2012•2013
FACULTY OF SCIENCES
Master of Statistics: Biostatistics

Masterproef

The impact of institution variability on patients outcome in a soft tissue sarcoma clinical trial

Promotor :
Prof. dr. Tomasz BURZYKOWSKI

Promotor :
Dr. SASKIA LITIERE

Susan Gachau

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2012•2013
FACULTY OF SCIENCES
Master of Statistics: Biostatistics

Masterproef

The impact of institution variability on patients outcome in a soft tissue sarcoma clinical trial

Promotor :
Prof. dr. Tomasz BURZYKOWSKI

Promotor :
Dr. SASKIA LITIERE

Susan Gachau
Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

universiteit hasselt | Maastricht University

Acknowledgments

First I thank the almighty God for giving me life and strength to successfully complete this thesis.
My sincere appreciation goes to my internal supervisor, Prof. Dr. Tomasz Burzykowski and external supervisor, Dr. Saskia Litière for all their guidance, time, constructive advice and suggestions that led to successful completion of this Thesis project.

I owe my deepest gratitude to the European Organization for Research and Treatment of Cancer (EORTC) for giving me the opportunity to work on such an interesting project using clinical trial data and permission to use their premises during thesis period.

I appreciate the financial support from the Flemish Interuniversity Council (VLIR) which has enabled me to pursue this valuable Master program through the scholarship grant. I would like to acknowledge my professors at Center for Statistics for imparting part of their statistical knowledge to me. Many thanks to Hasselt University staff particularly Martine Machiels and Marc Tholen for all the assistance prior to my coming and during my stay in Belgium.

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

Susan Gachau

September 20, 2013

ABSTRACT

Background: Soft-tissue sarcomas (STS) are a rare and heterogeneous group of tumors of mesenchymal origin and can start in any part of the body mainly in support and connective tissues. When performing large multicenter clinical trials, a lot of attention is paid to the standardization of the treatment procedures across the participating sites. This is done to ensure that the difference in effect observed between treatment arms, can be completely attributed to the difference in treatments. This is because heterogeneity between the participating centers may add noise to the estimation of the treatment effect.

Objective: The overall objective was to investigate the impact of institution variability on patient outcome in a soft tissue sarcoma clinical trial. Specifically, to compare parameter estimates from different statistical analysis methods as well as investigate the impact of frailty misspecification on parameters of interest.

Methods: In statistical analysis, semi parametric Cox marginal regression model and frailty models were used to analyze right censored data with Overall Survival (OS) and Progression-Free Survival (PFS) endpoints. A simulation study was also conducted to assess the impact of frailty distribution misspecification on treatment effect and on heterogeneity parameters estimates. Different settings in terms of the number of centers and true heterogeneity parameter were considered.

Results: The parameter estimates obtained from frailty models and marginal model were very close particularly for Overall Survival. Furthermore, the heterogeneity parameters in both endpoints were small yielding insignificant center effect. From the simulation study, the correctly specified gamma frailty model resulted in less biased estimates for treatment effect and heterogeneity parameters compared to misspecified models.

Conclusions: In the absence of center effect, marginal and frailty models were close in parameter estimates and either model could be adopted for statistical inference. From simulation study results, the heterogeneity parameter was more sensitive to misspecification of the frailty distribution and the choice of initial parameters compared to the treatment parameter estimate.

Key words: *Absolute relative bias, Gamma frailty, Lognormal frailty, Overall Survival, Progression-Free Survival.*

Contents

List of Abbreviations

CI	:	Confidence Interval
EORTC	:	European Organization for Research and Treatment of Cancer
EM	:	Expectation- Maximization
HR	:	Hazard Ratio
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

CHAPTER 1: INTRODUCTION

1.1 Back ground

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.

When performing large multicenter clinical trials, a lot of attention is paid to the standardization of the treatment procedures across the participating sites. This is done to ensure that the difference in effect observed between treatments arms can be completely attributed to the difference in treatments. Nevertheless, country-specific regulation, experience with the disease, experience with the treatment under investigation among other factors may introduce heterogeneity between the participating centers and may add noise to the estimation of the treatment effect. Moreover, Ha *et al.* (2012) noted that such heterogeneity may alter the interpretation and reporting of the treatment effect.

According to Duchateau *et al*. (2002), heterogeneity decreases the power to detect clinically important treatment differences, but on the other hand, more heterogeneous trials lead to more general conclusions as they are based on a wider patient population. Legrand *et al*. (2006) suggested that if the differences in outcome exist between centers, it is important to find out what

factors have caused the heterogeneity as it might serve to improve the quality of patient care. These factors include both patient-specific factors and center-specific factors.

A natural framework for estimating the unexplained variability is through a frailty model. Frailty models were introduced by Vaupel *et al.* (1979) as a generalization of the Cox's proportional hazards model allowing for random effects as a result of unobserved heterogeneity of each individual or group. In this model, the unobserved frailty shared by individual members in a cluster acts multiplicatively as a factor on the hazard function and is typically modelled parametrically (Li *et al.*, 2007; Ha and Gilbert, 2010; Legrand *et al.*, 2006; Govindarajulu *et al.*, 2009).

Several frailty distributions amongst them the lognormal and the power variance function family comprising of gamma, compound Poisson, inverse Gaussian and positive stable 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.

Duchateau and Janssen (2008) noted that the choice of the frailty distribution is crucial to obtain correct estimates of the dependence structure but the researcher often has no prior information with which to choose among the distributions. Furthermore, due to the latent nature of the frailty term, it can be difficult to determine an appropriate frailty distribution for a particular data set. Thus, misspecification 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). It is therefore useful in practice to examine to what extent misspecification of the frailty distribution affects the validity of the regression coefficients and heterogeneity parameter estimates (Li *et al.*, 2007).

1.2 Objectives

The primary objective of this Thesis is to investigate the impact of institution variability on patient outcome in a STS clinical trial. In order to achieve the overall objective, the problem will be broken down into specific sub-sections as follows;

1. Model time to event endpoints using marginal and frailty models, compare the parameter estimates and investigate the heterogeneity parameter.
2. Study the impact of frailty distribution misspecification on the parameters of interest i.e. treatment log hazard effect and heterogeneity parameter.
3. Assess sensitivity of parameter estimates in terms of bias with respect to different simulation parameter settings.

1.3 Thesis overview

This thesis is organized as follows: In chapter 2, the study design and variables in the data set are discussed. Chapter 3 discusses the statistical methods applied in the analysis and a description of the simulation scheme. In chapter 4, results from the applied statistical methods and simulations are presented while chapter 5 provides the discussion and conclusion of the study. The last sections present the references and appendix respectively.

CHAPTER 2: STUDY DESIGN

The data analyzed in this report came from a randomized phase III, open label, multicenter study conducted jointly by European Organization for Research and Treatment of Cancer (EORTC)-Soft Tissue and Bone Sarcoma Group and 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 in 9 different countries. Patients enrolled were between 18 and 63 years of age and had histological evidence of high grade STS with advanced unresectable or metastatic disease. Eligible patients were randomized to receive 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 patients' 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.

2.1 Data description

The variables in the data and their coding are presented in Table 2.1. The data are right censored and all the variables considered were categorical except age which was continuous.

Table 2.1: *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

CHAPTER 3: METHODS

3.1 Exploratory Data Analysis (EDA)

Exploratory data analysis was conducted to gain insight into the data. This was achieved by using summary statistics and frequencies to evaluate the distribution of patients' baseline characteristics and the distribution of patients across the centers by treatment arm. Kaplan-Meier survival curves were also obtained for OS and PFS by treatment and by center respectively.

3.2 Statistical models for survival data

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. In clinical trials, the investigators follow patients until they reach a pre-specified endpoint for example, death or tumor 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. In this Thesis, the focus is on semi parametric methods of survival analysis as discussed in the following sections.

3.2.1 Marginal Cox regression model

There are several types of models used in modeling survival data and the Cox's proportional hazards model proposed by Cox (1972) is the most popular. For this model, the hazard rate is expressed as

$$\lambda\left(t \mid \mathbf{x}\right) = \lambda_0\left(t\right)\exp\left(\mathbf{x}_j^T\boldsymbol{\beta}\right) \tag{3.1}$$

where $\lambda_0\left(t\right)$ is the baseline hazard function at time t, $\mathbf{x}_j^T$ is the vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of unknown regression coefficients. According to Collet (1994), this model is also referred to as a semi parametric proportional hazards model because no parametric form is imposed on $\lambda_0\left(t\right)$ (non-parametric part of the model) but assumes parametric form for the effect of the predictors on the hazard (parametric part of model). 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, not having to impose a parametric form on the

baseline hazard serves as a remedial virtue against misspecification (Keele, 2007). For this marginal model, 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.

3.2.1 Frailty models (Random effects models)

The frailty model is an extension of Cox's proportional hazards model allowing for a random effect. In this Thesis, the focus is on the multivariate frailty which assumes the center random effect operates at a group level. In this regard, a random effect is introduced for each center so that patients from one center are more alike than patients treated in different centers. The random effect describes the unobserved influences common to all patients of that particular center (Legrand *et al.*, 2006) and the variance of these random effects is a measure of the heterogeneity in the outcome between centers. The hazard rate for the j^{th} patient $\left(j=1,...,n_i\right)$ in the i^{th} center $\left(i=1,...,\mathrm{G}\right)$ is given by

$$\lambda_{ij}\left(t\right)=\lambda_0\left(t\right)\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}+w_i\right) \tag{3.2}$$

where $\lambda_0\left(t\right)$ is the unspecified baseline hazard at time t, $\mathbf{x}_{ij}^T$ is the vector of patient specific covariates and $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients (unknown parameters). w_i is the random effect for center i. Though the random effects $w_i's,\ i=1,...,\mathrm{G}$ are unobserved, it is assumed that they are independent and identically distributed from a density $f_W\left(\bullet\right)$. The corresponding frailty model can be re-written as follows

$$\begin{aligned}\lambda_{ij}\left(t\right)&=\lambda_0\left(t\right)\exp\left(w_i\right)\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}\right)\\&=\lambda_0\left(t\right)\mathrm{u}_i\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}\right)\end{aligned} \tag{3.3}$$

where $u_i=\exp\left(w_i\right)$ is known as the frailty and acts multiplicatively on the hazard rate for the j^{th} patient in the i^{th} center (Nguti, 2003; Duchateau and Janssen, 2008). For this model, regression coefficients are interpreted conditional on the center random effect.

3.3 Frailty distributions

In this Thesis, gamma and lognormal frailty distributions were considered. Heterogeneity and regression parameters were the parameters of interest.

3.3.1 Gamma distribution

Assuming that the frailties come from a one-parameter gamma density with mean 1 and variance γ, the frailty density is given by

$$f_U(u) = \frac{u^{1/\gamma - 1}\exp(-u/\gamma)}{\gamma^{1/\gamma}\Gamma(1/\gamma)} \tag{3.4}$$

The variance γ of the frailty term represents the heterogeneity among centers 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). 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 can be derived (see Duchateau and Janssen, 2008 for more details). 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).

3.3.2 Lognormal distribution

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

$$f_U(u) = \frac{1}{u\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\log u)^2}{2\sigma^2}\right) \qquad ,\gamma > 0 \tag{3. 5}$$

The mean and variance are expressed as

$$E(U) = \exp\left(\sigma^2/2\right)$$

$$Var(U) = \exp\left(\sigma^2\right)\left(\exp\left(\sigma^2\right)-1\right)$$

It is worth noting that mean and variance of the lognormal frailty density are both functions of 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 lognormal density, the heterogeneity parameter is typically described by $Var(w) = \sigma^2$, whereas for the gamma density it is often described by $Var(\mathrm{U}) = \gamma$ (Duchateau *et al.,* 2002). However, in this Thesis, θ was used to denote the heterogeneity parameter for both lognormal and gamma frailty models.

3.4 Estimation methods

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 information. However, the execution of the EM algorithm is computer intensive and slow. 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. 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, $\boldsymbol{\beta}$ and $\mathbf{w}$ (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. On the other hand, the outer loop of a gamma frailty distribution is based on the maximization of a profiled version of marginal likelihood and therefore, a REML estimate is not available (Duchateau *et al.,* 2002; Duchateau and Janssen, 2008). For gamma frailty model, PPL and EM algorithm lead to the same estimates. 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). PPL approach is preferred over

EM algorithm since it is faster and is implemented in most standard software. A detailed review of these estimation methods is covered by Duchateau *et al.* (2002) and Duchateau and Janssen (2008).

3.5 Heterogeneity parameter

In this Thesis, θ was estimated to get an idea on heterogeneity in the outcome between centers. When θ is large and differs significantly from zero; it reflects heterogeneity between centers and a strong association among patients in the same center. On the other hand, when θ is equal to zero, the frailties are identically equal to one which implies that the center 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 was 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).

3.6 Simulation study

A simulation study was undertaken to evaluate the performance and robustness of semi parametric frailty models with respect to the bias around the trues values of the treatment log hazard ($\hat{\beta}_1$) and the heterogeneity parameter ($\hat{\theta}$) estimate when the frailty distribution is misspecified. The simulated data was designed to reflect data obtained from the real clinical trial considering the PFS endpoint. The next section presents the details of the simulation study.

3.6.1 Simulation scheme

Assuming a fixed constant event rate λ_0, time to event outcome (survival time) for each patient was randomly generated from an exponential distribution expressed by

$$T_{ij} = -\frac{\log(U)}{\lambda_0 u_i \exp(\beta_1 * trt)} \qquad (3.6)$$

where U is a random variable following a uniform distribution in the interval [0,1] (Bender *et al.*, 2005). β_1 , the true treatment log hazard was estimated from the real trial data. 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 by Morden *et al.* (2011). Further, a follow-up period of 24 months was considered. Time at risk for a particular patient consisted of time at risk before the end of accrual period added to the follow-up time. A patient *j* in center *i* with time to event T_{ij} longer than time at risk was censored with time to censoring equal to time at risk such that $X_{ij} = \min\left(T_{ij}, C_{ij}\right)$ where C_{ij} is the censoring time independent of T_{ij} and $\delta_{ij} = I\left(C_{ij} > T_{ij}\right)$ is the censoring indicator as described by Duchateau *et al.* (2002). The frailties u_i were generated from three distributions. First, a one-parameter gamma with mean 1 and variance θ as discussed in section 3.4.1 was considered. The second frailty distribution considered was a transformation of lognormal distribution discussed in section 3.4.2. According to Duchateau and Janssen (2008), the transformed frailty density is expressed as

$$f_U(u) = \frac{1}{u\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\log u - \mu\right)^2}{2\sigma^2}\right) \quad , \gamma > 0 \tag{3.7}$$

Mean and variance are given by

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

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

Where $\mu = -\log(\theta + 1)/2$ and $\sigma^2 = \log\left(\theta + 1\right)$ 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 θ, x_1 and x_2 were obtained by solving the following set of constrained mean and variance equations

$$E(X) = \sum_{i=1}^{2} p(x_i) x_i = 1$$

$$Var(X) = \sum_{i=1}^{2} p(x_i)\left(x_i - E(X)\right)^2 = \theta$$

Though not a frailty distribution, the discrete distribution was considered so as to study the impact of extreme misspecification 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. 1000 datasets were generated for each parameter setting $(N, c_i, \lambda_0, \theta_i, \beta_1)$ where N is the fixed sample size and c is the number of centers under consideration. Frailty model (3.3) was fitted to the simulated data. The mean, median, per cent relative bias (RB%), standard deviation (SD) and the mean of the standard error (SE) were determined to describe the spread and the bias of $\hat{\beta}_1$ around β_1. The RB % and SD for $\hat{\beta}_1$ are respectively defined as

$$RB\% = \left| \frac{\bar{\beta}_1 - \beta_1}{\beta_1} \right| \times 100 \quad \text{and} \quad SD = \left\{ \sum_i \left(\hat{\beta}_1^{(i)} - \beta_1 \right)^2 \Big/ 999 \right\}^{1/2}$$

where $\bar{\beta}_1 = \sum_i \hat{\beta}_1^{(i)} \Big/ 1000$ is the mean of $\hat{\beta}_1^{(i)}$'s and $\hat{\beta}_1^{(i)}$ is the estimate of β_1 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 θ accordingly. To test for the need of center effect, 1000 data sets were again simulated under the null hypothesis i.e. assuming $\theta = 0$. Marginal log-likelihood estimates were obtained in all iterations and plotted against θ. Based on this plot, a 95% confidence interval for the heterogeneity parameter would be determined by taking two values of θ for which the marginal profile log-likelihood lies 1.92 units below the maximum profile likelihood (Morgan, 1992).

3.7 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. *Coxph* function in *Survival* Package was used to fit frailty models on simulated data. All simulations were conducted in R. For semi parametric frailty models in both SAS and R, the frailty distribution is restricted to one parameter gamma with mean 1 and lognormal frailty models with mean not equal to 1. Additionally, in both SAS and R, the default estimation method (PPL) was used. All statistical tests were conducted at 5% level of significance and 95% confidence intervals computed where necessary.

CHAPTER 4: RESULTS

4.1 Exploratory Data Analysis

A total of 450 patients were enrolled across 36 centers. However, 12 centers with less than 2 patients were dropped from the analysis to avoid estimation-related problems particularly when fitting frailty models. 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 B.1 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. Patients' baseline characteristics were well balanced across the treatment arms as observed in Table 4.1. 400 (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.

Table 4.1: *Patients baseline characteristics.*

	Treatment		**Total**
	A (n=212)	B (n=215)	(n=427)
	n (%)	n (%)	n (%)
Performance status			
0	122 (57.55)	116 (53.95)	238 (55.74)
1	90 (42.45)	99 (46.05)	189 (44.26)
Age group			
< 50 years	118 (55.66)	126 (58.60)	244 (57.14)
≥ 50 years	94 (44.34)	89 (41.40)	183 (42.86)
Histological type			
1 (Leiomyosarcoma)	49 (23.11)	55(25.58)	104 (24.36)
2 (Synovialsarcoma)	36 (16.98)	26 (12.09)	62 (14.52)
3 (Liposarcoma)	24 (11.32)	28 (13.02)	52 (12.18)
4 (Others)	103 (48.58)	106 (49.30)	209 (48.95)
Presence of Liver metastases at baseline			
0 (No)	179 (84.43)	179 (83.26)	358 (83.84)
1 (Yes)	33 (15.57)	36 (16.74)	69 (16.16)
Tumor grade			
2 (Intermediate)	102 (48.11)	106 (49.30)	208 (48.71)
3 (High)	110 (51.89)	109 (50.70)	219 (51.29)

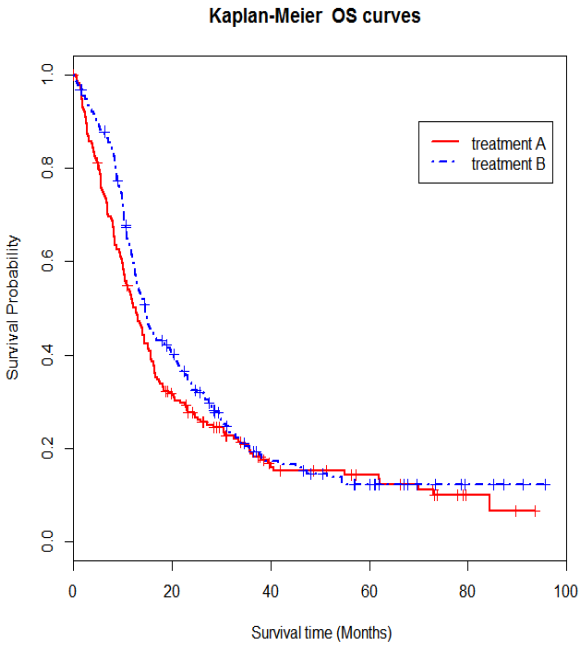

Figure 4.1: *Kaplan Meier OS curves by treatment.*

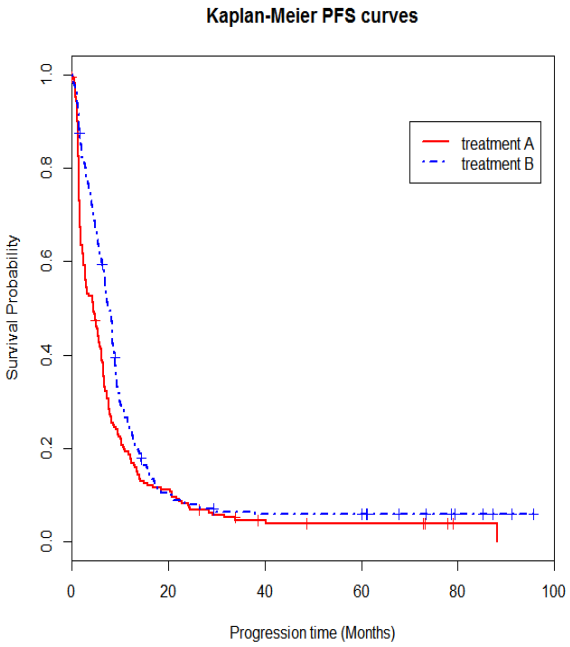

Figure 4. 2: *Kaplan Meier PFS curves by treatment.*

Figure 4.1 shows Kaplan–Meier OS curves by treatment. The estimated median OS time was 12.7 months with 95% confidence interval (CI) [10.4, 14.4] in treatment arm A and 14.3 months in treatment arm B with a 95% CI [12.7, 16.8]. Similarly, as observed from Figure 4.2, 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.

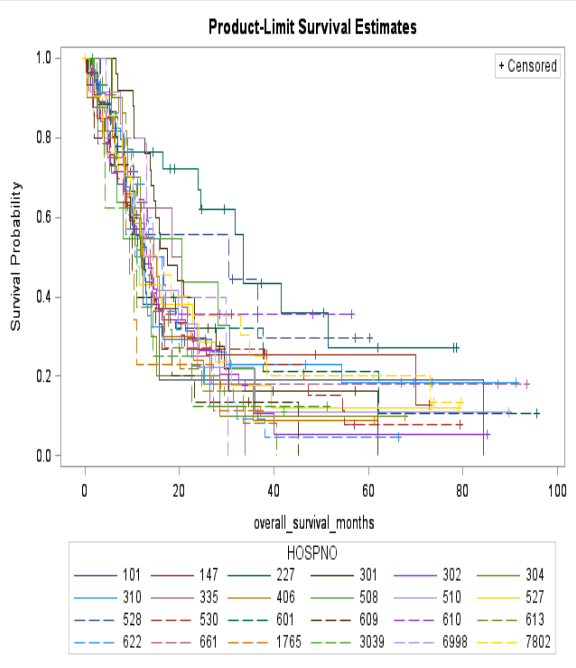

Figure 4.3: *Kaplan Meier curves for OS Stratified by centers.*

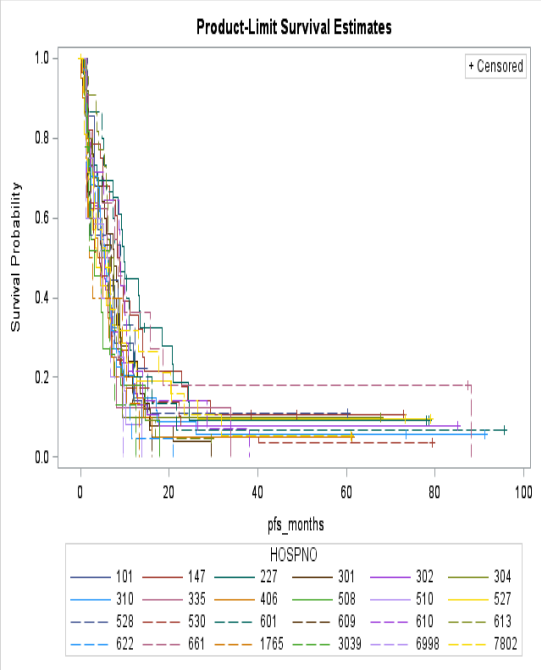

Figure 4.4: *Kaplan Meier curves for PFS stratified by centers*.

Figure 4.3 presents Kaplan–Meier OS curves stratified by centers. From the plot, there seems to be variability in the outcome between centers. Similar observations were made from PFS curves stratified by centers presented in Figure 4.4. However, due to the large number of curves corresponding to the centers, it was difficult to interpret such plots. Moreover, the precision with which the curves are estimated depends on the number of events observed in each center which should be taken into account as suggested by Legrand *et al.* (2006). Based on a classical log rank test, the P-values were 0.711 and 0.344 for OS and PFS endpoints respectively. Hence, there was no evidence to reject the null hypothesis. This suggests that there may not be substantial heterogeneity in outcome among the centers for the two endpoints.

4.2 Statistical Analysis

4.2.1 Marginal and shared frailty models

Table 4.2: *Overall survival Hazard Ratio (95% C I) for 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: 2	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 ≥ 50	1.157	(0.925, 1.447)	1.156	(0.924, 1.448)	1.158	(0.927, 1.447)

Table 4.2 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; therefore, there is a significant difference between the treatment arms. All other parameter estimates can be interpreted in a similar manner.

The similarity between the marginal and shared frailty models could be further attributed to the fact that for 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 (results omitted). 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; we therefore concluded that events were independent within and across centers for both frailty models.

Table 4.3: *PFS Hazard Ratio (95% C I) for frailty and marginal Cox models.*

	Gamma frailty		**Lognormal frailty**		**Marginal Cox model**	
Variable	**HR**	**(95 % CI)**	**HR**	**(95 % CI)**	**HR**	**(95 % CI)**
Treatment: B	0.675	(0.551, 0.826)	0.673	(0.550, 0.823)	0.686	(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 4.3 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, as observed in Table B.3 in the Appendix, 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 (results omitted). 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 B.2 and B.3 in the Appendix) 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 θ 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 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.

4.3 Simulation Results

4.3.1 Comparison of gamma and lognormal generated frailties

Figure 4.5 presents the histograms of generated frailties under gamma and lognormal frailty densities with mean 1 and variances (θ) 0.2 and 2 respectively over 1000 iterations. It is observed that for a given variance, the two distributions have approximately similar shapes but deviate from each other. 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 θ=2, the range of lognormal was two times the range of gamma frailties (Table B.4 in the Appendix). When the number of centers was increased to 25, a similar trend was observed but the range reduced accordingly for each θ.

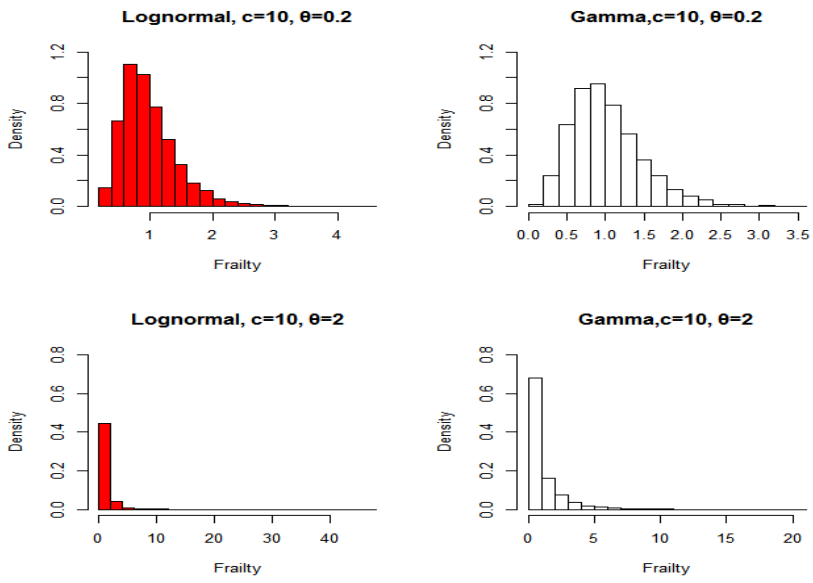

Figure 4.5: *Lognormal and gamma distributed frailties over 1000 iterations with mean 1 and variance 0.2 and 2 respectively*.

4.3.2 Regression coefficient

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 effect (log hazard) $\beta_1 = -0.353$ was estimated from a proportional hazard model with treatment as the only covariate. Allocation of each patient to receive either treatment A or B was generated from a binomial distribution with success probability 0.5 while a constant event rate, $\lambda_0 = 0.180$ was chosen such that approximately 8 % of patients were censored. 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 *Coxph* function has a mean $E\left(\hat{U}\right) \neq 1$ while the mean of generated lognormal frailties was constrained to 1. Moreover, as noted earlier, the mean and variance of lognormal distribution are linked. This difference led to inflated bias.

Table 4.4: *Simulation results for true β_1 = -0.353 from correctly specified gamma and misspecified 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: gamma						
Fitted: gamma frailty						
Mean	-0.356	-0.355	-0.352	-0.353	-0.356	-0.352
median	-0.354	-0.352	-0.352	-0.350	-0.354	-0.348
RB %	0.943	0.529	0.151	0.063	0.682	0.181
SD	0.097	0.102	0.114	0.099	0.102	0.114
SE	0.097	0.098	0.112	0.097	0.100	0.115
True frailty distribution: lognormal						
Fitted : gamma frailty						
Mean	-0.357	-0.358	-0.356	-0.356	-0.357	-0.356
median	-0.354	-0.358	0.356	-0.357	-0.356	-0.346
RB %	1.383	1.587	1.060	0.722	1.111	1.094
SD	0.099	0.100	0.090	0.099	0.101	0.103
SE	0.097	0.098	0.102	0.097	0.100	0.104
True frailty distribution: Discrete						
Fitted : gamma frailty						
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

Table 4.4 presents the simulation results for the estimated $\hat{\beta}_1$ obtained from correctly specified gamma frailty model (gamma frailty model fitted to clustered data generated from a gamma distribution). Generally, the mean and median of the estimated $\hat{\beta}_1$ were close to true β_1 with a 0.06 % to 0.9 % bias range. 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 θ =0.02 where a decrease was observed. Considering a10 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 θ. This agreement between SE and SD showed that the estimated SE well estimated.

To investigate sensitivity to misspecification for the frailty distribution, a gamma frailty model was first fitted to clustered data generated from a lognormal distribution. From the results presented in Table 4.4, it was observed that the mean and median of the estimated $\hat{\beta}_1$ were very similar. In general, the RB % ranged between 0.7 % and 1.58 %. A cross the two center settings, the RB % corresponding to θ=0.02 and θ=0.2 decreased when the centers increased from 10 to 25. On the other hand, a slight decrease was observed for true θ=2. Furthermore, the bias from this misspecified model and the correctly specified model did not vary substantially. This suggests robustness of the gamma frailty with respect to lognormal distribution (Duchateau and Janssen, 2008).

When the gamma frailty model was fitted to the data generated from a discrete distribution, the estimated β_1 was somewhat sensitive to misspecification as observed in Table 4.4. Generally, the bias range was between 1.24 % and 17.46%. Within a particular center scenario i.e. either 10 or 25, the RB% increased with an increase in size of the true heterogeneity parameter. 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. Is should be noticed that under misspecified models, the SE and SD tend to be smaller compared to correctly specified model. This could be due to downward bias observed particularly for discrete frailty producing estimated that were too optimistic.

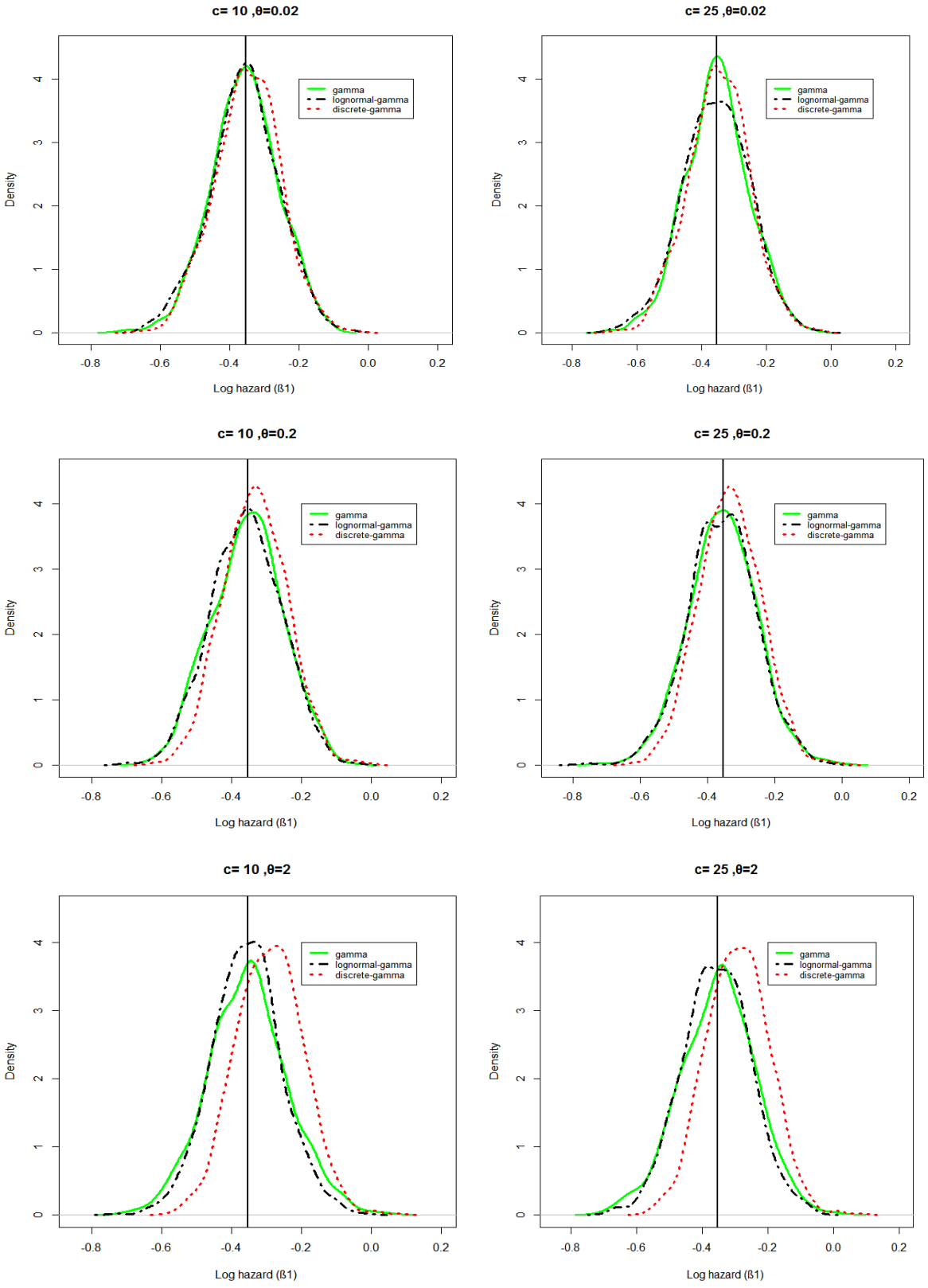

Figure 4.6: *Density curves of the estimated* β_1 *over different simulation settings.*

Figure 4.6 presents density plots of the estimated β_1 under different simulation settings with respect to number of centers and true heterogeneity parameters. The vertical line represents β_1 = -0.353. It is observed that for θ=0.02, the estimated $\hat{\beta}_1$ from different models were very close irrespective of the number of centers under consideration. However, differences among the fitted models and deviations from the true treatment log hazard effect were more pronounced for larger θ i.e. θ=0.2 and θ = 2 respectively. These observations are consistent with results discussed in Table 4.4 above.

4.3.3 Heterogeneity parameter

From Table 4.5, simulation results for estimated $\hat{\theta}$ obtained under correctly specified gamma frailty model are presented. It was observed that the RB% range was between 0.2 % and 7.26 %. 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. The marginal profile log likelihood plot for θ (Figure B.1 in the Appendix) suggested no center effect. Besides the 95% confidence interval range contained the value 0.

Similarly, sensitivity to misspecification of the frailty distribution was assessed with respect to the estimated θ. From Table 4.5, moderate to high RB% was observed for each of the assumed true θ. The RB% was much higher for the two misspecified 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% range between 64% 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 10.03% and 54.05 %. For θ =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 θ=0.02 and θ=0.2, the range of frailties for the two distributions were close whereas for θ=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 misspecified 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 heterogeneity parameter.

Table 4.5: *Simulation results for θ from correctly specified gamma and misspecified 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: gamma						
Fitted: gamma frailty						
Mean	0.019	0.179	2.005	0.019	0.201	1.988
median	0.002	0.160	1.984	0.009	0.187	1.944
RB %	7.265	6.513	0.266	4.681	0.473	0.570
SD	0.029	0.068	0.308	0.021	0.089	0.507
True frailty distribution: lognormal						
Fitted: gamma frailty						
Mean	0.018	0.266	0.969	0.017	0.163	0.919
median	0.013	0.244	0.941	0.009	0.154	0.908
RB %	11.89	32.83	51.56	10.03	18.664	54.05
SD	0.018	0.168	0.280	0.019	0.068	0.218
True frailty distribution: Discrete						
Fitted: gamma frailty						
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

CHAPTER 5: DISCUSSION AND CONCLUSION

5.1 Discussion

The primary objective of this Thesis was to investigate the impact of institution variability on patient outcome in a soft tissue sarcoma clinical trial. In the analysis attention was restricted to centers with at least 5 patients for avoid estimation-related problem in statistical analysis. The patients' baseline characteristics were well balanced between the treatment groups as expected.

In statistical analysis, parameter estimates (HR) and corresponding 95% confidence intervals from the frailty models and marginal Cox regression model were close particularly for the OS. This similarity was attributed to the fact that none of the center random effect was significant. Moreover, the corresponding heterogeneity parameter estimates from gamma and lognormal frailty models were very small and insignificant hence, no sufficient evidence to reject the null hypothesis of no center effect. Therefore, we concluded that events were independent within and across centers for both endpoints. The formal test of center effect was based on the mixture of chi-square likelihood ratio test with 0 and 1degrees of freedom. This is because the null hypothesis for θ was at the boundary of the parametric space and using a chi-square with one degree of freedom would be inappropriate.

A simulation study was conducted with an aim to investigate the impact of frailty distribution misspecification 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 misspecification of the frailty distribution, two scenarios were considered. First, a gamma frailty model was fitted to clustered data generated under lognormal distribution (misspecified 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 misspecification did not greatly affect the regression coefficient estimate despite the fact that different frailty distributions can lead to noticeably different association structures. From a previous study examining the gamma frailty model in multicenter clinical trial (cohort study), Glidden and Vittinghoff (2004) found by simulation that regression coefficient estimates were minimally affected by frailty misspecification similar to the findings here. However, their assumed true frailty distribution was inverse Gaussian.
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 misspecified 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 misspecification. The bias was much larger compared to other fitted models particularly for large 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 extreme and inappropriate.

5.2 Conclusion

With no sufficient evidence to reject the null hypothesis of homogeneity between centers in frailty models, we concluded that events were independent within and across centers in this study. Furthermore, parameter estimates from frailty models considered were almost identical to those of a marginal model. Therefore, in this particular study, either of the models could be used for statistical inference. However, this may not hold in other studies and 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.

From simulation study results, assuming a gamma frailty distribution when the underlying frailty is lognormal, the regression coefficient (treatment effect) estimate was minimally affected in terms of relative bias. In this case, we concluded that gamma frailty distribution was more robust with respect to lognormal distribution compared to discrete distribution. On the hand, for 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. Overall, heterogeneity parameter was more sensitive to misspecification of the frailty distribution and choice of initial parameters compared to regression parameter estimate.

5.3 Limitations and Recommendations

Our study was limited to investigating only the center random effect since the software used did not allow for more than one random effect. This limitation has been previously been acknowledged by Glidden and Vittinghoff, (2004) 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 frailty 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.

REFERENCES

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

Cance.net. (2013). Sarcoma. Accessed on 10/08/2013.Available at : http://www.cancer.net/cancer-types/sarcoma.

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.

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.

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.*

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.

Ha, I.D., Sylvester, R., Legrand, C., and MacKenzie, G. (2011). Frailty Modelling for Survival Data from Multi-Center 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.

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

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 mis-specified. *Statist. Med*. 26: 4657– 4678.

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.

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

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

APPENDIX A

Selected Codes

R codes

```
#overall survival curves by treatment
library(survival)

fit.hosp=survfit(Surv(osurv,censur)~as.factor(Trt1), data=sarcoma,type="kaplan-
meier",conf.type="none")
plot(fit.hosp, conf=T,main="",xlab=" Overall Survival (Months) ",ylab="Survival
Probability",lty=c(1,4),col=c(2,4),lwd=2.5,xlim=c(0,100))
legend(70,0.9,c("treatment A","treatment B"),lty=c(1,4),lwd=2.5,col=c(2,4))

#pfs survival curves by treatment
fit.hosp=survfit(Surv(pfsy,cenpfs)~as.factor(Trt1), data=sarcoma,type="kaplan-
meier",conf.type="none")
plot(fit.hosp, conf=T,main="",xlab=" PFS (Months) ",ylab="Survival
Probability",lty=c(1,4),col=c(2,4),lwd=2.5,xlim=c(0,100))
legend(70,0.9,c("treatment A","treatment B"),lty=c(1,4),lwd=2.5,col=c(2,4))
```

SAS codes

```
/*survival curves by hospital*/
 proc lifetest data=work.proj;
 time  timesur*censur(1);
 strata hospno ;
  run;quit;

/*gamma frailty  frailty*/
PROC PHREG DATA=work.proj plots(cl)=survival;
class hisloc trt1 (param=ref ref="1") tumorgrd  livermeta perfmsts agecat(param=ref ref="1")
hospno;
MODEL timesur*censur(1)= trt1 hisloc tumorgrd  livermeta perfmsts agecat /rl;
random hospno/ DIST=gamma solution;
RUN;

/*lognormal   frailty*/
   goptions reset=all;
ods graphics on;
ods rtf file="F:\uhasselt year2\Second sem\THESIS\draft\lognorm.rtf";
PROC PHREG DATA=work.proj plots(cl)=survival;
class hisloc trt1 (param=ref ref="1") tumorgrd  livermeta perfmsts (param=ref ref="1") hospno;
model timesur*censur(1)= trt1 hisloc tumorgrd  livermeta perfmsts agecat/rl;
random hospno/dist=LOGNORMAL solution;
RUN;
```

APPENDIX B

Additional Analysis Results

Table B.1: *Distribution of patients in centers by treatment received*

	treatment		
Center No.	**A n (%)**	**B 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 B.2: *Parameter Estimates(SE) for Overall survival*

	Gamma frailty model	**Lognormal frailty model**	**Marginal effects**
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 B.3: *Parameter Estimates (SE) for PFS*

	Gamma frailty model	**Lognormal frailty model**	**Marginal model**
parameter	Estimate (SE)	Estimate (SE)	Estimate (SE)
Treatment B	-0.394 (0.103)	-0.396 (0.103)	-0.377 (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 B.4: *Summary statistics for the generated frailties under gamma and lognormal distributions over 1000 iterations. (25000 observations)*

	10 centers											
	0.02				**0.2**				**2**			
	mean	**var**	**min**	**max**	**mean**	**var**	**min**	**max**	**mean**	**var**	**min**	**max**
Lognormal	0.99	0.0199	0.54	1.8	0.99	0.194	0.22	4.7	0.97	1.86	0.007	47.43
Gamma	0.99	0.0196	0.54	1.54	1.01	0.205	0.10	3.46	0.99	2.00	0.00	20.69
	25 centers											
	0.02				**0.2**				**2**			
	mean	**var**	**min**	**max**	**mean**	**var**	**min**	**max**	**mean**	**var**	**min**	**max**
Lognormal	0.99	0.02	0.59	1.82	0.99	0.197	0.17	4.72	1.01	2.10	0.01	32.87
Gamma	0.99	0.02	0.49	1.63	1.00	0.19	0.04	3.87	1.00	2.03	0.00	18.94

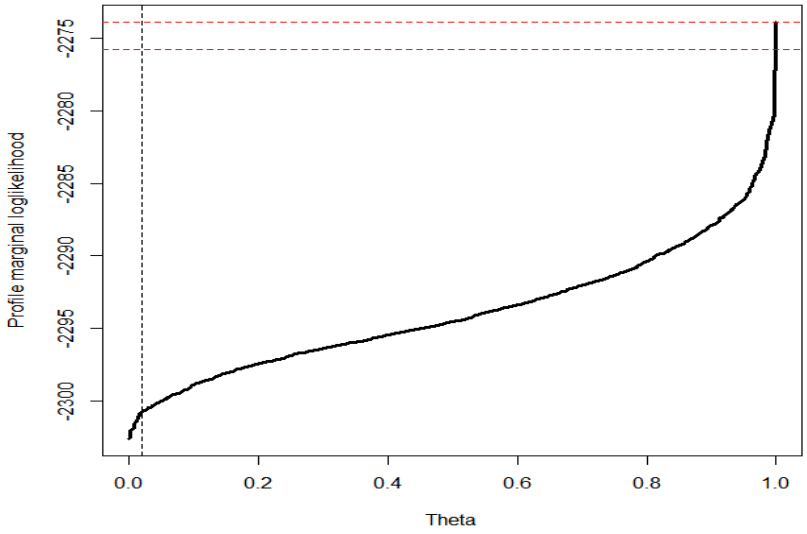

Figure B.1: *Profile marginal likelihood for the heterogeneity parameter with a horizontal lines at maximum and 1.92 units below.*

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
The impact of institution variability on patients outcome in a soft tissue sarcoma clinical trial

Richting: **Master of Statistics-Biostatistics**
Jaar: **2013**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Gachau, Susan

Datum: **11/09/2013**