SEQUENTIAL ANALYSIS OF CLUSTERED SURVIVAL DATA BY MARGINAL METHODS

BY BO HOU

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics

Written under the direction of

Professor Zhiliang Ying

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2008

ABSTRACT OF THE DISSERTATION

Sequential Analysis of Clustered Survival Data by Marginal Methods

by Bo Hou

Dissertation Director: Professor Zhiliang Ying

Clustered survival data are a type of multivariate survival data with naturally formed clusters so that event times within a cluster are parallel to each other and correlated. Lee, Wei and Amato (1992) introduced a semiparametric model for the analysis of clustered survival data that assumes event times follow a proportional hazards model.

Sequential analysis of clustered survival data arises in clinical studies in which patients are followed over time and interim analyses are performed. This thesis studies the sequential analysis of clustered survival data with staggered patient entry by adapting Lee, Wei and Amato's approach. It is shown that the two-parameter score process converges to a Gaussian random field irrespective of the correlation within clusters and staggered patient entry. The regression parameter estimator obtained at each time point has the desired properties including consistency and asymptotic normality. A consistent estimator of the baseline cumulative hazard function is also given. More importantly, we propose a novel optimal weighting strategy. We show that the resulting score process not only produces a more efficient estimator, but also has the important property of (asymptotically) independent increments. The latter can be used in conjunction with the well-known error-spending functions to construct proper boundaries in group sequential testing. Finally a sample size calculation formula is given for

designing clinical trials with clustered survival time as the endpoint.

Acknowledgements

I would like to express my deepest gratitude to my advisor Professor Zhiliang Ying for his thoughtful and insightful directions through years. Without his continuous encouragement and support, this thesis would not be possible.

I would like to thank colleagues and students at University of Science and Technology of China, Professor Lincheng Zhao, Dr. Wenquan Cui, Han Zhang, Zhanfeng Wang. I would especially like to thank Dr. Yaning Yang, whose assistance and encouragement have been generous and invaluable.

I would like to thank Professor Harold Sackrowitz, Professor Kesar Singh, Professor David Tyler, Professor Cunhui Zhang and Professor John Kolassa who have been very kind to me during my study at Rutgers and after I left Rutgers. Without their support, this thesis would not be possible. I would also like to thank Angela Klein for her support all these years.

I also would like to thank the members of my committee, Professor Harold Sackrowitz, Professor Cunhui Zhang, Professor Minge Xie and Prof. Mengling Liu for their kind advice and helpful suggestions.

Finally I would like to thank my husband, Sen Hu, and my children, Maomao and Mimi, for their understanding and support.

Dedication

This document is dedicated to my parents

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The clustered survival data represent such a type of multivariate survival data that there are naturally formed clusters and, within each cluster, multi-parallel event times are observed. The event times within each cluster may be correlated due to the nature of clusters.

1.1 Examples of clustered survival data

A typical example of clustered survival data is as follows. In a clinical trial experiment, a medicine to help healing of wounds caused by type II diabetic mellitus is compared with a standard wound care medicine. Instead of following one major wound per subject, the trial tracks all eligible wounds of every subject until their healing or the end of the study. The main focus of the trial is to compare the effect of the new medicine with that of the standard wound care medicine on the healing time of wounds. Furthermore, identifying factors that highly influence the healing of wounds is another major interest. In this example each subject forms a cluster. There may be two or more event times within each cluster if the subject has experienced more than one wound at the beginning of the study or during the study. These event times are parallel to each other instead of having natural time order between them. Another characteristic of the trial is that every subject receives only one randomly assigned treatment for all his wounds, it is not intended to compare the treatment effects within each cluster.

More examples of clustered survival data can be found in literature. In [25], Lee, Wei and Amato presented an example of a diabetic retinopathy study, in which, patient visual loss was studied. Patients took either oral sorbinil, or placebo, and, the time from the admission of medicine to visual loss was observed for each patient, one for

each eye. Here again each patient forms a cluster.

More generally, survival data from clinical trials with subjects recruited from multiple study centers can also be viewed as clustered survival data, with each center being a cluster. Group life insurance of employees sharing a common environment at their workplace, survival studies about married couples can all be viewed as clustered survival data.

1.2 Modeling of clustered survival data

As a type of common and important data, clustered survival data have been studied by many researchers. A major difficulty in analyzing clustered survival data is the modeling of the within cluster dependency. Two major methodologies regarding modeling of clustered survival data exist. One is based on the introduction of a frailty to each cluster, known as frailty model or mixed effects model. Another method is marginal modeling approach which models the marginal distribution of each event time.

The frailty model specifies the within cluster dependency directly by assuming that event times within each cluster share a common, cluster-specific risk. It is further assumed that this common risk is the only factor governing the dependency: given this risk, the event times in a cluster are independent. This common risk is called the frailty of this cluster. Although we assume that frailties exist, they are at most times unobservable. So it is usually assumed that frailties are i.i.d random variables, and are independent of their corresponding covariates. Let ξ_i represent the frailty for cluster i. Then given ξ_i, for a proportional hazards model, the hazard function $\lambda_{ij}(t|\xi_i)$ for the jth event time in cluster i is assumed to be:

$$\lambda_{ij}(t|\ \xi_i) = \xi_i \lambda_0(t) \exp(\beta' Z_{ij}), \quad t \geq 0, \tag{1.1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function. In the setting of frailty model, the random effects are assumed to have a known distribution. Vaupel et al. [39] studied the frailty model for univariate case and gave the model the name. Clayton [5] and Oakes [29] studied the bivariate model for frailty having gamma distribution.

Clayton and Cuzick [4] generalized the above result to include covariates. Hougaard [18], Hougaard [19] and Oakes [30] studied the bivariate case with frailty stable distributed. Crowder [6] studied the PVF model. Whitmore and Lee [40] suggested the multivariate model with inverse Gaussian frailty and exponential conditional distributions. Lu and Bhattacharyya [23] presented several stable-Weibull model extensions.

The marginal proportional hazards model for survival data was introduced by Wei, Lin and Weissfeld [41] in 1989 for multivariate survival data that are naturally ordered. In 1992 Lee, Wei and Amato [25] introduced the marginal model to the clustered survival data that all events are of the same type. This model was further applied by Spiekerman and Lin [37] in 1998 to clustered survival data that include different types of event times as well as multi-observations of each type. The marginal modeling method models the marginal distribution of the survival times by assuming that the marginal distributions of the survival times follow proportional hazards model. That is, given $Z_{ij} = z_{ij}$, the marginal method assumes that the hazard function of T_{ij} is of the form

$$\lambda_{ij}(t \mid z_{ij}) \;=\; \lambda_0(t)e^{\beta' z_{ij}},$$

where $\lambda_0(t)$ is an unspecified baseline hazard function. In this modeling, no specification is done on the dependence structure of T_{ij} within clusters. The partial likelihood is computed by assuming that all $\{T_{ij}\}$ are independent, and the regression parameter estimator is the maximum partial likelihood estimator. So the regression parameter estimator is obtained as if all $\{T_{ij}\}$ are independent. The dependence between $\{T_{ij}\}$ is dealt with by adjusting the covariance matrix of the maximum partial likelihood estimator to include the correlation. It has been shown that thus obtained regression parameter estimator and the estimator of the covariance of the regression parameter vector are all consistent estimator of the corresponding parameter.

The major difference between the marginal proportional hazard model and the frailty model is that the additional term of frailty is included to the hazard function in the frailty model. The frailty term describes the cluster effect, and changes the parameter estimation procedure and the interpretation of the results. In a frailty model, the regression parameters represent the effects of covariate difference between two members

of the same cluster, while in the marginal proportional hazard model, they represent effects of covariate difference between two members in the whole population.

1.3 Group sequential analysis for clustered survival data

One common feature for controlled clinical trials, at least for Phase III clinical trials, is to monitor the data periodically. That is, data up to some intermediate time points are analyzed. The major method to accomplish these interim analyses is the method of group sequential analysis.

A group sequential test is basically a sequence of repeated significance tests. It requires the knowledge of the joint distribution of the test statistics so that the critical values of the tests can be determined and the overall type I error can be well controlled. According to these requirements there are various ways to decide the test boundaries and the type I error allowed at each separate analysis.

Pocock (1977) [32] and O'Brian and Fleming (1979) [31] have made major impacts on this field. These two papers introduce group sequential two-sided tests that are easy to implement and can be applied to many response distributions. Pocock's test suggested to use a constant critical value for all analyses. O'Brian and Fleming proposed a test in which the critical values decrease as the study proceeds.

In 1982, Slud and Wei [36] introduced a method which, theoretically, can be applied to all situations that the joint distribution of the sequence of test statistics is known. Assume that the interim analyses are to be performed at time points $t_1 < t_2 < \cdots < t_K$ and the pre-specified significance levels are $\alpha_1, \cdots, \alpha_K$ satisfying $\sum_{i=1}^{K} \alpha_i = \alpha$. Let $W_{t_1}, \cdots, W_{t_K}$ be the test statistics with known joint distribution, then Slud and Wei suggested that at time t_l, the boundary point d_l is determined as follows:

$$\begin{aligned} & P\left(|W_{t_1}| \geq d_1\right) = \alpha_1 \\ & P\left(|W_{t_1}| \leq d_1, |W_{t_2}| \geq d_2\right) = \alpha_2 \\ & \vdots \\ & P\left(|W_{t_1}| \leq d_1, |W_{t_2}| \leq d_2, ..., |W_{t_{K-1}}| \leq d_{K-1}, |W_{t_K}| \leq d_K\right) = \alpha_K. \end{aligned} \tag{1.2}$$

Due to the minimum requirement, the Slud and Wei's method has been applied widely

in practice. But, on the other hand, when the joint distribution is complicated the method is very computing-intensive.

Slud and Wei's method requires that the maximum number of analyses and the significance levels at all separate analyses are fixed in advance. However in many studies, especially in survival studies, the increments of information are usually unequal and unpredictable, so we may prefer to adjust the number of analyses or the Type I error spent at analyses according to the actual accrual of information. Lan and DeMets [22] introduced the error-spending method which allows the flexibility of choosing the number of analyses and significance levels. Assume that a clinical trial is conducted on $[0, t]$. Let $f(t)$ be a pre-specified non-decreasing function which satisfies that $f(0) = 0$ and $f(t) = \alpha$. This $f(t)$ specifies the cumulative Type I error spent at time t. Suppose that the first interim analysis is done at $t_1 \in [0, t]$, then the Type I error spent at t_1 is $f(t_1)$. The second analysis at t_2 will spend Type I error $f(t_2) - f(t_1)$ and so on. Once the significance levels are determined, the boundaries of analyses are calculated by Slud and Wei's method.

In this thesis, we will study clustered survival data with covariates and staggered patient entry. Researches on survival data with covariates and staggered patient entry have been done by Tsiatis [38], Sellke and Siegmund [35] and Billias, Gu and Ying [2]. Tsiatis in 1981 studied the distributional properties of the score function $U(\beta, t)$ derived from a Cox partial likelihood function under the null hypothesis $\beta = 0$. He proved that $n^{-1/2}U(0, t)$ converges to a limiting Gaussian process with independent increments. Hence the joint distribution of standardized $n^{-1/2}U(0, t)$ at different time points is the same as the distribution of a sequence of normalized partial sums of independent and identically distributed standard normal random variables. Therefore standard results for repeated significance testing of a Brownian motion can be used. Sellke and Siegmund [35] suggested a transformation of time scale. Under the new time scale, they showed that the score process converges weakly to a time scaled Brownian motion. Billias, Gu and Ying [2] studied the distribution properties of the two-parameter score function $U(\beta, t, s)$ with t representing calender time and s survival time. Their results can be applied to obtain distributional approximation to various testing statistics, under both

the null an alternative hypotheses.

1.4 Lee, Wei and Amato's model

Our thesis is based on Lee, Wei and Amato [25] model for clustered survival data and is abbreviated as LWA model throughout this thesis. To be specific, suppose there are n clusters with k_i event times in cluster i. Let T_{ij} be the jth event time of the ith cluster associated with the $p \times 1$ covariate vector Z_{ij} and the noncensoring indicator δ_{ij}. The LWA model is then as follows. Given $Z_{ij} = z_{ij}$, the marginal hazard function of T_{ij} is of the form

$$\lambda_{ij}(t \mid z_{ij}) \;=\; \lambda_0(t)e^{\beta' z_{ij}},$$

where $\lambda_0(t)$ is an unspecified baseline hazard function.

This model gives only the marginal distribution of each T_{ij}. To make inference for β by Cox's partial likelihood, Lee, Wei and Amato ignored the correlation between T_{ij} by adopting the working assumption that T_{ij} are independent. Under the working assumption, the Cox's partial likelihood

$$L(\beta) = \prod_{i=1}^{n}\prod_{j=1}^{k_i}\left(\frac{\exp\{\beta' Z_{ij}\}}{\sum_{(l,m)\in R(i,j)}\exp\{\beta' Z_{lm}\}}\right)^{\delta_{ij}} \tag{1.3}$$

can be calculated. Here $R(i,j)$ is the risk set for jth member in cluster i. The score function

$$U(\beta) = \frac{\partial log L(\beta)}{\partial \beta} \tag{1.4}$$

has the representation

$$U(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^\infty (Z_{ij} - \bar{Z}(\beta,t))dM_{ij}(\beta,t) \tag{1.5}$$

with

$$M_{ij}(\beta,t) = N_{ij}(t) - \int_0^t Y_{ij}(u)\lambda_0(u)\exp(\beta' Z_{ij})du.$$

This score process $U(\beta)$ is no longer a martingale due to the staggered patient entry and correlation between survival times within clusters. The usual martingale method can not be applied here. Lee, Wei and Amato showed that, at the true regression

parameters vector β_0, $n^{-1/2}U(\beta_0)$ is asymptotically equivalent to $n^{-1/2}\tilde{U}(\beta_0)$ with

$$\tilde{U}(\beta_0) = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^{\infty}(Z_{ij} - \mu(\beta_0, t))dM_{ij}(\beta_0, t). \tag{1.6}$$

Here $\mu(\beta_0, t)$ is the the limit of $\bar{Z}(\beta_0, t)$ as the sample size tends to infinity. Let

$$U_i(\beta) = \sum_{j=1}^{k_i}\int_0^{\infty}(Z_{ij} - \mu(\beta, t))dM_{ij}(\beta, t).$$

Since $\mu(\beta, t)$ is a deterministic function, the replacement of $\bar{Z}(\beta, t)$ by $\mu(\beta, t)$ enables $U_1(\beta), \cdots, U_n(\beta)$ are independent. More regularity conditions guarantee that $U_1(\beta_0), \cdots, U_n(\beta_0)$ are independent and identically distributed. By multivariate central limit theorem, $n^{-1/2}U(\beta_0)$ is normal distributed with the limit variance-covariance matrix

$$\Sigma(\beta_0) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E(\sum_{j=1}^{k_i}\int_0^{\infty}(Z_{ij} - \mu(\beta_0, t))dM_{ij}(\beta_0, t))^{\otimes 2}. \tag{1.7}$$

Let $\hat{\beta}$ be the solution to the equation $U(\beta) = 0$. Although $\hat{\beta}$ is calculated under the assumption that T_{ij} are independent, Lee, Wei and Amato has showed that under regularity conditions, $\hat{\beta}$ is asymptotically normally distributed and a consistent estimator of β_0.

From the consistency of $\hat{\beta}$, Lee, Wei and Amato proposed the following consistent estimator of $\Sigma(\beta_0)$:

$$\hat{\Sigma} = n^{-1}\sum_{i=1}^{n}\sum_{j,k}^{k_i}(\int_0^{\infty}(Z_{ij} - \mu(\hat{\beta}, s))d\hat{M}_{ij}(\hat{\beta}, s))(\int_0^{\infty}(Z_{ik} - \mu(\hat{\beta}, s))d\hat{M}_{ik}(\hat{\beta}, s))^{\top}$$

where

$$\hat{M}_{ij}(\beta, s) = N_{ij}(s) - \int_0^{\infty}\frac{Y_{ij}(s)}{\sum_{i=1}^{n}\sum_{j=1}^{k_i}Y_{ij}(s)\exp(\hat{\beta}'Z_{ij})}d\sum_{i=1}^{n}\sum_{j=1}^{k_i}N_{ij}(s).$$

The limiting covariance matrix $W(\beta_0)$ of $\hat{\beta}$ is calculated by the so-called sandwiched matrix

$$W(\beta_0) = \lim_{n\to\infty} nVar(\hat{\beta}) = A^{-1}\Sigma(\beta_0)(A^{-1})^{\top}, \tag{1.8}$$

where

$$A = \lim_{n\to\infty} n^{-1}\frac{\partial^2 logL(\beta^*)}{\partial^2\beta},$$

and β^* is between $\hat{\beta}$ and β_0.

1.5 Thesis outline

This thesis studies issues related with sequential analysis of clustered survival data, based on the marginal proportional hazards model proposed by Lee, Wei and Amato [25]. Based on the working assumption that all survival times are independent and assume that they obey the marginal proportional hazards model, we obtain the score process $U(\beta, t, s)$ for $0 \leq s \leq t \leq \tau$ with t representing calender time, s survival time and τ a boundary of t to satisfy stability conditions. An estimate of the regression coefficient vector in the marginal model can be obtained from the following estimating equation

$$U(\beta, t, t) = 0$$

for any fixed $t \in [0, \tau]$. The score process $U(\beta, t, s)$ is no longer a martingale with respect to survival time s due to the staggered patient entry and correlations between survival times within clusters. More sophisticated method is needed in order to obtain the asymptotic distribution of the score process and its limiting covariance matrix function.

In Chapter 2, we investigate the asymptotic distribution properties of the score process $U(\beta, t, s)$. We will show that under regularity conditions, $U(\beta, t, s)$ converges to a Gaussian random field with mean vector 0, continuous sample paths and explicit covariance function. With the asymptotic distribution of the score process known we show in section 3 that, for every fixed t, the solution to the estimating equation is a consistent estimator of the true regression vector. In section 4, we propose a consistent estimate of the cumulative hazard function $\Lambda_0(t)$.

Chapter 3 investigates group sequential analysis of clustered survival data. In section 1 we show that, using Slud and Wei's method, group sequential analysis can be done based on sequence of standardized estimator $\hat{\beta}_t$ calculated on chosen analysis time points. In section 2, we introduce a partition method which produces estimators of regression parameters that are more efficient than LWA estimators. The main idea behind the approach is to break the score function into small blocks by chopping the real line into pieces. The covariance matrix for these blocks can be estimated using

a method similar to that of Wei *et al.* [41]. Then they are summed up reweighting by the product of their second derivative and covariance matrix to form a new score. We will show that this new score has the nice property of independent increments in its limit. So this score process can be regarded as a time rescaled Brownian motion , for which standard group sequential procedures are readily applicable. In section 4, extensive simulations are done on the independent increments property, on the accuracy of the estimators produced by partition method and the power of the sequential analysis method using the new estimators.

In chapter 4, we derive a sample size formula for clinical trials designed with clustered survival data. Discussions on the formula are given. Simulation results on the type I error and power using sample size calculated from the formula are also presented.

Chapter 2

Asymptotic distribution theory related to marginal proportional hazards model

In this chapter, we study the asymptotic distribution properties related to the marginal proportional hazards model. We will show in section 2 that under certain regularity conditions, the partial likelihood score $U(\beta, t, s)$ converges to a Gaussian random field. The asymptotic distribution of the maximum partial likelihood estimator $\hat{\beta}$ is derived in section 3. We also derive the asymptotic normality of the estimator of the baseline cumulative hazard function.

2.1 Introduction

Consider clustered event time data with n clusters. For cluster i, suppose there are k_i subjects, $i = 1, \cdots, n$. For the jth subject in the ith cluster, let R_{ij} be the study entry time, T_{ij} be the survival after entry, C_{ij} be the time elapsed from entry to censoring, and Z_{ij} be the $p \times 1$ vector of covariates. We shall assume throughout this thesis that $\{Z_{ij}\}$ are time independent.

For validity of the proposed methods, we make the following assumptions. We assume that $\{R_{ij}\}$ are i.i.d. random variables defined on a finite interval $[0, r],\ 0 < r < \infty$ and Z_{ij} are bounded by a nonrandom constant, say B. Within cluster i, $\{T_{ij},\ j = 1, \cdots, k_i\}$ may be correlated. We assume independence between clusters, i.e., for $i \neq j$, T_{ik} is independent of T_{jm} for all $k = 1, ..., k_i$ and $m = 1, .., k_j$. On entry times, R_{ij} are assumed to be independent of Z_{ij} and C_{ij}. Survival times T_{ij} are independent of censoring times C_{ij} conditional on covariates Z_{ij} and R_{ij}. k_i, the number of subjects in cluster i, $i = 1, \cdots, n$ are assumed to be constant with bound K. Throughout the sequel, we shall assume, without loss of generality, that R_{ij} For simplicity, we require

that $K_1 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} k_i$ and $K_2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} k_i^2$ exist.

Let t represent the calender time. At time t, T_{ij} is not only censored by C_{ij}, but also by $(t - R_{ij})^+$, as the failure has not occurred if $T_{ij} > t - R_{ij}$. Let $X_{ij}(t) = \min\{T_{ij}, C_{ij}, (t - R_{ij})^+\}$, then $X_{ij}(t)$ is the event time we observe at time t. Define the failure indicator variable $\delta_{ij}(t)$ as $\delta_{ij}(t) = 1$ if $X_{ij}(t) = T_{ij}$ and $\delta_{ij}(t) = 0$ otherwise. Let $R_{ij}(t) = \{(l, m) : 1 \le l \le n, m = 1, \cdots, k_l, X_{lm}(t) \ge X_{ij}(t)\}$.

Two stochastic processes need to be defined. For $x \le t$, let $N_{ij}(t, x) = I(X_{ij}(t) \le x, \delta_{ij}(t) = 1)$ and $Y_{ij}(t, x) = I(X_{ij}(t) \ge x)$. $N_{ij}(t, x) = 1$ means that the subject has experienced his failure before t and the survival is no greater than x. $Y_{ij}(t, x) = 1$ represents that observed at time t, the subject has stayed in trial at least x units of time. Notice that here we concern only information accumulated until time t. Any information occurred after time t will be analyzed at later periods. So we always have $X_{ij}(t, x) \le t$.

In this thesis, we assume that the marginal hazard function for each event time has a proportional hazards form, (see Lee, Wei and Amato [25]), that is, given the covariate vector $Z_{ij} = z_{ij}$ the hazard function for subject j in cluster i is,

$$\lambda_{ij}(t|z_{ij}) = \lambda_0(t) \exp(\beta' z_{ij}),\ t \ge 0. \tag{2.1}$$

Here $\lambda_0(t)$ is an unspecified baseline hazard function and $\beta = (\beta_1, ..., \beta_p)^\top$ is the regression parameter vector.

Model (2.1) does not specify the joint distribution of $\{T_{ij}\}$. Employing the working assumption proposed by Lee, *et al.* [25] that all T_{ij} are independent, the Cox partial likelihood function at calender time t takes form:

$$\begin{aligned} L(\beta, t) &= \prod_{i=1}^{n}\prod_{j=1}^{k_i}\left[\frac{\exp(\beta' Z_{ij})}{\sum_{(l,m)\in R_{ij}(t)} \exp(\beta' Z_{lm})}\right]^{\delta_{ij}(t)} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{k_i}\prod_{s\le t}\left[\frac{\exp(\beta' Z_{ij})}{\sum_{l=1}^{n}\sum_{m=1}^{k_l} Y_{lm}(t, s)\exp(\beta' Z_{lm})}\right]^{\Delta N_{ij}(t,s)}. \end{aligned} \tag{2.2}$$

Taking into consideration of both survival time s and calender time t, define a two-parameter score process

$$U(\beta, t, s) = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^s \left(Z_{ij} - \bar{Z}(\beta, t, u)\right) N_{ij}(t, du), \tag{2.3}$$

where

$$\bar{Z}(\beta,t,u)=\frac{\sum_{l=1}^{n}\sum_{m=1}^{k_l}Y_{lm}(t,u)\exp(\beta' Z_{lm})Z_{lm}}{\sum_{l=1}^{n}\sum_{m=1}^{k_l}Y_{lm}(t,u)\exp(\beta' Z_{lm})}.$$

Then

$$U(\beta,t,t)=\frac{\partial logL(\beta,t)}{\partial\beta}.$$

It is the score function at time t. In fact, if we generalize (2.2) to

$$L(\beta,t,s)=\prod_{i=1}^{n}\prod_{j=1}^{k_i}\prod_{u\le s}\left[\frac{\exp\left(\beta' Z_{ij}\right)}{\sum_{l=1}^{n}\sum_{m=1}^{k_l}Y_{lm}(t,u)\exp\left(\beta' Z_{lm}\right)}\right]^{\Delta N_{ij}(t,u)}.$$

Then

$$U(\beta,t,s)=\frac{\partial logL(\beta,t,s)}{\partial\beta}.$$

Simple calculation shows that (2.3) can be written as

$$U(\beta,t,s)=\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^s\left(Z_{ij}-\bar{Z}(\beta,t,u)\right)M_{ij}(\beta,t,du), \tag{2.4}$$

where

$$M_{ij}(\beta,t,s)=N_{ij}(t,s)-\int_0^s\lambda_0(u)Y_{ij}(t,u)\exp(\beta' Z_{ij})du.$$

Due to the staggered patient entry and the intra-cluster correlation, (2.4) is no longer a martingale. So the usual martingale method cannot be used here. In the next section, we will show that with intra-cluster dependence and staggered patient entry, $U(\beta,t,s)$ still converges in distribution to a Gaussian random field with mean vector 0 and continuous sample pathes.

2.2 The asymptotic distribution of the score process

Throughout this thesis, we use the following notation. Let β_0 denote the true value of the regression parameter vector in model (2.1). For a vector a, denote $a^{\otimes 0}=1, a^{\otimes 1}=a, a^{\otimes 2}=aa^{\tau}$. For a vector y or matrix Y let $\|y\|=\max_i|(y)_i|$ and $\|Y\|=\max_{i,j}|(Y)_{ij}|$. For any vector y, $|y|$ will denote the Euclidean norm $|y|=(\sum_{i=1}^{n}y_i^2)^{1/2}$.

Let

$$\tau=\sup\{t:\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}P\{(R_{ij}+T_{ij})\wedge(R_{ij}+C_{ij})\ge t\ge R_{ij}\}>0.$$

Defined this way, τ is the largest calendar time at which there is a positive proportion of clusters which have at least one member that is under observation. We will restrict the calender time t to be within the interval $[0, \tau]$.

For any $0 < t \leq \tau$, especially when t is large, it is possible that $\sum_{i=1}^{n} \sum_{j=1}^{k_i} Y_{ij}(t, s) = 0$ for some $s \leq t$. This means that observed at time t, no subject's observed survival is longer than s. By the definition of τ, this does not mean that all subjects have failed before time t. It simply represents the situation that many subjects enter the study late, so their failures are not observed yet. When $\sum_{i=1}^{n} \sum_{j=1}^{k_i} Y_{ij}(t, s) = 0$, no information on survival is provided by $\{Y_{ij}(t, s)\}$. It also causes difficulty in theoretical investigation. These thoughts lead us to the following definition:

Let $d > 0$ is a fixed small number. Define if $s > t$, $Y_{ij}(t, s) = Y_{ij}(t, t)$. Then define

$$s^* = \sup\{s : \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k_i} EY_{ij}(t, s^*) \geq d, \text{ for all } t \in [0, \tau]\}. \tag{2.5}$$

Such s^* is a survival time that at least a d proportion of subjects stay in trial longer than it. When d is very small, most of the data will satisfy (2.5). So from now on, we will restrict our study on $D^* = \{(t, s) : s \leq t, s \leq s^*, t \leq \tau.\}$

We begin with the following regularity conditions:

(2.1) $\lambda_0(t)$ is bounded on $[0, \tau]$.

(2.2) Let

$$S_n^{(d)}(\beta, t, s) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{k_i} Y_{ij}(t, s) \exp(\beta' Z_{ij}) Z_{ij}^{\otimes d}, \quad d = 0, 1, 2.$$

Then for all $(t, s) \in D^*$ and $\beta \in \mathcal{B}$, there exits $s^{(d)}(\beta, t, s)$ such that

$$s^{(d)}(\beta, t, s) = \lim_{n\to\infty} ES_n^{(d)}(\beta, t, s),$$

and

$$\sup_{(t,s)\in D^*, \beta\in\mathcal{B}} \|S_n^{(d)}(\beta, t, s) - s^{(d)}(\beta, t, s)\| \to 0, \ a.s, \quad d = 0.1, 2.$$

(2.3) Let

$$H_n((t_1,s_1),(t_2,s_2)) =$$
$$\frac{1}{n}\sum_{i=1}^{n} E[(\sum_{j=1}^{k_i}\int_0^{s_1}(Z_{ij}-\bar{Z}(\beta_0,t_1,u))dM_{ij}(\beta_0,t_1,u))(\sum_{j=1}^{k_i}\int_0^{s_2}(Z_{ij}-\bar{Z}(\beta_0,t_2,u))dM_{ij}(\beta_0,t_2,u))'].$$

Then $H_n((t_1,s_1),(t_2,s_2))$ converges uniformly on D^* as $n\to\infty$.

(2.4) Let

$$\Gamma(\beta,t)=\int_0^t v(\beta,t,x)s^{(0)}(\beta,t,x)\lambda_0(x)dx, \tag{2.6}$$

where

$$v(\beta,t,x)=\frac{s^{(2)}(\beta,t,x)}{s^{(0)}(\beta,t,x)}-\left(\frac{s^{(1)}(\beta,t,x)}{s^{(0)}(\beta,t,x)}\right)^2. \tag{2.7}$$

Then $\Gamma(\beta_0,t)$ is positive definite on $[0,\tau]$.

Remark 2.2.1 *From (2.5), for $(t,s)\in D^*$, there is a $N>0$ such that when $n>N$,*

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}EY_{ij}(t,s)\geq\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}EY_{ij}(t,s^*)\geq d/2.$$

Since β is bounded on $\mathcal{B}$, and Z_{ij} are bounded, so there is a constant c_1 such that $\exp(\beta'Z_{ij})\geq c_1$ for all $\beta\in\mathcal{B}$. So

$$ES_n^{(0)}(\beta,t,s)\geq\frac{dc_1}{2},\qquad (t,s)\in D^*,\ \beta\in\mathcal{B},$$

So

$$s^{(0)}(\beta,t,s)\geq\frac{dc_1}{2},\qquad (t,s)\in D^*,\ \beta\in\mathcal{B}. \tag{2.8}$$

That is, $s^{(0)}(\beta,t,s)$ is bounded away from 0 on $\mathcal{B}\times D^*$.

Remark 2.2.2 *For $d=0,1,2$*

$$\begin{aligned}
& ES^{(d)}(\beta,t,x)\\
= & \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}EZ_{ij}^{\otimes d}Y_{ij}(t,x)\exp(\beta'Z_{ij})\\
= & \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}E(Z_{ij}^{\otimes d}\exp(\beta'Z_{ij})E(Y_{ij}(t,x)|Z_{ij}))\\
= & P((t-R_{11})^+\geq x)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}E(Z_{ij}^{\otimes d}\exp(\beta'Z_{ij})P(T_{ij}\geq x,C_{ij}\geq x|Z_{ij})).
\end{aligned}$$

So

$$\begin{aligned} & s^{(d)}(\beta,t,x) \\ = \;& P((t-R_{11})^+ \geq x) \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k_i} E(Z_{ij}^{\otimes d} \exp(\beta' Z_{ij}) P(T_{ij} \geq x, C_{ij} \geq x | Z_{ij})). \end{aligned}$$

Therefore $s^{(d)}(\beta,t,x),\ d=0,1,2$ *depend on* t *through the common factor* $P((t-R_{11})^+ \geq x)$. *Hence*

$$\mu(\beta,t,x) = \frac{s^{(1)}(\beta,t,x)}{s^{(0)}(\beta,t,x)}$$

and $v(\beta,t,x)$ *in (2.7) are in fact not dependent on the calender time* t. *So from now on, we write* $\mu(\beta,t,x)$ *and* $v(\beta,t,x)$ *as* $\mu(\beta,x)$ *and* $v(\beta,x)$.

Because of the intra-cluster correlation and staggered patient entry, there is not a common filtration with which all $\{M_{ij}(\beta_0,t,s), i=1,\cdots,n, j=1,\cdots,k_i\}$ are martingales. But for every fixed $t>0$ and (i,j), $M_{ij}(\beta_0,t,s)$, as a function of s, is still a martingale related to a filtration specific to (i,j). This property can be seen from the following representation

$$M_{ij}(\beta,t,s) = \int_0^\infty I_{\{s\wedge(t-R_{ij})^+\wedge C_{ij}\geq u\}}\, dM_{ij}^0(\beta,u), \tag{2.9}$$

where

$$M_{ij}^0(\beta,u) = I_{(T_{ij}\leq u)} - \int_0^u I_{(T_{ij}\geq x)} \exp(\beta' Z_{ij}) \lambda_0(x) dx.$$

If β_0 is the the true regression parameter vector, then $M_{ij}^0(\beta_0,t)$ is a martingale with respect to $\mathcal{F}_{(ij)t} = \sigma(Z_{ij}, C_{ij}, R_{ij},\ I_{(T_{ij}\leq t)}, I_{(T_{ij}>t)}),\ 0\leq t<\infty$. The indicator function $I_{\{s\wedge(t-R_{ij})^+\wedge C_{ij}\geq u\}}$ is $\mathcal{F}_{(ij)u}$ predictable. So $M_{ij}(\beta_0,t,s)$ is a martingale with respect to $\mathcal{F}_{(ij)s}$. This "local" martingale property makes subsequent computation and proof much easier.

Our following effort will be devoted to show that the score process $U(\beta_0,t,s)$ converges weakly to a Gaussian random field on D^*. To this end, we first show that $U(\beta_0,t,s)$ is asymptotically equivalent to $\tilde{U}(\beta_0,t,s)$ which is obtained by replacing $\bar{Z}(\beta,t,x)$ by $\mu(\beta,t,x)$. This replacement eliminates the complexity caused by $\bar{Z}(\beta,t,x)$ because every term in $\tilde{U}(\beta,t,s)$ is a function of only one member. It is then easier to

prove the weak convergence of $\tilde{U}(\beta, t, s)$. The weak convergence to a Gaussian random field of $U(\beta_0, t, s)$ comes then from the asymptotic equivalence of it to $\tilde{U}(\beta, t, s)$.

Lemma 2.2.1 *Let*

$$\begin{aligned} & \tilde{H}_n((t_1, s_1), (t_2, s_2)) \\ = \; & \frac{1}{n} E[\sum_{i=1}^{n} (\sum_{j=1}^{k_i} \int_0^{s_1} (Z_{ij} - \mu(\beta_0, u)) dM_{ij}(\beta_0, t_1, u)) (\sum_{j=1}^{k_i} \int_0^{s_2} (Z_{ij} - \mu(\beta_0, u)) dM_{ij}(\beta_0, t_2, u))']. \end{aligned}$$

Then $\tilde{H}_n((t_1, s_1), (t_2, s_2))$ *converges uniformly on* D^* *to* $\lim_{n\to\infty} H_n((t_1, s_1), (t_2, s_2))$.

Proof. Notice that

$$\begin{aligned} & \tilde{H}_n((t_1, s_1), (t_2, s_2)) - H_n((t_1, s_1), (t_2, s_2)) \\ = \; & \frac{1}{n} \sum_{i=1}^{n} E[(\sum_{j=1}^{k_i} \int_0^{s_1} (Z_{ij} - \bar{Z}(\beta_0, t_1, u)) dM_{ij}(\beta_0, t_1, u)) \otimes \\ & (\sum_{j=1}^{k_i} \int_0^{s_2} (\mu(\beta_0, u) - \bar{Z}(\beta_0, t_2, u)) dM_{ij}(\beta_0, t_2, u)) + \\ & (\sum_{j=1}^{k_i} \int_0^{s_1} (\mu(\beta_0, u) - \bar{Z}(\beta_0, t_1, u)) dM_{ij}(\beta_0, t_1, u)) \otimes \\ & (\sum_{j=1}^{k_i} \int_0^{s_2} (Z_{ij} - \bar{Z}(\beta_0, t_2, u)) dM_{ij}(\beta_0, t_2, u))]. \end{aligned}$$

We would like to prove that $\tilde{H}_n((t_1, s_1), (t_2, s_2)) - H_n((t_1, s_1), (t_2, s_2))$ converges uniformly to 0 on D^*.

To this end, it is sufficient to show that

$$\begin{aligned} D_n((t_1, s_1), (t_2, s_2)) \; = \; & \frac{1}{n} \sum_{i=1}^{n} E[(\sum_{j=1}^{k_i} \int_0^{s_1} (Z_{ij} - \bar{Z}(\beta_0, t_1, u)) dM_{ij}(\beta_0, t_1, u)) \otimes \\ & (\sum_{j=1}^{k_i} \int_0^{s_2} (\mu(\beta_0, u) - \bar{Z}(\beta_0, t_2, u)) dM_{ij}(\beta_0, t_2, u)] \end{aligned}$$

converges uniformly to 0 on D^*. The convergence of the other part can be proved similarly.

From Cauchy Inequality,

$$\begin{aligned}
& |D_n((t_1,s_1),(t_2,s_2))|^2 \\
\leq\ & [\frac{1}{n}\sum_{i=1}^{n} E(\sum_{j=1}^{k_i}\int_0^{s_1}(Z_{ij}-\bar{Z}(\beta_0,t_1,u))dM_{ij}(\beta_0,t_1,u))^2] \\
& [\frac{1}{n}\sum_{i=1}^{n} E(\sum_{j=1}^{k_i}\int_0^{s_2}(\mu(\beta_0,u)-\bar{Z}(\beta_0,t_2,u))dM_{ij}(\beta_0,t_2,u))^2] \\
\leq\ & [\frac{1}{n}\sum_{i=1}^{n} k_i(\sum_{j=1}^{k_i} E(\int_0^{s_1}(Z_{ij}-\bar{Z}(\beta_0,t_1,u))dM_{ij}(\beta_0,t_1,u))^2)] \\
& [\frac{1}{n}\sum_{i=1}^{n} k_i(\sum_{j=1}^{k_i} E\int_0^{s_2}(\mu(\beta_0,u)-\bar{Z}(\beta_0,t_2,u))dM_{ij}(\beta_0,t_2,u))^2].
\end{aligned}$$

Since $\{Z_{ij}\}, \lambda_0(x)$ are all bounded, so it can be shown that there is a constant $C>0$ such that

$$|E(\int_0^{s_1}(Z_{ij}-\bar{Z}(\beta_0,t_1,u))dM_{ij}(\beta_0,t_1,u))^2| \leq C, \quad \text{for all } (t_1,s_1)\in D^*.$$

From Remark 2.2.1, there is a constant c such that when n is large enough,

$$s^{(0)}(\beta_0,t,s)\geq c \ \text{ for all } (t,s)\in D^*.$$

From this fact and Condition (2.2), for any $\varepsilon>0$, there is a N, when $n>N$,

$$\sup_{(t,s)\in D^*}\|\bar{Z}(\beta_0,t,s)-\mu(\beta_0,t,s)\|\leq\varepsilon.$$

So when $n>N$,

$$\begin{aligned}
& |D_n((t_1,s_1),(t_2,s_2))|^2 \\
\leq\ & \frac{K^2C\varepsilon}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}(EN_{ij}(t_2,s_2)+\int_0^{s_2}\lambda_0(x)EY_{ij}(t_2,x)\exp(\beta_0'Z_{ij})dx).
\end{aligned}$$

So there is a constant C_1 such that

$$|D_n((t_1,s_1),(t_2,s_2))|^2\leq\frac{K^2C_1\varepsilon}{n}.$$

Thus

$$|D_n((t_1,s_1),(t_2,s_2))|\to 0 \ \text{ uniformly on } D^*.$$

This completes the proof.

Theorem 2.2.1 *Under conditions (2.1) and (2.2), for every fixed* $(t,s) \in D^*$, $lim_{n\to\infty} P(\frac{1}{\sqrt{n}} U(\beta_0,t,s) \neq \frac{1}{\sqrt{n}} \tilde{U}(\beta_0,t,s)) = 0$.

Proof. It comes directly from the proof of Lemma 2.2.1.

Theorem 2.2.2 *Assume that Conditions (2.1), (2.2) and (2.3) are satisfied. Then* $\{n^{-1/2}\tilde{U}(\beta_0,t,s),\ (t,s) \in D^*\}$ *converges weakly to a Gaussian random field* ξ *that have mean vector 0, continuous sample paths and covariance function*

$$\Sigma((t_1,s_1),(t_2,s_2)) = \lim_{n\to\infty} H_n((t_1,s_1),(t_2,s_2)). \tag{2.10}$$

Proof. We use Theorem 10.6 of Pollard [32] to show the weak convergence. Without loss of generality, we assume that $p = 1$.

Let d be the usual Euclidean metric on D^*. Then (D^*, d) is a metric space.

Let

$$f_{ni}(t,s) = n^{-1/2} \sum_{j=1}^{k_i} \int_0^s (Z_{ij} - \mu(\beta_0,s)) M_{ij}(\beta_0,t,ds).$$

Then $f_{n1}(t,s)$, $f_{n2}(t,s)$, ..., $f_{nn}(t,s)$ are independent. We will show that $f_{ni}(t,s)$ satisfy the five conditions of Theorem 10.6 of Pollard.

Rewrite $f_{ni}(t,s)$ as

$$\begin{aligned} & f_{ni}(t,s) \\ = \ & n^{-1/2} \sum_{j=1}^{k_i} (Z_{ij} N_{ij}(t,s) - \mu(\beta_0,T_{ij}) N_{ij}(t,s) + \\ & \int_0^s Z_{ij} \lambda_0(x) \mu(\beta_0,x) Y_{ij}(t,x) \exp(\beta_0' Z_{ij}) dx - \int_0^s \lambda_0(x) \mu(\beta_0,x) Y_{ij}(t,x) \exp(\beta_0' Z_{ij}) dx). \end{aligned}$$

By Lemma A.1 of Bilias, Gu and Ying [2] which indicates that a finite sum of measurable functions that have common envelop is still manageable, we need to show that for every (i,j), each term within the sum parentheses is manageable. Since $Z_{ij} = Z_{ij}^+ - Z_{ij}^-$ and $\mu(\beta_0,x) = \mu(\beta_0,x)^+ - \mu(\beta_0,x)^-$, by Lemma A.1 of Bilias, Gu and Ying [2] again, we may assume that $Z_{ij},\ i = 1,\cdots,n, j = 1,\cdots,k_i$ and $\mu(\beta_0,x)$ is nonnegative.

For each fixed (i,j), $Z_{ij}N_{ij}(t,s) = \min\{Z_{ij} I_{(T_{ij}\le C_{ij}, T_{ij} \le (t-R_{ij})^+)}, Z_{ij} I_{(T_{ij}\le C_{ij})} I_{(T_{ij}\le s)}\}$. The term $Z_{ij} I_{(T_{ij}\le C_{ij}, T_{ij} \le (t-R_{ij})^+)}$ is a nondecreasing function of t, so it has pseudodimension at most 1, whereas $Z_{ij} I_{(T_{ij}\le C_{ij})} I_{(T_{ij}\le s)}$ is a nondecreasing function of s, so it

has pseudodimension at most 1. From Lemma 5.1 of Pollard, $Z_{ij}N_{ij}(t,s)$ has pseudodimension at most 10. So it is Euclidean, and hence is manageable.

We can prove the manageability of $\mu(\beta_0, T_{ij})N_{ij}(t,s)$ similarly.

Notice that

$$\begin{aligned}
&\int_0^s \lambda_0(x) Z_{ij} Y_{ij}(t,x) \exp(\beta_0' Z_{ij}) dx \\
= \quad & \min\{\int_0^{s\wedge T_{ij}\wedge C_{ij}} \lambda_0(x) \exp(\beta_0' Z_{ij}) Z_{ij} dx, \int_0^{(t-R_{ij})^+\wedge T_{ij}\wedge C_{ij}} \lambda_0(x) \exp(\beta_0' Z_{ij}) Z_{ij} dx\}.
\end{aligned}$$

The term $\int_0^{s\wedge T_{ij}\wedge C_{ij}} \lambda_0(x) \exp(\beta_0' Z_{ij}) Z_{ij} dx$ and $\int_0^{(t-R_{ij})^+\wedge T_{ij}\wedge C_{ij}} \lambda_0(x) \exp(\beta_0' Z_{ij}) Z_{ij} dx$ are a nondecreasing function of s and t respectively. So both have pseudodimension at most 1. Therefore $\int_0^s \lambda_0(x) Z_{ij} Y_{ij}(t,x) \exp(\beta_0' Z_{ij}) dx$ is manageable.

By the same decomposition we can prove that $\int_0^s \lambda_0(x)\mu(\beta_0, x) Y_{ij}(t,x) \exp(\beta_0' Z_{ij}) dx$ is manageable.

To verify (ii), notice that $Ef_{ni}(t,s) = 0, (t,s) \in D^*$. Furthermore,

$$\begin{aligned}
&E[(\sum_{i=1}^n f_{ni}(t_1,s_1))(\sum_{i=1}^n f_{ni}(t_2,s_2))] \\
= \quad & \frac{1}{n} E[(\sum_{i=1}^n \sum_{j=1}^{k_i} \int_0^{s_1} (Z_{ij} - \mu(\beta_0,u)) dM_{ij}(\beta_0, t_1, u))(\sum_{i=1}^n \sum_{j=1}^{k_i} \int_0^{s_2} (Z_{ij} - \mu(\beta_0,u)) dM_{ij}(\beta_0, t_2, u))] \\
= \quad & \frac{1}{n} \sum_{i=1}^n E[(\sum_{j=1}^{k_i} \int_0^{s_1} (Z_{ij} - \mu(\beta_0,u)) dM_{ij}(\beta_0, t_1, u))(\sum_{j=1}^{k_i} \int_0^{s_2} (Z_{ij} - \mu(\beta_0,u)) dM_{ij}(\beta_0, t_2, u))].
\end{aligned}$$

From Condition (2.3), $\lim_{n\to\infty} E[(\sum_{i=1}^n f_{ni}(t_1,s_1))(\sum_{i=1}^n f_{ni}(t_2,s_2))]$ exists for every $(s_1,t_1), (s_2,t_2) \in D^*$.

Let $F_{ni} = KZB^*/\sqrt{n}$ for some positive constant B^*. Then F_{ni} is the an envelop of $\{f_{ni}\}$ which satisfy (iii) and (iv) of Theorem 10.6 of Pollard.

To prove that condition (v) of Theorem 10.6 of Pollard is satisfied, define, for any $(t_1,s_1) \in D^*$ and $(t_2,s_2) \in D^*$,

$$\rho_n((t_1,s_1),(t_2,s_2)) = \left(\sum_{i=1}^n E|f_{ni}(t_1,s_1) - f_{ni}(t_2,s_2)|^2\right)^{1/2}.$$

Then

$$\begin{aligned} & \rho_n^2((t_1,s_1),(t_2,s_2)) \\ = & \sum_{i=1}^{n} Ef_{ni}^2(t_1,s_1) + \sum_{i=1}^{n} Ef_{ni}^2(t_2,s_2) - 2\sum_{i=1}^{n} Ef_{ni}(t_1,s_1)f_{ni}(t_2,s_2) \\ = & H_n((t_1,s_1),(t_1,s_1)) + H_n((t_2,s_2),(t_2,s_2)) - \frac{2}{n}H_n((t_1,s_1),(t_2,s_2)). \end{aligned}$$

From Condition (2.3), $\rho_n^2((t_1,s_1),(t_2,s_2))$ converges uniformly on D^*. So $\rho_n((t_1,s_1),(t_2,s_2))$ converges.

Suppose $\{(t_m,s_m)\}$ and $\{(u_m,v_m)\}$ are two deterministic sequences such that $\rho((t_m,s_m),(u_m,v_m)) \to 0$. Then for any small $\epsilon > 0$, there is a $N_1 > 0$, such that when $n > N_1$,

$$\rho((t_n,s_n),(u_n,v_n)) < \epsilon. \tag{2.11}$$

On the other hand, since ρ_n converges uniformly to ρ, for the same ϵ, there is a $N_2 > 0$, such that when $n > N_2$,

$$|\rho_n((t_1,s_1),(t_2,s_2)) - \rho((t_1,s_1),(t_2,s_2)| < \epsilon, \tag{2.12}$$

for all $(t_1,s_1) \in D^*$ and $(t_2,s_2) \in D^*$. Therefore, let $N = max(N_1,N_2)$. When $n > N$, from (2.11) and (2.12),

$$\rho_n((t_n,s_n),(u_n,v_n) \le |\rho_n((t_n,s_n),(u_n,v_n) - \rho((t_n,s_n),(u_n,v_n)| + \rho((t_n,s_n),(u_n,v_n) \le \epsilon + \epsilon = 2\epsilon.$$

So $\rho_n((t_n,s_n),(u_n,v_n) \to 0$. This proves that the Condition (v) of Theorem 10.6 of Pollard is true.

Corollary 2.2.1 *Under Condition (2.1) and (2.2), $U(\beta_0,t,s)$ converges in distribution to a Gaussian random field with continuous sample paths, mean vector 0 and covariance function (2.10).*

2.3 Convergence of the regression parameter estimator

Let β_0 be the true regression parameter vector. Let $\hat{\beta}_t$ be a solution to the equation

$$U(\beta,t,t) = 0,\ t \in [0,\tau]. \tag{2.13}$$

We will show that under some regularity conditions, $\hat{\beta}_t$ is consistent and asymptotically normal.

Theorem 2.3.1 *Suppose that Conditions (2.1) and (2.2) are satisfied. Then there is a solution $\hat{\beta}_t$ to equation (2.13) such that $\hat{\beta}_t$ converges to β_0 in probability.*

Proof. Let

$$\begin{aligned} X_n(\beta,t) &= \frac{1}{n}(log(L(\beta,t)) - log(L(\beta_0,t))) \\ &= n^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^t [(\beta-\beta_0)'Z_{ij} - log\{\frac{S^{(0)}(\beta,t,s)}{S^{(0)}(\beta_0,t,s)}\}]dN_{ij}(t,s)\right]. \end{aligned}$$

Let

$$A_n(\beta,t) = n^{-1}\left[\int_0^t [(\beta-\beta_0)'S^{(1)}(\beta_0,t,u) - log\{\frac{S^{(0)}(\beta,t,u)}{S^{(0)}(\beta_0,t,u)}\}S^{(0)}(\beta_0,t,u)]\lambda_0(u)du\right].$$

Then

$$\begin{aligned} & X_n(\beta,t) - A_n(\beta,t) \\ = \; & n^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^t [(\beta-\beta_0)'Z_{ij} - log\{\frac{S^{(0)}(\beta,t,u)}{S^{(0)}(\beta_0,t,u)}\}]dM_{ij}(\beta_0,t,u)\right] \\ = \; & n^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^t [(\beta-\beta_0)'Z_{ij} - log\{\frac{s^{(0)}(\beta,t,u)}{s^{(0)}(\beta_0,t,u)}\}]dM_{ij}(\beta_0,t,u)\right] + \\ & n^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^t [log\{\frac{s^{(0)}(\beta,t,u)}{s^{(0)}(\beta_0,t,u)}\} - log\{\frac{S^{(0)}(\beta,t,u)}{S^{(0)}(\beta_0,t,u)}\}]dM_{ij}(\beta_0,t,u)\right] \\ = \; & I + II. \end{aligned}$$

Since $(\beta-\beta_0)'Z_{ij}$ is predictable with $\mathcal{F}_{(ij)u}$ for any $0 \le u \le \tau$, and $\log\{\frac{s^{(0)}(\beta,t,u)}{s^{(0)}(\beta_0,t,u)}\}$ is a deterministic function, we have that

$$\int_0^s [(\beta-\beta_0)'Z_{ij} - log\{\frac{s^{(0)}(\beta,t,u)}{s^{(0)}(\beta_0,t,u)}\}]dM_{ij}(\beta_0,t,u)$$

is a martingale. So

$$E\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^t [(\beta-\beta_0)'Z_{ij} - log\{\frac{s^{(0)}(\beta,t,u)}{s^{(0)}(\beta_0,t,u)}\}]dM_{ij}(\beta_0,t,u) = 0.$$

It is similar to the proof of Theorem 2.2.2 that we can show the sequence $\{f_i(t)\}$ which is defined as

$$f_i(t) = \sum_{j=1}^{k_i} \int_0^t [(\beta - \beta_0)' Z_{ij} - log\{\frac{s^{(0)}(\beta, t, u)}{s^{(0)}(\beta_0, t, u)}\}] dM_{ij}(\beta_0, t, u)$$

is independent and manageable with some constant envelop C. So from Theorem 8.3 of Pollard, we have that

$$I = \frac{1}{n} \sup_{(t,s)\in D^*} |[\sum_{i=1}^{n} \sum_{j=1}^{k_i} \int_0^t [(\beta - \beta_0)' Z_{ij} - log\{\frac{s^{(0)}(\beta, t, u)}{s^{(0)}(\beta_0, t, u)}\}] dM_{ij}(\beta_0, t, u)]| \to 0, \ a.s.$$

For $\beta \in \mathcal{B}$, from Condition (2.2) and Remark 2.2.1, we have

$$\sup_{\beta\in\mathcal{B},(t,s)\in D^*} \left| log\{\frac{s^{(0)}(\beta, t, u)}{s^{(0)}(\beta_0, t, u)}\} - log\{\frac{S^{(0)}(\beta, t, u)}{S^{(0)}(\beta_0, t, u)}\} \right| \to 0, \ a.s.$$

From the boundedness of $\lambda_0(x)$, we can show that

$$II = \frac{1}{n} \sup_{0\le t\le\tau,\beta\in\mathcal{B}} | \sum_{i=1}^{n} \sum_{j=1}^{k_i} \int_0^t [log\{\frac{s^{(0)}(\beta, t, u)}{s^{(0)}(\beta_0, t, u)}\} - log\{\frac{S^{(0)}(\beta, t, u)}{S^{(0)}(\beta_0, t, u)}\}] dM_{ij}(\beta_0, t, u)|$$
$$\to 0, \ a.s.$$

Therefore we have showed that $X_n(\beta, t)$ converges almost surely to the same limit of $A_n(\beta, t)$ for all $\beta \in \mathcal{B}$. By Condition (2.2), $A_n(\beta, t)$ converges to

$$A(\beta, t) = \int_0^t [(\beta - \beta_0)' s^{(1)}(\beta_0, t, u) - log\{\frac{s^{(0)}(\beta, t, u)}{s^{(0)}(\beta_0, t, u)}\} s^{(0)}(\beta_0, t, u)] \lambda_0(u) du.$$

It is easy to see that $A(\beta, t)$ is a concave function of β with a unique maximum at $\beta = \beta_0$ and $X_n(\beta, t)$ is a concave function of β with a unique maximum $\hat{\beta}_t$. So from Lemma 8.3.1 of Fleming and Harrington [12], $\hat{\beta}_t \to \beta_0$ in probability as $n \to \infty$.

Theorem 2.3.2 *Let $\Sigma(t, t)$ represent the limiting variance of $\frac{1}{\sqrt{n}} U(\beta_0, t, t)$. Then under Conditions (2.1) and (2.2),*

$$\sqrt{n}(\hat{\beta}_t - \beta_0) \xrightarrow{d} N(0, W_t), \ t \in [0, \tau]$$

where

$$W_t = \Gamma(\beta_0, t)^{-1} \Sigma(t, t) \Gamma(\beta_0, t)^{-1}. \tag{2.14}$$

Proof. From the Taylor series expansion of $U(\hat{\beta}_t, t, t)$ around β_0 it yields that

$$\frac{1}{\sqrt{n}}U(\beta_0, t, t) = \frac{1}{n}\frac{\partial U(\beta, t)}{\partial \beta}|_{\beta=\beta^*}\sqrt{n}(\hat{\beta}_t - \beta_0), \tag{2.15}$$

where β^* is on line segment between $\hat{\beta}_t$ and β_0. Theorem 2.2.2 implies that $\frac{1}{\sqrt{n}}U(\beta_0, t, t)$ is asymptotically normal with covariance matrix $\Sigma(t, t)$. Since $\hat{\beta}_t$ is consistent, $\beta^* \to \beta_0$ in probability as $n \to \infty$.

By definition of $U(\beta, t, t)$,

$$\frac{\partial U(\beta, t, t)}{\partial \beta} = \sum_{i=1}^{n}\sum_{m=1}^{k_i}\int_0^t \left(\frac{S^{(2)}(\beta, t, x)}{S^{(0)}(\beta, t, x)} - \{\frac{S^{(1)}(\beta, t, x)}{S^{(0)}(\beta, t, x)}\}^{\otimes 2}\right) dN_{im}(\beta, t, x).$$

Let

$$\mathcal{I}_n(\beta, t) = -\frac{1}{n}\frac{\partial U(\beta, t, t)}{\partial \beta}.$$

Similar to the proof in Lemma 2.2.1, we can show that

$$\sup_{0\le x\le t}\left|\left(\frac{S^{(2)}(\beta^*, t, x)}{S^{(0)}(\beta^*, t, x)} - \{\frac{S^{(1)}(\beta^*, t, x)}{S^{(0)}(\beta^*, t, x)}\}^{\otimes 2}\right) - \left(\frac{s^{(2)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)} - \{\frac{s^{(1)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)}\}^{\otimes 2}\right)\right| \to 0, \ a.s..$$

So $\mathcal{I}(\beta^*, t)$ has the same limit with

$$A_n(t) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^t \left(\frac{s^{(2)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)} - \{\frac{s^{(1)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)}\}^{\otimes 2}\right) dN_{ij}(t, s).$$

Now since $M_{ij}(\beta_t, t, x)$ is a martingale for each (i, j) and

$$\frac{s^{(2)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)} - \{\frac{s^{(1)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)}\}^{\otimes 2}$$

is a deterministic function,

$$\int_0^t \left(\frac{s^{(2)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)} - \{\frac{s^{(1)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)}\}^{\otimes 2}\right) dM_{ij}(\beta_0, t, x)$$

is also a martingale. Therefore

$$\begin{aligned} & E\int_0^t \left(\frac{s^{(2)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)} - \{\frac{s^{(1)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)}\}^{\otimes 2}\right) dN_{ij}(t, x) \\ = \ & E\int_0^t \left(\frac{s^{(2)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)} - \{\frac{s^{(1)}(\beta_0, t, x)}{s^{(0)}(\beta_0, t, x)}\}^{\otimes 2}\right) \lambda_0(t) Y_{ij}(t, x)\exp(\beta_0' Z_{ij}) dt. \end{aligned}$$

So,

$$\begin{aligned}
&\frac{1}{n}E\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_0^t\left(\frac{s^{(2)}(\beta_0,t,x)}{s^{(0)}(\beta_0,t,x)}-\{\frac{s^{(1)}(\beta_0,t,x)}{s^{(0)}(\beta_0,t,x)}\}^{\otimes 2}\right)dN_{ij}(t,x)\\
=&\int_0^t\left(\frac{s^{(2)}(\beta_0,t,x)}{s^{(0)}(\beta_0,t,x)}-\{\frac{s^{(1)}(\beta_0,t,x)}{s^{(0)}(\beta_0,t,x)}\}^{\otimes 2}\right)\lambda_0(x)\ E\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}Y_{ij}(t,x)\exp\left(\beta_0'Z_{ij}\right)\right)dx\\
=&\int_0^t\left(\frac{s^{(2)}(\beta_0,t,x)}{s^{(0)}(\beta_0,t,x)}-\{\frac{s^{(1)}(\beta_0,t,x)}{s^{(0)}(\beta_0,t,x)}\}^{\otimes 2}\right)\lambda_0(x)ES^{(0)}(\beta_0,t,x)dx.
\end{aligned}$$

Therefore,

$$EA_n(\beta_0,t)\to\int_0^t\left(\frac{s^{(2)}(\beta_0,t,x)}{s^{(0)}(\beta_0,t,x)}-\{\frac{s^{(1)}(\beta_0,t,x)}{s^{(0)}(\beta_0,t,x)}\}^{\otimes 2}\right)\lambda_0(x)s^{(0)}(\beta_0,t,x)dx=\Gamma(\beta_0,t)\ \ a.s.$$

From Condition (2.4), $\Gamma(\beta_0,t)$ is positive definite on $[0,\tau]$. So for fixed $t\in[0,\tau]$, when n is large enough, $A_n(\beta_0,t)$ is positive definite. Since $\hat{\beta}_t$ is a consistent estimator of β_0, we conclude that when n is large enough, $\mathcal{I}_n(\beta^*,t)$ is positive definite. Hence,

$$\sqrt{n}(\hat{\beta}_t-\beta_0)=\mathcal{I}_n(\beta^*,t)^{-1}\frac{1}{\sqrt{n}}U(\beta_0,t,t). \tag{2.16}$$

The asymptotic distribution of $\sqrt{n}(\hat{\beta}_t-\beta_0)$ comes then from the asymptotic distribution of $\frac{1}{\sqrt{n}}U(\beta_0,t,t)$. This completes the proof.

The following theorem gives another form of the covariance of $\hat{\beta}_t$ which will be useful in the next chapter.

Theorem 2.3.3 *The covariance matrix of $\hat{\beta}_t$ can be rewritten as*

$$W(t)=\Gamma(\beta_0,t)^{-1}+\Gamma(\beta_0,t)^{-1}H^*(t)\Gamma(\beta_0,t)^{-1}, \tag{2.17}$$

where

$$H^*(t)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq k}^{k_i}E(\int_0^t(Z_{ij}-\mu(\beta_0,x))dM_{ij}(\beta_0,x)\int_0^t(Z_{ik}-\mu(\beta_0,x))dM_{ik}(\beta_0,x)).$$

Proof. The sum

$$\frac{1}{n}\sum_{i=1}^{n}E(\sum_{j=1}^{k_i}\int_0^t(Z_{ij}-\mu(\beta_0,x))dM_{ij}(\beta_0,x))^{\otimes 2}$$

can be decomposed as the sum of

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i} E(\int_0^t (Z_{ij}-\mu(\beta_0,x))dM_{ij}(\beta_0,x))^{\otimes 2}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq k}^{k_i} E(\int_0^t (Z_{ij}-\mu(\beta_0,x))dM_{ij}(\beta_0,x)\otimes\int_0^t (Z_{ik}-\mu(\beta_0,x))dM_{ik}(\beta_0,x)).$$

But

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i} E(\int_0^t (Z_{ij}-\mu(\beta_0,x))dM_{ij}(\beta_0,x))^{\otimes 2}\\ =\;& E\int_0^t [S^{(2)}(\beta_0,x)-2\mu(\beta_0,x)\otimes S^{(1)}(\beta_0,x)+\mu(\beta_0,x)^{\otimes 2}S^{(0)}(\beta_0,x)]dx.\end{aligned}$$

So

$$\begin{aligned}&\Sigma(t,t)\\ =\;& \lim_{n\to\infty} E\int_0^t [S^{(2)}(\beta_0,x)-2\mu(\beta_0,x)\otimes S^{(1)}(\beta_0,x)+\mu(\beta_0,x)^{\otimes 2}S^{(0)}(\beta_0,x)]dx+\\ & \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq k}^{k_i} E(\int_0^t (Z_{ij}-\mu(\beta_0,x))dM_{ij}(\beta_0,x)\otimes\int_0^t (Z_{ik}-\mu(\beta_0,x))dM_{ik}(\beta_0,x))\\ =\;& \Gamma(\beta_0,t)+H^*(t).\end{aligned}$$

This completes the proof.

2.4 The estimation of the baseline cumulative hazard function

Let

$$\Lambda_0(s)=\int_0^s \lambda_0(x)dx$$

be the baseline cumulative hazard function. For fixed time $t\in[0,\tau]$, we define the Breslow's estimator of $\Lambda_0(s)$ as

$$\hat{\Lambda}_0(t,s)=\int_0^s \left\{\sum_{i=1}^{n}\sum_{j=1}^{k_i} Y_{ij}(t,u)\exp(\hat{\beta}_t' Z_{ij})\right\}^{-1} d\bar{N}(t,u), \tag{2.18}$$

where

$$\bar{N}(t,u)=\sum_{i=1}^{n}\sum_{j=1}^{k_i} N_{ij}(t,u).$$

Theorem 2.4.1 *Assume that condition (2.1) - (2.4) hold. Then* $\sqrt{n}(\hat{\Lambda}_0(t,s) - \Lambda_0(s))$ *converges weakly to a Gaussian random field* ξ *with mean zero.*

Proof. Without loss of generality, we assume that $p = 1$. Notice that

$$\hat{\Lambda}_0(t,s) = \int_0^s \frac{d\bar{M}(t,u)}{nS^{(0)}(\hat{\beta}_t,t,u)} + \int_0^s \frac{S^{(0)}(\hat{\beta}_t,t,u)}{S^{(0)}(\beta_0,t,u)}\lambda_0(u)du.$$

So

$$\sqrt{n}(\hat{\Lambda}_0(t,s) - \Lambda_0(s)) = \frac{1}{\sqrt{n}}\int_0^s \frac{d\bar{M}(t,u)}{S^{(0)}(\hat{\beta}_t,t,u)} + \sqrt{n}\int_0^s \left(\frac{S^{(0)}(\hat{\beta}_t,t,u)}{S^{(0)}(\beta_0,t,u)} - 1\right)\lambda_0(u)du.$$

From the Taylor series expansion of $S^{(0)}(\hat{\beta}_t,t,u)$ at $\beta = \beta_0$, we have

$$S^{(0)}(\hat{\beta}_t,t,u) = S^{(0)}(\beta_0,t,u) + S^{(1)}(\beta^*,t,u)(\hat{\beta}_t - \beta_0),$$

where β^* is on the line segment between $\hat{\beta}_t$ and β_0. So

$$\sqrt{n}\int_0^s \left(\frac{S^{(0)}(\hat{\beta}_t,t,u)}{S^{(0)}(\beta_0,t,u)} - 1\right)\lambda_0(u)du = \sqrt{n}(\hat{\beta}_t - \beta_0)\int_0^s \frac{S^{(1)}(\beta^*,t,u)}{S^{(0)}(\beta_0,t,u)}\,\lambda_0(u)du$$

From Condition (2.2) and the consistency of $\hat{\beta}_t$, we have

$$\frac{S^{(1)}(\beta^*,t,u)}{S^{(0)}(\beta_0,t,u)} \to \frac{s^{(1)}(\beta_0,t,u)}{s^{(0)}(\beta_0,t,u)},\ a.s.$$

From Theorem 2.3.2, $\sqrt{n}(\hat{\beta}_t - \beta_0)$ converges in distribution to a normal random variable. So

$$\sqrt{n}\int_0^s \left(\frac{S^{(0)}(\hat{\beta}_t,t,u)}{S^{(0)}(\beta_0,t,u)} - 1\right)\lambda_0(u)du \xrightarrow{d} \xi\int_0^s \frac{s^{(1)}(\beta_0,t,u)}{s^{(0)}(\beta_0,t,u)}\lambda_0(u)du,$$

where $\xi \sim N(0, W_t)$.

As for the first term, from Condition (2.2), we can show that

$$\frac{1}{\sqrt{n}}\int_0^s \frac{d\bar{M}(t,u)}{S^{(0)}(\hat{\beta}_t,t,u)}$$

is asymptotically equivalent to

$$\frac{1}{\sqrt{n}}\int_0^s \frac{d\bar{M}(t,u)}{s^{(0)}(\beta_0,t,u)},$$

which is a sum of independent zero-mean random variables. It can be shown similarly that the process is tight. Combining this with the classical multivariate central limit theorem gives the desired weak convergence.

Chapter 3

Group sequential analysis of clustered survival data

3.1 Group Sequential analysis based on LWA estimator

Let $0 < t_1 < t_2 < ... < t_N = \tau$ be the predetermined analysis time points. We assume that the type I errors, $\alpha_1, \alpha_2, \cdots, \alpha_N$, spent at interim analyses which satisfy

$$\sum_{i=1}^{N} \alpha_i = \alpha$$

are also determined in advance. Let $\hat{\beta}_k,\ k = 1, ..., N,$ be the LWA estimate of the parameter vector $\beta = (\beta_1, \cdots, \beta_p)^\top$ of model (2.1) at successive analyses. From Theorem 2.3.2,

$$\sqrt{n}(\hat{\beta}_k - \beta_0) \xrightarrow{d} N(0, W_k),\ t \in [0, \tau]$$

where

$$W_k = \Gamma(\beta_0, t_k)^{-1}\Sigma(t_k, t_k)\Gamma(\beta_0, t_k)^{-1}.$$

Since each $\hat{\beta}_k$ is a linear combination of the score function $U(\beta, t, t)$ which is asymptotically a Gaussian random field, it is not hard to postulate that $\hat{\beta}_1, \cdots, \hat{\beta}_N$ follow the multivariate normal distribution.

Theorem 3.1.1 *Under model (2.1), the vectors* $\sqrt{n}(\hat{\beta}_1 - \beta_0), \cdots, \sqrt{n}(\hat{\beta}_N - \beta_0)$ *follow the multivariate normal distribution given by*

$$\begin{cases} \sqrt{n}(\hat{\beta}_k - \beta_0) \sim N(0, W_k), \\ \lim_{n\to\infty} Cov(\sqrt{n}(\hat{\beta}_{k_1} - \beta_0), \sqrt{n}(\hat{\beta}_{k_2} - \beta_0)) = \\ \Gamma(\beta_0, t_{k_1})^{-1}\Sigma((t_{k_1}, t_{k_1}), (t_{k_2}, t_{k_2}))\Gamma(\beta_0, t_{k_2})^{-1},\ k_1 \le k_2. \end{cases} \tag{3.1}$$

Proof. From expression

$$\sqrt{n}(\hat{\beta}_{k_1} - \beta_0) = n^{-1}\mathcal{I}(\beta_1^*, t_{k_1})\frac{1}{\sqrt{n}}U(\beta, t_{k_1}, t_{k_1}),$$
$$\sqrt{n}(\hat{\beta}_{k_2} - \beta_0) = n^{-1}\mathcal{I}(\beta_2^*, t_{k_2})\frac{1}{\sqrt{n}}U(\beta, t_{k_2}, t_{k_2}),$$

where β_1^* is between $\hat{\beta}_{k_1}$ and β_0, and β_2^* is between $\hat{\beta}_{k_2}$ and β_0, it follows that

$$\begin{aligned}
&\lim_{n\to\infty} Cov(\sqrt{n}(\hat{\beta}_{k_1} - \beta_0), \sqrt{n}(\hat{\beta}_{k_2} - \beta_0)) \\
= &\lim_{n\to\infty} n^{-1}\mathcal{I}(\beta_1^*, t_{k_1}) \lim_{n\to\infty} Cov(n^{-1/2}U(\beta_0, t_{k_1}, t_{k_1}), n^{-1/2}U(\beta_0, t_{k_2}, t_{k_2})) \lim_{n\to\infty} n^{-1}\mathcal{I}(\beta_2^*, t_{k_2}) \\
= &\ \Gamma(\beta, 0, t_{k_1})^{-1}\Sigma((t_{k_1}, t_{k_1}), (t_{k_2}, t_{k_2})\Gamma(\beta_0, t_{k_2})^{-1}.
\end{aligned}$$

Assume that the null hypothesis we wish to test is $H_0 : c^\top\beta = \gamma$ for a given $p \times 1$ vector c and scalar constant γ. Let

$$\hat{\xi}_k = \sqrt{n}(c^\top\hat{\beta}_k - \gamma) \tag{3.2}$$

be the test statistic for the kth interim analysis. Then under null hypothesis,

$$\hat{\xi}_k = \sqrt{n}c^\top(\hat{\beta}_k - \beta).$$

So from Theorem 3.1.1, $(\hat{\xi}_1, \cdots, \hat{\xi}_N)$ are asymptotically jointly normal distributed with the covariance matrix

$$\begin{aligned}
&\lim_{n\to} cov(\hat{\xi}_k, \hat{\xi}_l) \\
= &\ c^\top \lim_{n\to} \frac{1}{n} cov((\hat{\beta}_k - \beta_0), (\hat{\beta}_l - \beta_0))c \\
= &\ c^\top \left[\Gamma(\beta_0, t_k)^{-1}\Sigma((t_k, t_k), (t_l, t_l))\Gamma(\beta_0, t_l)^{-1}\right] c,\ k \le l.
\end{aligned} \tag{3.3}$$

Hence Slud & Wei's method [36] can be adopted here to calculate the boundaries of the successive tests. At time t_l, $1 \le l \le N$, the boundary d_l can be determined by the following equations:

$$\begin{cases}
P(|\hat{\xi}_1| \ge d_1) = \alpha_1, \\
P(|\hat{\xi}_1| \le d_1, |\hat{\xi}_2| \ge d_2) = \alpha_2, \\
\vdots \\
P(|\hat{\xi}_1| \le d_1, |\hat{\xi}_2| \le d_2, ..., |\hat{\xi}_N| \ge d_N) = \alpha_N.
\end{cases} \tag{3.4}$$

There are no explicit expressions for values $d_2, .., d_N$. We need numerical calculation to compute these values.

3.2 Group sequential analysis based on partition method

As before we assume that a clinical trial has been planned with n clusters and $k_1, ..., k_n$ members. Let

$$0 < t_1 < \cdots < t_{N-1} < t_N = \tau$$

be some time points of the whole study interval $[0, \tau]$. For $1 \leq k \leq N$, denote the partition of time interval $[0, t_k]$

$$0 = t_0 < t_1 < \cdots < t_{k-1} < t_k \tag{3.5}$$

as Π_k.

For a particular partition Π_k, break the score function $U(\beta, t_k, t_k)$ into k pieces:

$$U_k(\beta) = \begin{pmatrix} U_k^{(1)}(\beta) \\ \vdots \\ U_k^{(k)}(\beta) \end{pmatrix}, \tag{3.6}$$

where

$$U_k^{(l)}(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{k_i} \int_{t_{l-1}}^{t_l} \left(Z_{ij} - \bar{Z}(\beta, t_k, s) \right) dM_{ij}(t_k, s),\ l = 1, \cdots, k.$$

The limiting covariance matrix of $\frac{1}{\sqrt{n}} U_k(\beta)$ is

$$\Sigma_k(\beta) \quad = \quad \begin{pmatrix} \Sigma_k^{(11)}(\beta) & \cdots & \Sigma_k^{(1k)}(\beta) \\ \vdots & \cdots & \vdots \\ \Sigma_k^{(k1)}(\beta) & \cdots & \Sigma_k^{(kk)}(\beta) \end{pmatrix}_{pk \times pk},$$

where

$$\Sigma_k^{(ll')}(\beta) \quad = \quad \lim_{n \to \infty} \frac{1}{n}\, E U_k^{(l)}(\beta) U_k^{(l')}(\beta)^\top = \left(\sigma_{ss'}^{(ll')} \right),$$

and

$$\sigma_{ss'}^{(ll')}(\beta) =$$
$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{u,v}^{k_i} E \int_{t_{l-1}}^{t_l} (Z_{ius} - \bar{Z}_s(\beta, t_k, x)) dM_{iu}(t_k, x) \int_{t_{l'-1}}^{t_{l'}} (Z_{ivs'} - \bar{Z}_{s'}(\beta, t_k, x)) dM_{iv}(t_k, x).$$

Here Z_{ijs} and $\bar{Z}_s(\beta, t_k, u)$ are the sth components of Z_{ij} and $\bar{Z}(\beta, t_k, u)$ respectively.

Let $\hat{\beta}_k$ be the LWA estimator of the true parameter β_0 using data until time t_k. Then we can take an estimate of Σ_k as

$$\hat{\Sigma}_k = \begin{pmatrix} \hat{\Sigma}_k^{(11)} & \cdots & \hat{\Sigma}_k^{(1k)} \\ \vdots & \cdots & \vdots \\ \hat{\Sigma}_k^{(k1)} & \cdots & \hat{\Sigma}_k^{(kk)} \end{pmatrix}_{pk \times pk}, \tag{3.7}$$

where

$$\hat{\Sigma}_k^{(ll')} = \left(\hat{\sigma}_{ss'}^{(ll')}\right),$$

and

$$\hat{\sigma}_{ss'}^{(ll')} = \frac{1}{n}\sum_{i=1}^{n}\sum_{u,v}^{k_i}\int_{t_{l-1}}^{t_l}(Z_{ius} - \bar{Z}_s(\hat{\beta}_k, t_k, x))d\hat{M}_{iu}(t_k, x)\int_{t_{l'-1}}^{t_{l'}}(Z_{ivs'} - \bar{Z}_{s'}(\hat{\beta}_k, t_k, x))d\hat{M}_{iv}(t_k, x).$$

Here

$$\hat{M}_{ij}(t_k, x) = N_{ij}(t_k, x) - \int_0^x Y_{ij}(t_k, x)\exp(\hat{\beta}_k' Z_{ij})\hat{\lambda}_0(x)dx,$$

with

$$\hat{\lambda}_0(x)dx = (\sum_{i=1}^{n}\sum_{j=1}^{k_i} Y_{ij}(t_k, x)\exp(\hat{\beta}_k' Z_{ij}))^{-1} d\bar{N}(t_k, x).$$

Let

$$\hat{\Psi}_k(\beta) = -\frac{1}{n}\frac{\partial U_k(\beta)}{\partial \beta} = \begin{pmatrix} \hat{\psi}_k^{(1)}(\beta) \\ \vdots \\ \hat{\psi}_k^{(k)}(\beta) \end{pmatrix}_{pk \times p}, \tag{3.8}$$

where

$$\hat{\psi}_k^{(l)}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\int_{t_{l-1}}^{t_l}\left(\frac{S_n^{(2)}(\beta, t_k, u)}{S_n^{(0)}(\beta, t_k, u)} - \frac{(S_n^{(1)}(\beta, t_k, u))^{\otimes 2}}{(S_n^{(0)}(\beta, t_k, u))^2}\right)dN_{ij}(t_k, u),\ l = 1, ..., k.$$

Similar to the proof of Theorem 2.3.2, it can be shown that

$$\hat{\psi}_k^{(l)}(\hat{\beta}_k) \to \psi_k^{(l)}(\beta_0), \quad a.s.,$$

where

$$\psi_k^{(l)}(\beta) = \int_{t_{l-1}}^{t_l}\left(\frac{s^{(2)}(\beta, t_k, x)}{s^{(0)}(\beta, t_k, x)} - \frac{(s^{(1)}(\beta, t_k, x))^{\otimes 2}}{(s^{(0)}(\beta, t_k, x))^2}\right)s^{(0)}(\beta, t_k, x)\lambda_0(x)dx.$$

Let

$$\Psi_k(\beta) = \begin{pmatrix} \psi_k^{(1)}(\beta) \\ \vdots \\ \psi_k^{(k)}(\beta) \end{pmatrix}_{pk \times p}, \tag{3.9}$$

then $\hat{\Psi}_k(\hat{\beta}_k)$ is a consistent estimator of $\Psi_k(\beta_0)$.

We assume that $\Sigma_k(\beta_0)$ is positive definite. Since $\hat{\beta}_k$ is a consistent estimator of β_0, $\hat{\Sigma}_k(\hat{\beta}_k)$ is a consistent estimator of $\Sigma_k(\beta_0)$. So when n is large enough, $\Sigma_k(\hat{\beta}_k)^{-1}$ exits. Define the estimating equation

$$\xi(\beta) \;=\; \hat{\Psi}_k(\hat{\beta}_k)^\top \hat{\Sigma}_k(\hat{\beta}_k)^{-1} U_k(\beta) = 0. \tag{3.10}$$

This equation is formed by cutting the score function $U(\beta, t_k, t_k)$ into k pieces and then combining them with standardizing coefficients. We will show in the following that there exists a solution $\tilde{\beta}_k$ to (3.10) which is asymptotically normal distributed. It will be further shown that estimators generated from (3.10) are better estimators in the sense that they are more efficient than the corresponding LWA estimator.

Theorem 3.2.1 *Under the Conditions (2.1) and (2.2), for every* $1 \le k \le N$, *there exists a solution to equation (3.10),* $\tilde{\beta}_k$, *such that*

$$\sqrt{n}(\tilde{\beta}_k - \beta_0) \xrightarrow{d} N(0, W_k), \tag{3.11}$$

where

$$W_k = \Big(\Psi_k(\beta_0)^\top \Sigma_k(\beta_0)^{-1} \Psi_k(\beta_0)\Big)^{-1}. \tag{3.12}$$

Proof. Similar to the proof of Theorem 2.2.1, we can show that

$$\frac{1}{\sqrt{n}} U_k^{(l)}(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{k_i} \int_{t_{l-1}}^{t_l} \left(Z_{ij} - \mu(\beta_0, t_k, u)\right) dM_{ij}(t_k, u) + o_p(1). \tag{3.13}$$

Note that for every fixed i, $\sum_{j=1}^{k_i} \int_{t_{l-1}}^{t_l} \left(Z_{ij} - \mu(\beta_0, t_k, u)\right) dM_{ij}(t_k, u)$ has finite variance, so the first term in (3.13) is a sum of independent random variables with finite variances. By the multivariate central limit theorem, we have

$$\frac{1}{\sqrt{n}} U_k(\beta_0) \xrightarrow{d} N(0, \Sigma_k).$$

Since $\hat{\Psi}_k(\hat{\beta}_k)$ and $\hat{\Sigma}_k(\hat{\beta}_k)$ are respectively consistent estimates for $\Psi_k(\beta_0)$ and $\Sigma_k(\beta_0)$, we have

$$\frac{1}{\sqrt{n}}\hat{\Psi}_k(\hat{\beta}_k)^\top\hat{\Sigma}_k(\hat{\beta}_k)^{-1}U_k(\beta_0) \xrightarrow{d} N(0, \Psi_k(\beta_0)^\top\Sigma_k(\beta_0)^{-1}\Psi_k(\beta_0)). \tag{3.14}$$

In addition, for $1 \le k \le N$ we have

$$\hat{\Psi}_k(\hat{\beta}) = -\frac{1}{n}\frac{\partial U_k}{\partial \beta} \to \Psi_k(\beta_0), \quad \text{in probability.}$$

In fact, the above convergence holds uniformly over $\{\beta^* : \|\beta^* - \beta_0\| \le \epsilon_n\}$ for any $\epsilon_n \to 0$. Then

$$\hat{\Psi}_k(\hat{\beta})^\top\hat{\Sigma}_k(\hat{\beta})^{-1}\left(-\frac{1}{n}\frac{\partial U_k}{\partial \beta}\Big|_{\beta=\beta^*}\right) \to \Psi_k^\top\Sigma_k^{-1}\Psi_k, \quad \text{in probability.} \tag{3.15}$$

Since ψ_{kl}, Σ_k are assumed to be nondegenrate, we have $\Psi_k^\top\Sigma_k^{-1}\Psi_k > 0$. It is standard to show that there exists a solution to the equation in any small and fixed neighborhood of the true parameter β_0, we denote it by $\tilde{\beta}_k$. By Taylor expansion we have,

$$\begin{aligned}\frac{1}{\sqrt{n}}\hat{\Psi}_k(\hat{\beta}_k)^\top\hat{\Sigma}_k(\hat{\beta}_k)^{-1}U(\beta) &= \hat{\Psi}_k(\hat{\beta}_k)^\top\hat{\Sigma}_k(\hat{\beta}_k)^{-1}\left(-\frac{1}{n}\frac{\partial U_k}{\partial \beta}\Big|_{\beta=\beta^*}\right)\sqrt{n}\left(\tilde{\beta}_k - \beta_0\right) \\ &= \hat{\Psi}_k(\hat{\beta}_k)^\top\hat{\Sigma}_k(\hat{\beta}_k)^{-1}\hat{\Psi}_k(\beta^*)\sqrt{n}\left(\tilde{\beta}_k - \beta_0\right), \end{aligned} \tag{3.16}$$

for some β^* lies between β_0 and $\tilde{\beta}_k$. Therefore from (3.16), (3.14) and (3.15) we have

$$\sqrt{n}(\tilde{\beta}_k - \beta_0) \xrightarrow{d} N(0, (\Psi_k^\top\Sigma_k^{-1}\Psi_k)^{-1}).$$

The proof is completed.

Theorem 3.2.2 *Let Q_k be the asymptotic variance-covariance matrix of the LWA estimator $\hat{\beta}_k$. Then*

$$W_k \le Q_k$$

for all $1 \le k \le N$, where $A \le B$ for two matrices means $A - B$ is non-positive definite.

Proof. Let

$$J = \begin{pmatrix} I_{p\times p} & I_{p\times p} & \cdots & I_{p\times p} \end{pmatrix}_{p\times pk}.$$

Then

$$U(\beta_0, t_k, t_k) = JU_k(\beta_0) \quad \text{and} \quad \Gamma(\beta_0, t_k, t_k) = J\Psi_k(\beta_0). \tag{3.17}$$

So

$$\lim_{n\to\infty} \frac{1}{n} Cov(U(\beta_0, t_k, t_k)) = \lim_{n\to\infty} \frac{1}{n} EJU_k(\beta_0)U_k(\beta_0)^\top J^\top = J\Sigma_k(\beta_0)J^\top.$$

Hence

$$\begin{aligned} Q_k &= \Gamma(\beta_0, t_k, t_k)^{-1} \lim_{n\to\infty} \frac{1}{n} Cov(U(\beta_0, t_k, t_k))\Gamma(\beta_0, t_k)^{-1} \\ &= \Gamma(\beta_0, t_k, t_k)^{-1} J\Sigma_k(\beta_0)J^\top(\Gamma(\beta_0, t_k)^{-1})^\top. \end{aligned}$$

So

$$Q_k^{-1} = \Gamma(\beta_0, t_k, t_k)^\top (J\Sigma_k(\beta_0)J^\top)^{-1}\Gamma(\beta_0, t_k, t_k) = \Psi_k(\beta_0)^\top J^\top (J\Sigma_k(\beta_0)J^\top)^{-1} J\Psi_k(\beta_0).$$

The right hand side of above equality can be rewritten as

$$\Psi_k(\beta_0)^\top \Sigma_k(\beta_0)^{-\frac{1}{2}} \Sigma_k(\beta_0)^{\frac{1}{2}} J^\top (J\Sigma_k(\beta_0)J^\top)^{-1} J\Sigma_k(\beta_0)^{\frac{1}{2}} \Sigma_k(\beta_0)^{-\frac{1}{2}} \Psi_k(\beta_0).$$

But

$$\Sigma_k(\beta_0)^{\frac{1}{2}} J^\top (J\Sigma_k(\beta_0)J^\top)^{-1} J\Sigma_k(\beta_0)^{\frac{1}{2}}$$

is a projection matrix. Hence

$$Q_k^{-1} \leq \Psi_k(\beta_0)^\top \Sigma_k(\beta_0)^{-1} \Psi_k(\beta_0) = W_k^{-1}.$$

So

$$W_k \leq Q_k.$$

This completes the proof.

For every fixed $1 \leq k \leq N$, define

$$\tilde{U}_k(\beta) = \hat{\Psi}_k(\hat{\beta}_k)^\top \hat{\Sigma}_k(\hat{\beta}_k)^{-1} U_k(\beta). \tag{3.18}$$

Here $\hat{\beta}_k$ is the estimator based on LWA model with data accumulated till time t_k.

Many group sequential methods, including Pocock's method, O'Brien and Fleming's method and Slud and Wei's method, require that the score process to be approximately

a Brownian Motion, i.e., Gaussian process with independent increments. The next result shows that such an independent increments property holds to the rescaled score process when the method of partitioning is applied and when there are simultaneous entry times.

Theorem 3.2.3 *Let β_0 be the true regression parameter vector. Assuming that $R_{ij} = 0$ for all $1 \le j \le k_i,\ 1 \le i \le n$. Then asymptotically, $\frac{1}{\sqrt{n}}\tilde{U}_1(\beta_0),\ \frac{1}{\sqrt{n}}(\tilde{U}_2(\beta_0) - \tilde{U}_1(\beta_0)), \cdots, \frac{1}{\sqrt{n}}(\tilde{U}_N(\beta_0) - \tilde{U}_{N-1}(\beta_0))$ are independent, i.e.,*

$$\lim_{n\to\infty} cov(\frac{1}{\sqrt{n}}(\tilde{U}_{k_1}(\beta_0) - \tilde{U}_{k_1-1}(\beta_0)), \frac{1}{\sqrt{n}}(\tilde{U}_{k_2}(\beta_0) - \tilde{U}_{k_2-1}(\beta_0))) = 0.$$

Proof. Let $\tilde{U}_0(\beta_0) = 0$. We will show that for any $1 \le k_1 < k_2 \le N$, $\frac{1}{\sqrt{n}}(\tilde{U}_{k_1}(\beta_0) - \tilde{U}_{k_1-1}(\beta_0))$ is independent of $\frac{1}{\sqrt{n}}(\tilde{U}_{k_2}(\beta_0) - \tilde{U}_{k_2-1}(\beta_0))$.

When $R_{ij} = 0,\ 1 \le j \le k_i,\ 1 \le i \le n$, $Y_{ij}(t,s)$ and $N_{ij}(t,s)$ become

$$\begin{aligned} Y_{ij}(t,s) &= I(T_{ij} \ge s, C_{ij} \ge s)I(t \ge s), \\ N_{ij}(t,s) &= I(T_{ij} \le s, T_{ij} \le C_{ij})I(T_{ij} \le t). \end{aligned}$$

So for $t_{k_1} < t_{k_2}$, it is true that $U^{(l)}_{k_1}(\beta_0) = U^{(l)}_{k_2}(\beta_0)$ for all $1 \le l \le k_1$. Therefore, let

$$J = \begin{pmatrix} I_{p\times p} & I_{p\times p} & \cdots & I_{p\times p} \end{pmatrix}_{p\times pk_1}.$$

Then

$$U_{k_1}(\beta_0) = JU_{k_2}(\beta_0) \quad \text{and} \quad \Psi_{k_1}(\beta_0) = J\Psi_{k_2}(\beta_0). \tag{3.19}$$

So

$$\begin{aligned} \frac{1}{n}cov(\tilde{U}_{k_1}(\beta_0)\tilde{U}_{k_2}(\beta_0)') &= \hat{\Psi}^\top_{k_1}(\hat{\beta})\hat{\Sigma}^{-1}_{k_1}(\hat{\beta})U_{k_1}(\beta_0)U_{k_2}(\beta_0)'\hat{\Sigma}^{-1}_{k_2}(\hat{\beta})\hat{\Psi}_{k_2}(\hat{\beta}) \\ &= \hat{\Psi}^\top_{k_1}(\hat{\beta})\hat{\Sigma}^{-1}_{k_1}(\hat{\beta})JU_{k_2}(\beta_0)U_{k_2}(\beta_0)'\hat{\Sigma}^{-1}_{k_2}(\hat{\beta})\hat{\Psi}_{k_2}(\hat{\beta}). \end{aligned}$$

Hence

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}cov(\tilde{U}_{k_1}(\beta_0)\tilde{U}_{k_2}(\beta_0)') &= \Psi^\top_{k_1}(\beta_0)\Sigma^{-1}_{k_1}(\beta_0)J\Sigma_{k_2}(\beta_0)\Sigma^{-1}_{k_2}(\beta_0)\Psi_{k_2}(\beta_0) \\ &= \Psi^\top_{k_1}(\beta_0)\Sigma^{-1}_{k_1}(\beta_0)J\Psi_{k_2}(\beta_0) \\ &= \Psi^\top_{k_1}(\beta_0)\Sigma^{-1}_{k_1}(\beta_0)\Psi_{k_1}(\beta_0). \end{aligned}$$

Consequently,

$$\begin{aligned}
& \lim_{n\to\infty} cov(\frac{1}{\sqrt{n}}(\tilde{U}_{k_1}(\beta_0) - \tilde{U}_{k_1-1}(\beta_0)), \frac{1}{\sqrt{n}}(\tilde{U}_{k_2}(\beta_0) - \tilde{U}_{k_2-1}(\beta_0))) \\
= & \lim_{n\to\infty} \frac{1}{n}[cov(\tilde{U}_{k_1}(\beta_0), \tilde{U}_{k_2}(\beta_0)) - cov(\tilde{U}_{k_1}(\beta_0), \tilde{U}_{k_2-1}(\beta_0)) - \\
& \qquad cov(\tilde{U}_{k_1-1}(\beta_0), \tilde{U}_{k_2}(\beta_0)) + cov(\tilde{U}_{k_1-1}(\beta_0), \tilde{U}_{k_2-1}(\beta_0))] \\
= & W_{k_1}^{-1} - W_{k_1}^{-1} - W_{k_1-1}^{-1} + W_{k_1-1}^{-1} \\
= & 0.
\end{aligned}$$

This completes the proof.

Remark 3.2.1 *In Phase III clinical trials, it is seldom the case that patients enter a study at the same time. It would be desirable that the above result be extended to the more general stagger entry case. Our simulation results in the next section do show this independent increments property when the study entry time is uniformly distributed.*

Sequential analysis can be done based on the standardized estimator $\tilde{\beta}_k$ and Slud & Wei's method [36]) which according to Theorem 3.2.2, improves the efficiency of the test.

Theorem 3.2.4 *For* $1 \le k_1 \le k_2 \le N$,

$$\lim_{n\to\infty} Cov(\sqrt{n}(\tilde{\beta}_{k_1} - \beta_0), \sqrt{n}(\tilde{\beta}_{k_2} - \beta_0)) = W_{k_2}, \tag{3.20}$$

where W_k *is given by (3.12).*

Proof. It can be verified that for every $1 \le k \le N$,

$$\sqrt{n}(\tilde{\beta}_k - \beta_0) = W_k \frac{1}{\sqrt{n}} \tilde{U}_k(\beta_0) + o_p(1). \tag{3.21}$$

So

$$\begin{aligned}
\lim_{n\to\infty} Cov(\sqrt{n}(\tilde{\beta}_{k_1} - \beta_0), \sqrt{n}(\tilde{\beta}_{k_2} - \beta_0)) &= W_k \lim_{n\to\infty} \frac{1}{n} Cov(\tilde{U}_{k_1}(\beta_0), \tilde{U}_{k_2}(\beta_0)) W_{k_2} \\
&= W_{k_1} W_{k_1}^{-1} W_{k_2} = W_{K_2}.
\end{aligned}$$

Corollary 3.2.1 *The vectors* $\sqrt{n}(\tilde{\beta}_1 - \beta_0), \cdots, \sqrt{n}(\tilde{\beta}_N - \beta_0)$ *follow asymptotically the multivariate normal joint distribution with*

$$\lim_{n\to\infty} Cov(\sqrt{n}(\tilde{\beta}_k - \beta_0), \sqrt{n}(\tilde{\beta}_l - \beta_0)) = W_{l\vee k}, \tag{3.22}$$

where W_k *is given by (3.12).*

3.3 Simulation

We evaluate the proposed sequential method by extensive simulations. Parameter estimates and their accuracy, type I error and power are assessed by empirical studies and are compared with existing methods. Results show that the proposed sequential method based on partition gain efficiency in parameter estimating and hypothesis testing.

In our simulation, we consider the case that there are two members in each cluster (i.e., $k_i = 2$) and the two marginal survival times follow the same model, i.e., the two survival times share the same covarites and the covariates have the same effects on the marginal distributions. The shared covariate within cluster is taken to be a binary variable. Let $\{(T_{i1}, T_{i2}), i = 1, \cdots, n\}$ be n paired failure times. They are generated from the bivariate Frank copula family [13], that is, for each fixed i, (T_{i1}, T_{i2}) has the joint distribution function

$$F_\rho(t_1, t_2) = \left(\frac{1}{F_1(t_1)^\rho} + \frac{1}{F_2(t_2)^\rho} - 1 \right)^{-\frac{1}{\rho}}, \ \rho > 0. \tag{3.23}$$

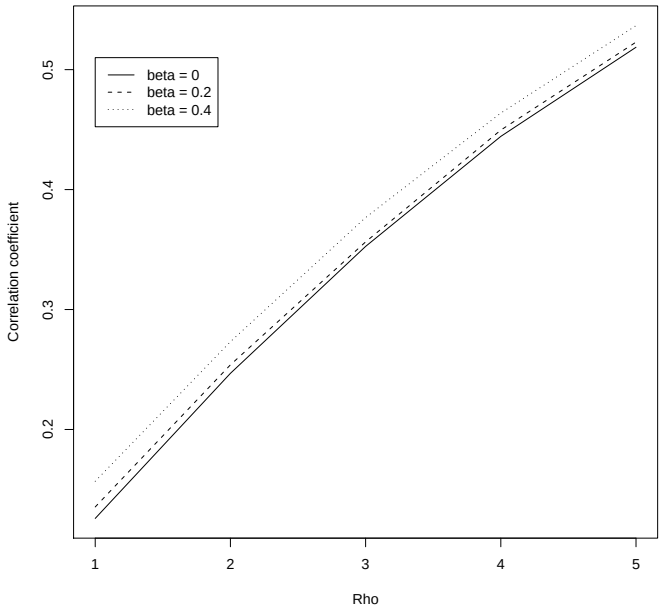
We denote the common covariate of T_{i1} and T_{i2} by z_i. let $\lambda_i = e^{\beta z_i}$. Then T_{i1} and T_{i2} marginally have exponential distributions

$$T_{i1} \sim F_1(t_1) = 1 - e^{-\lambda_i t_1}$$

$$T_{i2} \sim F_2(t_2) = 1 - e^{-\lambda_i t_2},$$

The tuning parameter ρ governs the dependence of the two survival times. The larger ρ is, the more dependent the two variables are. In the simulation studies, we take two values of it, $\rho = 1$ and $\rho = 5$, to represent a situation with small correlation and a large correlation between the two survival times within one cluster. Figure 3.1 demonstrates the role of ρ in characterizing the dependence of the two marginal variables.

Figure 3.1: *Correlation coefficient of* T_1 *and* T_2



To apply our method, we partition the interval $(0, \infty)$ into two intervals $(0, t_1]$ and $(t_1, t_2]$ with $t_1 = 1, t_2 = 8$. Failure times greater than 8 are rare. The censoring variables are generated from uniform distribution $U(0, t_2)$. The members in the same cluster have the same entry time R which follows an uniform distributed $U(0, 2)$.

Table 3.1 lists the parameter estimates, their empirical standard deviations (in parentheses) and their estimates based on asymptotic variances. For this table, when $\rho = 1$ for the Frank's copula, the correlation coefficient of the two survival times within each cluster is small (between 0.128 and 0.191, on average); for $\rho = 5$, the correlation coefficient is larger (between 0.516 and 0.551 on average). From this table, the partial likelihood estimates of the regression coefficient at both stage 1 and stage 2 are consistent, though the estimates at the first stage have much larger variance or standard deviation due to the smaller number of samples used in the first stage. In addition, the empirical variances are greater than the variance estimates based on large sample property. Our proposed partitioning method provides consistent parameter estimates at stage 2. The empirical variances are slightly larger than the ones based on marginal

approach. This is because the sample size, $n = 100$, is not large enough. But our asymptotic variance estimates are smaller than those for the marginal method, indicating an efficiency gain by applying the partitioning method.

Table 3.1: *Parameter estimates and their standard deviations*

		Stage 1		Stage 2: Marginal approach		Stage 2: Partitioning method	
	β	$\hat{\beta}_1$ **	$\hat{\sigma}_1$ *	$\hat{\beta}_2$	$\hat{\sigma}_2$ *	$\tilde{\beta}_2$	$\tilde{\sigma}_2$ *
	-0.6	-0.600 (0.434)	0.415	-0.609 (0.177)	0.170	-0.619 (0.179)	0.166
	-0.4	-0.421 (0.412)	0.392	-0.397 (0.169)	0.165	-0.402 (0.171)	0.162
	-0.2	-0.220 (0.395)	0.378	-0.208 (0.165)	0.163	-0.211 (0.168)	0.159
$\rho = 1$	0.0	0.001 (0.376)	0.362	0.012 (0.162)	0.161	0.013 (0.167)	0.158
	0.2	0.196 (0.372)	0.350	0.202 (0.158)	0.161	0.207 (0.162)	0.157
	0.4	0.405 (0.353)	0.338	0.408 (0.161)	0.161	0.418 (0.167)	0.158
	0.6	0.620 (0.349)	0.332	0.606 (0.162)	0.163	0.619 (0.167)	0.160
	-0.6	-0.622 (0.521)	0.490	-0.616 (0.202)	0.199	-0.626 (0.205)	0.192
	-0.4	-0.439 (0.492)	0.465	-0.411 (0.206)	0.194	-0.418 (0.210)	0.187
	-0.2	-0.197 (0.471)	0.447	-0.203 (0.189)	0.191	-0.209 (0.193)	0.184
$\rho = 5$	0.0	-0.014 (0.455)	0.429	-0.008 (0.195)	0.188	-0.007 (0.200)	0.181
	0.2	0.196 (0.372)	0.350	0.202 (0.158)	0.161	0.207 (0.162)	0.157
	0.4	0.404 (0.409)	0.405	0.415 (0.195)	0.188	0.425 (0.201)	0.181
	0.6	0.617 (0.392)	0.392	0.607 (0.191)	0.191	0.621 (0.194)	0.183

* $\hat{\sigma}_k$ and $\tilde{\sigma}_2$ are estimated asymptotic standard deviations.
** Standard deviations are in parentheses.

We have showed in section 3 of this chapter the independence increment property for the scaled scores with simultaneous entry times. Simulation results show that this property hold with staggered entry times. Figure A.1 shows that the scores are pretty close to a normal distribution, and therefore the parameter estimates are asymptotically normal. Figure A.2 illustrates scatter plot of $\tilde{U}_2$ versus $\tilde{U}_1$, a linear trend relationship of them can be spotted implying the independent increment property is roughly true. This can also be seen from Figure A.3 in which there is no linear relationship between $\tilde{U}_1$ and $\tilde{U}_2 - \tilde{U}_1$. In fact, for this figure, the correlation coefficient of $\tilde{U}_1$ and $\tilde{U}_2 - \tilde{U}_1$ is shown to be 0.018. As another example, we present A.4 - A.6 corresponding figures for $\rho = 5$, consistency, asymptotic normality and independent increment can be seen to be true from these figures.

As a tool to verify the independent increment property of the scaled score process, for two stage partition, we have checked the type I error by using Pocock's boundaries and O'Brien and Fleming boundaries. We have used data generated from sample size $n = 100$, $n = 200$ and $n = 400$ population. It can be seen from Table 3.2 the results for $n = 200$ and $n = 400$ populations are quite close to intended type I error. This reflects that $\tilde{U}_1$ and $\tilde{U}_2$ exhibits the property of scores when data are normal and independent.

Table 3.2: *Empirical type I error of the proposed sequential test when the critical values are determined by Pocock's method and O'Brien-Fleming's method*

		significance level (α)		
n	Methods	0.01	0.05	0.1
100	Pocock	0.016	0.082	0.130
	O'Brien-Fleming	0.017	0.070	0.116
200	Pocock	0.011	0.056	0.115
	O'Brien-Fleming	0.015	0.055	0.121
400	Pocock	0.010	0.055	0.107
	O'Brien-Fleming	0.011	0.057	0.098

At the two stages, a normal z-test is applied. For testing $H_0 : \beta = \beta_0$, the Wald-type test statistics are

$$T_k = \frac{\tilde{\beta}_k - \beta_0}{\hat{\sigma}_k}, \quad k = 1, 2,$$

where $\tilde{\beta}_k$ is the estimate of cox regression coefficient by the partitioning method and $\hat{\sigma}_k$ is the estimated standard deviation of $\tilde{\beta}_k$ at stage k. The overall type I error is taken to be $\alpha = 0.05$ and the significance levels are α_1 and $\alpha_2 = \alpha - \alpha_1$. In our simulation, $\alpha_1 = 0.02$, $\alpha_2 = 0.03$, but other spending function of the overall significance level is also investigated. Critical values, C_{α_1} and C_{α_2}, of the sequential test are determined by the joint normal distribution of (T_1, T_2), utilizing the independent increment property of $\tilde{U}_1$ and $\tilde{U}_2$ by noting the relationship of T_k with $\tilde{U}_k$:

$$T_k = \hat{\sigma}_k \tilde{U}_k.$$

Namely, we assess value of C_{α_k} by Monte Carlo method. We first generate a batch of independent normal variates Z_{ki} from $N(0, \hat{\sigma}_k), k = 1, 2$ and set $\tilde{U}_{i1} = Z_{i1}, \tilde{U}_{i2} = Z_{i1} + Z_{i2}$ and let $T_{i1} = \tilde{U}_{i1}/\hat{\sigma}_1$ and $T_{i2} = \tilde{U}_{i2}/\hat{\sigma}_2$, then C_{α_k} is obtained as the $(1 - \alpha_k) \times 100\%$ quantile of T_{ik}.

Power of the proposed partitioning method are assessed for small ($\rho = 1$) and large ($\rho = 5$) within-cluster correlation. The sequential procedure are as described below. The null hypothesis is $H_0 : \beta = 0.2$. For overall significance level $\alpha = 0.05$ and significance levels, $\alpha_1 = 0.02$ and $\alpha_2 = 0.03$, at the two stages, the critical values are obtained by 10 million Monte Carlo computations under the null. For $\rho = 1$, $C_{\alpha_1} = 2.326$, $C_{\alpha_2} = 2.137$; for $\rho = 5$, $C_{\alpha_1} = 2.326$, $C_{\alpha_2} = 2.142$. Figures A.7 and A.8 are the power of the sequential test based on our partitioning method and the marginal method. It can be seen that there is power gain by applying the partitioning method. The power gain is small for small sample size. We also conducted simulation for larger sample sizes. For $n = 200$, the power gain is more evident, especially so when the two survival times are more dependent.

Chapter 4

Sample size calculation for clustered survival data

4.1 Introduction

Sample size calculation is a crucial step in the design of a clinical trial. Its appropriate estimation ensures that the study has enough power to detect a significant treatment effect. In this chapter we will study how to estimate sample size for clinical trials designed with clustered event data.

4.2 Formula for sample size calculation

In this chapter, we are dealing with clustered survival data that associated with each T_{ij}, there is only one covariate Z_{ij}. We assume that Z_{ij} are identically distributed with mean m and variance σ^2. We assume that the censoring variables $\{C_{ij}\}$ are also identically distributed. We use $Y_{ij}(x), N_{ij}(x)$ and $M_{ij}(\beta, x)$ to represent respectively $Y_{ij}(t, x)$, $N_{ij}(t, x)$ and $M_{ij}(\beta, t, x)$ at $t = \infty$. Let $U(\beta)$ represent $U(\beta, t, t)$ at $t = \infty$.

Let β be the true regression coefficient for the only co-variate in the model. The null hypothesis is:

$$H_0 : \beta = 0.$$

Under this hypothesis, the marginal proportional hazards model becomes

$$\lambda_{ij}(t) = \lambda_0(t),\ i = 1, .., n,\ j = 1, .., k_i.$$

So all T_{ij} have same marginal distribution.

Let $\Sigma(\beta)$ be $\Sigma(\beta, t, s)$ at $t = s = \infty$. The test statistic we use for the null hypothesis is $\hat{\beta}$ which is the solution of $U(\beta) = 0$. It has been shown by Spiekerman and Lin [37] that $\hat{\beta}$ is asymptotically normal and a consistent estimator of β. It can be shown that

the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta)$ is

$$W(\beta)=\frac{1}{\Gamma(\beta)^2}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E(\sum_{j=1}^{k_i}\int_0^{\infty}(Z_{ij}-\mu(\beta,x))dM_{ij}(\beta,x))^2, \tag{4.1}$$

where

$$\Gamma(\beta)=\int_0^{\infty}v(\beta,x)\lambda_0(x)s^{(0)}(x)dx. \tag{4.2}$$

We see from (4.1) that $W(\beta)$ changes when β is different. So here we don't have the situation of a constant variance.

Let α be the significance level of the test for the null hypothesis. The two-sided level α test of $H_0:\beta=0$ against $H_1:\beta\neq 0$ is:

$$\phi(\hat{\beta})=\begin{cases}\frac{\sqrt{n}\ |\hat{\beta}|}{\sqrt{W(0)}}\geq z_{\alpha/2} & reject\ \ H_0\\ otherwise & accept\ \ H_0,\end{cases}$$

here $z_{\alpha/2}$ is the upper $\alpha/2$ profile of the standard normal distribution. Let $\beta_\phi(\beta_1)$ denote the power of ϕ at any $H_1:\beta=\beta_1$, then

$$\begin{aligned}\beta_\phi(\beta_1) &= P_{\beta_1}(\frac{\sqrt{n}\ |\hat{\beta}|}{\sqrt{W(0)}}\geq z_{\alpha/2})\\ &= P_{\beta_1}(\frac{\sqrt{n}\ (\hat{\beta}-\beta_1)}{\sqrt{W(\beta_1)}}\geq z_{\alpha/2}\sqrt{\frac{W(0)}{W(\beta_1)}}-\frac{\sqrt{n}\ \beta_1}{\sqrt{W(\beta_1)}})+\\ &\quad P_{\beta_1}(\frac{\sqrt{n}\ (\hat{\beta}-\beta_1)}{\sqrt{W(\beta_1)}}\leq -z_{\alpha/2}\sqrt{\frac{W(0)}{W(\beta_1)}}-\frac{\sqrt{n}\ \beta_1}{\sqrt{W(\beta_1)}})\\ &= 1-\Phi(z_{\alpha/2}\sqrt{\frac{W(0)}{W(\beta_1)}}-\frac{\sqrt{n}\ \beta_1}{\sqrt{W(\beta_1)}})+\Phi(-z_{\alpha/2}\sqrt{\frac{W(0)}{W(\beta_1)}}-\frac{\sqrt{n}\ \beta_1}{\sqrt{W(\beta_1)}}).\end{aligned}$$

If $\beta_1>0$, neglect the term $\Phi(-z_{\alpha/2}\sqrt{\frac{W(0)}{W(\beta_1)}}-\frac{\sqrt{n}\ \beta_1}{\sqrt{W(\beta_1)}})$, then

$$\beta_\phi(\beta_1)\ \ \geq\ \ 1-\Phi(z_{\alpha/2}\sqrt{\frac{W(0)}{W(\beta_1)}}-\frac{\sqrt{n}\ \beta_1}{\sqrt{W(\beta_1)}}).$$

Let β^* be the power we would like to attain at the alternative hypothesis. Then in order to have

$$\beta_\phi(\beta_1)\geq\beta^*, \tag{4.3}$$

we must have

$$z_{\alpha/2}\sqrt{\frac{W(0)}{W(\beta_1)}} - \frac{\sqrt{n}\ \beta_1}{\sqrt{W(\beta_1)}} \leq -z_{1-\beta^*},$$

thus

$$\sqrt{n} \geq \frac{z_{\alpha/2}\sqrt{W(0)} + z_{1-\beta^*}\sqrt{W(\beta_1)}}{\beta_1}. \tag{4.4}$$

Similarly if $\beta_1 < 0$, we will have

$$\sqrt{n} \geq \frac{z_{\alpha/2}\sqrt{W(0)} + z_{1-\beta^*}\sqrt{W(\beta_1)}}{-\beta_1}. \tag{4.5}$$

Letting n be the minimum positive integer that satisfy (4.4) or (4.5), we get that the no. of clusters required for an experiment with a clustered event data to satisfy the power requirement is:

$$n = \frac{(z_{\alpha/2}\sqrt{W(0)} + z_{1-\beta^*}\sqrt{W(\beta_1)})^2}{\beta_1^2}. \tag{4.6}$$

In practice, β_1 usually represents the minimum treatment effect that is of clinical meaning. Using the average cluster size K_1 we get the total sample size

$$N = nK_1. \tag{4.7}$$

4.3 Estimation of $W(0)$ and $W(\beta_1)$

In formula (4.4), α, β_1 and β^* are pre-determined and known. So the major task in estimating a sample size is to estimate $W(0)$ and $W(\beta_1)$ as accurately as possible.

Let us assume that for a clinical trial with clustered survival data, some pilot studies has been done. If these pilot studies don't show any sign of treatment effect, then this clinical trial will terminate and no further study will be planned. Therefore it is reasonable to assume when we plan a new study, that pilot studies do show clinical efficacy. That is, the data from pilot studies constitutes a population from $\beta = \beta_1 \neq 0$. By this reasoning, data from pilot studies can be used to estimate $W(\beta_1)$ using the following formula:

$$\hat{W}(\beta_1) = \frac{\frac{1}{n}\sum_{j=1}^{n}(\sum_{j=1}^{k_i}\int_0^\infty (Z_{ij} - \bar{Z}(\hat{\beta}, x))d\hat{M}_{ij}(\hat{\beta}, x))^2}{\hat{\Gamma}(\hat{\beta})^2} \tag{4.8}$$

with

$$\hat{\Gamma}(\beta_1) = \frac{1}{n}\sum_{j=1}^{n}\sum_{j=1}^{k_i}\int_0^{\infty}(S^{(2)}(\hat{\beta},x) - \frac{S^{(1)}(\hat{\beta},x)}{S^{(0)}(\hat{\beta},x)})\hat{\lambda}_0(x)dx. \tag{4.9}$$

Usually only when a pilot study shows a treatment effect then a clinical trial continues. So the data of a pilot study represent a data population away from the null hypothesis. It is therefore reasonable to use a pilot study data to estimate $W(\beta_1)$. On the other hand, irrespective of the true regression coefficient of the treatment effect, data from a pilot study reflects the basic data structure, such as the correlatedness of event times. Hence we use (4.8) to estimate $W(0)$ by inserting $\hat{\beta} = 0$.

4.4 Simulation

In our simulation, we consider again the clustered survival data generated from a Frank copula family with n clusters and 2 members in every cluster. In each cluster the two survival times share same covariate and same censoring variable. The covariate is taken to be a binary variable with probability 0.5 to be one or 0.

Let $\{(T_{i1}, T_{i2}), i = 1, \cdots, n\}$ be n paired failure times. Then for each fixed i, (T_{i1}, T_{i2}) has the joint distribution function

$$F_\rho(t_1, t_2) = \left(\frac{1}{F_1(t_1)^\rho} + \frac{1}{F_2(t_2)^\rho} - 1\right)^{-\frac{1}{\rho}},\ \rho > 0. \tag{4.10}$$

If we denote the common covariate of T_{i1} and T_{i2} by z_i, then T_{i1} and T_{i2} have marginal distributions

$$T_{i1} \sim F_1(t_1) = 1 - e^{-\lambda_i t_1}$$
$$T_{i2} \sim F_2(t_2) = 1 - e^{-\lambda_i t_2},$$

where $\lambda_i = e^{\beta z_i}$. The parameter ρ characterizes the dependence of the two survival times. The larger ρ is, the stronger the correlation would be. In simulations, we take two values of ρ, 1 and 5, to represent a situation with small correlation and a large correlation between the two survival times within one cluster.

In the following simulation, we take the alternative regression parameter β_1 to be 0.1, 0.3, 0.5, 0.7 and 0.9 respectively, the type I error to be 0.05 and 0.1, and the power

intended to be 80%, 85% and 90%. At each iteration, we take number of clusters to be $n = 100$ to emulate the real situation at which a pilot study usually does not have a large sample size. For each fixed type I error, β_1 and the intended power β^*, we run 150 iterations. The mean sample size obtained from 150 iterations is then presented in the following two tables.

Table 4.1: *Sample size with rho* $= 1$

		β_1				
	Power	0.1	0.3	0.5	0.7	0.9
	0.8	2009	225	82	43	28
$\alpha = 0.05$	0.85	2312	259	93	49	32
	0.9	2714	298	109	58	38
	0.8	1586	175	64	34	22
$\alpha = 0.10$	0.85	1852	205	72	39	25
	0.9	2200	242	89	47	30

Table 4.2: *Sample size with rho* $= 5$

		β_1				
	Power	0.1	0.3	0.5	0.7	0.9
	0.8	2739	302	108	56	35
$\alpha = 0.05$	0.85	3149	349	124	65	41
	0.9	3688	405	145	77	48
	0.8	2158	237	85	44	28
$\alpha = 0.10$	0.85	2533	278	100	52	33
	0.9	2989	332	120	61	39

Tables A.1 and A.2 list the powers achieved by the sample sizes for $\beta_1 = 0.1, 0.5$ and 0.9.

4.5 Discussion

When Z_{ij} and C_{ij} are all identically distributed, so are $Y_{ij} \exp(\beta' Z_{ij})$ and $Y_{ij} Z_{ij} \exp(\beta' Z_{ij})$. So for $d = 0, 1, 2$

$$s^{(d)}(\beta, x) = K_1 E Z_{11}^d Y_{11}(x) \exp(\beta' Z_{11}).$$

Therefore

$$\mu(\beta, x) = \frac{EZ_{11}Y_{11}(x)\exp(\beta' Z_{11})}{EY_{11}(x)\exp(\beta' Z_{11})}$$

and

$$\Gamma(\beta, x) = K_1 \int_0^\infty (EZ_{11}^{\otimes 2} Y_{11}(x)\exp(\beta' Z_{11}) - \frac{(EZ_{11}Y_{11}(x)\exp(\beta' Z_{11}))^2}{EY_{11}(x)\exp(\beta' Z_{11})})\lambda_0(x)dx.$$

Let

$$\Gamma_1(\beta, x) = \int_0^\infty (EZ_{11}^2 Y_{11}(x)\exp(\beta' Z_{11}) - \frac{(EZ_{11}Y_{11}(x)\exp(\beta' Z_{11}))^2}{EY_{11}(x)\exp(\beta' Z_{11})})\lambda_0(x)dx.$$

Then

$$\Gamma(\beta, x) = K_1\Gamma_1(\beta, x).$$

If we assume that for all $1 \le j \ne k \le k_i, i = 1, ..., n$

$$\begin{aligned} & E(\int_0^\infty (Z_{ij} - \mu(\beta, x))dM_{ij}(\beta, x) \int_0^\infty (Z_{ik} - \mu(\beta, x))dM_{ik}(\beta, x)) \\ = \; & E(\int_0^\infty (Z_{11} - \mu(\beta, x))dM_{11}(\beta, x) \int_0^\infty (Z_{12} - \mu(\beta, x))dM_{12}(\beta, x)), \end{aligned}$$

then

$$W(\beta) = \frac{K_1\Gamma_1(\beta, x) + (K_2 - K_1)E(\int_0^\infty (Z_{11} - \mu(\beta, x))dM_{11}(\beta, x) \int_0^\infty (Z_{12} - \mu(\beta, x))dM_{12}(\beta, x))}{(K_1\Gamma_1(\beta, x))^2}.$$

For $\beta = 0$, Z_{ij} is independent of Y_{ij}, so $\mu(0, x) = m$, $v(0, x) = \sigma^2$ and $s^{(0)}(0, x) = EY_{11}(x)$. Hence

$$\Gamma(0) = K_1\Gamma_1(0) = K_1\sigma^2 E \int_0^\infty EY_{11}(x)\lambda_0(x)dx = K_1\sigma^2 EN_{11}(\infty) = K_1\sigma^2 P(T_{11} \le C_{11}),$$

and

$$E \int_0^\infty (Z_{11} - m)dM_{11}(0, x) \int_0^\infty (Z_{12} - m)dM_{12}(0, x) = cov(Z_{11}, Z_{12})EM_{11}(0, \infty)M_{12}(0, \infty).$$

So

$$\begin{aligned} W(0) &= \frac{K_1\sigma^2 P(T_{11} \le C_{11}) + r(K_2 - K_1)EM_{11}(0, \infty)M_{12}(0, \infty)}{(K_1\sigma^2 P(T_{11} \le C_{11}))^2} \\ &= = \frac{1}{K_1\sigma^2 P(T_{11} \le C_{11})} + \frac{r(K_2 - K_1)EM_{11}(0, \infty)M_{12}(0, \infty)}{(K_1\sigma^2 P(T_{11} \le C_{11}))^2}, \end{aligned}$$

where $\sigma^2 = var(Z_{11})$ and $r = cov(Z_{11}, Z_{12})$.

When $k_i = 1$ for all $i = 1, .., n$, (4.6) reduces to

$$n_1 = \frac{(\frac{z_{\alpha/2}}{\sqrt{\Gamma_1(0)}} + \frac{z_{\beta*}}{\sqrt{\Gamma_1(\beta_1)}})^2}{\beta_1^2}. \tag{4.11}$$

Here n_1 is the no of clusters required for experiments with one subject per cluster. In order to have

$$n \leq n_1,$$

we must have

$$z_{\alpha/2}\sqrt{W(0)} + z_{\beta^*}\sqrt{W(\beta_1)} \leq \frac{z_{\alpha/2}}{\sqrt{\Gamma_1(0)}} + \frac{z_{\beta*}}{\sqrt{\Gamma_1(\beta_1)}},$$

which will be satisfied when

$$W(0) \leq \frac{1}{\Gamma_1(0)} \quad \text{and} \quad W(\beta_1) \leq \frac{1}{\Gamma_1(\beta_1)}. \tag{4.12}$$

To satisfy (4.12), we must have

$$\frac{cov(Z_{11}, Z_{12})E(M_{11}(0,\infty)M_{12}(0,\infty))}{\Gamma_1(0)} \leq \frac{K_1^2 - K_1}{K_2 - K_1}, \tag{4.13}$$

and

$$\frac{E(\int_0^\infty (Z_{11} - \mu(\beta_1, x))dM_{11}(\beta_1, x) \int_0^\infty (Z_{12} - \mu(\beta_1, x))dM_{12}(\beta_1, x))}{\Gamma_1(\beta_1)} \leq \frac{K_1^2 - K_1}{K_2 - K_1}. \tag{4.14}$$

The right hand side of (4.13) and (4.14) are related with the relationship between K_1 and K_2, while the left hand side deals with the relationship between covariates, and event times. The two sides are not intrinsically related. Depending on how large the relatedness between covariates and event times is, either side of the inequalities may be bigger than the other side.

Some special cases are:

1. All $Z_{ij}, i = 1, .., n, j = 1, .., k_i$ are independent.

In this case, the left hand of (4.13) and (4.14) are zero, and

$$w(0) = \frac{1}{K_1\Gamma_1(0)} \quad \text{and} \quad W(\beta_1) = \frac{1}{K_1\Gamma_1(\beta_1)}.$$

So

$$n = \frac{(\frac{z_{\alpha/2}}{\sqrt{\Gamma_1(0)}} + \frac{z_{\beta*}}{\sqrt{\Gamma_1(\beta_1)}})^2}{K_1\beta_1^2} = \frac{n_1}{K_1}.$$

2. k_i are constant, $k_i = K, i = 1, .., n$.

When the condition is true, $K_2 = K_1^2 = K^2$, so the right hand sides of (4.13) and (4.14) are 1. It is easy to show, under our assumptions, that the left hand sides of (4.13) and (4.14) are less than 1. So in this case, we have

$$n \leq n_1.$$

3. k_i are not all equal.

In this case, $K_1^2 < K_2$. So the right hand sides of (4.13) and (4.14) are strictly less than 1.

When $Z_{ij} = Z_{i1}\ a.s.,\ j = 1, .., k_i$, that is, all members in the same cluster get the same treatment, we have $Cov(Z_{11}Z_{12}) = \sigma^2$. If the treatment is the primary factor that affects the failure of patients, then $T_{ij} \approx T_{i1}$, so

$$EM_{11}(0)M_{12}(0) \approx EM_{11}(0)^2 = P(T_{11} \leq C_{11}),$$

and

$$E(\int_0^\infty (Z_{11} - \mu(\beta_1, x))dM_{11}(\beta_1, x) \int_0^\infty (Z_{12} - \mu(\beta_1, x))dM_{12}(\beta_1, x)) = \Gamma_1(\beta_1).$$

So the left hand sides of (4.13) and (4.14) are equal to 1. In this situation, we have

$$n \approx \frac{K_2}{K_1^2} n_1 > n_1.$$

This is a strange result. From intuitive thinking, since within a cluster, every event time is a representative of the others, we have a case equivalent to that of one event time per subject. So the sample size required in this case should be no more than that in one event time per subject case.

This phenomenon can be explained as this. Suppose $\{X_{ij}, i = 1, ..., n, j = 1, ..., k_i\}$ is a sequence of random variables which satisfies that $X_{ij} = X_{i1},\ a.s,\ i = 1, ..., n$, and

X_{ik} is independent of X_{jl} when $i \neq j$. $k_1, ..., k_n$ is a sequence of constants. Each X_{i1} has mean μ and variance σ^2.

Define

$$\bar{X}_1 = \frac{1}{k_1 + k_2 + ... + k_n} \sum_{i=1}^{n} \sum_{j=1}^{k_i} X_{ij},$$
$$\bar{X}_2 = \frac{1}{n} \sum_{i=1}^{n} X_{i1}.$$

Then

$$E\bar{X}_1 = \mu, \quad var\bar{X}_1 = \frac{k_1^2 + k_2^2 + ... + k_n^2}{(k_1 + ... + k_n)^2}\sigma^2,$$
$$E\bar{X}_2 = \mu, \quad var\bar{X}_2 = \frac{\sigma^2}{n}.$$

From Jensen's Inequality, when k_i are not all equal,

$$(k_1 + ... + k_n)^2 < n(k_1^2 + k_2^2 + ... + k_n^2),$$

we get

$$var\bar{X}_1 = \frac{k_1^2 + k_2^2 + ... + k_n^2}{(k_1 + ... + k_n)^2}\sigma^2 > \frac{1}{n}\sigma^2 = var\bar{X}_2.$$

This is a situation that is unlikely to occur in reality, since usually treatment is not the only factor that affect the failure of a subject.

References

[1] Andersen, P. K., Borgan, Q., Gill, R. D. and Keiding, N., *Statistical Models Based on Counting Processes*, Springer, 1997

[2] Billias, Y., Gu, G. and Ying, Z., Towards a general asymptotic theory for Cox model with staggered entry, *The Annals of Statistics*, Vol. 25, No.2, 662-682, 1997

[3] Jennison C. and Turnbull B., *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall/CRC, 1999

[4] Clayton, D.G. and Cuzick, J., Multivariate generalizations of the proportional hazards model (with discussion), *J.R. Statist. Soc.* A 148, 82-117, 1985

[5] Clayton, D.G., 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, 1978

[6] Crowder, M., A multivariate distribution with Weibull connections, *Journal of the Royal Statistical Society*, Ser.B, 51, 93-107, 1989

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

[8] Cox, D.R., Partial likelihood, *Biometrika*, 62, 269-276, 1974

[9] Cox, D.R. and Oakes, D.A., *Analysis of Survival Data*, New York: Chapman & Hill, 1984

[10] Cui, W. Q., PhD Dissertation, University of Science and Technology of China, 2005. Unpublished.

[11] Cheng, S.C., Wei, L.J. and Ying, Z., Analysis of transformation models with censored data, *Biometrika*, 82,4, 835-845, 1995

[12] Fleming, T.R. and Harrington, D.P., *Counting processes and survival analysis*, John Wiley & Sons, 1991

[13] Frank, M. T., On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequations Math.* 19, 194-226, 1979

[14] Glidden, D.V. and Vittinghoff, E., Modelling clustered survival data from multicentre clinical trials, *Statistics in medicine*, 23, 369-388, 2004

[15] Gu, M. G. and Lai, T. L., *Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials*, Ann. Statist., 19, 1403-1433, 1991

[16] Hougaard, P., *Analysis of Multivariate Survival Data*, Springer, 2000

[17] Hougaard, P., Survival models for heterogeneous populations derived from stable distributions, *Biometrika*, 73, 2, 387-396, 1986

[18] Hougaard, P., A class of multivarite failure time distributions, *Biometrika*, 73, 671-678, 1986 b, (cCorrection, 75, 395)

[19] Hougaard, P., Fitting a multivariate failure time distribution . *IEEE Trans. on Reliability*, 38, 444-448, 1989

[20] Jones, D., Sequential forms of the log rank and modified Wilcoxon tests for censored Data, *Biometrika*, 66, 105-113, 1979

[21] Klein, J., Semiparametric estimation of random effects using the Cox model based on the EM algorithm, *Biometrics*, 48, 795-806, September, 1992

[22] Lan, K. K. G. and DeMets, D., Discrete sequential boundaries for clinical trials, *Biometrika*, 70, 659-663, 1983

[23] Lu, J.-C. and Bhattacharyya, G.K., Some new constructions of bivariate Weibull models, *Ann. Inst. Statist. Math.*,42, 543-559, 1990

[24] Lin, D. Y., Cox Regression analysis of multivariate failure time data: the marginal approach, *Statistics in Medicine*, Vol 13, 2233-2247, 1994

[25] Lee, E.W., Wei, L.J., and Amato, D.A., Cox-Type regression analysis for large number of strata groups of correlated failure time observations, *J.P.Klein and P.K.Goel(eds), Survival Analysis: State of the Arts*, 237-247, 1992

[26] Manatunga, A. K., and Oakes, D., Efficiency of indenpendence working analysis of correlated survival data with general cluster size, *Statistics Neerlandica*, Vol 58, nr. 3, 313-321, 2004

[27] Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sorensen, T.I.A. *A Counting process approach to maximum likelihood estimation in frailty models*, Scand. J. Statist. 19. 25-34, 1992

[28] Oakes, D., A Concordance test for independence in the presence of censoring, *Biometrics*, 38, 451-455, 1982a

[29] Oakes, D., Bivariate survival models induced by frailties, *J. Am. Statist. Asso.*, 84, 487-493, 1989

[30] Oakes, D., Multivariate survival distributions. *Nonparametric Statistics*, 3, 343-354, 1994

[31] O'Brien, P.C. and Fleming, T.R., A multiple testing procedure for clinical trials. *Biometrics,* 35, 549-556, 1979.

[32] Pollard, D., *Empirical Processes: Theory and Applications,* IMS, Hayward, CA, 1990

[33] Prentice, R.L. and Cai, J., Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, 79, 495-512, 1978

[34] Pocock, S. J., Group sequential methods in the design and analysis of clinical trials, *Biometrika*, 64, 2, 191-199, 1977

[35] Sellke, T. and Siegmund, D., Sequential analysis of the proportional hazards model, *Biometrika*, 70, 315-326, 1983

[36] Slud, W. and Wei, L. J., Two-Sample repeated significance tests based on the modified Wilcoxon statistic, *Journal of the American Statistical Association*, 77, 862-868, 1982

[37] Spiekerman, C. F., and Lin, D. Y., Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93, 1164-1175, 1998

[38] Tsiatis, A.A., The asympototic joint distribution of the efficiet scores test for the proportional hazards model calculated over time. *Biometrika*, 68, 1, 311-315, 1981

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

[40] Whitmore, G.A. and Lee, M.-L.T., A multivariate survival distributions generated by an inverse Gaussian mixture of exponentials, *Technometrics*, 33, 39-50, 1991

[41] Wei, L.J., Lin, D.Y. and Weissfeld, L., Regression analysis of multivariate incomplete failure time data by modeling marginal distributions *Journal of American Statistical Assocition*, 84, 1065-1073, 1989

[42] Yang, Y. N., PhD Dissertation, Rutgers University, 2000. Unpublished.

Appendix A

Tables and Figures

Table A.1: *Power achieved by sample size with rho* $= 1$

		β_1								
		0.1	0.1	0.1	0.5	0.5	0.5	0.9	0.9	0.9
$\alpha = 0.05$	Power intended	0.8	0.85	0.90	0.8	0.85	0.90	0.8	0.85	0.90
	Power achieved	0.812	0.858	0.907	0.813	0.852	0.90	0.812	0.856	0.904
$\alpha = 0.10$	Power intended	0.8	0.85	0.90	0.8	0.85	0.90	0.8	0.85	0.90
	Power achieved	0.797	0.869	0.905	0.792	0.834	0.906	0.793	0.854	0.897

Table A.2: *Power achieved by sample size with rho* $= 5$

		β_1								
		0.1	0.1	0.1	0.5	0.5	0.5	0.9	0.9	0.9
$\alpha = 0.05$	Power intended	0.8	0.85	0.9	0.8	0.85	0.90	0.8	0.85	0.90
	Power achieved	0.797	0.857	0.897	0.795	0.835	0.898	0.792	0.874	0.91
$\alpha = 0.10$	Power intended	0.8	0.85	0.9	0.8	0.85	0.90	0.8	0.85	0.90
	Power achieved	0.808	0.873	0.897	0.788	0.845	0.908	0.795	0.838	0.914

Figure A.1: *Distributions of scores* $\tilde{U}_1$ *and* $\tilde{U}_2$ *(*$\rho = 1$*)*

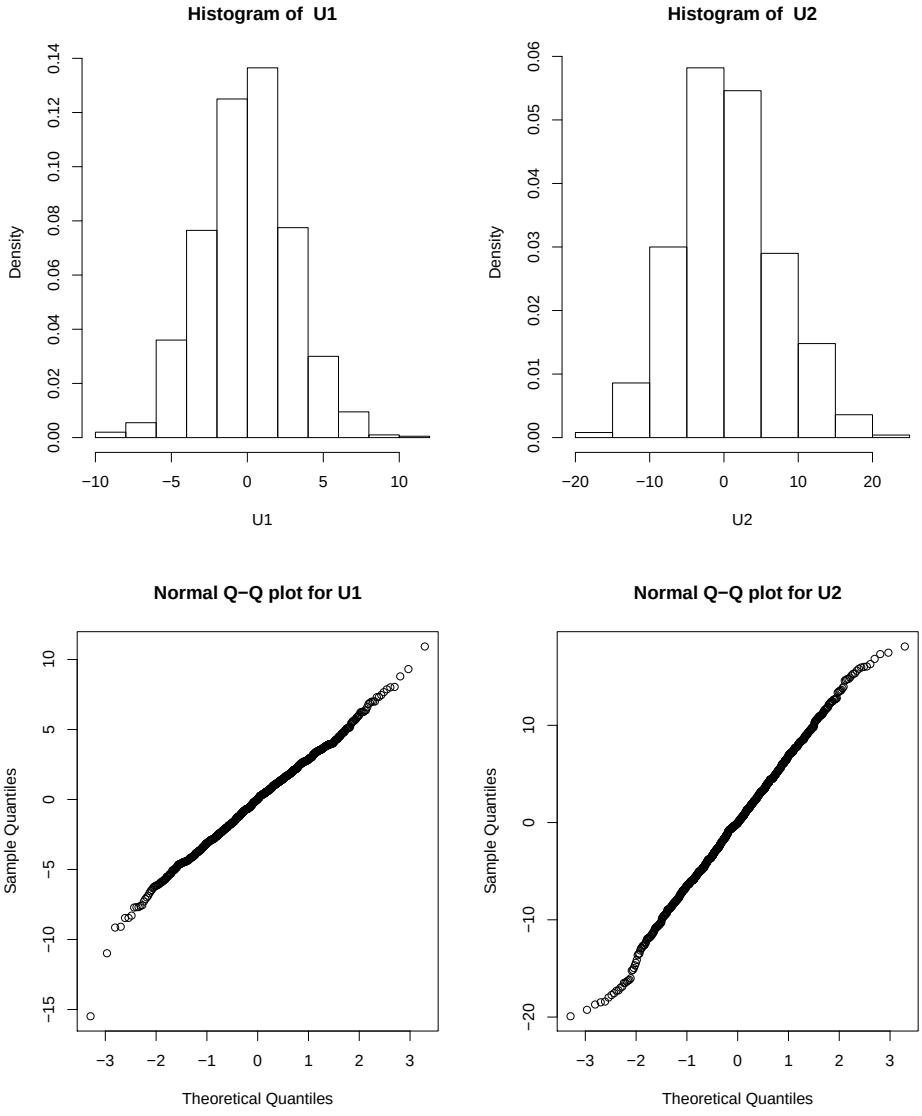


Figure A.2: *Scatter plot of* $\tilde{U}_2$ *versus* $\tilde{U}_1$ *(*$\rho = 1$*)*

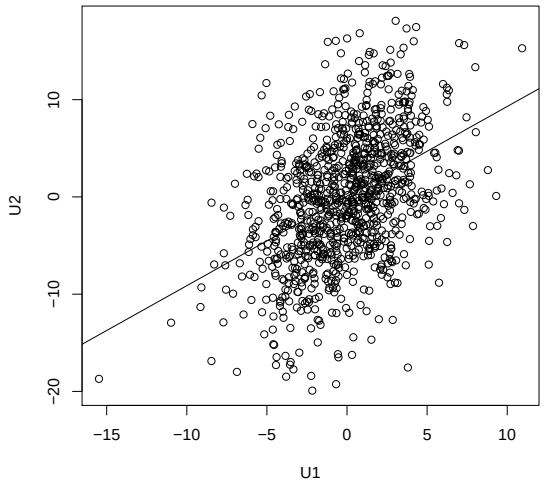


Figure A.3: *Scatter plot of* $\tilde{U}_2 - \tilde{U}_1$ *versus* $\tilde{U}_1$ *(*$\rho = 1$*)*

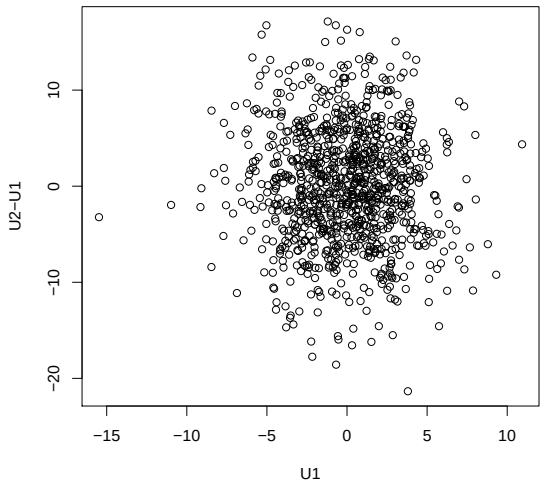


Figure A.4: *Distributions of scores* $\tilde{U}_1$ *and* $\tilde{U}_2$ *(*$\rho = 5$*)*

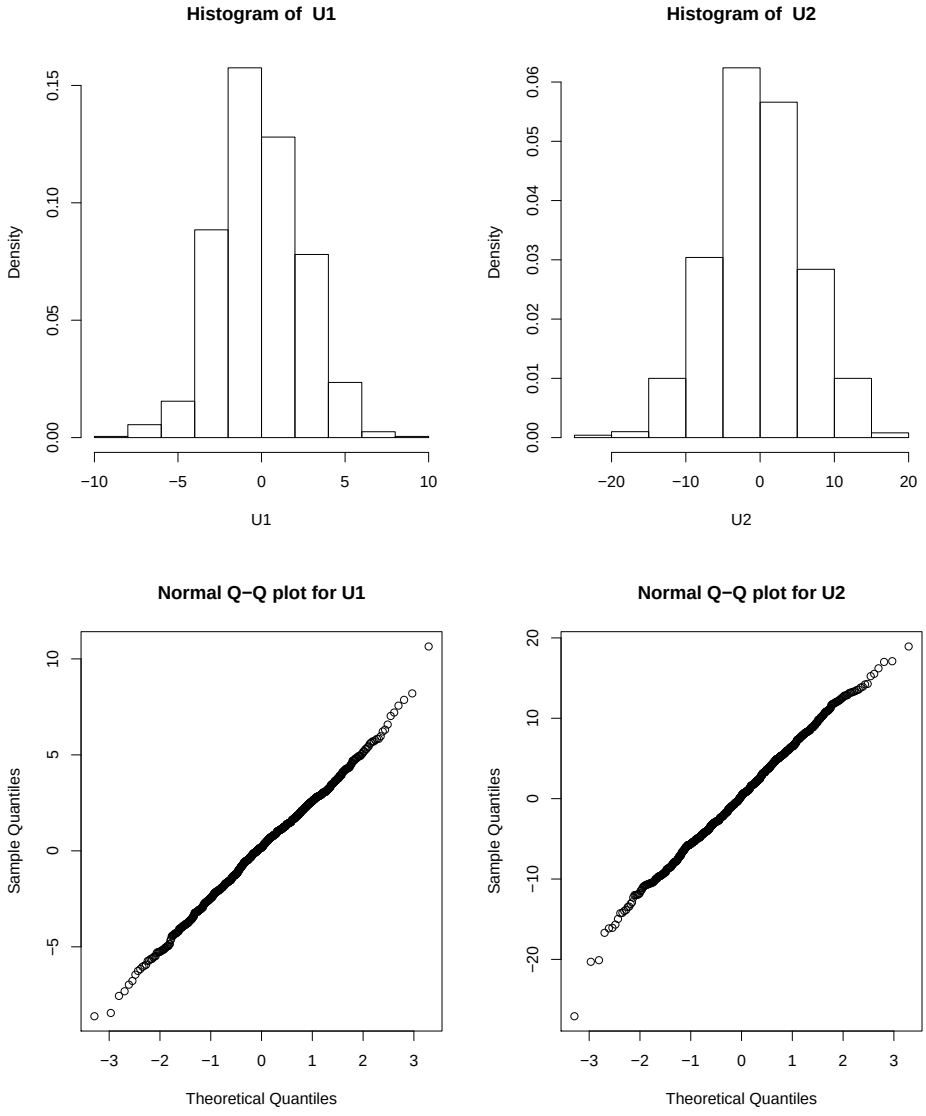


Figure A.5: *Scatter plot of* $\tilde{U}_2$ *versus* $\tilde{U}_1$ *(*$\rho = 5$*)*

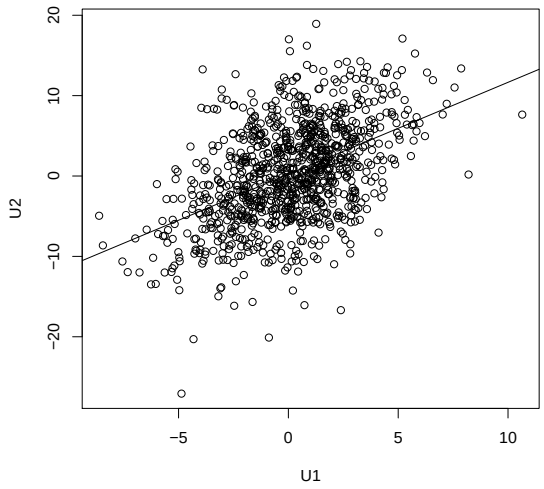


Figure A.6: *Scatter plot of* $\tilde{U}_2 - \tilde{U}_1$ *versus* $\tilde{U}_1$ *(*$\rho = 5$*)*

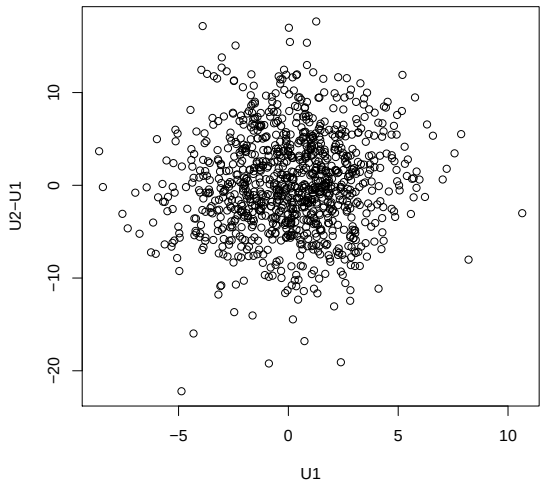


Figure A.7: *Power by marginal method and by partition method ($\rho = 1$)*

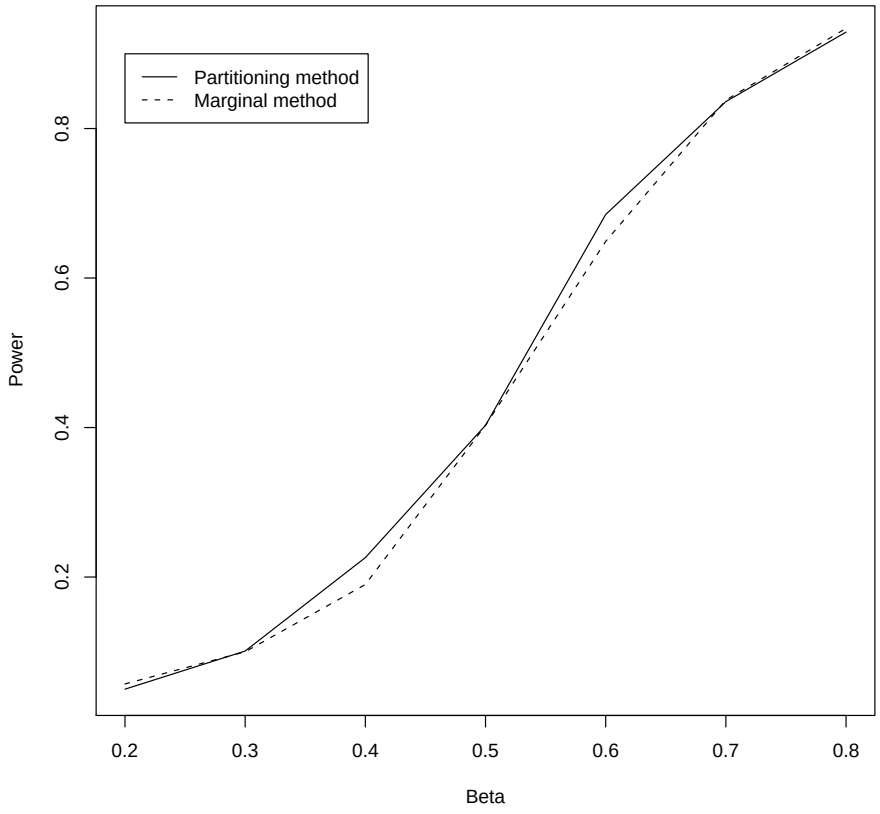
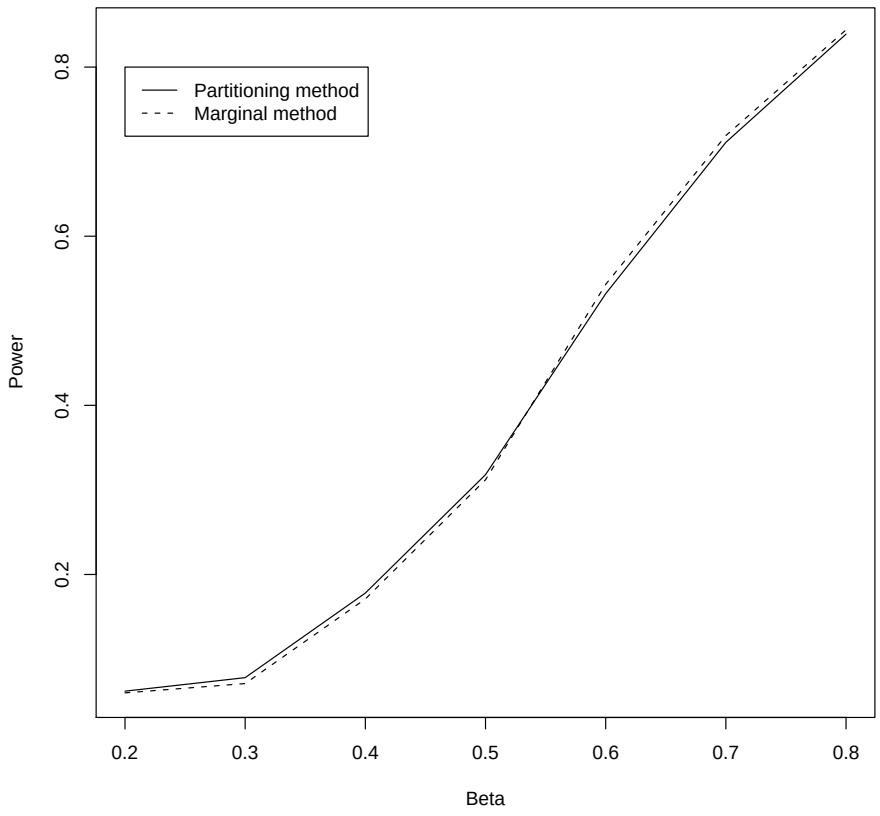


Figure A.8: *Power by marginal method and by partition method ($\rho = 5$)*



Vita

Bo Hou

1978-83 University of Science and Technology of China, Hefei, Anhui, P.R.China. Majored in Mathematics.

1983 B.S., University of Science and Technology of China, Hefei, Anhui, P.R.China.

1983-86 Institute of Applied Mathematics, Academia Sinica, Beijing, P.R.China. Majored in Operational Research.

1986 M.S., Institute of Applied Mathematics, Academia Sinica, Beijing, P.R.China.

1986-88 Teaching Assistant, Department of Mathematics, Tsinghua University, Beijing, P.R.China.

1988-89 Fellowship, Department of Civil Engineering, Princeton University.

1989-1995 Teaching Assistant, Graduate School, Rutgers University.

1995-1996 Consultant Programmer,

1996-1998 Programmer, New Jersey,

1998-1999 Biostatistician, Johnson & Johnson, New Jersey.

1999-2002 Biostatistician, Johnson & Johnson, New Jersey.

2002-now Instructor, Department of Statistics and Finance, University of Science and Technology of China, Hefei, Anhui Province, P.R.China.