true θ, the RB% was quite similar for exponential and Weibull distributions for each of the mis-specified models with no substantial variation across the two center settings.

Chapter 8

Discussion

From Expolatory Data Anaysis (EDA), the patients' baseline characteristics in the observed data were well balanced between the treatment groups as expected. This was important in ensuring there was no allocation bias to influence the treatment outcome. It was also observed that some centers accrued less than 5 patients. This small centers were dropped to avoid estimation-related problem in subsequent statistical analysis.

Exploring the survival function, the Kaplan-Meier OS and PFS survival curves by treatment were crossing suggesting violation of proportional hazard assumptions. In this regard, alternative methods such as the Accelerate Failure time (AFT) model which do not require proportion hazard assumptions to hold can be used.

Additionally, the Kaplan-Meier survival curves for both OS and PFS stratified by centers suggested some variability in outcome between centers. These plots

were similar to those from a previous study on Heterogeneity in disease free survival between centers: lessons learned from an EORTC breast cancer trial conducted by Legrand et al. (2006). From this previous study, such plots are difficult to interpret, first due to the large number of curves and second because the precision in the estimation of each curve, which depends on the number of events observed in each center, should also be taken into account (for example, through confidence bands).

From the statistical analysis results, parameter estimates (HR) and corresponding 95% confidence intervals from the frailty models and marginal Cox PH model were close. This similarity was attributed to the fact that none of the center random effect was significant. Furthermore, based on the mixture of chi-square likelihood ratio test with 0 and 1 degrees of freedom, the heterogeneity parameter estimates for both gamma and lognormal frailty models were very small and insignificant hence, there was no sufficient evidence to reject the null hypothesis of no center effect.

The PPL estimation method was used in the analysis of observed data using semi-parametric frailty models both OS and PFS. From results, it was observed that the standard error for the estimated heterogeneity parameter was available for lognormal frailty model and missing for gamma frailty model. This was due to the difference in the outer loop for the two frailty distributions i.e. a REML estimate is available for in the case of lognormal density whereas such an estimate is not available for gamma frailty distribution (Duchateau and Janssen, 2008).

Furthermore, a comparison of the parameter estimates between these two frailty models is not straightforward due to the differences in means and variances.

A simulation study was conducted with an aim to investigate the impact of frailty distribution mis-specification on estimated regression and the heterogeneity parameter. PFS was the endpoint of interest and several settings with respect to number of centers and true heterogeneity parameter were considered. From the results of correctly specified gamma, the estimated mean and median of the treatment log hazard were very close to the true treatment effect. As a result, the RB% was small with no major discrepancies with respect to number of centers or true heterogeneity parameter considered. On the other hand, low to moderate percentage RB was observed in the estimation of the heterogeneity parameter.

To investigate the impact of mis-specification of the frailty distribution, two scenarios were considered. First, a gamma frailty model was fitted to clustered data generated under lognormal distribution (mis-specified model). The estimated mean and median of the treatment log hazard were very similar. Additionally, the RB% was relatively small and comparable to that of correctly specified model. This indicated that frailty distribution mis-specification did not greatly affect the regression coefficient estimate despite the fact that different frailty distributions can lead to noticeably different association structures.

These findings were similar to those rom a previous study examining the gamma frailty model in multi-center clinical trial (cohort study). Glidden and

Vittinghoff (2004) found by simulation that regression coefficient estimates were minimally affected by frailty mis-specification. However, their assumed true frailty distribution was inverse Gaussian. Their study also differed from this one in terms of center size and initial simulation parameters.

For the heterogeneity parameter, the RB% was somewhat large and more pronounced as compared to correctly specified model. Besides, this bias was increasing with an increase in magnitude of true heterogeneity parameter but was less affected by the number of centers. These study findings were consistent with results from a perioperative breast cancer clinical trial study whereby Duchateau and Janssen, (2008) which investigated the robustness of the gamma frailty distribution assumptions with respect to the lognormal distribution. Results revealed downward bias of the variance estimator in the misspecified model (gamma frailty model fitted to clustered data generated from the lognormal frailty model). However, their study differed from this one in terms of true θ considered as well as number of centers.

In the second scenario of mis-specified models, gamma frailty model was fitted to data simulated from a discrete distribution. From the results, it was evident that the regression coefficient was somewhat sensitive to the extreme frailty mis-specification. The bias was much larger compared to other fitted models particularly for large true heterogeneity parameters.

Likewise, large RB% was observed for the heterogeneity parameter. Specifically, the relative bias was more pronounced for large θ and slightly influenced

by the number of centers considered. These results showed lack of fit of the continuous gamma frailty distribution approximation for the discrete frailties. This clearly indicated that a discrete frailty distribution was an extreme and inappropriate choice.

For the three baseline hazards distributions considered in this study, there was no substantial difference in the percent relative bias particulary for the regression coefficient. Therefore, any of them could be appropriate for inference in this particular study.

Chapter 9

Conclusion

The primary objective of this project was to review various survival models for clustered data with focus on frailty models and their properties. Specifically, semi-parametric gamma and lognormal frailty models and marginal models were fitted to observed clinical trial data with an objective to compare the parameter estimates and assess the estimated heterogeneity parameters. Based on study results, the parameter estimates from the two frailty models considered were almost identical to those estimated from a marginal Cox PH model. Furthermore, with no sufficient evidence to reject the null hypothesis of homogeneity between the centers in frailty models, we concluded that events were independent within and across centers. Therefore, for this particular study, either of the marginal or frailty models could be used for statistical inference. However, this may not hold in other studies and the choice of model should be driven by the scientific objectives of the study i.e. a marginal model should be used when population

average risk is of interest whereas a frailty model would be more appropriate when interest lies on center specific risk.

This study also sought to assess the impact of frailty distribution mis-specification on the parameters of interest i.e. treatment log hazard and the heterogeneity parameter as well as assess sensitivity of these parameter estimates in terms of bias with respect to varying baseline hazard distributions and choice of initial parameters (center size and true heterogeneity parameters). Assuming a gamma frailty distribution when the true frailty distribution is lognormal, the regression coefficient (treatment effect) estimate was minimally affected in terms of relative bias. On the other hand, results showed lack of fit of the continuous gamma frailty distribution approximation for the discrete distributed frailties. This was a clear indication that the discrete distribution was an extreme and inappropriate choice. For the heterogeneity parameter, assuming a gamma distribution when the true frailty distribution is either lognormal or discrete, robustness was an issue particularly for large values of true θ. There was no substantial difference in the percent relative bias for the three baseline hazards particulary for the regression coefficient therefore, any of them could be used for inference in this particular study.

From the simulation study, we concluded that the heterogeneity parameter was more sensitive to mis-specification of the frailty distribution and choice of initial parameters (center size and true heterogeneity parameters) compared to regression parameter estimate. In this regard, the gamma frailty model can be

a practical choice in real data analysis when the regression parameters are of primary interest, as in multi-center clinical trial with survival data when the choice of underlying frailty distribution is not straightforward.

9.1 Limitations and recommendations for further research

This project was limited to investigating only the center random effect since the software used did not allow for more than one random effect i.e. some gaps remain, especially in the use of frailty models for treatment-by-center interaction. Thus development of computation and theory for such extended multifrailty models is a useful area for future development. Furthermore, in the simulation study, treatment was the only covariate and therefore we recommend future testing of the frailty models with baseline hazard adjusted for other patient-specific covariates so as to evaluate the models in more details.

In this study only the gamma and the lognormal frailty densities have been used. It might be of interest to consider other frailties distributions in future studies. The simulation study was conducted under the assumptions that all center had equal number of patients was considered for ease of computation; however this is not often the case in real life multicenter clinical trials. In this regard, we recommend an extension of this study considering centers with unequal number of patients.

Although the findings from the simulations are not necessarily limited to the parameter settings studied, the conclusions are highly relevant for Soft-Tissue Sarcoma clinical trials. Therefore, for other types of tumors and diseases, different parameter settings might be more relevant.

In this project, only frequentists survival analysis techniques were applied for analysis and estimation of the parameters of interest. In this regard, this project can be extended in future by considering methods such as Bayesian survival analysis techniques which incorporate prior information available into the study to estimate the parameters of interest.

References

Allison, P. (2010). Survival Analysis Using SAS: A Practical Guide. Cary, NC, USA: SAS Institute Inc.

Balakrishnan, N. and Peng, Y.(2006) Generalized gamma frailty model.*Stat. Med.* **25**:2797-2816

Bear, R.E. (1959). Note on some mathematical mortality models. The life span of animals.Ciba Foundation colloquium on aging. Little,Brown, Boston 302-311

Bender, R., Augustin, T. and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazard models. *Stat. Med.* **24**:1713-1723.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics.* **30**,89-99.

Cance.net. (2013). Sarcoma. Accessed on 10/04/2014.
Available at : http://www.cancer.net/cancer-types/sarcoma.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease

incidence. Biometrika,**65**:141-151.

Collett, D. (1994).*Modelling Survival Data in Medical Research.* Chapman and Hall, London.

Cox, D.R (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society.* **34**:187-220

Demissie,S.M (2009)Investigating center effects in a multi-center clinical trial study using a parametric proportional hazards meta-analysis model

Duchateau, L., Janssen, P., Lindsey, P., Lengrand, C., Nguti, R. and Sylvester, R. (2002).The shared frailty model and the power for heterogeneity test in multicenter trials. *Comput. Stat.* **40**: 603-620.

Duchateau, L. and Janssen, P. (2008). *The Frailty Model.* New York, Springer.

Efron, B. (1977). The efficiency of Cox's likelihood function of censored data. *Journal of the American Statistical Association.* 72, 557-565.

Genest, C. and MacKay, J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *Am. Stat.* **40**, 280-283.

Glidden, D.V. and Vittinghoff, E. (2004). Modelling clustered survival data from multicenter clinical trials. *Stat. Med.* **23**: 369-388.

Govindarajulu, U.S., Lin, H., Lunetta, K.L., and D'Agostino, R.B. (2011). Frailty Models:Application to biomedical and genetic studies. *Stat. Med.* **30(22)**:2754-64.

Garcia , C.R., Supko, J. G., Manola, J., Seiden, M.V., Harmon, D., Ryan, D.P., Quigley, M.T., Merriam, P., Canniff, J., Goss, G., Matulonis, U., Maki, R.G., Lopez, T., Puchalski, T. A., Sancho, M.A., Gomez, J., Guzman, C., Jimeno, J., and Demetri, G.D. (2004). Phase II and Pharmacokinetic Study of Ecteinascidin 743 in Patients With Progressive Sarcomas of Soft Tissues Refractory to Chemotherapy. *Journal of clinical oncology.* Vol **22**:1480-1490.

Ha, I.D., Sylvester, R., Legrand, C., and MacKenzie, G. (2011). Frailty Modelling for Survival Data from Multicenter Clinical Trial. *Stat. Med.* **30**:2144-2159.

Ha, I. D. and Mackenzie, G. (2010). Robust frailty modelling using non-proportional hazard models. *Statistical modelling* **10**: 315-332.

Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**: 387-396.

Jenkins, Stephen P. (2008). Survival Analysis. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester.

Johansen, S.(1978). The product limit estimate as a maximum likelihood estimate. *Scandinavian Journal of Statistics* **5**

Keele, L. (2007). Cross validation tests for frailty models. *Technical report*, Ohio State University.

Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika* **30**,81-93.

Klein, J. P. and Moeschberger, L. M. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, Second Edition. Springer-Verlag, New York.

Legrand, C., Duchateau, L., Sylvestera, R., Janssen, P., Hage, J.A., Velded, C. and Therassea, P. (2006). Heterogeneity in disease free survival between centers: lessons learned from an EORTC breast cancer trial. *Clinical trials.* **3**:10-18.

Li, H., Malka, G. and Kathleen, M. (2007). On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is misspecified. *Stat. Med.* **26**:4657-4678.

McGilchrist, C.A. and Aisbett, C.W. (1991). Regression with Frailty in Survival Analysis. Biometrics, **47**:461-466.

McGilchrist, C. A. (1993). REML estimation for survival models with frailty. *Biometrics* **49**:221-225.

Morden, J.P., Lambert, P.C., Latimer N., Abrams, K.R and Wailoo, A. J. (2011). Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. BMC Medical Research Methodology. Vol 11:4 doi: **10**.1186/1471-2288.

Moreno, E.S.S. (2008). Nonparametric frailty models for clustered survival data. Ph.D. Dissertation. Cornell University 2008.

Morgan, B.J.T. (1992). *Analysis of quantal response data.* Chapman and Hall, London.

Nguti, R. (2003). Random effects survival models applied to animal breeding data. Ph.D. Dissertation. Limburgs Universitair Centrum.

Oakes, D. (1989). Bivariate survival models induced by frailties. Journal of the American Statistical Association, **84**:487-493.

Qi,J,(2009). Comparison of Proportional Hazards and Accelerated Failure Time Models. Masters dissertaion,University of Saskatchewan.

Rodriguez ,G. (2010) Parametric Survival Models.

Therneau, T. M. and Grambsch, P. M.(2000). *Modeling Survival data*: Extending the Cox model. Springer, New York.

Tundo, J. (2009). Frailty models for the between center variation in survival following rectum cancer diagnosis. Masters Dissertaion, Ghent University,Belgium.

Ulviya, A. (2013).Frailty Models for modelling heterogeneity. masters dissertation at McMaster University.

Vaupel, J.W., Manton, K.G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**: 439-454.

Wienke, A. (2003). Frailty Models. Max Planck Institut fr demograsche Forschung,

Working Paper 2003-032.

Wienke, A. (2011). *Frailty Models in Survival data.* Chapman and Hall, New York, 1 edition.

Appendix

Table A1: *Distribution of patients in centers by treatment received*

	treatment A	treatment B	
Center No.	**n (%)**	**n (%)**	**Total**
101	3 (42.86)	4 (57.14)	7
147	15(53.57)	13 (46.43)	28
227	14 (53.85)	12 (46.15)	26
301	12 (48.00)	13 (52.00)	25
302	20 (52.63)	18 (47.37)	38
304	4 (40.00)	6 (60.00)	10
310	16 (47.06)	18 (52.94)	34
335	5 (62.50)	3 (37.50)	8
406	11 (55.00)	9 (45.00)	20
508	7 (63.64)	4 (36.36)	11
510	5 (41.67)	7(58.33)	12
527	7 (33.33)	14 (66.67)	21
528	6 (66.67)	3 (33.33)	9
530	16 (48.48)	17 (51.52)	33
601	7 (46.67)	8 (53.33)	15
609	6 (40.00)	9 (60.00)	15
610	8 (57.14)	6 (42.86)	14
613	12 (54.55)	10 (45.45)	22
622	9 (40.91)	13 (59.09)	22
661	5 (45.45)	6 (54.55)	11
1765	5 (50.00)	5 (50.00)	10
3039	5 (55.56)	4(44.44)	9
6998	3 (60.00)	2 (40.00)	5
7802	11 (50.00)	11 (50.00)	22

Table A2: *Parameter Estimates(SE) for Overall survival*

	Gamma frailty	**Lognormal frailty**	**Marginal model**
parameter	**Estimate(SE)**	**Estimate (SE)**	**Estimate (SE)**
Treatment: B	-0.242 (0.109)	-0.242 (0.109)	-0.241(0.109)
Hisloc :1	-0.173(0.138)	-0.174(0.139)	-0.172 (0.137)
Hisloc: 2	-0.085(0.164)	-0.088(0.164)	-0.080 (0.163)
Hisloc: 3	-0.546(0.187)	-0.549 (0.188)	-0.540 (0.187)
Tumorgrade :2	-0.269(0.111)	-0.269 (0.111)	-0.268 (0.109)
Livermeta: 0	-0.334(0.149)	-0.335(0.149)	-0.332 (0.149)
Perform status:0	-0.572(0.109)	-0.574(0.109)	-0.569(0.109)
Age $\geq$ 50	0.146 (0.114)	0.145(0.115)	0.146 (0.114)
Hospno (θ)	0.005(-)	0.008 (0.022)	-

Table A3: *Parameter Estimates(SE) for PFS survival*

	Gamma frailty	**Lognormal frailty**	**Marginal model**
parameter	**Estimate(SE)**	**Estimate (SE)**	**Estimate (SE)**
Treatment B	-0.353 (0.103)	-0.356 (0.103)	-0.357 (0.109)
Hisloc: 1	-0.031 (0.130)	-0.034 (0.131)	-0.072 (0.134)
Hisloc :2	-0.081 (0.156)	-0.085 (0.157)	-0.045 (0.163)
Hisloc: 3	-0.357 (0.166)	-0.360 (0.167)	-0.509 (0.186)
Tumorgrade :2	-0.204 (0.104)	-0.203 (0.104)	-0.318 (0.109)
Liver meta: 0	-0.267 (0.141)	-0.266 (0.141)	-0.344 (0.145)
Perform status:0	-0.344 (0.103)	-0.345 (0.103)	-0.394 (0.108)
Age $\geq$50	-0.105 (0.108)	-0.105 (0.109)	-0.034 (0.114)
Hospno (θ)	0.023 (-)	0.029 (0.027)	-

Table A4: *OS hospital specific random effects :gamma frailty model*

HOSP NO.	Estimate	Exponentiated estimate	Exp 95% CI
101	0.00548	1.005	[0.879, 1.151]
147	0.1129	0.982	[0.861, 1.120]
227	-0.0644	0.938	[0.821, 1.071]
301	-0.00893	0.991	[0.870, 1.130]
302	0.0147	1.015	[0.893, 1.153]
304	0.0135	1.014	[0.887, 1.159]
310	-0.0187	0.981	[0.862, 1.118]
335	0.0118	1.012	[0.885, 1.157]
406	0.00887	1.009	[0.884, 1.151]
508	0.0100	1.010	[0.883, 1.155]
510	-0.00129	0.999	[0.874, 1.142]
527	0.00511	1.005	[0.881, 1.146]
528	-0.0109	0.989	[0.864, 1.132]
530	0.0195	1.020	[0.896, 1.160]
601	0.000249	1.000	[0.876, 1.143]
609	0.0201	1.020	[0.894, 1.165]
610	-0.00975	0.990	[0.866, 1.132]
613	0.0201	1.020	[0.895, 1.163]
622	0.00852	1.009	[0.885, 1.149]
661	-0.0190	0.981	[0.858, 1.122]
1765	0.0116	1.012	[0.885, 1.157]
3039	0.000387	1.000	[0.874, 1.144]
6998	0.00309	1.003	[0.876, 1.148]
7802	-0.00546	0.995	[0.872, 1.135]

Table A5: *OS hospital specific random effects: lognormal frailty model*

HOSP NO.	Estimate	Exponentiated estimate	Exp 95% CI
101	0.00982	1.010	[0.842, 1.211]
147	-0.0315	0.969	[0.817, 1.149]
227	-0.1061	0.899	[0.759, 1.065]
301	-0.0154	0.985	[0.830 ,1.168]
302	0.0241	1.024	[0.867 ,1.210]
304	0.0245	1.025	[0.855, 1.228]
310	-0.0315	0.969	[0.820, 1.146]
335	0.0214	1.022	[0.852 ,1.225]
406	0.0153	1.015	[0.852, 1.210]
508	0.0180	1.018	[0.850, 1.220]
510	-0.00255	0.997	[0.834, 1.193]
527	0.00855	1.009	[0.847 ,1.200]
528	-0.0196	0.981	[0.819 ,1.174]
530	0.0330	1 1.034	[0.872, 1.225]
601	0.000229	1.000	[0.838 ,1.194]
609	0.0359	1.037	[0.867 ,1.239]
610	-0.0174	0.983	[0.823, 1.174]
613	0.0352	1.036	[0.870 ,1.234]
622	0.0144	1.015	[0.854 ,1.206]
661	-0.0334	0.967	[0.810, 1.154]
1765	0.0210	1.021	[0.852 ,1.224]
3039	0.000552	1.001	[0.835, 1.199]
6998	0.00556	1.006	[0.838 ,1.207]
7802	-0.00981	0.990	[0.833, 1.178]

Table A6: *pfs hospital specific random effects :gamma frailty model*

HOSP NO.	Estimate	Exponentiated estimate	Exp 95% CI
101	0.0155	1.016	[0.770, 1.340]
147	-0.2117	0.809	[0.631 ,1.038]
227	-0.1965	0.822	[0.639 ,1.057]
301	0.0131	1.013	[0.796 ,1.290]
302	0.0120	1.012	[0.804 ,1.275]
304	0.0133	1.013	[0.772 ,1.330]
310	-0.00030	1.000	[0.791, 1.263]
335	0.0246	1.025	[0.779, 1.349]
406	0.0821	1.086	[0.845, 1.394]
508	0.0834	1.087	[0.832 ,1.420]
510	0.0555	1.057	[0.811 ,1.378]
527	0.0208	1.021	[0.794 ,1.313]
528	-0.0216	0.979	[0.743 ,1.289]
530	0.0568	1.058	[0.838 ,1.336]
601	-0.0939	0.910	[0.700 ,1.184]
609	0.0537	1.055	[0.814 ,1.367]
610	-0.0496	0.952	[0.732 ,1.237]
613	-0.00206	0.998	[0.778, 1.280]
622	0.0975	1.102	[0.862 ,1.410]
661	-0.1250	0.883	[0.671, 1.161]
1765	0.0453	1.046	[0.797, 1.373]
3039	0.0394	1.040	[0.790, 1.369]
6998	0.0518	1.053	[0.794, 1.396]
7802	-0.0380	0.963	[0.748, 1.240]

Table A7: *pfs hospital specific random effects :lognormal frailty model*

HOSP NO.	Estimate	Exponentiated estimate	Exp 95% CI
101	0.0186	1.019	[0.750, 1.384]
147	-0.2249	0.799	[0.622, 1.026]
227	-0.2099	0.811	[0.629 ,1.045]
301	0.0156	1.016	[0.783 ,1.318]
302	0.0143	1.014	[0.793, 1.298]
304	0.0163	1.016	[0.754 ,1.370]
310	-0.00005	1.000	[0.779 ,1.283]
335	0.0301	1.031	[0.760 ,1.397]
406	0.0990	1.104	[0.835, 1.459]
508	0.1039	1.109	[0.820, 1.502]
510	0.0679	1.070	[0.796, 1.439]
527	0.0242	1.025	[0.779 ,1.347]
528	-0.0254	0.975	[0.723, 1.314]
530	0.0658	1.068	[0.829, 1.376]
601	-0.1050	0.900	[0.684, 1.185]
609	0.0647	1.067	[0.801 ,1.421]
610	-0.0566	0.945	[0.715, 1.249]
613	-0.00219	0.998	[0.764, 1.304]
622	0.1176	1.125	[0.855, 1.480]
661	-0.1394	0.870	[0.655, 1.155]
1765	0.0554	1.057	[0.780, 1.432]
3039	0.0485	1.050	[0.773 ,1.426]
6998	0.0650	1.067	[0.775 ,1.469]
7802	-0.0433	0.958	[0.732, 1.253]

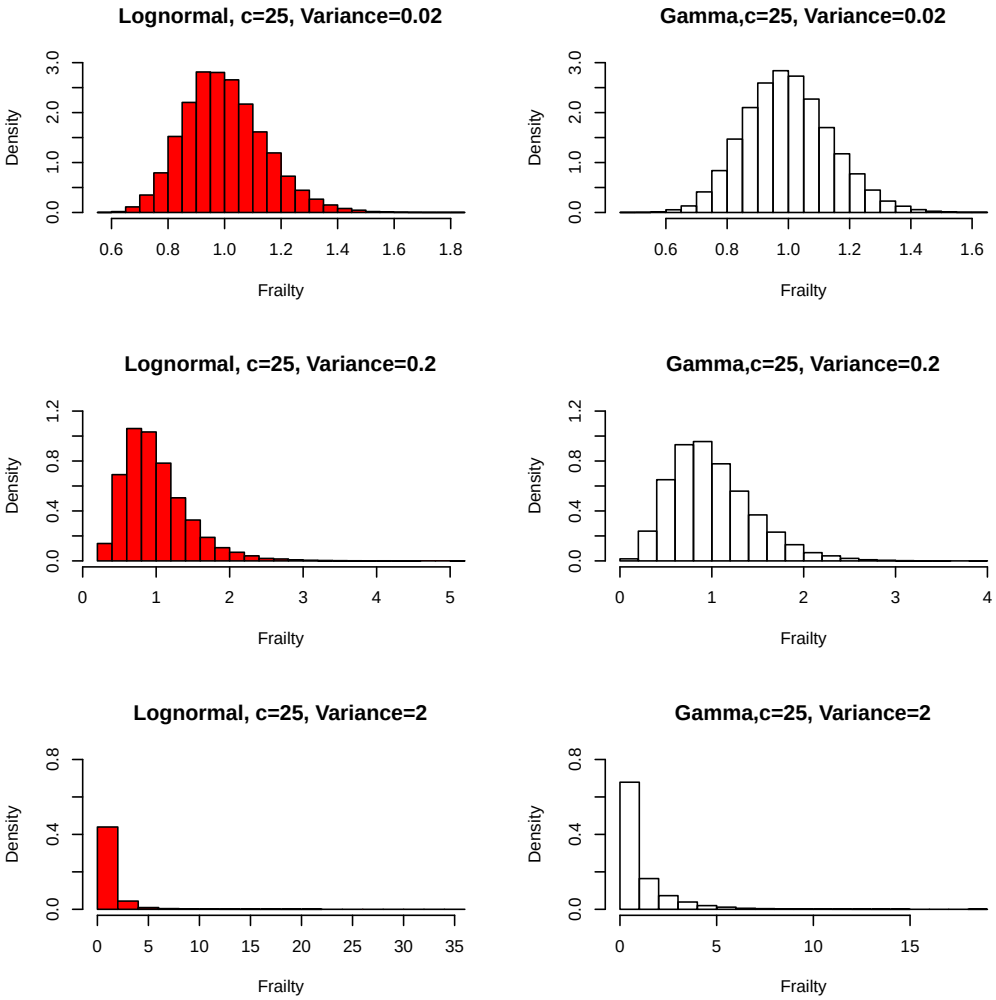

Figure 9.1: *Lognormal and gamma distributed frailties in 25 centers over 1000 iterations*