Statistical Inference for the Treatment Effect in Cancer Clinical Trials

by

Shan Jiang

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
May 2011

Dedicated to Ran and Our Parents

Abstract

Randomized clinical trials provide the best evidence on the effect of treatment studied. There are different types of measures on the treatment effect, depending on the endpoints of the trials. For a given measure, based on the data from clinical trials, various statistical procedures are available for the inference of the treatment effect in terms of this measure.

In a cancer clinical trial with a time to an event as the endpoint, hazard ratio is a popular measure for the relative difference between treatment groups. Most current statistical inference procedures for hazard ratio rely on the proportional hazard assumption, which may not be applicable to practice when it does not hold. Nonparametric confidence intervals for the hazard ratio have been proposed based on the asymptotic normality of the kernel estimate for the hazard ratio, but they were found not very satisfactory in the simulation studies. In the first part of this thesis, the empirical likelihood method is used to construct the confidence interval for the time-dependent hazard ratio. The asymptotic distribution of the empirical likelihood ratio is derived and simulation studies are conducted to evaluate the proposed method.

It was also argued that the measure of the relative treatment effect based on the hazard ratio may be difficult to understand by clinicians. An alternative measure called probabilistic index was suggested and the C-index was proposed to estimate this

index. However, it was pointed out recently that the expected value of the estimate based on the C-index may be far removed from the true index. In the second part of this thesis, assuming a semi-parametric density ratio model, two new estimates based on respectively the conditional likelihood and weighted empirical likelihood are proposed. Associated confidence intervals are also derived based on the bootstrap re-sampling method. The proposed inference procedures are evaluated by Monte-Carlo simulations and applied to the analysis of data from a clinical trial on early breast cancer.

After primary analysis including all patients is completed in clinical trials, analysis by subgroups defined based on covariates of patients is often of interest to assess the homogeneity of treatment effects over these subgroups. The treatment-covariate interaction is usually used for this assessment. In the last part of this thesis, a non-parametric measure is used to quantify the interaction between treatments and binary covariates in the presence of censoring. Asymptotic distribution of the interaction estimates are derived and the bootstrap method is applied to construct the confidence intervals. The proposed approaches are also evaluated and compared by Monte-Carlo simulations and applied to a real data set from clinical trial.

Acknowledgments

First and foremost, I would like to give thanks to my Ph. D. supervisor, Prof. Dongsheng Tu. During my Ph. D. career, Prof. Tu has provided me much more than I had expected, not only knowledge, advices and research ideas as a teacher gives to a student, but also support, guidance and encouragement as a father gives to a son. This thesis is also full of his efforts. Without the time and energy he has devoted, this thesis would not have been finished.

Many thanks to my Ph. D. committee members: Prof. B. J. Diak, Prof. F. Alajaji, Prof. G. K. Takahara, Dr. C. J. O'Callaghan and Prof. H. Liang. Without the valuable advices from them, this thesis would not have been nearly as good.

I am very grateful to Prof. Paul Peng and Prof. Bingshu Chen for their comments and suggestions on my thesis writing. I also want to take this opportunity to thank Prof. Devon Lin and Prof. Wenyu Jiang, for the encouragement during a particular period of my study.

I also would like to express my sincere gratitude to Dr. Xiao Li, who patiently helped me with my research and had shared with me many thoughts and experiences during my tenure as a doctoral student. I am indebted to my student colleagues, Hui Song, Yi Niu, Huan Jin, Tracy Zhang, Xiaoqun Sun and Liting Zhu, for the great research environment they created as well as the assistance and support they provided

in many different ways.

Finally, I would especially like to acknowledge my parents, Houping Jiang and Mei Ke, for their endless love, support and encouragement. My family has always been a joyful inspiration for completion of my thesis and throughout my life.

Statement of Originality

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Table of Contents

Abstract . . . . . **ii**

Acknowledgments . . . . . **iv**

Statement of Originality . . . . . **vi**

Table of Contents . . . . . **vii**

List of Tables . . . . . **x**

List of Figures . . . . . **xii**

Chapter 1 Introduction **1**

1.1 Clinical Trials and Treatment Effect Assessment 1

1.2 Inference on Treatment Effects in Cancer Clinical Trials 2

1.3 Assessment of the Heterogeneity of Treatment Effects 4

1.4 Organization of This Thesis 4

Chapter 2 Measures of Treatment Effect and Their Inferences in Cancer Clinical Trials **6**

2.1 The Hazard Ratio and Its Inferences 7

2.2 The Probabilistic Index and Its Inferences 12

2.3 Non-parametric Assessment of the Treatment-Covariate Interaction . 15

Chapter 3 Empirical Likelihood Procedures for Survival Data · · · 20

3.1 Empirical Likelihood Confidence Interval of Survival Function for Censored Data 20

3.2 Empirical Likelihood Confidence Intervals for Hazard and Density Functions under Right Censorship 23

3.3 Empirical Likelihood Confidence Band for the Difference of Two Survival Functions under Right Censorship 26

3.4 Simultaneous Confidence Bands for Ratios of Survival Functions via Empirical Likelihood 28

Chapter 4 The Density Ratio Model and Its Inference · · · · · 30

4.1 Inference for the Treatment Effect with the Density Ratio Model . . . 31

4.2 Inference on the Density Ratio Model under the Random Censorship via a Conditional Empirical Likelihood Method 34

4.3 Inference on the Density Ratio Model under the Random Censorship via a Weighted Empirical Likelihood Method 38

Chapter 5 Empirical Likelihood Confidence Intervals for the Hazard Ratio Function · · · · · 41

5.1 Introduction 41

5.2 Empirical Likelihood Ratio and Confidence Interval for the Hazard Ratio 43

5.3 Simulation Studies 49

5.4 Application to a Data Set from Cancer Clinical Trial 51

5.5 Conclusions and Discussions 54

5.6 Proof of Theorem 5.1 56

Chapter 6 Inference of Treatment Effect Based on the Probabilistic Index 66

6.1 Introduction 66

6.2 Estimation and Confidence Interval of θ 71

6.3 Simulation Studies 77

6.4 Application to Data from a Clinical Trial 83

6.5 Conclusions and Discussions 85

Chapter 7 Inference on Treatment-Covariate Interactions in Cancer Clinical Trials 88

7.1 Introduction 88

7.2 Non-parametric Assessment of the Treatment-Covariate Interaction . 91

7.3 Comparison of Treatment-Covariate Interactions 94

7.4 Simulation Studies 97

7.5 Application to Data Set from a Clinical Trial 98

7.6 Conclusions and Discussions 108

7.7 Proof of (7.2) and (7.9) 109

Chapter 8 Summary and Future Work 116

List of Tables

5.1 Comparison of Estimated Coverage Probabilities of Confidence Intervals for Hazard Ratio ($h_1(t) = \lambda$) 52

5.2 Comparison of Estimated Lengths of Confidence Intervals for Hazard Ratio ($h_1(t) = \lambda$) 53

5.3 Estimates of the Hazard Ratio of Treatment CEF over CMF and associated 95% Confidence Intervals 54

6.1 Configurations of Parameters for Tables 6.2-6.4 79

6.2 Bias, Standard Deviation and Mean Square Error of Estimates $\hat{\theta}_c$, $\hat{\theta}_w$ and $\hat{\theta}_e$ 80

6.3 Comparison of Estimated Coverage Probabilities of Confidence Intervals for θ 81

6.4 Comparison of Estimated Lengths of Confidence intervals for θ 82

7.1 Simulation Numbers and Distribution Assumptions for Tables 7.2 and 7.3 99

7.2 Comparison of Estimated Coverage Probabilities and Interval Lengths among I_{Ae} I_{Ac} and I_{Aw} 100

7.3 Comparison of Estimated Coverage Probabilities and Interval Lengths among I_e I_c and I_w 101

7.4 Comparison of Estimation and 95% Confidence Interval among Three Different Approaches 107
7.5 Comparison of P-values among Four Different Approaches 107

List of Figures

6.1 Survival curves for the CMF treatment group based on respectively the Kaplan-Meier method and density ratio model 86

7.1 Kaplan-Meier Curves in HER2 Over-expressed Patients 103

7.2 Kaplan-Meier Curves in HER2 Normal Patients 104

7.3 Kaplan-Meier Curves in TOP2A Alteration Patients 105

7.4 Kaplan-Meier Curves in TOP2A Normal Patients 106

Chapter 1

Introduction

1.1 Clinical Trials and Treatment Effect Assessment

Randomized clinical trials provide the best evidence in the assessment of the effect of a new treatment procedure or intervention against a standard treatment or placebo. The treatment effect is usually quantified by the difference (absolute or relative) in a predefined outcomes between the new and standard treatment groups. Because subjects are randomized to receive treatments under comparison in a clinical trial, treatment groups are balanced in terms of all the characteristics of the subjects. Therefore, the difference in outcomes observed between treatment groups can be attributed only to the different treatments subjects received.

There are many different measures which can be used to assess the treatment effect. For example, when outcomes are continuous, such as the reduction of blood pressure, the difference in mean outcomes between treatment groups is commonly

used as a measure of treatment effect. In cases, where outcomes are binary, such as disappearance of a symptom, the treatment effect can be measured by the difference or the ratio of the event rates. In clinical trials with a time to an event as the outcome, such as disease-free or overall survival, difference in event rates at a given time, difference of median survival times or hazard ratio are often used as the measures of treatment effect.

Since the true treatment effect is unknown, statistical procedures have to be used to make inference on the measure of the treatment effect, which include estimation, construction of confidence interval, testing related hypothesis. In the next, I will describe some measures of the treatment effect in cancer clinical trials and their statistical inference procedures.

1.2 Inference on Treatment Effects in Cancer Clinical Trials

In cancer clinical trials, the outcomes are usually the time to an event in nature. For example, overall survival or disease-free survival are the frequently used endpoints in cancer clinical trials. As mentioned previously, the hazard ratio is one of the popular measures which can be used to quantify the treatment effect. The standard approach to make the statistical inference on the treatment effect based on this measure is to use Cox regresssion models to estimate the hazard ratio as well as to construct the associated confidence intervals. However, Cox regression models require the proportional hazard assumption be satisfied, which may not be true in many circumstances. For example, in clinical trial MA. 5 conducted by the NCIC Clinical Trial Group

(NCIC CTG), it was shown that the proportional hazard assumption was not satisfied by the overall survival, which was defined as one of the clinical endpoints in this study.

Tu (2007) derived two types of under-smoothed kernel confidence intervals for the hazard ratio at a given time point t. It was found that, in terms of the coverage probability, both under-smoothed confidence intervals performed reasonably well when the proportional hazard assumption was violated. However, these procedures were still not found very satisfactory, because when the sample size is small, the true coverage probability was still far away from the nominal level. In the first part of this thesis, I will apply the empirical likelihood method to construct the confidence interval for the hazard ratio, which may have a higher accuracy in the coverage probability.

Although it is a popular measure, the hazard ratio may be difficult to understand by clinicians and patients as pointed out by Moser and McCann (2008). They proposed another measure of the treatment effect for clinical trials with the time to an event endpoint, which is called the probabilistic index and may allow better communication between the statisticians and the clinicians. Moser and McCann (2008) also proposed to estimate this index by generalizing the Wilcoxon statistic under random censoring. Koziol and Jia (2009) pointed out, however, that the expected value of this type of estimate is dependent upon the underlying censoring distributions, hence it may be far removed from the true value. They suggested an alternative approach based on an estimate proposed by Efron (1967) and showed by simulation study that its mean was closer to the true value with considerable smaller variance.

In the second part of this thesis, I will explore the applications of the empirical likelihood method under a density ratio model to estimate and construct confidence

interval for the probabilistic index.

1.3 Assessment of the Heterogeneity of Treatment Effects

Although the major interest in clinical trials is to assess treatment effect over all subjects included in the trial, it is also often of interest to assess the heterogeneity of treatment effect across subgroups defined by baseline characteristics of subjects to identify whether treatment effects in certain subgroup is larger than that in other subgroups. The standard approach used for this assessment is the test of interaction between treatments and covariate.

Based on the probabilistic index, Patel and Hoel (1973) proposed a non-parametric measure for the treatment-covariate interaction. Schemper (1988) proposed an estimate of this measure of interaction based on a generalized Wilcoxon statistic. This approach suffers the same problem as mentioned before, since it was based on the C-index. In the last part of this thesis, I will apply both approaches based on Efron's estimate and the empirical likelihood method with density ratio model to derive some new estimates and confidence interval procedures for this non-parametric measure of interaction.

1.4 Organization of This Thesis

A review of the current statistical inference procedures for the measure of the treatment effect in cancer clinical trials is presented first in the Chapter 2. I will then

provide a review for the applications of the empirical likelihood method for survival data and statistical inference procedures for the density ratio model in respectively Chapter 3 and Chapter 4. Chapter 5 presents a new empirical likelihood based procedure for construction of the confidence interval for a time-dependent hazard ratio function. Inferences of the probabilistic index under a density ratio model are presented in Chapter 6. Chapter 7 discusses the issues of statistical assessment of treatment-covariate interactions. A summary of the results in this thesis is presented in Chapter 8.

Chapter 2

Measures of Treatment Effect and Their Inferences in Cancer Clinical Trials

As mentioned in Chapter 1, the hazard ratio is a commonly used relative measure for treatment effect when the outcomes are survival times, which is usually the case in cancer clinical trials. Moser and McCann (2008) proposed an alternative measure, the probabilistic index, which may allow for better communications between the statisticians and clinicians. In this Chapter, I will give detailed definitions about these two measures and review the current approaches for their statistical inferences. I will also introduce a non-parametric definition of treatment-covariate interaction based on the probabilistic index and review procedures for its inferences based on both complete and censored data.

2.1 The Hazard Ratio and Its Inferences

Let T_1 and T_2 be two independent survival times in two treatment groups. For $j = 1,\ 2$, write f_j and F_j as respectively the density and cumulative distribution functions of T_j. The hazard function $h_j(t)$ of T_j is defined as the event rate at time t conditional on survival until time t or later (that is, $T_j \geq t$)

$$h_j(t)dt = \Pr(t \leq T_j < t + dt | T_j \geq t) = \frac{f_j(t)dt}{1 - F_j(t)} = \frac{F_j'(t)dt}{1 - F_j(t)}.$$

The hazard ratio function, defined by the ratio of two hazard functions, can be written as:

$$\rho(t) = \frac{h_1(t)}{h_2(t)}.$$

The hazard ratio is commonly used when presenting results in clinical trials involving survival data. For example, if the hazard ratio is less than 1, we may conclude that patients in the first treatment group tend to have a lower hazard of event than those in the second treatment group. In particular, when the distribution of T_j is exponential with parameter λ_j, ρ equals also λ_1/λ_2.

The Cox proportional hazard model (Cox, 1972) has been the most widely used procedure over many years to estimate the hazard ratio as well as construct the confidence interval. It is a semi-parametric regression method, based on the assumption that the hazard ratio is a constant function of time. This assumption is called the proportional hazard assumption, which, however, may not be satisfied by the data from epidemiologic studies or clinical trials, see the example provided by Stablein et al. (1981). Tu and Pater (1999) showed that, if proportional hazard assumption is not satisfied, the analysis based on the Cox's proportional hazard model may have very low power in testing the potential difference between treatment groups. Thus,

other methods are desired if the proportional hazard assumption is violated.

Tu (2007) derived two types of under-smoothed kernel confidence intervals for the hazard ratio at a given time point t and found that, in terms of the coverage probability, both under-smoothed confidence intervals performed reasonably well when the proportional hazard assumption was violated. In the following, I will give a detailed review of these procedures.

2.1.1 The Confidence Interval Based on the Cox Proportional Hazard Model

As mentioned before, the Cox proportional hazard model has been the standard method to derive the confidence interval for the hazard ratio in the analysis of data from clinical trials for many years. In the context of two treatment groups with $h_j(t)$ as the hazard rate in the j-th group, $j = 1,\ 2$, the Cox model can be expressed as:

$$h_1(t) = h_2(t)\exp(z\beta), \tag{2.1}$$

where

$$z = \begin{cases} 1, \text{ if the subject is in the first group} \\ 0, \text{ if the subject is in the second group} \end{cases}$$

and β is an unknown regression coefficient.

This model can be fitted by using a partial likelihood procedure based on censored data to obtain an estimate for the regression coefficient β. With the estimated coefficient $\hat{\beta}$, the hazard ratio is estimated by:

$$\hat{\rho}_{cox} = \exp(\hat{\beta}).$$

Based on the asymptotic normality of $\hat{\beta}$, which can be proved under some conditions, (see, e.g. Fleming and Harrington, 1991), a $100(1-\alpha)\%$ confidence interval for the hazard ratio can be calculated by:

$$I_{cox} = \left(\exp\left\{ \hat{\beta} - z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\beta})} \right\},\ \exp\left\{ \hat{\beta} + z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\beta})} \right\} \right), \tag{2.2}$$

where $\hat{V}(\hat{\beta})$ is the estimated variance of $\hat{\beta}$, $z_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile point of the standard normal distribution.

2.1.2 Under-smoothed Kernel Confidence Intervals for Hazard Ratio

Let T_1 and T_2 be two independent survival times in respectively two treatment groups. For $j = 1,\ 2$, write f_j and F_j as respectively the density and cumulative distribution function of T_j. Because of the potential censoring, we can not observe T_1 or T_2 for all subjects, therefore a random censoring model is assumed. Denote $C_j,\ j = 1,\ 2$ as two independent censoring times for subjects in respectively two groups with density and distribution function respectively $g_j(t)$ and $G_j(t)$. C_1 and C_2 are also assumed to be independent of T_1 and T_2. The data that we would observe from a subject in the j-th group are:

$$\begin{cases} X_j &= \min(T_j,\ C_j) \\ \delta_j &= I(T_j \le C_j) \end{cases}. \tag{2.3}$$

Here and thereafter, $I(A)$ stands for the indicator function of A.

For $(T_j,\ C_j,\ X_j,\ \delta_j),\ j = 1,\ 2$, defined above, let $(T_{ji},\ C_{ji},\ X_{ji},\ \delta_{ji}),\ i = 1,\ 2, ..., n_j$, be independent values of $(T_j,\ C_j,\ X_j,\ \delta_j)$, for the i-th subject in j-th group. We write $0 \le X_{j(1)} \le X_{j(2)} \le \cdots \le X_{j(n_j)} < \infty$ as the ordered statistics of

sample $\{X_{ji}\}$ and $\delta_{j(i)}$ the concomitant of $X_{j(i)}$ for $i = 1,\ 2, \cdots,\ n_j$ and $j = 1,\ 2$. Also denote

$$r_{ji} = \sum_{k=1}^{n_j} I(X_{jk} \geq X_{j(i)}) = n_j - i + 1$$

as the number of subjects that are still at risk just before $X_{j(i)}$.

With the above notations, the kernel estimate proposed by Tu (2007) for the hazard ratio $\rho(t)$ can be expressed as:

$$\hat{\rho}_{tu}(t) = \frac{\hat{h}_{1,a_1}(t)}{\hat{h}_{2,a_2}(t)},$$

where

$$\hat{h}_{j,a_j}(t) = \frac{1}{a_j} \sum_{i=1}^{n_j} K_j \left(\frac{t - X_{j(i)}}{a_j} \right) \frac{\delta_{ji}}{r_{ji}}$$

is the kernel estimate of $h_j(t)$ based on the kernel function $K_j(x)$ and smoothing bandwidth a_j.

Under the under-smoothing conditions which requires $a_j \to 0$, $n_j a_j \to \infty$ and $n_j a_j^5 \to 0$, and also assuming that the bandwidth a_1 and a_2 satisfy that:

$$\frac{n_1 a_1}{n_1 a_1 + n_2 a_2} \to \lambda, \text{ for some } 0 < \lambda < 1 \text{ as } n_j \to \infty (j = 1,\ 2),$$

Tu (2007) proved that

$$\frac{\hat{\rho}_{tu}(t) - \rho(t)}{\sqrt{V(t)}} \to N(0,\ 1) \text{ in distribution,}$$

where

$$V(t) = \frac{V_1(t) + \rho^2(t) V_2(t)}{h_2^2(t)}$$

with

$$V_j(t) = \frac{1}{n_j a_j} \left(\int K_j^2(x) dx \right) \frac{h_j(t)}{1 - H_j(t)},\ j = 1, 2.$$

Based on this result, a $100(1-\alpha)\%$ confidence interval for $\rho(t)$ can be defined as:

$$I_D = \left(\hat{\rho}(t) - z_{1-\alpha/2}\sqrt{\hat{V}(t)},\ \hat{\rho}(t) + z_{1-\alpha/2}\sqrt{\hat{V}(t)} \right), \tag{2.4}$$

where $\hat{V}(t)$ is an estimator for $V(t)$.

Tu (2007) also introduced another way to construct the confidence interval for the hazard ratio through Fieller's Theorem (Fieller, 1954). Under the same conditions given above, we have

$$\hat{h}_{1,a_1} - \rho(t)\hat{h}_{2a_2} \longrightarrow N(0,\ V_1(t) + \rho^2 V_2(t)),\ \text{in distribution},$$

which implies that

$$\Pr\left\{ \frac{[\hat{h}_{1,a_1} - \rho(t)\hat{h}_{2a_2}]^2}{\hat{V}_1(t) + \rho(t)^2\hat{V}_2(t)} \leq z^2_{\alpha/2} \right\} \approx 1-\alpha.$$

From this result, another asymptotic $100(1-\alpha)\%$ level confidence interval for $\rho(t)$, called Fieller type confidence interval, can be expressed as

$$\begin{aligned} I_F = \ & \left(\frac{1}{(1-g_{1a}(t))} \left[\hat{\rho}_{tu}(t) - \frac{z_{1-\alpha/2}}{\hat{h}_{2,a_2}} \sqrt{\rho^2_{tu}(t)\hat{V}_2(t) + (1-g_{1a}(t))\hat{V}_1(t)} \right], \right. \\ & \left. \frac{1}{(1-g_{1a}(t))} \left[\hat{\rho}_{tu}(t) + \frac{z_{1-\alpha/2}}{\hat{h}_{2,a_2}} \sqrt{\rho^2_{tu}(t)\hat{V}_2(t) + (1-g_{1a}(t))\hat{V}_1(t)} \right] \right), \end{aligned}$$

where

$$g_{1a}(t) = \frac{z^2_{1-\alpha/2}\hat{V}_2(t)}{\hat{h}^2_{2,a_2}},$$

and $\hat{V}_j(t)$ is an estimator of $V_j(t)$ $(j = 1, 2)$.

Simulation studies were conducted by Tu (2007) to compare the estimated coverage probabilities between proposed confidence intervals (I_D and I_F) and confidence intervals based on the Cox proportional hazard model (I_{cox}). The results showed that, in terms of the coverage probability, both types of under-smoothed confidence

intervals performed reasonably well, but were still not perfectly satisfactory. The simulation showed also the poor performance of confidence intervals based on the Cox proportional hazard model when the proportional hazard assumption is not true. Both of these procedures require, however, estimation on the variance of the hazard ratio estimator, which may be tedious in practice and lead to a lower accuracy of the confidence intervals. In Chapter 5, we will derive an alternative confidence interval for the hazard ratio based on empirical likelihood approaches.

2.2 The Probabilistic Index and Its Inferences

Although the hazard ratio is a popular measure of treatment effect in oncology literature, Moser and McCann (2008) argued that, based on their experience, it may be difficult for clinicians and patients to interpret and understand. They instead proposed another measure θ, which may allow a better communication between statisticians and clinicians. This θ is referred as probabilistic index, which is defined as the probability that a patient given one treatment will have an event earlier than if the same patient were given another treatment. They also suggested an estimate of θ by using Gehan's generalization of Wilcoxon statistic (Gehan, 1965), which is also called C-index or C statistic according to Harrell et al. (1982).

Their proposal was supported by Buyse (2008), who stated that this probabilistic index would express beneficial and harmful treatment effects on comparable scales, so that a direct comparison of these effects is possible. As mentioned by Buyse (2008), this measure has been used for quite some time in many fields including, for example, psychology under the name "probabilistic index" (Acion et al., 2006) and reliability as a measure for stress-strength (Zhou, 2008). This measure is also identical

to the area under the receiver operator characteristic (ROC) curve (Qin and Zhang, 2003; Brumback et al., 2006), which is a commonly used global index for measuring diagnostic accuracy in the medical diagnostic testing.

C-index has been widely used to assess the separation of two survival distributions. However, this index might not be consistent beyond the supports of the censored observation as pointed out by Bassiakos et al. (1991). Recently, Koziol and Jia (2009) pointed out again that the expected value of the estimate based on the C-index is dependent upon the underlying censoring distributions, hence may be far removed from the true θ. Buyse (2008) also showed that, without adjustment for censoring, considerable bias may be incurred when estimating the hazard ratio using the C-index based on the relationship derived by Moser and McCann (2008) between the hazard ratio and θ. Koziol and Jia (2009) suggested an alternative estimate of θ based on an estimator proposed by Efron (1967) and showed in their simulation study that it is closer to the true θ with considerable smaller variance than the estimate based on the C-index. In the next, I give the details about these two estimation approaches.

2.2.1 Estimating Probabilistic Index through C Index (C Statistic)

When there is no censoring in the data, the probability $\theta = \Pr(T_1 \leq T_2)$ can be estimated using the Wilconxon statistic, which was shown to be a consistent estimator for θ and has been repeatedly utilized for the non-parametric comparisons of two distribution functions. Gehan (1965) generalized the Wilconxon statistic to the setting where data may be randomly censored. Specifically, if we use the same notation as (2.3) and let $(T_{ji},\ C_{ji},\ X_{ji},\ \delta_{ji})$ be the independent replications of $(T_j,\ C_j,\ X_j,\ \delta_j)$,

$i = 1, 2, \cdots, n_j,\ j = 1, 2$, then Gehan's generalized statistic (also called the C index or C statistic) can be written as:

$$U = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} U_{ik}, \tag{2.5}$$

where

$$U_{ik} = \begin{cases} 0 & \text{if } X_{2i} < X_{1k},\ \delta_{2i} = 0 \\ 1 & \text{if } X_{1i} < X_{2k},\ \delta_{1i} = 0 \\ 1/2 & \text{otherwise} \end{cases} .$$

Let $F_j(t)$ and $G_j(t)$ be the distribution functions of the true survival time T_j and true censoring time C_j, respectively, $j = 1, 2$. Assuming $n_j/(n_1 + n_2) \to \rho_j > 0$, Efron (1967) showed that, under the null hypothesis that $F_1(t) = F_2(t)$,

$$\frac{1}{n_1 + n_2}(U - \frac{1}{2}) \to N\left(0, \frac{1}{12}\left(\frac{\sigma_1^2}{\rho_1} + \frac{\sigma_2^2}{\rho_2}\right)\right),$$

where

$$\begin{aligned} \sigma_1^2 &= -\int (1 - G_1(t))(1 - G_2(t))^2 d(1 - F_1(t))^3, \\ \sigma_2^2 &= -\int (1 - G_1(t))^2 (1 - G_2(t)) d(1 - F_1(t))^3. \end{aligned}$$

2.2.2 Estimating Probabilistic Index through Efron's Estimator

For U defined in (2.5), Efron (1967) showed that

$$\begin{aligned} EU &= \int \bar{F}_1(t)\bar{G}_1(t)\bar{G}_2(t)dF_2(t) \\ &\quad + \frac{1}{2}\left[\int \bar{F}_1(t)\bar{F}_2(t)\bar{G}_2(t)dG_1(t) + \int \bar{F}_1(t)\bar{F}_2(t)\bar{G}_1(t)dG_2(t)\right], \end{aligned}$$

where for any distribution function $F(t)$, $\bar{F}(t) = 1 - F(t)$. Koziol and Jia (2009) further showed in a simulation study that, the expected value of U is dependent upon

the underlying censoring distributions, hence may be far removed from the true θ. They suggested an alternative estimate:

$$\hat{\theta} = \int \hat{\bar{F}}_1^{KM}(t) d\hat{\bar{F}}_2^{KM}(t), \tag{2.6}$$

where $\hat{\bar{F}}_j^{KM}(t)$ is the Kaplan-Meier estimate of $\bar{F}_j(t)$, $j = 1, 2$. This estimate was actually first proposed in Efron (1967), therefore is called Efron's estimator in this thesis. Koziol and Jia (2009) also suggested that an alternative estimate for θ may be obtained when other estimates of $F_1(t)$ and $F_2(t)$ are available, by replacing them with these estimates.

2.3 Non-parametric Assessment of the Treatment-Covariate Interaction

In the analysis by subgroup defined based on baseline covariates in clinical trials with time to an event endpoints, Cox proportional hazard model was usually the standard method to test the interaction between the treatment and these covariates. When the proportional hazard assumption is not satisfied, non-parametric method may provide a better alternative. In the following, I will first present the definition of a non-parametric measure of the treatment-covariate interaction proposed by Patel and Hoel (1973) as well as its inference based on complete data, and then review a generalization of this inference procedure to censored data.

2.3.1 A Non-parametric Measure of Interaction in a Two-by-Two Factorial Design

Patel and Hoel (1973) introduced the following non-parametric measure of interaction in the context of a two-by-two factorial experiment.

$$\begin{aligned} \mu &= \Pr(X_{12} \leq X_{11}) - \Pr(X_{22} \leq X_{21}) \\ &:= \theta_1 - \theta_2, \end{aligned} \tag{2.7}$$

where X_{ij} is the response at the i-th level of factor A and j-th level of factor B.

Let X_{ijk} be the response of k-th replicate of the i-th level of factor A and j-th level of factor B. Patel and Hoel suggested the following consistent, unbiased, asymptotically normal estimator of μ:

$$\hat{\mu} = \hat{\theta}_1 - \hat{\theta}_2, \tag{2.8}$$

where

$$\hat{\theta}_i = \sum_{j=1}^{n_{i1}} \sum_{k=1}^{n_{i2}} \psi(X_{i1j}, X_{i2k})/n_{i1}n_{i2}, \ i = 1, 2,$$

with

$$\psi(X_{ijk}, X_{i'j'k'}) = \begin{cases} 0 & \text{if } X_{ijk} < X_{i'j'k'} \\ 1/2 & \text{if } X_{ijk} = X_{i'j'k'} \\ 1 & \text{if } X_{ijk} > X_{i'j'k'} \end{cases} .$$

Let $F_{ij}(x)$ be the distribution function of X_{ij}, for $i = 1, 2$, $j = 1, 2$, and $\nu = \min\{n_{11}, n_{12}, n_{21}, n_{22}\}$, it was shown that, as $\nu \to \infty$,

$$\frac{\nu^{\frac{1}{2}}(\hat{\mu} - \mu)}{\sigma} \to N(0, 1), \text{ in distribution,} \tag{2.9}$$

where $\sigma^2 = \sigma_1^2 + \sigma_2^2$, with

$$\begin{aligned}\sigma_i^2 &= \frac{\nu}{n_{i1}n_{i2}}\left\{\int F_{i2}dF_{i1} + (n_{i1}-1)\int (1-F_{i1})^2 dF_{i2}\right. \\ &\left. +(n_{i2}-1)\int F_{i2}^2 dF_{i1} - (n_i 1 + n_{n2} - 1)\left(\int F_{i2}dF_{i1}\right)^2\right\}, i = 1, 2.\end{aligned}$$

Therefore to test the following hypothesis about the existence of the interaction between factor A and B,

$$H_0 : \ \mu = 0 \text{ v.s. } H_1 : \mu \neq 0,$$

we may replace σ^2 by a consistent estimate $\hat{\sigma}^2$, and reject H_0 at α level if

$$\left|\frac{\nu^{\frac{1}{2}}\hat{\mu}}{\hat{\sigma}}\right| \geq z_{1-\alpha/2},$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of standard normal distribution.

2.3.2 Inference of the Non-parametric Measure of Treatment-Covariate Interaction in the Presence of Censoring

Schemper (1988) derived an estimate of Patel-Hoel's non-parametric measure of the interaction in the setting where there are censored data. We consider only the case where there are two treatment groups and the baseline covariate is dichotomous.

Let T_{ja} and C_{ja} be the true survival and censoring time, respectively, for a patient in treatment group j and with covariate level a, for $j = 1, 2,\ a = 0, 1$. The non-parametric measure of the interaction introduced by Patel and Hoel (1973) can be expressed as

$$\mu = \Pr(T_{11} \leq T_{21}) - \Pr(T_{10} \leq T_{20}).$$

However, because of censoring, usually the data observed from a cancer clinical trial

are

$$\begin{cases} X_{ja} &= \min(T_{ja}, C_{ja}) \\ \delta_{ja} &= I(X_{ja} = T_{ja}) \end{cases} \quad j = 1, 2, a = 0, 1.$$

In this setting, Schemper (1988) proposed to estimate μ by the C-index, which is the sum of U-scores generalized from Wilcoxon statistics under a random censoring mechanism. Let (X_{jai}, δ_{jai}), $i = 1, 2, \cdots, n_{ja}$ be the independent replications of (X_{ja}, δ_{ja}), then in the presence of censoring, one of the scoring scheme could be

$$U_a = \sum_{i=1}^{n_{1a}} \sum_{i'=1}^{n_{2a}} u(X_{2ai}, X_{1ai'}),\ a = 0, 1,$$

where

$$u(X_{jai}, X_{j'a'i'}) = \begin{cases} -1 & \text{if } X_{jai} < X_{j'a'i'},\ \delta_{jai} = 1 \\ 1 & \text{if } X_{jai} > X_{j'a'i'},\ \delta_{j'a'i'} = 1 \\ 0 & \text{if otherwise} \end{cases} .$$

If we denote the number of positive $u(X_{2ai}, X_{1ai'})$ scores as PR_a and the number of negative $u(X_{2ai}, X_{1ai'})$ scores as IN_a for $a = 0, 1$, then $U_a = PR_a - IN_a$ and an estimate of $\theta_a = \Pr(T_{1a} \leq T_{2a})$ could be:

$$\hat{\theta}_a = \frac{PR_a}{PR_a + IN_a},\ a = 0, 1,$$

which leads to an estimate of μ as:

$$\hat{\mu} = \hat{\theta}_1 - \hat{\theta}_0.$$

A non-parametric test of interaction would by obtained based on the jackknife variance estimate of $\hat{\theta}_a$ defined as:

$$s_a^2 = (n_{1a} + n_{2a} - 1) \left[\sum_{q=1}^{n_{1a}+n_{2a}} \frac{\hat{\theta}_{aq}^2}{n_{1a} + n_{2a}} - \left(\sum_{q=1}^{n_{1a}+n_{2a}} \frac{\hat{\theta}_{aq}}{n_{1a} + n_{2a}} \right)^2 \right],$$

where $\hat{\theta}_{aq}$ is the proportion of positive scores in the covariate level a without the q-th

observation. Under the hypothesis of no interaction, the test statistic defined as

$$H = \sum_{a=0}^{1} \frac{(\hat{\theta}_a - \bar{\theta})^2}{s_a^2},$$

where

$$\bar{\theta} = \frac{\sum_{a=0}^{1} \hat{\theta}_a s_a^{-2}}{\sum_{a=0}^{1} s_a^{-2}}$$

would follow a Chi-square distribution with one degree of freedom, and therefore, we would reject

$$H_0 : \ \mu = 0 \text{ v.s. } H_1 : \ \mu \neq 0$$

at α level if $H > \chi_1^2(1-\alpha)$, where $\chi_1^2(1-\alpha)$ is the $1-\alpha$ percentile of a Chi-square distribution with one degree of freedom.

Chapter 3

Empirical Likelihood Procedures for Survival Data

After its introduction by Owen (1988), the empirical likelihood method has been applied to many statistical problems. A summary of these applications may be found in the book by Owen (2001). In this Chapter, I will review only applications of the empirical likelihood to some problems with censored data, which include constructions of confidence intervals for the survival function, hazard function, density function and the difference and ratio of two survival functions.

3.1 Empirical Likelihood Confidence Interval of Survival Function for Censored Data

Before Owen (1988) formally introduced the concept of empirical likelihood, Thomas and Grunkemeier (1975) used this method to construct the confidence interval for a

survival function and compared it with the confidence interval based on the normal approximation and Greenwood formula for the variance estimate of Kaplan-Meier estimator.

Let $T_1,\ T_2, ..., T_n$ be true survival times which are i.i.d. with a distribution function $F(t)$ and continuous density function $f(t)$, and $C_1,\ C_2, ..., C_n$ true censoring time which are also i.i.d. with a distribution function G(t). Assume T's and C's are independent. For $i = 1,\ 2, ...,\ n$, denote the observed data as

$$\begin{cases} X_i &= \min(T_i,\ C_i) \\ \delta_i &= I(T_i \leq C_i) \end{cases}.$$

Let $X_{(i)}$ be an ordered sequence of X_i, $\delta_{(i)}$ the concomitant of $X_{(i)}$ and r_i the number of subjects at risk before time $X_{(i)}$, that is,

$$r_i = \sum_{j=1}^{n} I(X_j > X_{(i)}).$$

Then the survival function $\bar{F}(t) = 1 - F(t)$ of the true survival time can be estimated by the Kaplan-Meier method (Kaplan and Meier, 1958) as

$$\hat{\bar{F}}(t) = \prod_{i:\ X_{(i)} \leq t} \left(1 - \frac{\delta_{(i)}}{r_i}\right),\ 0 \leq t < \infty. \tag{3.1}$$

Kaplan and Meier (1958) showed in two steps that $\hat{\bar{F}}(t)$ is a nonparametric maximum likelihood estimate of $\bar{F}(t)$ over the infinite-dimensional parameter space

$$\Theta \equiv \{\text{all survival functions on } [0,\ \infty)\}.$$

That is, it maximizes the likelihood function

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_i-))^{\delta_i} \bar{F}(X_i)^{1-\delta_i}. \tag{3.2}$$

In the first step, they proved that maximizing $L(F)$ is equivalent to maximizing $L(F)$

over the subspace

$$\Theta_T \equiv \{\text{all discrete survival functions supported on } X_i \text{ with } \delta_i = 1\}.$$

Let

$$\lambda_i = \frac{F(X_{(i)}) - F(X_{(i)}-)}{1 - F(X_{(i)}-)},\ i = 1, 2, ..., n.$$

Next they showed that for each $\bar{F} \in \Theta_T$, the likelihood can be represented as

$$L(F) = \prod_{i=1}^{n} \lambda_i^{\delta_{(i)}} (1 - \lambda_i)^{(r_i - \delta_{(i)})}, \tag{3.3}$$

and Kaplan-Meier estimator (3.1) maximizes $L(F)$ with

$$\hat{\lambda}_i = \frac{\delta_{(i)}}{r_i}.$$

Thomas and Grunkemeier (1975) proposed that a $1 - \alpha$ confidence set for $\bar{F}(t)$ could be formed as all values p for which the hypothesis $H_0 :\ \bar{F}(t) = p$ is not rejected by a nonparametric likelihood ratio test at level α. Define the empirical likelihood ratio test statistic as

$$R(p) = \frac{\sup\left\{\prod_{i=1}^{n} \lambda_i^{\delta_{(i)}} (1 - \lambda_i)^{(r_i - \delta_{(i)})} :\ \prod_{i:\ X_{(i)} \le t} (1 - \lambda_i) = p \text{ and } 0 < \lambda_i \le 1\right\}}{\prod_{i=1}^{n} \left(\frac{\delta_{(i)}}{r_i}\right)^{\delta_{(i)}} \left(1 - \frac{\delta_{(i)}}{r_i}\right)^{(r_i - \delta_{(i)})}}.$$

They suggested that

$$\{p :\ \ -2 \ln R(p) \le C_\alpha\}$$

can be defined as a confidence set for $\bar{F}(t)$ with approximate level $1-\alpha$, where C_α satisfies $P(\chi_1^2 \le C_\alpha) = 1 - \alpha$. Their simulation results indicated that, for small samples, the actual coverage probability of the proposed confidence interval is closer to the nominal confidence level than that of confidence intervals obtained from the normal approximation to the distribution for the Kaplan-Meier estimate with Greenwood's formula for the estimation of its variance. However, Thomas and Grunkemeier (1975)

did not give any theoretical justifications of their method. This has been done by Li (1995) who proved that assuming that $F(t)$ is continuous, then as $n \to \infty$,

$$-2 \ln R(p) \to \chi^2_1, \quad \text{in distribution,}$$

which provided the rigorous justification for this likelihood ratio method.

3.2 Empirical Likelihood Confidence Intervals for Hazard and Density Functions under Right Censorship

Shen and He (2008) used smoothed empirical likelihood methods to construct confidence intervals for hazard and density functions with censored data.

We consider first the confidence interval for a hazard function. Let Θ be the space of all survival functions defined on $[0, \infty)$. For $\bar{F} \in \Theta$, its likelihood function $L(F)$ based on one-sample data from a random censoring model is given by (3.2). From Li (1995), we know that this likelihood $L(F)$ can be rewritten as:

$$L(F) = \prod_{i=1}^{n} \lambda_i^{\delta_{(i)}} (1 - \lambda_i)^{(r_i - \delta_{(i)})},$$

and by maximizing $L(F)$, we get

$$\hat{\lambda}_i = \frac{\delta_{(i)}}{r_i}.$$

Assume the hazard function $h(t)$ could be estimated from one of estimators in the following family of kernel estimators:

$$\hat{h}(t|\lambda) = -\sum_{i=1}^{n} \ln(1 - \lambda_i) K_i(t), \tag{3.4}$$

where

$$K_i(t) = \frac{1}{a} K\left(\frac{t - X_{(i)}}{a}\right).$$

Different selection of λ_i gives different estimators in this family. The estimators in this family are slightly different from the kernel estimator used by Tu (2007).

With this assumption, the empirical likelihood ratio for the hazard function $h(t)$ can be defined as:

$$R(h(t),\ t) := \frac{\sup_{\lambda_1,\ldots,\lambda_n}\{L(F) : \sum_{i=1}^{n} \ln(1-\lambda_i)K_i(t) + h(t) = 0\}}{\sup_{\lambda_1,\ldots,\lambda_n} L(F)}.$$

By using Lagrange's method, it can be shown that

$$\ln R(h(t),\ t) = \sum_{i=1}^{n} \left\{ (r_i - \delta_{(i)}) \ln\left(1 + \frac{\mu K_i(t)}{r_i - \delta_{(i)}}\right) - r_i \ln\left(1 + \frac{\mu K_i(t)}{r_i}\right)\right\},$$

where the Lagrange multiplier μ satisfies

$$\sum_{i=1}^{n} \ln\left(1 - \frac{\delta_{(i)}}{r_i + \mu K_i(t)}\right) K_i(t) + h(t) = 0.$$

Under some mild conditions and assuming a smoothing bandwidth a, it is shown that for each fixed $t \in [\tau_1,\ \tau_2]$, where $0 < \tau_1 < \tau_2 < \infty$ are two constants, as $n \longrightarrow \infty$,

$$-2 \ln R(h,\ t) \longrightarrow \chi_1^2,\ \text{in distribution.} \tag{3.5}$$

Based on (3.5), a confidence interval for $h(t)$ with an asymptotic coverage probability $1 - \alpha$ can be defined as

$$I_{n,\alpha}(h,\ t) := \{h :\ \ -2 \ln R(h,\ t) \le C_\alpha\},$$

where C_α is given by $P(\chi_1^2 \le C_\alpha) = 1 - \alpha$.

Next I will take a look at the construction of empirical likelihood confidence interval for the density function. Note that $f(t) = h(t)\bar{F}(t)$. Adding another constraint

$\bar{F}(t) = \eta$, we have $h(t) = f(t)/\eta$. Define the empirical likelihood ratio for f and η as

$$R(f,\ \eta,\ t) := \frac{\sup_{\lambda_1,\ldots\lambda_n} \left\{ L(F) : \begin{array}{c} \sum_{i=1}^{n} \ln(1-\lambda_i) I(X_i \le t) = \ln \eta, \\ \sum_{i=1}^{n} \ln(1-\lambda_i) K_i(t) + \dfrac{f(t)}{\eta} = 0 \end{array} \right\}}{\sup_{\lambda_1,\ldots\lambda_n} L(F)}$$

and the empirical likelihood ratio for f as

$$R(f,\ t) := \sup_{\eta \in (0,1)} R(f,\ \eta,\ t).$$

Let

$$W_{ni} := (I(X_{(i)} \le t),\ K_i(t))^T.$$

By the Lagrange method we have

$$\ln R(f,\ t) = \sum_{i=1}^{n} \left\{ (r_i - \delta_{(i)}) \ln \left(1 + \frac{\mu^T W_{ni}}{r_i - \delta_{(i)}} \right) - r_i \ln \left(1 + \frac{\mu^T W_{ni}}{r_i} \right) \right\},$$

where the Lagrange multiplier $\mu = (\mu_1,\ \mu_2)^T$ and nuisance parameter η satisfy

$$\left\{ \begin{array}{l} \displaystyle\sum_{i=1}^{n} \ln \left(1 - \frac{\delta_{(i)}}{r_i + \mu^T W_{ni}} \right) I(X_{(i)} \le t) - \ln \eta = 0 \\ \displaystyle\sum_{i=1}^{n} \ln \left(1 - \frac{\delta_{(i)}}{r_i + \mu^T W_{ni}} \right) K_i(t) + \frac{f(t)}{\eta} = 0 \\ \mu_1(\eta,\ t)\eta + \mu_2(\eta,\ t) f(t) = 0 \end{array} \right. . \tag{3.6}$$

Then under the same condition as previous stated for the hazard rate function, for each $t \in [\tau_1,\ \tau_2]$, and with probability 1 for large n, the third equation of equation array (3.6) has a solution $\eta_E = \eta_E(t)$, such that $R(f,\ \eta,\ t)$ attains its maximum value at $\eta = \eta_E$, and as $n \to \infty$, we have

$$-2 \ln R(f,\ \eta_E,\ t) \to \chi^2_1,\ \text{in distribution.}$$

Similarly, an asymptotic $100(1-\alpha)\%$ confidence interval for $f(t)$ can be defined as:

$$I_{n,\alpha}(f,\ t) := \{ f :\ -2 \ln R(f,\ \eta_E,\ t) \le C_\alpha \}.$$

3.3 Empirical Likelihood Confidence Band for the Difference of Two Survival Functions under Right Censorship

Shen and He (2006) used the empirical likelihood method to construct the confidence band for the difference of two survival functions, defined as:

$$\begin{aligned} D_0(t) &:= \bar{F}_1(t) - \bar{F}_2(t) \\ &:= P(T_1 > t) - P(T_2 > t). \end{aligned}$$

A natural estimate for the difference $D_0(t)$ could be:

$$D_n(t) = \hat{\bar{F}}_1(t) - \hat{\bar{F}}_2(t),$$

where $\hat{\bar{F}}_1$, $\hat{\bar{F}}_2$ are Kaplan-Meier estimators of $\bar{F}_1$ and $\bar{F}_2$, respectively. Let Θ be the space of all survival functions defined on $[0,\ \infty)$. For $\bar{F}_1,\ \bar{F}_2 \in \Theta$, the likelihood function of F_1 and F_2 based on censored data can be defined as follows:

$$L(F_1,\ F_2) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} (F_j(X_{ji}) - F_j(X_{ji}-))^{\delta_{ji}} (1 - F_j(X_{ji}))^{(1-\delta_{ji})}.$$

Define the empirical likelihood ratio for D_0 and nuisance parameter η as

$$R(D_0,\ \eta,\ t) := \frac{\sup\{L(F_1,\ F_2) :\ \bar{F}_1 = \eta + D_0(t),\ \bar{F}_2 = \eta,\ \bar{F}_j \in \Gamma\}}{\sup\{L(F_1,\ F_2) :\ \bar{F}_j \in \Gamma\}}$$

Following the same method as in Li (1995), it can be shown that

$$\ln R(D_0,\ \eta,\ t) = \sum_{j=1}^{2} \sum_{X_{j(i)} \le t} \left\{ (r_{ji} - \delta_{j(i)}) \ln \left(1 + \frac{\mu_j}{r_{ji} - \delta_{j(i)}} \right) - r_{ji} \ln \left(1 + \frac{\mu_j}{r_{ji}} \right) \right\},$$

where the Lagrange multipliers μ_1, μ_2 and nuisance parameter η satisfy

$$\begin{cases} \sum\limits_{X_{j(i)}\le t} \ln(1-\dfrac{\delta_{1(i)}}{r_{1i}+\mu_1}) - \ln(D_0+\eta) = 0 \\ \sum\limits_{X_{j(i)}\le t} \ln(1-\dfrac{\delta_{2(i)}}{r_{2i}+\mu_2}) - \ln\eta = 0 \\ \dfrac{\mu_1(\eta,\ t)}{D_0(t)+\eta} + \dfrac{\mu_2(\eta,\ t)}{\eta} = 0 \end{cases} \tag{3.7}$$

Shen and He (2006) proved that for each $t \in [\tau_1,\ \tau_2]$, there exists a solution η_E to the third equation of equation array (3.7) almost surely for sufficiently large n , such that $R(D_0,\ \eta,\ t)$ attains its maximum value at $\eta = \eta_E(t)$, and

$$-2\ln R(D_0,\ \eta_E,\ .) \longrightarrow \frac{W^2(.)}{\sigma^2(.)} \text{ in distribution,} \tag{3.8}$$

where

$$W(t) = \frac{\bar{F}_1(t)W_1(\sigma_1^2(t))}{\sqrt{p_1}} + \frac{\bar{F}_2(t)W_2(\sigma_2^2(t))}{\sqrt{p_2}},$$

with W_1, W_2 independent standard Brownian motions and

$$\sigma_j^2(t) = \int_0^t \frac{dF_j(s)}{\bar{F}_j(s-)\bar{H}_j(s)},$$

$$\sigma^2(t) = \frac{\bar{F}_1^2(t)\sigma_1^2(t)}{p_1} + \frac{\bar{F}_2^2(t)\sigma_2^2(t)}{p_2}.$$

Based on (3.8), we can define the following asymptotic $100(1-\alpha)\%$ confidence band of $D_0(t)$ as

$$I_{n,\alpha} = \{(t,\ D) :\ -2\hat{\sigma}^2(t)\ln R(D,\ \eta_E,\ t) \le K_\alpha^2[\tau_1,\ \tau_2],\ t \in [\tau_1,\ \tau_2]\},$$

where

$$\hat{\sigma}^2(t) = \frac{\bar{F}_{1n}^2(t)\hat{\sigma}_1^2(t)n}{n_1} + \frac{\bar{F}_{2n}^2(t)\hat{\sigma}_2^2(t)n}{n_2},$$

$$\hat{\sigma}_j^2(t) = n_j \sum_{X_{j(i)}\le t} \frac{\delta_{j(i)}}{r_{ji}(r_{ji}-\delta_{j(i)})},$$

and $K_\alpha[\tau_1,\ \tau_2]$ is the upper α-quantile of the distribution of $\sup_{t\in[\tau_1,\ \tau_2]}|W(t)|$. In practice, Monte Carlo method can be used to simulate the distribution of $\sup_{t\in[\tau_1,\ \tau_2]}|W(t)|$.

Particularly, for each $t \in [\tau_1,\ \tau_2]$, we have

$$-2\ln R(D_0,\ \eta_E,\ t) \longrightarrow \chi_1^2 \text{ in distribution,}$$

Therefore, confidence interval for $D_0(t)$ with an asymptotically coverage probability $1-\alpha$ is defined by

$$I_{n,\alpha}(t) = \{D:\ \ -2\ln R(D_0,\ \eta_E,\ t) \le C_\alpha\},$$

where C_α is given by $P(\chi_1^2 \le C_\alpha) = 1-\alpha$.

3.4 Simultaneous Confidence Bands for Ratios of Survival Functions via Empirical Likelihood

A simultaneous confidence band for the ratio of two survival functions based on censored data using empirical likelihood method is derived by McKeague and Zhao (2002).

Let Θ be the space of all survival functions defined on $[0,\ \infty)$. For $\bar{F}_1,\ \bar{F}_2 \in \Theta$, the likelihood function for F_1 and F_2 based on censored data is defined as before:

$$L(F_1,\ F_2) = \prod_{j=1}^{2}\prod_{i=1}^{n_j}(F_j(X_{ji}) - F_j(X_{ji}-))^{\delta_{ji}}(1-F_j(X_{ji}))^{(1-\delta_{ji})}.$$

The empirical likelihood ratio for $\tilde{\theta}(t) = \bar{F}_1(t)/\bar{F}_2(t)$ at $\tilde{\theta}(t) > 0$ for a given $t \ge 0$ can be defined as:

$$R(\tilde{\theta}(t),\ t) = \frac{\sup\{L(\bar{F}_1,\ \bar{F}_2):\ \bar{F}_1(t)/\bar{F}_2(t) = \tilde{\theta}(t),\ (\bar{F}_1,\ \bar{F}_2 \in \Gamma\times\Gamma)\}}{\sup\{L(\bar{F}_1,\ \bar{F}_2):\ ((\bar{F}_1,\ \bar{F}_2 \in \Gamma\times\Gamma)\}}.$$

It can be shown using the Lagrange method that

$$-2\ln R(\tilde{\theta}(t),\ t) = -2\sum_{j=1}^{2}\sum_{X_{j(i)}\le t}\left\{(r_{ji}-\delta_{j(i)})\ln\left(1+\frac{\mu}{r_{ji}-\delta_{j(i)}}\right)-r_{ji}\ln\left(1+\frac{\mu}{r_{ji}}\right)\right\},$$

where the Lagrange multiplier μ satisfies the equation

$$\ln\prod_{X_{1(i)}\le t}\left(1-\frac{\delta_{1(i)}}{r_{1i}+\mu}\right)-\ln\prod_{X_{2(i)}\le t}\left(1-\frac{\delta_{2(i)}}{r_{2i}-\mu}\right)=\ln(\tilde{\theta}(t))$$

For further convenience, we define

$$\sigma_j^2(t)=\int_0^t\frac{dF_j(s)}{\bar{F}_j(s)\bar{H}_j(s-)}$$

and

$$\sigma^2(t)=\frac{\sigma_1^2(t)}{p_1}+\frac{\sigma_2^2(t)}{p_2},$$

which can be uniformly and consistently estimated by respectively

$$\hat{\sigma}_j^2(t)=n_j\sum_{X_{j(i)}\le t}\frac{\delta_{j(i)}}{r_{ji}(r_{ji}-\delta_{j(i)})}$$

and

$$\hat{\sigma}^2(t)=\frac{\hat{\sigma}_1^2(t)n}{n_1}+\frac{\hat{\sigma}_2^2(t)n}{n_2}.$$

McKeaguea and Zhao (2002) proved the following result which can be used to construct simultaneous confidence bands for $\theta(t)=\bar{F}_1(t)/\bar{F}_2(t)$ over the time span of interest $[\tau_1,\ \tau_2]$: $-2\hat{\sigma}^2(t)\ln R(\rho(t),\ t)$ converges in distribution to $U^2(t)$ in $[\tau_1,\ \tau_2]$, where $U(t)$ is a Gaussian martingale with mean zero and variance function $\sigma^2(t)$.

Simulation needs to be used to obtain the critical value of the Gaussian martingale, but if the critical value is replaced by a χ_1^2 critical value, it is shown that we can obtain a pointwise confidence interval.

Chapter 4

The Density Ratio Model and Its Inference

Let X_1 and X_2 be respectively responses observed in the two treatment groups and $f_1(x)$ and $f_2(x)$ their probability density functions. The density ratio model assumes that

$$f_1(x) = \exp(\beta_0 + \beta_1 h(x)) f_2(x),$$

where $h(x)$ is a known or partially known function. The density ratio model has a long history in its relationship with the logistic regression model. Qin and Zhang (1997) formed it as an equivalent to the logistical regression model for data from a case-control study and used it for checking assumptions underlying the logistic regression model. When there is no censoring in the data set, various aspects of inference on the density ratio model have been considered by many researchers. A sample of those references can be found in, for example, Qin (1998), Zhang (2000), Cheng and Chu (2004), Fokianos (2004), Kezioua and Leoni-Aubin (2008), and Kedem et al. (2009).

Density ratio model can be an alternative to the Cox model when the proportional hazard assumption does not hold, since Fokianos and Kaimi (2006) showed that the density ratio model is actually broader than the Cox proportional hazard models. When there is censoring in the data set, the density ratio model was studied only by a few authors, partly because the inference based on a full likelihood function of censored data may be computationally intensive.

In this Chapter, I will give a detailed review of some related aspects of the density ratio model and its inference procedures related, which include inferences of the treatment effect based on the density ratio model and complete data as well as the conditional and weighted empirical likelihood procedures for the inference of the density ratio model with censored data.

4.1 Inference for the Treatment Effect with the Density Ratio Model

Fokianos and Troendle (2007) considered the problem of estimating and testing the relative treatment effect between two treatment groups when the data in the two groups satisfy a density ratio model. Specifically, suppose that there are two independent samples $\{x_{11}, x_{12}, \cdots, x_{1n_1}\}$ and $\{x_{21}, x_{22}, \cdots, x_{2n_2}\}$. As mentioned above, a density ratio model is defined as:

$$\begin{aligned} x_{11}, x_{12}, \cdots, x_{1n_1} &\sim f_1(x) = \exp(\beta_0 + \beta_1 h(x)) f_2(x), \\ x_{21}, x_{22}, \cdots, x_{2n_2} &\sim f_2(x), \end{aligned} \tag{4.1}$$

where $h(x)$ is a known link function and $f_j(x)$ is unknown density function with the corresponding distribution function denoted as $F_j(x)$, $j = 1, 2$.

To quantify the difference between two treatment groups, Fokianos and Troendle (2007) proposed the following measure

$$\theta = \Pr(X_1 < X_2) + \frac{1}{2}\Pr(X_1 = X_2),$$

where X_j denotes a single random variable from the j-th sample, $j = 1, 2$. We can see that if $\theta > 1/2 (\theta < 1/2)$, then the random variable X_1 tends to be smaller (larger) than X_2.

Let p_{ji} be the size of jump at the observed datum x_{ji}, $j = 1, 2,\ i = 1, 2, \cdots, n_j$, and consider the following empirical likelihood of the parameters given the data:

$$L(\beta, F_2) = \left(\prod_{j=1}^{2}\prod_{i=1}^{n_j} p_{ji}\right)\left(\prod_{i=1}^{n_1}\exp(\beta_0 + \beta_1 h(x_{1i}))\right), \tag{4.2}$$

where $\beta = (\beta_0, \beta_1)$. Following Qin and Zhang (1997), the inference for the parameter β can be based on the following score equations:

$$\begin{cases} 0 &= n_1 - \displaystyle\sum_{j=1}^{2}\sum_{i=1}^{n_j} \frac{\rho\exp(\beta_0 + \beta_1 h(x_{ji}))}{1 + \rho\exp(\beta_0 + \beta_1 h(x_{ji}))}, \\ 0 &= \displaystyle\sum_{i=1}^{n_1} h(x_{1i}) - \sum_{j=1}^{2}\sum_{i=1}^{n_j} \frac{\rho\exp(\beta_0 + \beta_1 h(x_{ji}))h(x_{ji})}{1 + \rho\exp(\beta_0 + \beta_1 h(x_{ji}))}, \end{cases} \tag{4.3}$$

where $\rho = n_1/n_2$. It was shown that

$$\hat{p}_{ji} = \frac{1}{n_2(1 + \rho\exp(\hat{\beta}_0 + \hat{\beta}_1 h(x_{ji})))},\ j = 1, 2, i = 1, 2, \cdots, n_j$$

maximize $L(\beta, F_2)$ in (4.2) with $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ as the solution for the score equations (4.3).

As a result, estimators of $F_1(x)$ and $F_2(x)$ are derived respectively as:

$$\hat{F}_1(x) = \sum_{j=1}^{2}\sum_{i=1}^{n_j} \hat{p}_{ji}\exp(\hat{\beta}_0 + \hat{\beta}_1 h(x_{ji}))I(x_{ji} \le x)$$

and

$$\hat{F}_2(x) = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \hat{p}_{ji} I(x_{ji} \leq x)$$

With these two estimates, an estimate of θ can be given as:

$$\hat{\theta} = \int_R \hat{F}_1(y) d\hat{F}_2(y) = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \hat{F}_1(x_{ji}) \hat{p}_{ji}.$$

It was shown that, $\sqrt{n}(\hat{\theta} - \theta)$ has an asymptotic normal distribution under the assumptions that the density ratio model (4.1) holds and $n_1/n_2 \to \rho$ as $n \to \infty$. As a result, a semi-parametric test for the hypothesis:

$$H_0 : \ \theta = \frac{1}{2} \ v.s. \ H_1 : \ \theta \neq \frac{1}{2}$$

can be performed based on the asymptotic normality.

To take into account a possible mis-specification of the function $h(\cdot)$, Fokianos and Troendle (2007) also considered the case where $h(\cdot)$ belongs to the Box-Cox family of transformations with an unknown parameter r, i.e.

$$h(x, r) = \begin{cases} \dfrac{x^r - 1}{r} & \text{when } r \neq 0 \\ \log x & \text{when } r = 0 \end{cases} . \tag{4.4}$$

and derived a test procedure by estimating r through a profile likelihood method. The new test procedure was compared with some non-parametric competitors and found to have relatively high power across a wide variety of distributions. More importantly, it was shown in the simulation study that the flexibility of the Box-Cox transformation allows robustness outside of the density ratio family.

4.2 Inference on the Density Ratio Model under the Random Censorship via a Conditional Empirical Likelihood Method

Shen et al. (2007) used the density ratio model to quantify the potential change in the time to disease diagnosis between the treatment and placebo arms in a breast cancer prevention trial. They proposed an approach based on the estimating equations conditional on uncensored survival times to estimate the parameters in the model.

Assume a random censoring model, $X = \min(T, C)$ and $\delta = I(X = T)$, where T and C are respectively true time to disease diagnosis and time to the last follow up. In addition, define a treatment indicator Z, where $Z = 1$ indicates the subject receiving treatment, while $Z = 0$ placebo. Assume also that the censoring variable C is independent of T given Z. In the j-th treatment group, assume that $\pi_k = \Pr(D = 1|Z = k) = \Pr(T < \infty|Z = j)$ is the probability that the disease would occur if the follow-up time were not right-censored, where $D = 1$ indicates the disease status and $j = 0$ or 1.

The survival distribution of T in the placebo group was assumed to follow a mixture cure rate model:

$$\bar{F}_1(t) = \pi_1 \bar{F}_{1d}(t|D = 1) + (1 - \pi_1), \tag{4.5}$$

where $\bar{F}_1(t)$ is an improper distribution of time to disease diagnosis and $\bar{F}_{1d}(t|D = 1)$ is the proper survival function for the time to disease diagnosis, conditional upon the ultimate occurrence of the disease. Subsequently, the corresponding probability

density function of T is:

$$f_1(t) = \pi_1 f_{1d}(t),$$

where $f_{1d}(t)$ is the probability density function of T conditional on $D = 1$.

Similarly, in the treatment group, the unconditional survival function of the time to disease diagnosis follows:

$$\bar{F}_2(t) = \pi_2 \bar{F}_{2d}(t|D = 1) + (1 - \pi_2). \tag{4.6}$$

The probability density function and distribution functions of the time to disease diagnosis conditional on $D = 1$ in the treatment group are denoted by f_{2d} and $F_{2d} = 1 - \bar{F}_{2d}$.

Suppose f_{1d} and f_{2d} satisfy a density ratio model:

$$f_{2d}(t) = \exp(\alpha^* + \beta h(t)) f_{1d}(t), \tag{4.7}$$

where h is a specified function to transform the observed time t, and f_{1d} and f_{2d} are unspecified positive density functions.

Let $f_{j\delta}$ define the density function of T conditional on $\delta = 1$ in the j-th group, where $j = 1$ for placebo group and $j = 2$ for treatment group. That is:

$$f_{j\delta}(t) = \Pr(T \in [t, t + dt)|\delta = 1, Z = j) = \frac{\Pr(T \in [t, t + dt), C > t|Z = j)}{\Pr(T < C|Z = j)}.$$

Under the independent censoring assumption, the above formula can be expressed as:

$$f_{j\delta}(t) = \frac{\pi_j f_{jd} \bar{G}_j(t)}{\int_0^\infty \pi_j f_{jd}(v) \bar{G}_j(v) dv} := \frac{f_{jd}(t) \bar{G}_j(t)}{\mu(f_{jd}, G_j)},$$

where $\bar{G}_j(t)$ is the survival function of the censoring variable C in the j-th group and $\mu(f_{jd}, G_j) = \int_0^\infty \pi_j f_{jd}(v) \bar{G}_j(v) dv$ is a constant with unknown quantities f_{jd} and G_j.

Let $y_1, \cdots, y_{m_1}, y_{m_1+1}, \cdots, y_m$ denote the observed failure times among $x_1, \cdots, x_{n_1}$, $x_{n_1+1}, \cdots, x_n$, where $m = m_1 + m_2$ and m_j is the number of diagnosed diseased cases

in the j-th group. Under the model assumption (4.7) between f_{1d} and f_{2d}, $f_{1\delta}$ and $f_{2\delta}$ can be related by the following equation:

$$\frac{f_{2\delta}(y)}{f_{1\delta}(y)} = \frac{f_{2d}(y)\bar{G}_2(y)\mu(f_{1d}, G_1)}{f_{1d}(y)\bar{G}_1(y)\mu(f_{2d}, G_2)} := \exp(\alpha + \beta h(y) + \psi(y)), \tag{4.8}$$

where $\alpha = \alpha^* + \log(\mu(f_{1d}, G_1)) - \log(\mu(f_{2d}, G_2))$ and $\psi(y) = \log \bar{G}_2(y) - \log \bar{G}_1(y)$. Given the observed disease cases, the corresponding conditional log-likelihood is:

$$\begin{aligned} l_c(\alpha, \beta) &= \sum_{i=1}^{m_1} \log f_{1\delta}(y_i) + \sum_{i=m_1+1}^{m} (\log f_{1\delta}(y_i) + \alpha + \beta h(y_i) + \psi(y_i)) \\ &:= \sum_{i=1}^{m} \log p_i + \sum_{i=m_1+1}^{m} (\alpha + \beta h(y_i) + \psi(y_i)), \end{aligned}$$

where $p_i = f_{1\delta}(y_i)$ subject to the constraint that:

$$\sum_{i=1}^{m} p_i = 1,\ p_i \geq 0,\ \text{and}\ \sum_{i=1}^{m} p_i (\alpha + \beta h(y_i) + \psi(y_i)) = 0.$$

If ψ is known, the score equations for (α, β) are:

$$\begin{aligned} \frac{\partial l_c}{\partial \alpha} &= m_2 - \sum_{i=1}^{m} \frac{\rho \exp(\alpha + \beta h(y_i) + \psi(y_i))}{1 + \rho \exp(\alpha + \beta h(y_i) + \psi(y_i))}, \\ \frac{\partial l_c}{\partial \beta} &= \sum_{i=m_1+1}^{m} h(y_i) - \sum_{i=1}^{m} \frac{\rho h(y_i) \exp(\alpha + \beta h(y_i) + \psi(y_i))}{1 + \rho \exp(\alpha + \beta h(y_i) + \psi(y_i))}, \end{aligned} \tag{4.9}$$

where $\rho = m_2/m_1$. Under some mild regularity conditions, solving estimation equation (4.9) produces consistent estimators of (α, β).

With the estimated $(\hat{\alpha}, \hat{\beta})$, we can subsequently obtain the estimate for $F_{1\delta} = \int_0^t f_{1\delta}(v)dv$, which can be defined by $\hat{F}_{1\delta}(t) = \sum_{y_i<t} \hat{p}_i$, where

$$\hat{p}_i = \frac{1}{m_1(1 + \rho \exp(\hat{\alpha} + \hat{\beta} h(y_i) + \psi(y_i)))},$$

and estimate for $F_{2\delta} = \int_0^t f_{2\delta}(v)dv$, denoted by $\hat{F}_{2\delta}(t) = \sum_{y_i<t} \hat{q}_i$, where

$$\hat{q}_i = \hat{p}_i \exp(\hat{\alpha} + \hat{\beta} h(y_i) + \psi(y_i)).$$

As a consequence, the estimates of $F_{jd},\ j = 1, 2$, can be derived respectively as:

$$\begin{aligned} \hat{F}_{1d}(t) &= \hat{\mu}(f_{1\delta}, G_1) \sum_{i=1}^{m} \hat{\bar{G}}_1^{-1}(y_i)\hat{p}_i I(y_i \leq t), \\ \hat{F}_{2d}(t) &= \hat{\mu}(f_{2\delta}, G_2) \sum_{i=1}^{m} \hat{\bar{G}}_2^{-1}(y_i)\hat{q}_i I(y_i \leq t), \end{aligned} \tag{4.10}$$

where

$$\hat{\mu}(f_{1d}, G_1) = \left(\sum_{i=1}^{m} \frac{\hat{p}_i}{\hat{\bar{G}}_1(y_i)} \right)^{-1}, \quad \hat{\mu}(f_{2d}, G_2) = \left(\sum_{i=1}^{m} \frac{\hat{q}_i}{\hat{\bar{G}}_2(y_i)} \right)^{-1},$$

with $\hat{\bar{G}}_j(t)$ obtained from the Kaplan-Meier estimate for the survival function of the censoring time.

In addition, they proposed to estimate π_1 and π_2 in (4.5) and (4.6) by the non-parametric method of Maller and Zhou (1992):

$$\hat{\pi}_1 = 1 - \hat{F}_1^{KM}(x_{(n_1)}),\ \hat{\pi}_2 = 1 - \hat{F}_2^{KM}(x_{(n_2)}),$$

where $F_j^{KM}(t)$ is the Kaplan-Meier estimator for $F_j(t)$, $x_{(n_1)}$ and $x_{(n_2)}$ are the last observed times in two groups. Since, without the model assumption (4.7), F_{1d} and F_{2d} can be also estimated as

$$\tilde{F}_{jd}(t) = \hat{\pi}_j^{-1} F_j^{KM}(t),\ j = 1, 2, \tag{4.11}$$

by comparison of the semi-parametric estimates (4.10) and non-parametric estimates (4.11), Shen et al. (2007) derived a goodness of fit test for checking model (4.7) based on the bootstrap method.

After validating the model assumption, Shen et al. (2007) proposed to test the treatment effect based on the testing of the following hypothesis:

$$H_0 :\ \beta = 0 \ v.s.\ H_1 :\ \beta \neq 0.$$

If we assume the censoring schemes are the same in two treatment groups, which

implies $\psi(t) = 0$, then the likelihood ratio statistic

$$R_n = 2(l_c(\hat{\alpha}, \hat{\beta}) - l(0, 0))$$

was shown to converge to a Chi-square distribution with one degree of freedom. Simulations were conducted to examine the size and power properties of the goodness of fit test for model (4.7) and the conditional likelihood ratio test for β.

4.3 Inference on the Density Ratio Model under the Random Censorship via a Weighted Empirical Likelihood Method

Ren (2008) instead applied a weighted empirical likelihood method (Ren, 2001) to the inference of parameters in a density ratio model under various censoring models, which is computationally more efficient than the method based on maximizing the full likelihood function with an EM algorithm.

Assume the true survival times in two independent treatment groups satisfying the following density ratio model:

$$\begin{aligned} T_{11}, \ \cdots, \ T_{1n_1} &\sim f_1(x), \\ T_{21}, \ \cdots, \ T_{2n_2} &\sim f_2(x) = g(x, \beta) f_1(x), \end{aligned} \tag{4.12}$$

where f_1 and f_2 are unknown density functions with distribution functions F_1 and F_2 respectively, and $g(x, \beta)$ is a known monotone function of $x \in R$ with unknown parameter $\beta = (\beta_1, \beta_2) \in R^2$. A special cases of $g(x, \beta)$ is

$$g(x, \beta) = \exp(\beta_0 + \beta_1 x). \tag{4.13}$$

Ren (2008) studied model (4.12) in several cases where at least one of the two

samples is not completely observable due to censoring, of which random censoring is a special case. In the next, I will only consider this special case.

Assume the observed data under the random censorship is $(X_{ji}, \delta_{ji}),\ j = 1, 2, i = 1, 2, \cdots, n_j$. Let $T^*_{11} < T^*_{12} < \cdots < T^*_{1m_1}$ and $T^*_{21} < T^*_{22} < \cdots < T^*_{2m_2}$ be uncensored survival times observed respectively in two treatment groups. Let

$$\hat{p}_{ji} = \prod_{k=1}^{i-1} \left(1 - \frac{d_{jk}}{r_{jk}}\right) \frac{d_{jk}}{r_{jk}},\ j = 1, 2,\ i = 1, 2, \cdots, m_j,$$

be the jump of the Kaplan-Meier estimator of F_j at T^*_{ji}, where d_{ji} and r_{ji} are respectively the number of death and number of subjects at risk at time T^*_{ji}. Then the Kaplan-Meier estimator for $F_j,\ j = 1, 2$ can be written as

$$\hat{F}^{KM}_j(x) = \sum_{i=1}^{m_j} \hat{p}_{ji} I(T^*_{ji} \le x).$$

Also let

$$\begin{aligned}
(W_1, W_2, \cdots, W_m) &= (T^*_{11}, T^*_{12}, \cdots, T^*_{1m_1}, T^*_{21}, T^*_{22}, \cdots, T^*_{2m_2}), \\
(\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_m) &= (\hat{p}_{11}, \hat{p}_{12}, \cdots, \hat{p}_{1m_1}, \hat{p}_{21}, \hat{p}_{22}, \cdots, \hat{p}_{2m_2}), \\
(\omega_1, \omega_2, \cdots, \omega_m) &= (\rho_1\hat{p}_{11}, \rho_1\hat{p}_{12}, \cdots, \rho_1\hat{p}_{1m_1}, \rho_2\hat{p}_{21}, \rho_2\hat{p}_{22}, \cdots, \rho_2\hat{p}_{2m_2}),
\end{aligned}$$

where $m = m_1 + m_2,\ \rho_1 = n_1/n$ and $\rho_2 = n_2/n$. Then the weighted empirical likelihood function for model (4.12) is given by:

$$L(\beta, F) = \left(\prod_{i=1}^{m} p_i^{n\omega_i}\right) \left(\prod_{i=m_1+1}^{m} [g(W_i, \beta)^{n\omega_i}]\right),$$

where $p_i = F(W_i) - F(W_i-)$. It was shown that $L(\beta, F)$ can be maximized by solving the following equations:

$$\begin{cases}
0 &= \displaystyle\int_0^\infty \frac{g(x,\beta)}{\rho_1 + \rho_2 g(x,\beta)} d\hat{F}^{KM}_1(x) - \int_0^\infty \frac{1}{\rho_1 + \rho_2 g(x,\beta)} d\hat{F}^{KM}_2(x) \\
0 &= \displaystyle\int_0^\infty \frac{\partial g(x,\beta)/\partial\beta_1}{\rho_1 + \rho_2 g(x,\beta)} d\hat{F}^{KM}_1(x) - \int_0^\infty \frac{\partial g(x,\beta)/\partial\beta_1}{(\rho_1 + \rho_2 g(x,\beta))g(x,\beta)} d\hat{F}^{KM}_2(x)
\end{cases}. \tag{4.14}$$

Suppose $\hat{\beta}$ is the semi-parametric maximum likelihood estimate obtained from (4.14), we get the semi-parametric MLE $\hat{F}_j(t)$ for $F_j(t)$, $j = 1, 2$, as:

$$\begin{aligned} \hat{F}_1(t) &= \sum_{i=1}^{m} \frac{\omega_i I(W_i \leq t)}{\rho_1 + \rho_2 g(W_i, \hat{\beta})}, \\ \hat{F}_2(t) &= \sum_{i=1}^{m} \frac{\omega_i g(W_i, \hat{\beta}) I(W_i \leq t)}{\rho_1 + \rho_2 g(W_i, \hat{\beta})}. \end{aligned} \tag{4.15}$$

Assuming some regularity conditions, Ren (2008) proved under model (4.12), $\hat{\beta}$, $\hat{F}_j(t)$ are consistent estimates for β and $F_j(t)$, respectively. She also showed that $\hat{\beta}$ has an asymptotic normal distribution as $n \to \infty$, and $\sqrt{n}(\hat{F}_j(t) - F_j(t))$ weakly converges to a centered Gaussian process.

In addition, when $g(x, \beta)$ is of the exponential form (4.13), Ren (2008) also proposed the following goodness of fit test statistic to assess the validity of the model assumption:

$$T_n = \sqrt{n} \sup_{0 \leq t < \infty} |\hat{F}_1(t) - \hat{F}_1^{KM}(t)|.$$

The bootstrap method was applied to compute the p-value for the test statistic T_n.

Chapter 5

Empirical Likelihood Confidence Intervals for the Hazard Ratio Function

5.1 Introduction

Hazard ratio is the most used statistical measure to assess the difference between treatments (Scott, 2000). It is defined as the ratio of two hazard rate functions. For a subject in the j-th group with a survival time T_j, hazard rate function at time t is defined as:

$$h_j(t) = \frac{f_j(t)}{1 - F_j(t)},$$

where $F_j(t)$ and $f_j(t)$ are respectively the distribution function and density function of T_j. The hazard ratio function at time t is, therefore, defined as:

$$\rho(t) = \frac{h_1(t)}{h_2(t)}.$$

Several procedures have been proposed in the literature to construct confidence intervals for the hazard ratio based on data with potential censoring. The Cox proportional hazard model (Cox, 1972) has been the most widely used procedure over many years to estimate the hazard ratio as well as to construct associated confidence interval, but the crucial assumption behind this procedure, the proportional hazard assumption, may not be satisfied by the data from epidemiologic studies or clinical trials, see the example provided by Stablein et al. (1981). Tu (2007) derived two types of under-smoothed kernel confidence intervals for the hazard ratio at a given time point t: one based on directly the asymptotic normality of the kernel hazard ratio estimate and the other on the Fieller's transformation of the hazard ratio estimator. It was found that, in terms of coverage probability, both under-smoothed confidence intervals performed reasonably well when the proportional hazard assumption was violated. However, these procedures are still not very satisfactory, because when the sample size is small, the true coverage probability is still far from the stated nominal level. This was not improved by a linear transformation of the kernel estimate. The requirement of estimating variance for the hazard ratio estimator may be the reason for the low accuracy of confidence intervals based on the asymptotic normality.

In this Chapter, I will explore the applications of the empirical likelihood method for the construction of a confidence interval for the hazard ratio. The empirical likelihood ratio confidence interval was first introduced by Owen (1988) for a single functional. A comprehensive introduction of the empirical likelihood method can be found in Owen (2001). Based on a data-driven likelihood ratio function expressed through constraints, an empirical likelihood method does not need to estimate variance when constructing a confidence interval, which leads to very favorable small

sample properties in comparison with its competitors. This method has been applied to some statistical problems with censored data, for example, construction of confidence interval for a survival function (Li, 1995), density and hazard function (Shen and He, 2008), the difference and ratio of survival functions (McKeaguea and Zhao, 2002; Shen and He, 2006).

In this Chapter, an empirical likelihood ratio function is defined for hazard ratio and shown to have a Chi-square asymptotic distribution with one degree of freedom. The coverage probability of the confidence interval based on this result is evaluated through Monte-Carlo simulations.

This Chapter is organized as follows: the empirical likelihood ratio function and associated confidence interval for the hazard ratio are defined in Section 5.2. Section 5.3 presents results of simulation studies and Section 5.4 the application of the proposed method to data from a clinical trial. Conclusions and discussions are presented in Section 5.5. Proof of the major results is given in Section 5.6.

5.2 Empirical Likelihood Ratio and Confidence Interval for the Hazard Ratio

Denote T_{ji} and C_{ji} ($j = 1,\ 2;\ i = 1,\ 2,\ \cdots,\ n_j$) the true survival and censoring times of the subjects in two groups, respectively. The data we observe from a clinical trial or cohort study are the pairs $(X_{j1},\ \delta_{j1})$, $(X_{j2},\ \delta_{j2})$, $\cdots$, $(X_{jn_j},\ \delta_{jn_j})$, where

$$\begin{cases} X_{ji} &= \min(T_{ji},\ C_{ji}) \\ \delta_{ji} &= I(T_{ji} \leq C_{ji}) \end{cases}.$$

The total sample size from two groups is $n = n_1 + n_2$. Write $0 \leq X_{j(1)} \leq X_{j(2)} \leq \cdots \leq X_{j(n_j)} < \infty$ as the ordered statistics of sample $\{X_{ji}\}$ and $\delta_{j(i)}$ the concomitant of $X_{j(i)}$ for $i = 1,\ 2, \cdots,\ n_j$ and $j = 1,\ 2$. Let

$$r_{ji} = \sum_{k=1}^{n_j} I(X_{jk} \geq X_{j(i)}) = n_j - i + 1$$

be the number of subjects that are still at risk before $X_{j(i)}$.

Now we will make some assumptions on the distribution of the true survival and censoring times. Suppose that $\{T_{ji} :\ i = 1, 2, \cdots, n_j\}$ are independently distributed with distribution function $F_j(t)$. The survival function of T_{ji} is defined as: $\bar{F}_j(t) = 1 - F_j(t)$. We also assume that $F_j(t)$ has a continuous density $f_j(t)$. The hazard function of T_{ji} can be written as

$$h_j(t) = \frac{f_j(t)}{\bar{F}_j(t)}.$$

Suppose that $\{C_{ji} :\ i = 1, 2, \cdots, n_j\}$ are independently distributed with distribution function $G_j(t)$ and write

$$H_j(t) = 1 - (1 - F_j(t))(1 - G_j(t)),$$

then H_j is the distribution functions of $\{X_{ji} :\ i = 1, 2, \cdots, n_j\}$, $j = 1,\ 2$.

The likelihood function based on censored data (1) is defined as:

$$L(F_1,\ F_2) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} (F_j(X_{ji}) - F_j(X_{ji}-))^{\delta_{ji}} (1 - F_j(X_{ji}))^{(1-\delta_{ji})}.$$

From Li (1995), this likelihood function can be rewritten as:

$$L(F_1,\ F_2) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} \lambda_{ji}^{\delta_{j(i)}} (1 - \lambda_{ji})^{r_{ji} - \delta_{j(i)}},$$

where

$$\lambda_{ji} = \frac{F(X_{j(i)}) - F(X_{j(i)}-)}{1 - F(X_{j(i)}-)},\ i = 1,\ 2,\ \cdots,\ n_j,\ j = 1, 2.$$

Therefore, we may express the cumulative hazard function $\Lambda_j(t) = -\ln \bar{F}_j(t)$ in terms

of $\{\lambda_{ji} : \ i = 1, 2, \cdots, n_j, j = 1, 2\}$:

$$\Lambda_j(t) = -\sum_{i=1}^{n_j} \ln(1 - \lambda_{ji}) I(X_{j(i)} \leq t).$$

Let $K_j(t)$ be a kernel function and $a_j = a(n_j)$ a bandwidth parameter. By the kernel smoothing method, an estimator of the hazard function could be chosen from the following estimation family:

$$\tilde{h}_j(t) = -\sum_{i=1}^{n_j} \ln(1 - \lambda_{ji}) K_{ji}(t),$$

where

$$K_{ji}(t) = \frac{1}{a_j} K_j \left(\frac{t - X_{ji}}{a_j} \right).$$

Note that different $\{\lambda_{ji}\}$ will lead to different estimate of $h_j(t)$. It is easy to show that $L(F_1, F_2)$ can be maximized by choosing:

$$\hat{\lambda}_{ji} = \frac{\delta_{j(i)}}{r_{ji}},$$

and this $\hat{\lambda}$ will give one of the estimators from the estimation family $\tilde{h}_j(t)$ defined as:

$$\hat{h}_j(t) = -\sum_{i=1}^{n_j} \ln \left(1 - \frac{\delta_{j(i)}}{r_{ji}} \right) K_{ji}(t), \ j = 1, 2.$$

The hazard ratio $\rho(t) = h_1(t)/h_2(t)$ is then estimated by:

$$\hat{\rho}(t) = \frac{\hat{h}_1(t)}{\hat{h}_2(t)}.$$

Under the constraints $\eta = \tilde{h}_2(t)$ and $\eta\rho(t) = \tilde{h}_1(t)$, we can define the following empirical likelihood ratio for $\rho(t)$:

$$R(\rho(t), \ \eta, \ t) = \frac{\sup_{\lambda_{ji}} \left\{ L(F_1, \ F_2) : \ \rho(t)\eta - \tilde{h}_1(t) = 0, \ \eta - \tilde{h}_2(t) = 0 \right\}}{\sup_{\lambda_{ji}} \{ L(F_1, \ F_2) \}}.$$

Then the log likelihood can be written as:

$$\ln(R(\rho(t),\ \eta,\ t)) =$$

$$\sup_{\lambda_{ji}} \left\{ \sum_{j=1}^{2} \sum_{i=1}^{n_j} (\delta_{j(i)} \ln \lambda_{ji} + (r_{ji} - \delta_{j(i)}) \ln(1 - \lambda_{ji})) : \rho(t)\eta - \tilde{h}_1(t) = 0,\ \eta - \tilde{h}_2(t) = 0 \right\}$$

$$- \sum_{j=1}^{2} \sum_{i=1}^{n_j} \delta_{j(i)} \ln \left(\frac{\delta_{j(i)}}{r_{ji}} \right) + (r_{ji} - \delta_{j(i)}) \ln \left(1 - \frac{\delta_{j(i)}}{r_{ji}} \right).$$

By the Lagrange Multiplier Method, we can get:

$$\ln(R(\rho(t),\ \eta,\ t)) = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \left\{ (r_{ji} - \delta_{j(i)}) \ln \left(1 + \frac{\mu_j K_{ji}(t)}{r_{ji} - \delta_{j(i)}} \right) - r_{ji} \ln \left(1 + \frac{\mu_j K_{ji}(t)}{r_{ji}} \right) \right\},$$

where the Lagrange Multipliers $\mu_j,\ j = 1,\ 2$, should satisfy:

$$\rho(t)\eta + \sum_{i=1}^{n_1} \ln \left(1 - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)} \right) K_{1i}(t) = 0, \tag{5.1}$$

$$\eta + \sum_{i=1}^{n_2} \ln \left(1 - \frac{\delta_{2(i)}}{r_{2i} + \mu_2 K_{2i}(t)} \right) K_{2i}(t) = 0. \tag{5.2}$$

We denote the left-hand sides of equations (5.1) and (5.2) as $Q_{1n}(\eta,\ \mu_1,\ \mu_2,\ t)$ and $Q_{2n}(\eta,\ \mu_1,\ \mu_2,\ t)$, respectively. Define

$$L_j(t) = \max_i \left\{ \frac{\delta_{j(i)} - r_{ji}}{K_{ji}(t)} \right\}.$$

Assume τ_1, τ_2 are two numbers such that

$$c_{F_1} \vee c_{F_2} < \tau_1 < \tau_2 < d_{H_1} \vee d_{H_2},$$

where $c_{F_j} = \inf\{x :\ F_j(x) > 0\}$ and $d_{F_j} = \sup\{x :\ F_j(x) < 1\}$. We restrict t in interval $[\tau_1,\ \tau_2]$. The reason t has to be restricted to this interval is that the law of the iterated logarithm for Kaplan-Meier estimator (Csorgo and Horvath, 1983) and kernel hazard estimator (Xiang, 1994), major tool in the proof of Theorem 5.1 below, may not be valid outside this interval. It can be shown that for each $t \in [\tau_1,\ \tau_2]$, Q_{1n} is a strictly increasing function of μ_1 on interval $(L_1(t),\ \infty)$ for a fixed n_1. When μ_1

approaches $L_1(t)$, we can find Q_{1n} decreasing to $-\infty$; when μ_1 approaches ∞, the limit of Q_{1n} will be η, which is positive. Therefore equation (5.1) has a unique root, and we can write it as $\mu_1(\eta,\ t)$. Similarly, we can show that equation (5.2) has a unique root $\mu_2(\eta,\ t)$. By the implicit function theorem, we get:

$$\frac{\partial \mu_1(\eta,\ t)}{\partial \eta} = -\rho(t)\left(\sum_{i=1}^{n_1} \frac{\delta_{1(i)} K_{1i}^2(t)}{(r_{1i}+\mu_1 K_{1i}(t))(r_{1i}+\mu_1 K_{1i}(t)-\delta_{1(i)})}\right)^{-1},$$

$$\frac{\partial \mu_2(\eta,\ t)}{\partial \eta} = -\left(\sum_{i=1}^{n_2} \frac{\delta_{2(i)} K_{2i}^2(t)}{(r_{2i}+\mu_2 K_{2i}(t))(r_{2i}+\mu_2 K_{2i}(t)-\delta_{2(i)})}\right)^{-1}.$$

Therefore, the equation:

$$\begin{aligned}
&\frac{\partial \ln(R(\rho(t),\ \eta,\ t))}{\partial \eta} \\
&= \frac{\partial \ln(R(\rho(t),\ \eta,\ t))}{\partial \mu_1}\frac{\partial \mu_1}{\partial \eta} + \frac{\partial \ln(R(\rho(t),\ \eta,\ t))}{\partial \mu_2}\frac{\partial \mu_2}{\partial \eta} \\
&= -\sum_{j=1}^{2}\sum_{i=1}^{n_j} \frac{\delta_{j(i)} K_{ji}^2(t)\mu_j}{(r_{ji}-\delta_{j(i)}+\mu_j K_{ji}(t))(r_{ji}+\mu_j K_{ji}(t))}\frac{\partial \mu_j}{\partial \eta} \\
&= 0
\end{aligned}$$

can be simplified into

$$\rho(t)\mu_1(\eta,\ t) + \mu_2(\eta,\ t) = 0,$$

which is equivalent to

$$\frac{\rho(t)\mu_1(\eta,\ t)}{n_1 a_1 + n_2 a_2} + \frac{\mu_2(\eta,\ t)}{n_1 a_1 + n_2 a_2} = 0 \tag{5.3}$$

We will show in our main Theorem that the unique root η_E of equation (5.3) can be found so that the log likelihood ratio function $\ln(R(\rho(t),\ \eta,\ t))$ reaches its maximum.

Denote the left-hand side of equation (5.3) as $Q_{3n}(\eta,\ \mu_1,\ \mu_2,\ t)$. Define the following conditions for the kernel function, bandwidth and hazard function:

1. $K_j(t)$ $(j = 1,\ 2)$ are bounded functions with a compact support $[-c,\ c]$ such that:

$$\int_{-\infty}^{\infty} u^i K_j(u)du = \begin{cases} = 1, & \text{if } i = 0, \\ = 0, & \text{if } i = 1, \\ \neq 0, & \text{if } i = 2. \end{cases}$$

The first order derivative of $K_j(t)$ exists.

2. Assume that $h_1(t) > 0$ and $h_2(t) > 0$ hold for $t \in [\tau_1,\ \tau_2]$. The derivative $h_j'(t)$ of $h_j(t)$ exists and is continuous.

3. As $n_j \to \infty$, we have $a_j \to 0$, $n_j a_j \to \infty$, $n_j a_j^5 \to 0$, $\liminf_{n\to\infty} n^{1/3} a_j > 0$, $\ln a_j^{-1}/n_j a_j \to 0$, $\ln a_j^{-1}/\ln\ln n_j \to \infty$, and $n_j a_j/(n_1 a_1 + n_2 a_2) \to \rho_j > 0,\ j = 1,\ 2$.

Specifically, we have the following theorem.

Theorem 5.1. *Assuming conditions 1–3, for each $t \in [\tau_1,\ \tau_2]$, there exists a solution $\eta_E(t)$ to equation (5.3) almost surely as $n \to \infty$, such that $R(\rho(t),\ \eta,\ t)$ attains its maximum value at $\eta = \eta_E$, and we have for fixed t*

$$-2\ln R(\rho(t),\ \eta_E,\ t) \to \chi_1^2,\ \textit{in distribution}$$

Proof. In Section 5.6. □

Remark: It is very important to select a bandwidth in our kernel smoothing estimate. Cheng et al. (2006) proposed an under-smoothing kernel bandwidth for the construction of the confidence interval for a hazard function. This under-smoothing bandwidth alleviates estimation difficulties caused by the bias and is shown to minimize the coverage error of a confidence interval for the hazard rate function. This

bandwidth satisfies the condition 3 for Theorem 5.1 and can be used in practice, although any bandwidth of order $O(n^{-1/3})$ can also be used. We used this bandwidth in our simulation studies and applications to real data from clinical trials.

From Theorem 5.1, an empirical likelihood confidence interval for hazard ratio function $\rho(t)$ at a fixed $t \in [\tau_1,\ \tau_2]$ with an asymptotical coverage accuracy $1-\alpha$ can be defined as:

$$I_{n,\alpha}(t) = \{\rho(t) :\ \ -2\ln R(\rho(t),\ \eta_E,\ t) \leq C_\alpha\}$$

where C_α satisfies:

$$\mathrm{P}(\chi_1^2 \leq C_\alpha) = 1-\alpha.$$

5.3 Simulation Studies

Simulations are conducted following the same scenarios in Tu (2007). Specifically, true survival times are assumed in the first group from an exponential distribution with parameter λ and in the second group from respectively, exponential distribution with parameter λ and Weibull and Gamma distributions with the shape and scale parameters respectively γ and λ. The censoring distribution is assumed to be uniformly distributed over interval $[T_f,\ T_a + T_f]$, which corresponds to a clinical trial process with patients accrued uniformly into the study from time 0 to time T_a and all patients followed for at least T_f time unit before the end of the study. λ ranges from 0.075, 0.05, and 0.025 but γ is fixed at 2. In addition, we fix T_a and T_f respectively at 60 and 6, as λ varies from 0.075, 0.05 to 0.025, which gives us the censoring rate of respectively 14%, 23%, and 45% when the distribution of the survival time is exponential, 10%, 20%, and 48% when the distribution of the survival time is Weibull,

and 34%, 50%, and 77% when the distribution of the survival time is Gamma. For each parameter configuration, 3000 random samples of sizes $n_1 = 100$ and $n_2 = 100$ are generated. The proportion of confidence intervals covering the true hazard ratio over 3000 samples are used to estimate the coverage probability for each confidence interval, and the average length of confidence intervals to estimate the length of the proposed confidence interval. The nominal significant level α used in all simulations is 0.05 and the following kernel function is used for all kernel estimates:

$$K(x) = \frac{15}{16}(1 - x^2)^2 I(|x| \leq 1)$$

The results of simulations are presented in Tables 5.1 and 5.2 respectively for the true coverage probability and length of proposed confidence intervals. In these tables, C_{el}, C_{tu}, C_{cox} and l_{el}, l_{tu}, l_{cox} represent respectively the coverage probabilities and lengths of confidence intervals based on the empirical likelihood method, asymptotic normality and Cox proportional hazard model. It can be seen from these tables that empirical likelihood method improves the confidence interval based on the normal approximation in almost all cases and the lengths of these two intervals are also comparable at majority of cases.

A special scenario is $h_2 = 2\lambda^2 t$, $\lambda = 0.025$ and $t = 6$, where the coverage probability based on the empirical likelihood is only 0.856, which was worse than that based on the asymptotic normality. If we take a look at the interval lengths in this scenario, we found both the empirical likelihood interval and asymptotic normality interval were very long (114.150 and 2421.374, respectively). This could be explained that at time $t = 6$, the observed uncensored cases are very few and not enough to produce a stable confidence interval. However, in practice, we do not recommend making any inferences at very beginning or very end of a clinical trial.

5.4 Application to a Data Set from Cancer Clinical Trial

We also applied the proposed empirical likelihood method to the same data set from a randomized clinical trial considered by Tu (2007). This trial was designed to compare two chemotherapy regimens (CEF v.s. CMF) in women with early stage breast cancer. 710 pre-menopausal women with axillary node positive breast cancer were recruited in this trial with a median follow-up of 8.8 years for all patients at the end of trial.

Table 5.3 presents confidence intervals for the hazard ratio of death at respectively 2, 4, 6 and 8 years after randomization based on respectively the normal approximation and empirical likelihood methods. "KHR" in Table 5.3 stands for the kernel hazard ratio estimate, while CI_{na} and CI_{el} represent the 95% confidence interval based on the normal approximation and empirical likelihood, respectively. From Table 5.3, we found the empirical likelihood confidence interval is slightly shorter except at 8 years from randomization. Although both methods would conclude that CEF is significantly better than CMF at 4 years after randomization, the upper endpoint of the empirical likelihood confidence interval is closer than 1, which confirms the results from the simulation study that the confidence interval based on the normal approximation may be more liberal than the empirical likelihood confidence interval.

Table 5.1: Comparison of Estimated Coverage Probabilities of Confidence Intervals for Hazard Ratio ($h_1(t) = \lambda$)

$h_2(t)$	λ	t	True Ratio	C_{el}	C_{tu}	C_{cox}
λ	0.075	6	1.000	0.941	0.925	0.953
		12	1.000	0.946	0.914	0.950
		24	1.000	0.950	0.901	0.950
	0.05	6	1.000	0.951	0.929	0.952
		12	1.000	0.952	0.932	0.951
		24	1.000	0.938	0.917	0.950
	0.025	6	1.000	0.947	0.930	0.942
		12	1.000	0.951	0.915	0.954
		24	1.000	0.955	0.908	0.944
$2\lambda^2 t$	0.075	6	1.111	0.945	0.928	0.699
		12	0.556	0.948	0.917	0.168
		24	0.278	0.948	0.907	0.000
	0.05	6	1.667	0.928	0.924	0.059
		12	0.833	0.944	0.929	0.884
		24	0.417	0.928	0.901	0.003
	0.025	6	3.333	0.856	0.916	0.002
		12	1.667	0.913	0.912	0.637
		24	0.833	0.935	0.910	0.507
$t\exp(-\lambda t)/\int_t^\infty u\exp(-\lambda u)du$	0.075	6	3.222	0.943	0.936	0.671
		12	2.111	0.945	0.922	0.825
		24	1.556	0.930	0.900	0.189
	0.05	6	4.333	0.950	0.927	0.312
		12	2.667	0.944	0.927	0.950
		24	1.833	0.925	0.907	0.417
	0.025	6	7.667	0.952	0.908	0.158
		12	4.333	0.939	0.921	0.865
		24	2.667	0.906	0.917	0.824

Table 5.2: Comparison of Estimated Lengths of Confidence Intervals for Hazard Ratio ($h_1(t) = \lambda$)

$h_2(t)$	λ	t		L_{el}	L_{tu}	L_{cox}
λ	0.075	6	1.000	1.441	1.603	0.617
		12	1.000	1.716	2.053	0.619
		24	1.000	3.549	32.083	0.618
	0.05	6	1.000	1.477	1.558	0.658
		12	1.000	1.597	1.901	0.657
		24	1.000	2.419	3.144	0.660
	0.025	6	1.000	1.853	1.692	0.788
		12	1.000	1.685	1.848	0.780
		24	1.000	2.060	2.580	0.778
$2\lambda^2 t$	0.075	6	1.111	1.455	1.715	0.558
		12	0.556	0.823	0.906	0.557
		24	0.278	109.996	12.984	0.555
	0.05	6	1.667	2.308	2.988	0.627
		12	0.833	1.135	1.325	0.625
		24	0.417	0.907	0.903	0.624
	0.025	6	3.333	5.926	12.938	0.979
		12	1.667	2.368	3.452	0.979
		24	0.833	1.681	1.611	0.979
$t\exp(-\lambda t)/\int_t^\infty u\exp(-\lambda u)du$	0.075	6	3.222	5.909	5.837	1.725
		12	2.111	3.410	3.635	1.729
		24	1.556	36.834	3.722	1.737
	0.05	6	4.333	10.583	9.742	2.101
		12	2.667	4.775	4.979	2.108
		24	1.833	4.338	3.995	2.109
	0.025	6	7.667	114.150	2421.374	3.856
		12	4.333	13.937	13.478	3.857
		24	2.667	8.472	6.379	3.824

Table 5.3: Estimates of the Hazard Ratio of Treatment CEF over CMF and associated 95% Confidence Intervals

Years from	Number at Risk		Hazard Rate		KHR	CI_{na}	CI_{el}
randomization	CEF	CMF	CEF	CMF			
2	323	326	0.0696	0.0897	0.78	0.44-1.11	0.46-1.07
4	284	265	0.0472	0.0782	0.60	0.28-0.92	0.39-0.99
6	251	239	0.0578	0.0437	1.32	0.55-2.09	0.80-2.12
8	216	207	0.0403	0.0537	0.75	0.28-1.21	0.45-1.52

5.5 Conclusions and Discussions

In this Chapter, we discussed how to apply the empirical likelihood method to construct a confidence interval for the hazard ratio function under the right censorship. The ratio of two kernel hazard estimates with an under-smoothed kernel bandwidth is used to estimate the hazard ratio. The empirical likelihood ratio is shown to be Chi-square distributed and the results of simulation studies find that the empirical likelihood method improves the coverage probabilities of the confidence intervals based on asymptotic normality. The proposed approach is applied to a real data set from clinical trials as illustrated in Section 5.4.

Generally speaking, when the censoring rate is very high, which occurs frequently in clinical trials assessing adjuvant treatments for early stage breast or other cancers, it is not recommended to make inferences on hazard ratio via either empirical likelihood method or approach based on asymptotic normality. Simulation studies have shown that when the number of uncensored cases is small, those two methods may produce very wide confidence intervals which may not provide useful information to guide clinical practice.

Just after the paper based on the work presented in this Chapter was published

(Jiang and Tu, 2010), I found that Zhao and Zhao (2011) also constructed confidence intervals for both ratio and difference of two hazard functions using the empirical likelihood methods. They have shown that the log empirical likelihood ratio of hazard difference and hazard ratio have an asymptotic chi-square distribution. Simulations were done and confirmed that the empirical likelihood approach outperformed asymptotic normality. In addition, Zhao and Zhao (2011) also proposed a bootstrap algorithm to obtain confidence bands for hazard ratio and hazard difference. The smoothing bandwidth used in Zhao and Zhao (2011) is, however, of order $o(n^{-1/5})$, while I considered an under-smoothing bandwidth of order $O(n^{-1/3})$. Also, while I used directly the formula of Cheng et al. (2006) to calculate bandwidth in simulation studies presented in this chapter, Zhao and Zhao (2011) used a bootstrap method to estimate the bandwidth in their simulation studies, which may be time consuming.

The confidence interval defined above is for a hazard ratio at a fixed time t. In practice, it may also be useful to have a simultaneous confidence interval over a given time interval. There is a technical difficulty to directly generalize the procedure developed in this Chapter to construct simultaneous confidence intervals since, as pointed out by Gilbert et al. (2002), the stochastic process defined by the kernel estimate of the hazard rate is not tight. For the density function, Hall and Owen (1993) derived empirical likelihood based simultaneous confidence intervals by following the technique used by Bickel and Rosenblatt (1973). Application of the same technique to construct simultaneous confidence intervals for a hazard ratio is an interesting problem for further investigation.

5.6 Proof of Theorem 5.1

In this Appendix, we assume the conditions of Theorem 5.1 are satisfied.

Lemma 5.1.

$$\hat{h}_j(t) - h_j(t) = O\left(\sqrt{\frac{\ln n_j}{n_j a_j}}\right),\ j = 1, 2.$$

Proof. Lemma 5.1 can be proved following the same arguments in the proof of Theorem 2.3 in Xiang (1994). □

Lemma 5.2. *As* $n \to \infty$,

$$\sqrt{n_j a_j}(h_j(t) - \hat{h}_j(t)) \to N(0, \sigma_j^2(t)),\ (j = 1, 2)\ \textit{in distribution,}$$

where

$$\sigma_j^2(t) = \frac{h_j}{\bar{H}_j} \int_{-c}^{c} K_j^2(t)dt.$$

Proof. Lemma 5.2 can be proved from Theorem 4.2 in Lo et al. (1989). □

Lemma 5.3. *Define* $\varepsilon_n = n^{-s}$, *with* $\frac{1}{3} < s < \frac{1}{2}$. *Let* $\eta_0 = h_2(t)$ *and assume that* $t \in [\tau_1, \tau_2]$, *then for any* η *satisfies* $|\eta - \eta_0| \leq a_1^{-1/2}\varepsilon_n$, *the solutions* $\mu_1(\eta,\ t)$ *and* $\mu_2(\eta,\ t)$ *of equations (5.1) and (5.2), respectively, satisfy:*

$$\frac{\mu_1(\eta,\ t)}{n_1} = O(a_1^{\frac{1}{2}}\varepsilon_n)\ \textit{and}\ \ \frac{\mu_2(\eta,\ t)}{n_2} = O(a_2^{\frac{1}{2}}\varepsilon_n)\ \textit{a.s.}$$

Proof. For $j = 1,\ 2$, define

$$\begin{cases} \hat{\sigma}_j^2(t) &= a_j n_j \sum_{i=1}^{n_j} \dfrac{\delta_{j(i)} K_{ji}^2}{r_{ji}(r_{ji} - \delta_{j(i)})}, \\ \tilde{\sigma}_j^2(t) &= a_j n_j \sum_{i=1}^{n_j} \dfrac{\delta_{j(i)} K_{ji}^2}{r_{ji}^2}. \end{cases} \tag{5.4}$$

Similar to the proof of Proposition 3.3.1 of Ramlau-Hansen (1983), we can show that

$$\hat{\sigma}_j^2(t) \to \sigma_j^2(t) \text{ a.s.}$$

$$\tilde{\sigma}_j^2(t) \to \sigma_j^2(t) \text{ a.s.}$$

Denote

$$\begin{cases} A_{1n}(\eta, t) &= \hat{h}_1(t) - \eta\rho(t) \\ A_{2n}(\eta, t) &= \hat{h}_2(t) - \eta \end{cases} . \tag{5.5}$$

Since we have from (5.1) and (5.2)

$$\begin{aligned} \eta\rho(t) &= -\sum_{i=1}^{n_1} \ln\left(1 - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)}\right) K_{1i}(t) \\ \eta &= -\sum_{i=1}^{n_2} \ln\left(1 - \frac{\delta_{2(i)}}{r_{2i} + \mu_2 K_{2i}(t)}\right) K_{2i}(t) \end{aligned} ,$$

using inequality $|\ln(1-x) - \ln(1-y)| \geq |x-y|$ for $x,\ y \in (0,\ 1)$, we can get

$$\begin{aligned} & \mu_1 A_{1n}(\eta,\ t) \\ = \ & \mu_1 \left[-\sum_{i=1}^{n_1} \ln\left(1 - \frac{\delta_{1(i)}}{r_{1i}}\right) K_{1i}(t) + \sum_{i=1}^{n_1} \ln\left(1 - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)}\right) K_{1i}(t) \right] \\ = \ & |\mu_1| \left| \sum_{i=1}^{n_1} K_{1i}(t) \left[\ln\left(1 - \frac{\delta_{1(i)}}{r_{1i}}\right) - \ln\left(1 - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)}\right) \right] \right| \\ \geq \ & |\mu_1| \sum_{i=1}^{n_1} K_{1i}(t) \left| \frac{\delta_{1(i)}}{r_{1i}} - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)} \right| \\ = \ & \mu_1^2 \sum_{i=1}^{n_1} K_{1i}^2(t) \frac{\delta_{1(i)}}{r_{1i}(r_{1i} + \mu_1 K_{1i}(t))} \\ \geq \ & \frac{\mu_1^2}{1 + \max_i \left(\frac{\mu_1 K_{1i}(t)}{r_{1i}}\right)} \sum_{i=1}^{n_1} \frac{K_{1i}^2(t)\delta_{1(i)}}{r_{1i}^2} \\ = \ & \frac{\mu_1^2 \tilde{\sigma}_1^2}{\left(1 + \max_i \left(\frac{\mu_1 K_{1i}(t)}{r_{1i}}\right)\right) n_1 a_1}. \end{aligned}$$

From condition 1, we have $|K_j(x)| \leq M$, for an $M > 0$ and $j = 1,\ 2$, which leads to

$$\mu_1 A_{1n}(\eta, t) \geq \frac{\mu_1^2 \tilde{\sigma}_1^2}{a_1 n_1 + M|\mu_1| \max_i(\frac{n_1}{r_{1i}})}.$$

Since for sufficiently large n_1 and n_2, we have almost surely ((4.6) in Shen and He (2008))

$$\max_i \left| \frac{n_j}{r_{ji}} \right| \leq \frac{2}{\bar{H}_j(\tau_2)},\ j = 1,\ 2,$$

and

$$\tilde{\sigma}_1^2(t) \geq \frac{1}{2}\sigma_1^2(\tau_1).$$

Therefore, we have

$$|A_{1n}(\eta,\ t)| \geq \frac{|\mu_1|\sigma_1^2(\tau_1)}{2\left(a_1 n_1 + 2M|\mu_1|\bar{H}_1^{-1}(\tau_2)\right)}. \tag{5.6}$$

On the other hand, from definition $h_2(t) = \eta_0$, we have by Lemma 5.1

$$\begin{aligned} A_{1n}(\eta,\ t) &= \hat{h}_1(t) - \rho(t)\eta_0 + \rho(t)\eta_0 - \rho(t)\eta \\ &= \hat{h}_1(t) - h_1(t) + \rho(t)(\eta_0 - \eta) \\ &\leq o(a_1^{-\frac{1}{2}}\varepsilon_n) + O(a_1^{-\frac{1}{2}}\varepsilon_n) \\ &= O(a_1^{-\frac{1}{2}}\varepsilon_n). \end{aligned} \tag{5.7}$$

Combining (5.6) and (5.7), we get

$$\frac{\mu_1(\eta,\ t)}{n_1} = O(a_1^{\frac{1}{2}}\varepsilon_n) \text{ a.s. for fixed } t \in [\tau_1,\ \tau_2].$$

Similarly, we can prove

$$\frac{\mu_2(\eta,\ t)}{n_2} = O(a_2^{\frac{1}{2}}\varepsilon_n) \text{ a.s. for fixed } t \in [\tau_1,\ \tau_2].$$

□

Lemma 5.4. *Almost surely, for large n_1 and n_2, equation (5.3) has a solution $\eta_E(t)$, such that $R(\rho(t),\ \eta,\ t)$ reaches its maximum value $R(\rho(t),\ t)$ at $\eta = \eta_E(t)$.*

Proof. For any pair $(j,\ i)$ which satisfies $X_{ji} < \tau_2$, we have almost surely for sufficiently large n

$$\frac{n_j}{r_{ji}} \leq \frac{n_j}{\sum_{k=1}^{n_j}(X_{jk} \geq \tau_2)} \leq \frac{2}{\bar{H}_j(\tau_2)}.$$

By Taylor Expansion and Lemma 5.3, we get

$$\begin{aligned}
&\ln\left[1-\frac{\delta_{j(i)}}{r_{ji}+\mu_j K_{ji}(t)}\right]K_{ji}(t)\\
&=\ln\left[1-\frac{\delta_{j(i)}}{r_{ji}}\left(1+\frac{\mu_j K_{ji}(t)}{r_{ji}}\right)^{-1}\right]K_{ji}(t)\\
&=K_{ji}(t)\ln\left[1-\frac{\delta_{j(i)}}{r_{ji}}\left(1-\frac{\mu_j K_{ji}(t)}{r_{ji}}+O\left(\frac{\mu_j^2K_{ji}^2(t)}{r_{ji}^2}\right)\right)\right]\\
&=K_{ji}(t)\ln\left[1-\frac{\delta_{j(i)}}{r_{ji}}+\frac{\mu_j\delta_{j(i)}K_{ji}(t)}{r_{ji}^2}-\delta_{j(i)}O\left(\frac{\mu_j^2K_{ji}^2(t)}{r_{ji}^3}\right)\right]\\
&=K_{ji}(t)\ln\left\{\left(1-\frac{\delta_{j(i)}}{r_{ji}}\right)\left[1+\left(1-\frac{\delta_{j(i)}}{r_{ji}}\right)^{-1}\delta_{j(i)}\left(\frac{\mu_jK_{ji}(t)}{r_{ji}^2}+O\left(\frac{\mu_j^2K_{ji}^2}{r_{ji}^3}\right)\right)\right]\right\}\\
&=K_{ji}(t)\ln\left(1-\frac{\delta_{j(i)}}{r_{ji}}\right)+K_{ji}(t)\ln\left[1+\frac{\delta_{j(i)}\mu_jK_{ji}(t)}{(r_{ji}-\delta_{j(i)})r_{ji}}+O\left(\frac{\mu_j^2K_{ji}^2(t)}{r_{ji}^2(r_{ji}-\delta_{j(i)})}\right)\right]\\
&=K_{ji}(t)\ln\left(1-\frac{\delta_{j(i)}}{r_{ji}}\right)+K_{ji}^2(t)\frac{\delta_{j(i)}\mu_j}{(r_{ji}-\delta_{j(i)})r_{ji}}+O\left(\frac{\mu_j^2K_{ji}^3(t)}{r_{ji}^3}\right)\\
&=K_{ji}(t)\ln\left(1-\frac{\delta_{j(i)}}{r_{ji}}\right)+K_{ji}^2(t)\frac{\delta_{j(i)}\mu_j}{(r_{ji}-\delta_{j(i)})r_{ji}}+O\left(\frac{\varepsilon_n^2}{na_j^2}\right).
\end{aligned}$$

Therefore, from (5.1), (5.4) and the above equation, we have almost surely

$$\begin{aligned}
\eta\rho(t) &= -\sum_{i=1}^{n_1}K_{1i}(t)\ln\left(1-\frac{\delta_{1(i)}}{r_{1i}+\mu_1K_{1i}(t)}\right)-\sum_{i=1}^{n_1}K_{1i}(t)\ln\left(1-\frac{\delta_{1(i)}}{r_{1i}}\right)\\
&\quad -\sum_{i=1}^{n_1}K_{1i}^2(t)\frac{\delta_{1(i)}\mu_1}{(r_{1i}-\delta_{1(i)})r_{1i}}+O\left(\frac{\varepsilon_n^2}{a_1^2}\right)\\
&= \hat{h}_1(t)-\frac{\mu_1(\eta,\ t)\hat{\sigma}_1^2}{n_1a_1}+O\left(\frac{\varepsilon_n^2}{a_1^2}\right).
\end{aligned}$$

Similarly, we can show that, almost surely

$$\eta = \hat{h}_2(t) - \frac{\mu_2(\eta,\ t)\hat{\sigma}_2^2}{n_2 a_2} + O\left(\frac{\varepsilon_n^2}{a_2^2}\right).$$

Hence, from (5.5), we get

$$\begin{aligned}
\mu_1(\eta,\ t) &= \frac{n_1 a_1(\hat{h}_1(t) - \eta\rho(t))}{\hat{\sigma}_1^2} + O\left(\frac{n_1\varepsilon_n^2}{a_1}\right) \\
&= \frac{a_1 n_1 A_{1n}(\eta,\ t)}{\hat{\sigma}_1^2} + O\left(\frac{n_1\varepsilon_n^2}{a_1}\right) \quad \text{a.s.,} \\
\mu_2(\eta,\ t) &= \frac{a_2 n_2 A_{2n}(\eta,\ t)}{\hat{\sigma}_2^2} + O\left(\frac{n_2\varepsilon_n^2}{a_2}\right) \quad \text{a.s.}
\end{aligned} \tag{5.8}$$

From Lemma 5.1, 5.3 and using Taylor Expansion again, we have

$$\begin{aligned}
& -2\ln(R(\rho(t),\ \eta,\ t)) \\
= & -2\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{(r_{ji} - \delta_{j(i)})\ln(1 + \frac{\mu_j K_{ji}(t)}{r_{ji} - \delta_{j(i)}}) - r_{ji}\ln(1 + \frac{\mu_j K_{ji}(t)}{r_{ji}})\right\} \\
= & -2\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{(r_{ji} - \delta_{j(i)})\left[\frac{\mu_j K_{ji}(t)}{r_{ji} - \delta_{j(i)}} - \frac{\mu_j K_{ji}(t)}{2(r_{ji} - \delta_{j(i)})^2} + O\left(\frac{\mu_j^3 K_{ji}^3(t)}{(r_{ji} - \delta_{j(i)})^3}\right)\right]\right. \\
& \left. -r_{ji}\left[\frac{\mu_j K_{ji}(t)}{r_{ji}} - \frac{\mu_j^2 K_{ji}^2}{2r_{ji}^2} + O\left(\frac{\mu_j^3 K_{ji}^3(t)}{r_{ji}^3}\right)\right]\right\} \\
= & 2\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{\frac{\mu_j^2 K_{ji}^2(t)}{2(r_{ji} - \delta_{j(i)})} + O\left(\frac{\mu_j^3 K_{ji}^3(t)}{(r_{ji} - \delta_{j(i)})^2}\right) - \frac{\mu_j^2 K_{ji}^2(t)}{2r_{ji}} - O\left(\frac{\mu_j^3 K_{ji}^3(t)}{r_{ji}^2}\right)\right\} \\
= & \sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{\frac{\delta_{j(i)} K_{ji}^2(t)\mu_j^2}{r_{ji}(r_{ji} - \delta_{j(i)})} + O\left(\varepsilon_n^3 n_j a_j^{-3/2}\right)\right\} \\
= & \frac{\mu_1^2(\eta,\ t)\hat{\sigma}_1^2}{n_1 a_1} + \frac{\mu_2^2(\eta,\ t)\hat{\sigma}_2^2}{n_2 a_2} + O\left(n\varepsilon_n^3 a_1^{-3/2}\right).
\end{aligned} \tag{5.9}$$

From (5.8) and (5.9)

$$-2\ln R(\rho(t),\ \eta,\ t) = \frac{a_1 n_1 A_{1n}^2(\eta,\ t)}{\hat{\sigma}_1^2} + \frac{a_2 n_2 A_{2n}^2(\eta,\ t)}{\hat{\sigma}_2^2} + O(n\varepsilon_n^3 a_1^{-3/2}). \tag{5.10}$$

If we write $\eta_n = \eta_0 + \Delta = h_2(t) + \Delta$, such that $\Delta \to 0$, $\Delta^2 a_1^{5/2}/\varepsilon_n^3 \to \infty$, and $\Delta^2 a_1 n_1 / \ln\ln n \to \infty$, using Taylor Expansion of $A_{jn}^2(\eta_n,\ t),\ j = 1,\ 2$ at η_0, we have almost surely

$$-2\ln R(\rho(t),\ \eta_n,\ t) = \frac{a_1 n_1}{\hat{\sigma}_1^2}(A_{1n}(\eta_0,\ t) - \rho(t)\Delta)^2 + \frac{a_2 n_2}{\hat{\sigma}_2^2}(A_{2n}(\eta_0,\ t) - \Delta)^2 + O(n\varepsilon_n^3 a^{-3/2}). \tag{5.11}$$

Since from Lemma 5.1, we have almost surely

$$\begin{aligned} A_{1n}(\eta_0,\ t) &= \hat{h}_1 - \eta_0\rho(t) = O(n^{-1/2}a_1^{-1/2}\sqrt{\ln n}), \\ A_{2n}(\eta_0,\ t) &= \hat{h}_2 - \eta_0 = O(n^{-1/2}a_1^{-1/2}\sqrt{\ln n}), \end{aligned} \tag{5.12}$$

from (5.11) and (5.12), we have almost surely

$$-2\ln R(\rho(t),\ \eta_n,\ t) = O(na\Delta^2).$$

On the other hand, for sufficiently large n_1 and n_2, we have from (5.10) almost surely

$$\begin{aligned} -2\ln R(\rho(t),\ \eta_0,\ t) &= O(\ln\ln n) + O(n\varepsilon_n^3 a^{-3/2}) \\ &= o(na\Delta^2) \ \text{ (from assumption on } \Delta.) \end{aligned}$$

Therefore, for sufficiently large n_1 and n_2, we have

$$-2\ln R(\rho(t),\ \eta_0 + \Delta,\ t) > -2\ln R(\rho(t),\ \eta_0,\ t),\ \text{a.s.}. \tag{5.13}$$

Similarly, we can obtain

$$-2\ln R(\rho(t),\ \eta_0 - \Delta,\ t) > -2\ln R(\rho(t),\ \eta_0,\ t),\ \text{a.s.}. \tag{5.14}$$

Combining (5.13) and (5.14), we know that $-2\ln R(\rho(t),\ \eta,\ t)$ attains its minimum in the region $(\eta_0 - \Delta,\ \eta_0 + \Delta)$, say at η_E.

□

Proof of Theorem 5.1

Proof. Denote $\nu_1 = \mu_1(\eta,\ t)/(n_1a_1)$, $\nu_2 = \mu_2(\eta,\ t)/(n_2a_2)$, $\nu_{1E} = \mu_1(\eta_E,\ t)/(n_1a_1)$ and $\nu_{2E} = \mu_2(\eta_E,\ t)/(n_2a_2)$. From (5.1)-(5.3), together with $n_1a_1/(n_1a_1+n_2a_2) \to p_1$ and $n_2a_2/n_1a_1 + n_2a_2 \to p_2$, we have

$$S_n(\eta,\ t) = \left.\frac{\partial(Q_{1n}, Q_{2n}, Q_{3n})}{\partial(\eta, \nu_1, \nu_2)}\right|_{(\eta,\nu_1,\nu_2,t)=(\eta,0,0,t)} = \begin{pmatrix} \rho(t) & \hat{\sigma}_1^2(t) & 0 \\ 1 & 0 & \hat{\sigma}_2^2(t) \\ 0 & \dfrac{n_1a_1\rho(t)}{n_1a_1+n_2a_2} & \dfrac{n_2a_2}{n_1a_1+n_2a_2} \end{pmatrix},$$

and

$$S_n(\eta,\ t) \to S(\eta,\ t) := \begin{pmatrix} \rho(t) & \sigma_1^2(t) & 0 \\ 1 & 0 & \sigma_2^2(t) \\ 0 & p_1\rho(t) & p_2 \end{pmatrix}$$

in probability. By Taylor expansion we get

$$\begin{aligned} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} Q_{1n}(\eta_E, \nu_{1E}, \nu_{2E}, t) \\ Q_{2n}(\eta_E, \nu_{1E}, \nu_{2E}, t) \\ Q_{3n}(\eta_E, \nu_{1E}, \nu_{2E}, t) \end{pmatrix} \\ &= \begin{pmatrix} Q_{1n}(\eta_0, 0, 0, t) \\ Q_{2n}(\eta_0, 0, 0, t) \\ 0 \end{pmatrix} + S_n(\eta_0,\ t) \begin{pmatrix} \eta_E - \eta_0 \\ \nu_{1E} \\ \nu_{2E} \end{pmatrix} + o(1). \end{aligned}$$

Therefore,

$$\begin{pmatrix} \eta_E - \eta_0 \\ \nu_{1E} \\ \nu_{2E} \end{pmatrix} \approx S_n^{-1}(\eta_0,\ t) \begin{pmatrix} \hat{h}_1(t) - h_1(t) \\ \hat{h}_2(t) - h_2(t) \\ 0 \end{pmatrix}$$

$$\approx \begin{pmatrix} -p_1\rho(t)\sigma_2^2(t) & -p_2\sigma_1^2(t) & * \\ -p_2 & p_2\rho(t) & * \\ p_1\rho(t) & -p_1\rho^2(t) & * \end{pmatrix} \begin{pmatrix} \hat{h}_1(t) - h_1(t) \\ \hat{h}_2(t) - h_2(t) \\ 0 \end{pmatrix} \frac{1}{\det(S(\eta_0,\ t))},$$

where

$$\det(S_n(\eta_0,\ t)) = -p_1\rho^2(t)\sigma_2^2(t) - p_2\sigma_1^2(t).$$

This leads to

$$\begin{aligned} \nu_{1E} &\approx \left(-p_1\rho^2(t)\sigma_2^2(t) - p_2\sigma_1^2(t)\right)^{-1} \Big(p_2(h_1(t) - \hat{h}_1(t)) - p_2\rho(t)(h_2(t) - \hat{h}_2(t))\Big) \\ &= \frac{p_2}{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)} \Big(\rho(t)(h_2(t) - \hat{h}_2(t)) - (h_1(t) - \hat{h}_1(t))\Big) \\ &= \frac{p_2}{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)} W(t), \end{aligned}$$

where

$$W(t) = \rho(t)(h_2(t) - \hat{h}_2(t)) - (h_1(t) - \hat{h}_1(t)).$$

From Lemma 5.2, we have $\sqrt{n_1 a_1}\ W(t)$ is asymptotically normal distributed with mean 0 and variance

$$\mathrm{var}(\sqrt{n_1 a_1}\ W(t)) = \frac{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)}{p_2}. \tag{5.15}$$

On the other hand, from (5.9) and (5.3), we can get

$$\begin{aligned}
-2\ln(R(\rho(t),\ \eta_E,\ t)) &\approx \frac{\mu_1^2(\eta_E,\ t)\sigma_1^2}{n_1 a_1} + \frac{\mu_2^2(\eta_E,\ t)\sigma_2^2}{n_2 a_2} \\
&= \frac{\mu_1^2(\eta_E,\ t)\sigma_1^2}{n_1 a_1} + \frac{\rho^2(t)\mu_1^2(\eta_E,\ t)\sigma_2^2(t)}{n_2 a_2} \\
&= \nu_{1E} n_1 a_1 \sigma_1^2(t) + \nu_{2E} n_2 a_2 \sigma_2^2(t) \\
&\approx \frac{n_1 a_1 p_2^2 \sigma_1^2(t)}{(p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t))^2} W^2(t) + \frac{\rho^2(t) n_1 a_1 p_1 p_2 \sigma_2^2(t)}{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)} W^2(t) \\
&= \frac{p_2}{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)} (\sqrt{n_1 a_1}\ W(t))^2.
\end{aligned}$$

Combining the above with (5.15), we have

$$-2\ln R(\rho,\ \eta_E,\ t) \to \chi_1^2,\ \text{in distribution.}$$

□

Chapter 6

Inference of Treatment Effect Based on the Probabilistic Index

6.1 Introduction

In a clinical trial with a time to an event, such as cancer recurrence or death, as the primary endpoint, the relative effect between two treatments under comparison are usually measured by a hazard ratio function and associated confidence interval (Scott, 2000). Without loss of generality, we call this time to an event variable as the survival time throughout this Chapter. Moser and McCann (2008) argued that the measure of relative treatment effect based on the hazard ratio may be difficult to understand by clinicians. They suggested the use of an alternative measure, defined as the probability that a patient receiving an experimental treatment has a better outcome than a patient receiving a control treatment, when a clinical trial with a time to an event endpoint is designed. They also demonstrated the connection between this measure and the hazard ratio.

Specifically, let T_1 and T_2 be two independent and true survival times for patients randomized respectively to two treatment groups. The measure suggested by Moser and McCann (2008) can be defined mathematically by:

$$\theta = \Pr(T_1 \leq T_2).$$

Under this definition, $\theta > 0.5$ implies that subjects in the second group are more likely to have a longer survival time than those in the first group, or the treatment for subjects in the second group is likely to be more effective compared to that for the subjects in the first group.

As mentioned by Buyse (2008), the proposed measure has been used for quite some time in many fields including, for example, psychology under the name "probabilistic index" (Acion et al., 2006) and reliability as a measure for stress-strength (Zhou, 2008). This measure is also identical to the area under the receiver operator characteristic (ROC) curve (Qin and Zhang, 2003; Brumback et al., 2006), which is a commonly used global index for measuring diagnostic accuracy in medical diagnostic testing.

In this Chapter, we are interested in the estimation and construction of confidence intervals for the θ based on potentially censored survival data $\{(X_{ji},\ \delta_{ji}) : i = 1, \cdots, n_j, j = 1, 2\}$ from clinical trials with two treatment groups, where

$$\begin{cases} X_{ji} &= \min(T_{ji},\ C_{ji}) \\ \delta_{ji} &= I(T_{ji} \leq C_{ji}) \end{cases},$$

with T_{ji} and C_{ji} are respectively true but potentially unobservable survival and censoring times for i-th subject in j-th group.

A common procedure to estimate the θ based on censored survival data, as mentioned by Moser and McCann (2008) , is to use the C-index or concordance C introduced by Harrell et al. (1982), which is based on the sum of the following scores generalized from the Wilcoxon statistic to censored data:

$$U_{kl} = \begin{cases} -1 & \text{if } X_{1k} < X_{2l} \text{ and } \delta_{1k} = 1 \\ 1 & \text{if } X_{1k} > X_{2l} \text{ and } \delta_{2l} = 1 \\ 0 & \text{otherwise} \end{cases} \quad k = 1, 2, \cdots, n_1,\ l = 1, 2, \cdots, n_2.$$

This estimator might not be consistent beyond the supports of the censored observation as pointed out by Bassiakos et al. (1991). Koziol and Jia (2009) pointed out again that the expected value of the estimate based on the C-index is dependent upon the underlying censoring distributions, hence may be far removed from the true θ. Buyse (2008) also showed that, without adjustment for censoring, considerable bias may be incurred when estimating the hazard ratio using the C-index based on the relationship derived by Moser and McCann (2008) between the hazard ratio and θ.

Since θ can be rewritten as:

$$\theta = \Pr(T_1 \leq T_2) = \int_0^\infty F_1(x) dF_2(x), \tag{6.1}$$

where $F_1(x)$ and $F_2(x)$ are respectively the distribution functions of T_1 and T_2, Koziol and Jia (2009) suggested that the θ be estimated by:

$$\hat{\theta}_E = \int \hat{F}_1^{KM}(x) d\hat{F}_2^{KM}(x), \tag{6.2}$$

where $\hat{F}_j^{KM}(x)$ is the Kaplan-Meier estimator (Kaplan and Meier, 1958) for $F_j(x)$, $j = 1, 2$. $\hat{\theta}_E$ was actually first proposed by Efron (1967) and shown in a simulation study of Koziol and Jia (2009) closer to the true θ with considerable smaller variance than the estimate based on the C-index. Koziol and Jia (2009) also suggested that an

alternative estimate for θ may be obtained when other estimates of $F_1(x)$ and $F_2(x)$ are available.

Zhou and Liang (2005) considered empirical likelihood inferences for a general measure of the treatment effect with θ as a special case under a parametric model for the distribution in one of the groups. In this Chapter, we consider the inference of θ when the distribution in one group is nonparametric but a relationship between $F_1(x)$ and $F_2(x)$ can be assumed. Specifically, we consider the following semi-parametric density ratio model which assumes that

$$\begin{aligned} &T_{11}, \ \cdots, \ T_{1n_1} \sim f_1(x), \\ &T_{21}, \ \cdots, \ T_{2n_2} \sim f_2(x) = \exp(\beta_0 + \beta_1 g(x)) f_1(x), \end{aligned} \tag{6.3}$$

where $f_1(x)$ and $f_2(x)$ are the corresponding density function of $F_1(x)$ and $F_2(x)$, respectively, β_0 and β_1 are unknown parameters, and $g(x)$ is a link function, which may depend on an unknown parameter. Better estimate and more accurate confidence interval for the θ may be obtained by combining data from two groups to estimate $F_1(x)$ and $F_2(x)$ under this model.

Model (6.3) has a long history in its relationship with the logistic regression model (see, for example, earlier references by Anderson (1979), Kay and Little (1987) and Cox and Ferry (1991)). Qin and Zhang (1997) formed it as an equivalent to the logistical regression model for data from a case-control study and used it for checking assumptions underlying a logistic regression model. They also pointed out that the density ratio model is also a biased sampling model with a specific but unknown weight. Qin (1998) also mentioned connections between this model and Cox proportional hazards model. As seen in Fokianos and Kaimi (2006), the density ratio model

is actually broader than the Cox proportional hazard models. For example, two log-normal distributions with different means and same variance satisfy the density ratio model but not the proportional hazards model. When there is no censoring in the data set, various aspects of the inference on the density ratio model have been considered by many researchers. A sample of recent references can be found in, for example, Zhang (2000), Cheng and Chu (2004), Fokianos (2004), Fokianos and Troendle (2007), Kezioua and Leoni-Aubin (2008), and Kedem et al. (2009). Specifically, Fokianos and Troendle (2007) derived an estimate of the θ under a density ratio model.

In contrast, when there is a censoring in the data set, model (6.3) was studied only by a few authors, partly because the inference based on a full likelihood function of censored data may be computationally intensive. Shen et al. (2007) used the density ratio model to quantify the potential change in the time to disease diagnosis between the treatment and placebo arms in a breast cancer prevention trial and proposed an approach based on the estimating equations conditional on uncensored survival times to estimate parameters in the model. Ren (2008) instead applied a weighted empirical likelihood method (Ren, 2001) to the inference of parameters in a density ratio model under various censoring models, which is computationally more efficient than the method based on maximizing the full likelihood function with an EM algorithm.

In this Chapter, when a density ratio model (6.3) with a specific but flexible link function $g(x)$ defined by the Box-Cox transformation (Box and Cox, 1964) is satisfied by the density functions of survival times in two treatment groups, we adopt both conditional and weighted likelihood methods to estimate and construct the confidence interval for the θ based on censored survival times. Simulations are performed to compare the estimates and confidence intervals from these two methods. The proposed

methods are also applied to a data set from a clinical trial on early breast cancer after deriving a goodness of fit test to check whether the density ratio model is satisfied by the data set.

The rest of the Chapter is organized as follows. In Section 6.2, we present the methods for estimating θ and constructing its confidence interval via conditional likelihood method and weighted empirical likelihood method. A summary of results from simulations evaluating proposed methods is provided in Section 6.3. In Section 6.4, the proposed methods are applied to analyze a data set from a clinical trial. The Chapter is completed with some conclusions and discussions in Section 6.5.

6.2 Estimation and Confidence Interval of θ

As mentioned above, throughout this Chapter, the following density ratio model is assumed for the true survival times from two treatment groups:

$$\begin{aligned} &T_{11}, \ \cdots, \ T_{1n_1} \sim f_1(x), \\ &T_{21}, \ \cdots, \ T_{2n_2} \sim f_2(x) = \exp(\beta_0 + \beta_1 h(x,r)) f_1(x), \end{aligned} \tag{6.4}$$

where β_0 and β_1 are unknown parameters, $f_1(x)$ is an unknown density function, and $h(x,r)$ is a Box-Cox transformation of x defined as:

$$h(x,r) = \begin{cases} \dfrac{x^r - 1}{r} & \text{when } r \neq 0 \\ \log x & \text{when } r = 0 \end{cases}, \tag{6.5}$$

with an unknown parameter r. The commonly used link functions, for example, $h(x,r) = \ln x$ and $h(x,r) = x - 1$, could be obtained by letting $r = 0$ and $r = 1$ in definition (6.5), respectively. With an unknown parameter r, model (6.5) could fit broader types of survival data observed from practise. For example, Fokianos and

Kaimi (2006) provided some other examples of link functions which would fit survival times from two log-normal distributions with different means but same variance, or two Gamma distributions with different scale parameters but same shape parameter. The use of $h(x, r)$ as the link function would also reduce the bias associated with estimating β_0 and β_1 when the link function is mis-specified as pointed out by Fokianos and Kaimi (2006) and Fokianos and Troendle (2007).

Based on censored survival times in two treatment groups, we may estimate unknown parameters β_0, β_1, and r and the distribution function associated with $f_1(x)$. Once these estimates are obtained, we can derive the estimate for the the distribution function associated with $f_2(x)$ from the density ratio model. An estimate of θ can then be obtained by plugging the estimates of distribution functions $F_1(x)$ and $F_2(x)$ into (6.1). Next we introduce both conditional and weighted likelihood methods for estimation of β_0, β_1, r, and $F_1(x)$ under the density ratio model (6.4).

Denote $T_{11}^* < T_{12}^* < \cdots < T_{1m_1}^*$ and $T_{21}^* < T_{22}^* < \cdots < T_{2m_2}^*$ as uncensored survival times observed respectively in two treatment groups. Let

$$\hat{p}_{ji} = \prod_{k=1}^{i-1} \left(1 - \frac{d_{jk}}{r_{jk}}\right) \frac{d_{jk}}{r_{jk}}, \ j = 1, 2, \ i = 1, 2, \cdots, m_j,$$

be the jump of the Kaplan-Meier estimator of F_j at T_{ji}^*, where d_{ji} and r_{ji} are respectively the number of death and number of subjects at risk at time T_{ji}^*. Then the Kaplan-Meier estimator for $F_j, \ j = 1, 2$ can be written as

$$\hat{F}_j^{KM}(x) = \sum_{i=1}^{m_j} \hat{p}_{ji} I(T_{ji}^* \leq x).$$

Define

$$\begin{aligned}(W_1, W_2, \cdots, W_m) &= (T^*_{11}, T^*_{12}, \cdots, T^*_{1m_1}, T^*_{21}, T^*_{22}, \cdots, T^*_{2m_2}),\\ (\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_m) &= (\hat{p}_{11}, \hat{p}_{12}, \cdots, \hat{p}_{1m_1}, \hat{p}_{21}, \hat{p}_{22}, \cdots, \hat{p}_{2m_2}),\\ (\omega_1, \omega_2, \cdots, \omega_m) &= (\rho_1\hat{p}_{11}, \rho_1\hat{p}_{12}, \cdots, \rho_1\hat{p}_{1m_1}, \rho_2\hat{p}_{21}, \rho_2\hat{p}_{22}, \cdots, \rho_2\hat{p}_{2m_2}),\end{aligned}$$

where $m = m_1 + m_2$, $\rho_1 = n_1/n$ and $\rho_2 = n_2/n$.

From Shen et al. (2007) and Ren (2008), we can define respectively the log conditional likelihood l_c and log weighted empirical likelihood l_w of parameters based on observed data as resepctively:

$$\begin{aligned}l_c(\beta, r, \mathbf{p}) &= \sum_{i=1}^{m} \log p_i + \sum_{j=m_1+1}^{m} (\beta_0 + \beta_1 h(W_j, r)),\\ l_w(\beta, r, \mathbf{p}) &= n\sum_{i=1}^{m} \omega_i \log p_i + n\sum_{j=m_1+1}^{m} \omega_j(\beta_0 + \beta_1 h(W_j, r)),\end{aligned}$$

where $\beta = (\beta_0, \beta_1)$ and $\mathbf{p} = (p_1, \cdots, p_m)$ with $p_i = F_1(W_i) - F_1(W_i-)$ $(i = 1, \cdots, m)$ satisfying the restrictions:

$$p_i \geq 0,\ \sum_{i=1}^{m} p_i = 1 \text{ and } \sum_{i=1}^{m} p_i \exp(\beta_0 + \beta_1 h(W_i, r_0)) = 1, \tag{6.6}$$

which guarantee respectively $\int f_1(x)dx = 1$ and $\int f_2(x)dx = 1$.

Let $l_u(\beta, r, \mathbf{p})$, $u = c$ or w, be either the log conditional or weighted empirical likelihood functions. The following algorithm can be used to estimate $\beta = (\beta_0, \beta_1)$, r, and $\mathbf{p} = (p_1, \cdots, p_m)$ based on each of these likelihood functions:

Algorithm 6.1.

1. For the fixed r and β, maximize first the profile log-likelihood $l_u(\beta, r, \boldsymbol{p})$ over all possible probability mass $\boldsymbol{p}$ satisfying the restrictions in (2.6). Denote respectively

$$PL_u(\beta, r) = \max_{\boldsymbol{p}} l_u(\beta, r, \boldsymbol{p})$$

and

$$\hat{\boldsymbol{p}}_u(\beta, r) = \arg\max_{\boldsymbol{p}} l_u(\beta, r, \boldsymbol{p}).$$

2. Obtain the maximum empirical likelihood estimator of β given r by maximizing the profile log likelihood $PL_u(\beta, r)$. Denote this estimator by:

$$\hat{\beta}_u(r) = (\hat{\beta}_{0u}(r), \hat{\beta}_{1u}(r)) = \arg\max_{\beta} PL_u(\beta, r).$$

3. Define a profile likelihood for r as

$$PL_u(r) = l_u(\hat{\beta}_u(r), r, \hat{\boldsymbol{p}}_u(\hat{\beta}_u(r), r)).$$

Obtain the final estimate of r by maximizing $PL_u(r)$ and denote it by:

$$\hat{r}_u = \arg\max_{r} PL_u(r).$$

4. Define the final estimates for $\beta = (\beta_0, \beta_1)$ and $\boldsymbol{p} = (p_1, \cdots, p_m)$ as respectively:

$$\hat{\beta}_u = \hat{\beta}_u(\hat{r}_u)$$

and

$$\hat{\boldsymbol{p}}_u = \hat{\boldsymbol{p}}_u(\hat{\beta}_u, \hat{r}_u).$$

For the conditional likelihood function and by the Lagrange's method, we have $\hat{\mathbf{p}}_c = (\hat{p}_{1c}, \cdots, \hat{p}_{mc})$, where

$$\hat{p}_{ic}(\beta, r) = \frac{1}{m_1 + m_2 \exp(\beta_0 + \beta_1 h(W_i, r))}, \ i = 1, 2, \cdots m.$$

It can also be shown that $\hat{\beta}_c(r)$ is the root of the following system of equations in respect of β_0 and β_1:

$$\begin{cases} 0 &= m_2 - \sum_{i=1}^{m} \dfrac{m_2 \exp(\beta_0 + \beta_1 h(W_i, r))}{m_1 + m_2 \exp(\beta_0 + \beta_1 h(W_i, r))} \\ 0 &= \sum_{j=m_1+1}^{m} h(W_j, r) - \sum_{i=1}^{m} \dfrac{m_2 h(W_i, r)}{m_1 + m_2 \exp(\beta_0 + \beta_1 h(W_i, r))} \end{cases}.$$

With $\hat{\mathbf{p}}_c(\beta, r)$ and $\hat{\beta}_c(r)$ defined above, we can calculate $\hat{r}_c$ by a numerical algorithm, such as the bisection method. The final maximum conditional empirical likelihood estimates $\hat{\beta}_c$ and $\hat{\mathbf{p}}_c$ for β and $\mathbf{p}$ can then be obtained. Consequently, we have the following estimates of F_1, F_2 and θ based on the conditional empirical likelihood method:

$$\begin{aligned} \hat{F}_{1c}(t) &= \sum_{i=1}^{m} \hat{p}_{ic} I(W_i \le t) \\ &= \sum_{i=1}^{m} \frac{I(W_i \le t)}{m_1 + m_2 \exp(\hat{\beta}_{0c} + \hat{\beta}_{1c} h(W_i, \hat{r}_c))}, \end{aligned} \tag{6.7}$$

$$\begin{aligned} \hat{F}_{2c}(t) &= \sum_{i=1}^{m} \hat{p}_{ic} \exp(\hat{\beta}_{0c} + \hat{\beta}_{1c} h(W_i, \hat{r}_c)) I(W_i \le t) \\ &= \sum_{i=1}^{m} \frac{\exp(\hat{\beta}_{0c} + \hat{\beta}_{1c} h(W_i, \hat{r}_c)) I(W_i \le t)}{m_1 + m_2 \exp(\hat{\beta}_{0c} + \hat{\beta}_{1c} h(W_i, \hat{r}_c))} \end{aligned} \tag{6.8}$$

and

$$\begin{aligned} \hat{\theta}_c &= \int_0^{\infty} \hat{F}_{1c}(t) d\hat{F}_{2c}(t) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\exp(\hat{\beta}_{0c} + \hat{\beta}_{1c} h(W_j, \hat{r}_c)) I(W_i \le W_j)}{(m_1 + m_2 \exp(\hat{\beta}_{0c} + \hat{\beta}_{1c} h(W_i, \hat{r}_c)))(m_1 + m_2 \exp(\hat{\beta}_{0c} + \hat{\beta}_{1c} h(W_j, \hat{r}_c)))}. \end{aligned}$$

Similarly, if we define $\Omega_0 = \sum_{j=1}^{m_1} \omega_j$ and $\Omega_1 = \sum_{j=m_1+1}^{m} \omega_j$, we can derive the following estimates of F_1, F_2 and θ based on the log weighted empirical likelihood function:

$$\begin{aligned} \hat{F}_{1w}(t) &= \sum_{i=1}^{m} \hat{p}_{iw} I(W_i \le t) \\ &= \sum_{i=1}^{m} \frac{\omega_i I(W_i \le t)}{\Omega_0 + \Omega_1 \exp(\hat{\beta}_{0w} + \hat{\beta}_{1w} h(W_i, \hat{r}_w))}, \end{aligned} \tag{6.9}$$

$$\begin{aligned}\hat{F}_{2w}(t) &= \sum_{i=1}^{m} \hat{p}_{iw} \exp(\hat{\beta}_{0w} + \hat{\beta}_{1w} h(W_i, \hat{r}_w)) I(W_i \le t) \\ &= \sum_{i=1}^{m} \frac{\omega_i \exp(\hat{\beta}_{0w} + \hat{\beta}_{1w} h(W_i, \hat{r}_w)) I(W_i \le t)}{\Omega_0 + \Omega_1 \exp(\hat{\beta}_{0w} + \hat{\beta}_{1w} h(W_i, \hat{r}_w))}\end{aligned} \tag{6.10}$$

and

$$\begin{aligned}\hat{\theta}_w &= \int_0^{\infty} \hat{F}_{1w}(t) d\hat{F}_{2w}(t) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\omega_i \omega_j \exp(\hat{\beta}_{0w} + \hat{\beta}_{1w} h(W_j, \hat{r}_w)) I(W_i \le W_j)}{(\Omega_0 + \Omega_1 \exp(\hat{\beta}_{0w} + \hat{\beta}_{1w} h(W_i, \hat{r}_w)))(\Omega_0 + \Omega_1 \exp(\hat{\beta}_{0w} + \hat{\beta}_{1w} h(W_j, \hat{r}_w)))},\end{aligned}$$

where $\hat{p}_{iw} = \hat{p}_{iw}(\hat{\beta}_w, \hat{r}_w)$ with

$$\hat{p}_{iw}(\beta, r) = \frac{\omega_i}{\Omega_0 + \Omega_1 \exp(\beta_0 + \beta_1 h(W_i, r))},\ i = 1, 2, \cdots, m,$$

and $\hat{\beta}_w = \hat{\beta}_w(\hat{r}_w)$ with $\hat{\beta}_w(r)$ as the root of the following system of equations:

$$\begin{cases} 0 &= \sum_{i=1}^{m} \frac{\omega_i [\exp(\beta_0 + \beta_1 h(W_i, r)) - 1]}{\Omega_0 + \Omega_1 \exp(\beta_0 + \beta_1 h(W_i, r))} \\ 0 &= \sum_{j=m_1+1}^{m} \omega_j h(W_j, r_0) - \Omega_1 \sum_{i=1}^{m} \hat{p}_{iw}(\beta, r) \exp(\beta_0 + \beta_1 h(W_i, r)) h(W_i, r) \end{cases}.$$

Under some mild conditions given in Shen et al. (2007) and Assumption (AS5) in Ren (2008) respectively, it can be shown that $\hat{\theta}_c$ and $\hat{\theta}_w$ are consistent and asymptotically normal distributed by using the similar approaches in Qin and Zhang (1997) and Ren (2008). However, their asymptotic variances are quite involved, which we will not write down the details here, and difficult to estimate directly. In the following, we present a bootstrap method (Shao and Tu, 1995) to estimate the variances of $\hat{\theta}_c$ and $\hat{\theta}_w$.

Define $S_j = \{(X_{ji}, \delta_{ji}),\ i = 1, 2, \cdots, n_j\},\ j = 1, 2$. We can generate B bootstrap samples S^*_{jb} $(b = 1, 2, \cdots, B)$ by re-sampling respectively n_j items with replacement from the original sample S_j $(j = 1, 2)$. For each of these bootstrap samples, we

can obtain the bootstrap estimate $\hat{\theta}^*_{bc}$ or $\hat{\theta}^*_{bw}$ $(b = 1, 2, \cdots, B)$ by following the same estimation procedures described above. With these B bootstrap estimates, we can define the bootstrap variance estimator for $var(\hat{\theta}_u)$ $(u = c$ or $w)$ as:

$$\widehat{var}^*(\hat{\theta}_u) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^*_{bu} - \bar{\hat{\theta}}^*_u)^2,$$

where

$$\bar{\hat{\theta}}^*_u = \sum_{b=1}^{B}\frac{\hat{\theta}^*_{bu}}{B}.$$

With this bootstrap variance estimator, $1-\alpha$ confidence interval for θ can be defined as

$$\left(\hat{\theta}_u + z_{\alpha/2}\widehat{sd}(\hat{\theta}_u), \hat{\theta}_u + z_{1-\alpha/2}\widehat{sd}(\hat{\theta}_u)\right),$$

where $\Phi(z_\alpha) = \alpha$, with Φ as the standard normal cumulative distribution function, and

$$\widehat{sd}(\hat{\theta}_u) = \sqrt{\widehat{var}^*(\hat{\theta}_u)}.$$

In the next section, we report results from Monte-Carlo simulations, which evaluate the performance of both estimation and confidence interval procedures based on the conditional and weighted empirical likelihood functions.

6.3 Simulation Studies

The following scenario of a clinical trial process was considered in our simulations. Patients were assumed accrued uniformly into a clinical trial from the time 0 to time T_a, and then followed for at least T_f time units before the end of the study. The total duration of the study is therefore $T_a + T_f$. The true survival time of patients in the first treatment group was assumed to have a Weibull distribution with a scale

parameter λ and shape parameter γ. The density ratio model with given β and r was used to generate the true survival times for patients in the second group. The survival times of patients were censored if the true survival times was longer than $(T_a - u) + T_f$, where u is the time when the patient entered the trial. In all simulations, T_a and T_f were fixed at respectively 1 and 1.6 time units but we investigated several configurations of λ, γ, β and r, which produced various values of θ (from 0.5 which indicates no treatment effect to 0.824 which indicates a very large treatment effect) and censoring rates (from as low as 9% in both groups, which may represent a clinical trial for patients with an advanced stage of the cancer, to as high as 35% in one group and 83% in the other, which may represent a clinical trial for patients with a newly diagnosed curable cancer, such as breast and prostate cancers). The details of these simulation configurations are given in Table 6.1.

The sample size was assumed the same in two treatment groups, which varied from 50, 100 to 150. For each configuration of parameters and the sample size, 3000 samples were simulated to evaluate the bias, defined as $\hat{\theta} - \theta$, standard error (SD) and mean squared error (MSE) of the two proposed estimates for θ and the estimate defined in (6.2), which is called as the Efron's estimator. The coverage probability and length of the bootstrap confidence intervals based on the proposed estimates for θ was also evaluated and compared with that based on the asymptotic normality of the Efron's estimator with its asymptotic variance estimated from a formula derived by Efron (1967). The nominal level varied from 90%, 95% to 99% and the number of bootstrap replications used to construct the confidence intervals was 100 for all configurations.

The results of simulations are presented in Tables 6.2-6.4. From Table 6.2, we can

Table 6.1: Configurations of Parameters for Tables 6.2-6.4

Configurations	λ	γ	r	β_0	β_1	θ	Censoring Rate 1	Censoring Rate 2
1	3	1.5	0	0	0	0.5	9%	9%
2	3	1.5	0.2	-3.81	1	0.608	9%	15%
3	1.5	1.5	0	0	0	0.5	20%	20%
4	1.5	1.5	0.3	-3.75	1	0.702	20%	44%
5	0.5	1.8	0	0	0	0.5	34%	34%
6	1	2	0.4	-5.32	1	0.824	35%	83%

see that the estimator based on the conditional likelihood has the smallest standard deviation in most of the cases. Its bias and MSE are also the smallest in almost all cases except when the censoring in one treatment group is very heavy. The MSE of the Efron's estimator is comparable to that based on the conditional likelihood but smaller than that based on the weighted empirical likelihood. Table 6.3 shows the coverage probability of confidence interval based on the conditional likelihood is closer to the nominal level except in configurations 5 and 6 where the censoring is moderate or heavy in both groups. The conditional likelihood method did not perform well when the censoring rate is high may be because it uses only information of uncensored survival times to define the likelihood. The coverage probability of the confidence interval based on the Efron's estimator became far below the nominal level as the censoring rate increases, which may be due to the difficulty in defining the Kaplan-Meier estimate after the last observed survival time which is censored. The average length of the confidence intervals based on the weighted empirical likelihood is always longer than that based on the conditional likelihood except in one case when the sample size is small.

Table 6.2: Bias, Standard Deviation and Mean Square Error of Estimates $\hat{\theta}_c$, $\hat{\theta}_w$ and $\hat{\theta}_e$

simulation	TRUE	$\hat{\theta}_c$			$\hat{\theta}_w$			$\hat{\theta}_e$		
configuration	θ	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
					$n_1 = n_2 = 50$					
1	0.5	0.006	0.063	0.0041	0.007	0.066	0.0046	0.001	0.065	0.0042
2	0.608	0.009	0.058	0.0034	0.018	0.061	0.004	0.013	0.064	0.0042
3	0.5	0.004	0.08	0.0064	-0.005	0.075	0.0057	-0.01	0.073	0.0054
4	0.702	-0.001	0.061	0.0037	0.031	0.065	0.0052	0.044	0.065	0.0062
5	0.5	0.007	0.069	0.0048	-0.008	0.084	0.0072	-0.025	0.088	0.0084
6	0.824	-0.058	0.078	0.0095	-0.041	0.091	0.01	0.066	0.051	0.0069
					$n_1 = n_2 = 100$					
1	0.5	0.003	0.044	0.002	-0.003	0.047	0.0022	-0.003	0.046	0.0021
2	0.608	-0.004	0.041	0.0017	0.014	0.043	0.002	0.01	0.045	0.0021
3	0.5	-0.002	0.05	0.0025	-0.002	0.054	0.0029	0.004	0.053	0.0028
4	0.702	-0.008	0.045	0.002	0.029	0.047	0.0031	0.029	0.048	0.0038
5	0.5	0.005	0.058	0.0033	-0.005	0.061	0.0037	0.002	0.062	0.0042
6	0.824	-0.068	0.055	0.0076	-0.035	0.067	0.0057	0.059	0.037	0.0049
					$n_1 = n_2 = 150$					
1	0.5	0.002	0.036	0.0013	0.002	0.038	0.0015	0	0.038	0.0014
2	0.608	-0.003	0.032	0.001	0.013	0.034	0.0013	0.01	0.038	0.0015
3	0.5	0.001	0.041	0.0017	0.001	0.046	0.002	-0.004	0.043	0.0019
4	0.702	0.009	0.035	0.0013	0.029	0.039	0.0023	0.03	0.039	0.003
5	0.5	0.003	0.046	0.0022	0.003	0.049	0.0024	0.013	0.052	0.0029
6	0.824	-0.07	0.046	0.007	-0.032	0.059	0.0045	0.058	0.031	0.0043

Table 6.3: Comparison of Estimated Coverage Probabilities of Confidence Intervals for θ

Simulation	$n_1 = n_2 = 50$			$n_1 = n_2 = 100$			$n_1 = n_2 = 150$		
Configuration	$\hat{\theta}_c$	$\hat{\theta}_w$	$\hat{\theta}_e$	$\hat{\theta}_c$	$\hat{\theta}_w$	$\hat{\theta}_e$	$\hat{\theta}_c$	$\hat{\theta}_w$	$\hat{\theta}_e$
Nominal coverage=0.99									
1	0.979	0.972	0.978	0.98	0.976	0.976	0.986	0.98	0.981
2	0.979	0.973	0.972	0.99	0.98	0.98	0.984	0.977	0.974
3	0.976	0.971	0.957	0.979	0.972	0.959	0.985	0.977	0.964
4	0.98	0.97	0.852	0.983	0.975	0.856	0.987	0.973	0.861
5	0.965	0.977	0.926	0.975	0.98	0.928	0.976	0.986	0.929
6	0.955	0.969	0.74	0.913	0.964	0.692	0.859	0.973	0.619
Nominal coverage=0.95									
1	0.926	0.91	0.931	0.928	0.915	0.926	0.933	0.927	0.931
2	0.93	0.91	0.914	0.947	0.926	0.925	0.943	0.919	0.924
3	0.908	0.897	0.872	0.92	0.904	0.884	0.929	0.916	0.887
4	0.929	0.905	0.745	0.939	0.913	0.739	0.938	0.904	0.739
5	0.891	0.911	0.827	0.904	0.928	0.828	0.897	0.926	0.83
6	0.864	0.886	0.61	0.767	0.896	0.532	0.671	0.922	0.451
Nominal coverage=0.9									
1	0.867	0.851	0.879	0.87	0.86	0.868	0.882	0.869	0.872
2	0.885	0.851	0.849	0.895	0.871	0.861	0.891	0.861	0.85
3	0.843	0.824	0.805	0.851	0.828	0.82	0.865	0.843	0.819
4	0.877	0.843	0.661	0.888	0.85	0.655	0.887	0.836	0.655
5	0.809	0.834	0.744	0.82	0.867	0.747	0.825	0.866	0.74
6	0.787	0.817	0.53	0.645	0.831	0.451	0.534	0.872	0.366

Table 6.4: Comparison of Estimated Lengths of Confidence intervals for θ

Simulation	$n_1 = n_2 = 50$			$n_1 = n_2 = 100$			$n_1 = n_2 = 150$		
Configuration	$\hat{\theta}_c$	$\hat{\theta}_w$	$\hat{\theta}_e$	$\hat{\theta}_c$	$\hat{\theta}_w$	$\hat{\theta}_e$	$\hat{\theta}_c$	$\hat{\theta}_w$	$\hat{\theta}_e$
Nominal coverage=0.99									
1	0.299	0.303	0.31	0.211	0.213	0.216	0.173	0.174	0.176
2	0.292	0.297	0.301	0.211	0.214	0.211	0.175	0.176	0.171
3	0.318	0.312	0.319	0.225	0.236	0.223	0.184	0.191	0.181
4	0.31	0.345	0.285	0.231	0.258	0.198	0.193	0.216	0.161
5	0.342	0.386	0.338	0.241	0.273	0.236	0.197	0.224	0.191
6	0.371	0.455	0.248	0.268	0.376	0.179	0.229	0.346	0.146
Nominal coverage=0.95									
1	0.228	0.231	0.293	0.161	0.162	0.204	0.131	0.133	0.166
2	0.222	0.226	0.285	0.161	0.162	0.199	0.133	0.134	0.162
3	0.242	0.252	0.301	0.171	0.178	0.21	0.14	0.145	0.171
4	0.236	0.263	0.269	0.175	0.196	0.187	0.147	0.164	0.152
5	0.26	0.291	0.32	0.183	0.208	0.223	0.15	0.17	0.181
6	0.282	0.346	0.235	0.204	0.286	0.169	0.174	0.263	0.138
Nominal coverage=0.9									
1	0.191	0.194	0.198	0.134	0.136	0.138	0.11	0.111	0.113
2	0.168	0.189	0.192	0.135	0.136	0.134	0.112	0.112	0.109
3	0.203	0.212	0.204	0.144	0.15	0.142	0.117	0.122	0.116
4	0.198	0.22	0.182	0.147	0.165	0.127	0.124	0.138	0.103
5	0.218	0.244	0.216	0.154	0.174	0.151	0.126	0.143	0.122
6	0.237	0.29	0.159	0.171	0.24	0.114	0.146	0.221	0.093

6.4 Application to Data from a Clinical Trial

We applied the proposed methods to a data set from a randomized clinical trial conducted by the National Cancer Institute of Canada (NCIC) Clinical Trial Group. This trial was designed to compare two chemotherapy regimens, Cyclophosphamide-Methotrexate-Fluorouracil (CMF) versus Cyclophosphamide-Epirubicin-Fluorouracil (CEF) in women with early stage breast cancer. 710 pre-menopausal women with the axillary node positive breast cancer were recruited in this trial with a median follow-up of 8.8 years for all patients at the time of analysis. The survival time of interest is the time from randomization to recurrence of breast cancer, which was termed as relapse-free survival in the original trial Levine et al. (2005). Around 49% of patients were censored for this survival time and the hazard ratio of CEF to CMF in the relapse-free survival, obtained from the Cox model, was 0.76 with a 95% confidence interval from 0.62 to 0.94.

We first checked the validity of the density ratio model using the following goodness of fit test statistic proposed by Ren (2008):

$$T = \sqrt{n} \sup_t |\psi(\hat{F}_1^{KM}(t), \hat{F}_2^{KM}(t)) - \psi(F_1, F_2)|,$$

where $\psi(\cdot,\cdot)$ is a functional such that $\psi(\hat{F}_1^{KM}(t), \hat{F}_2^{KM}(t)) = \hat{F}_1(t) - \hat{F}_1^{KM}(t)$ and $\psi(F_1(t), F_2(t)) = 0$. We used the following bootstrap method to obtain the p-value of T: we first generate bootstrapped true survival times $T_{11}^b, T_{12}^b, \cdots, T_{1n_1}^b$ and $T_{21}^b, T_{22}^b, \cdots, T_{2n_2}^b$ respectively from $\hat{F}_{1w}(t)$ and $\hat{F}_{2w}(t)$ for each $b = 1, 2, \cdots, B$ and then bootstrapped true censoring times $C_{j1}^b, C_{j2}^b, \cdots, C_{jn_j}^b$ $(b = 1, 2, \cdots, B,\ j = 1, 2)$ from the Kaplan-Meier estimate for the distribution of the censoring time in the j-th group. The b-th bootstrap sample is thus $S_b^* = \{(X_{ji}^b, \delta_{ji}^b)\ j = 1, 2, i = 1, 2, \cdots, n_j\}$

$(b = 1, 2, \cdots B)$. Based on each bootstrap sample, we can calculate the estimate $\hat{F}^*_{1wb}(t)$ of $F_1(t)$ based on the weighted empirical likelihood and also another estimate $\hat{F}^{KM*}_{1b}(t)$ based on the Kaplan-Meier method. The bootstrap version of the test statistic T based on the b-th bootstrap sample can thus be defined as:

$$T^*_b = \sqrt{n} \sup_t |\hat{F}^*_{1wb}(t) - \hat{F}^{KM*}_{1b}(t) - (\hat{F}_{1w}(t) - \hat{F}^{KM}_1(t))|.$$

The bootstrap estimate for the p-value associated with the test statistic T is defined as the percentage of T^*_b's that are greater than the test statistic T.

Based on the data from the clinical trial, we calculated that $T = 2.174$ and $(\hat{\beta}_{0c}, \hat{\beta}_{1c}, \hat{r}_c) = (0.217, -0.382, -0.64)$. With 100 bootstrap replications, the bootstrap estimate of the p-value associated with T was 0.57, which indicates the density ratio model may provide adequate fit of the data. We may also see this from Figure 6.1, which plots, for the patients in the CMF group, the survival curve estimated from the density ratio model for the CMF group against the Kaplan-Meier survival curve.

Under the density ratio model, the relative treatment effect θ, which is the probability that the relapse-free survival of patients treated by CEF is longer than those treated by CMF, was estimated as 0.526 by the conditional empirical likelihood method and 0.537 by the weighted empirical likelihood method. The 95% confidence interval with 100 bootstrap replications was respectively (0.488, 0.564) and (0.505, 0.569) based on the conditional and weighted empirical likelihood methods. The lower endpoints of both intervals were very close to 0.5 with that of intervals based on the weighted empirical likelihood just above 0.5. This indicates that the weighted empirical likelihood method is slightly liberal than the conditional likelihood method since one may only conclude that the probability that the patients treated by CEF took longer time to recur than those treated by CMF is significantly different from

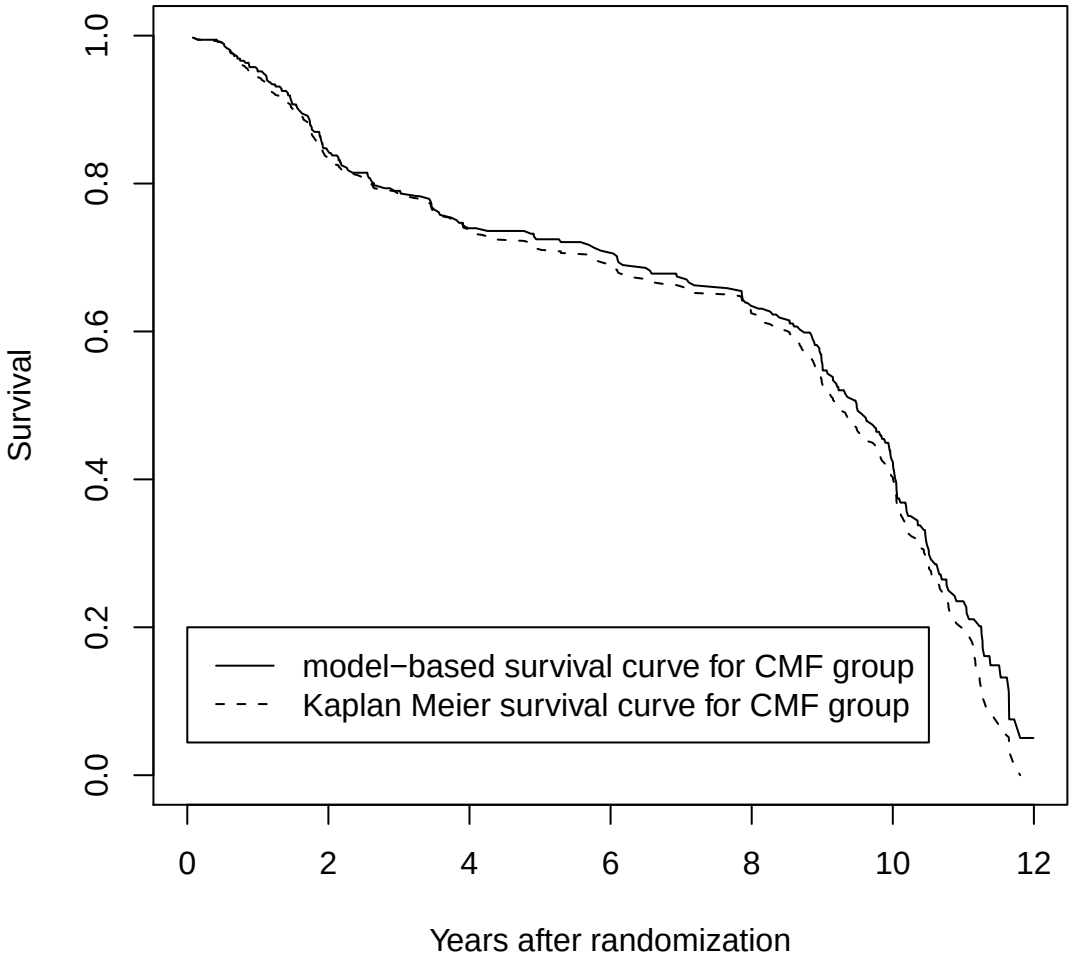
0.5 based on the weighted empirical likelihood confidence interval, which is consistent with the results from the simulation studies.

6.5 Conclusions and Discussions

In this Chapter, inference on an alternative measure of treatment effect in the clinical trial is discussed. We assumed that the density functions in two treatment groups satisfy a semi-parametric density ratio model. We used the Box-Cox transformation to define the density ratio model, which makes this model more flexible in fitting the data from clinical trials. Two methods are proposed to estimate and construct the confidence interval for this measure. Based on the results of simulation, the method based on the conditional likelihood may be used when the sample size is small or the censoring is not very heavy, while the weighted empirical likelihood method may be used when the censoring rate is high. Although the estimate proposed by Efron (1967) had comparable MSE with the estimate we defined, the coverage probability of the confidence interval based on this estimate was found to be below from the nominal level when the censoring is moderate or high.

In clinical trials where several important baseline covariates are available, a multiple Cox model may be used to derive an estimate and associated confidence interval for a hazard ratio adjusted for the effect of covariates, although in most of randomized clinical trials, covariates are usually balanced because of the randomization procedure and, therefore, adjustment may not be necessary (Tu et al., 2000). It remains an interesting topic for further research to take into account of other covariates in the definition and inference of θ when it is used for observational studies or clinical trials with unbalanced covariates. Recently, Wang and Lai (2011) considered the empirical

Figure 6.1: Survival curves for the CMF treatment group based on respectively the Kaplan-Meier method and density ratio model



likelihood inference of the treatment effect as measured by the difference in medians of a response variable, such as the survival time, between two non-randomized treatment groups by calibrating baseline covariate information effectively. The generalization of the procedures proposed in that paper to the situations where the response variable may be censored is also of interest.

Chapter 7

Inference on Treatment-Covariate Interactions in Cancer Clinical Trials

7.1 Introduction

In clinical trials, after a primary analysis including all patients is completed, analysis by subgroups defined based on baseline covariates of patients is often of interest to assess the treatment-covariate interaction, or in other words, the heterogeneity of treatment effects across the subgroups. For example, Muss et al. (2008) assessed the interaction between the letrozole treatment and age of the patients in NCIC CTG MA. 17 trial. Based on the insignificance of the interaction, they suggested both young and old patients should be considered for extended adjuvant therapy with letrozole. On the other hand, Karapetis et al. (2008) studied the interaction between the cetuximab treatment and K-ras mutation status in the NCIC CTG CO. 17 trial. Because patients

with mutated K-ras did not benefit from cetuximab, whereas patients with wild-type K-ras had a significant benefit and the interaction was highly significant, they recommended that cetuximab should only be given to patients with wild-type K-ras.

When the endpoint of a clinical trial is a time to and event with potential censoring, the traditional approach to assess the treatment-covariate interaction is based on the Cox proportional hazard model and to include the interaction term as one of the covariates. However, it is well known that the key assumption behind the Cox model, the proportional hazard assumption, may not be satisfied by the data from clinical trials. Schemper (1988) suggested a non-parametric test for the treatment-covariate interaction based on a non-parametric definition of the covariate-treatment interaction proposed by Patel and Hoel (1973). This test procedure is based on the C-index, which is the sum of U-scores generalized from the Wilcoxon statistics under a random censoring mechanism. As pointed out by Koziol and Jia (2009), the expected value of the C-index depends on the underlying censoring distributions, and therefore, it may be far removed from the true value. The censorship also adversely affects the precision of the estimator. They suggested an alternative approach based on an estimator proposed by Efron (1967) and showed in a simulation study that it has smaller bias and variance than the approach based on the C-index.

In this Chapter, I will first derive a non-parametric test for the treatment-covariate interaction based on Efron's estimator of probabilistic index. Since it was shown in the last Chapter that the inference procedure based on the Efron's estimator performs well when the censoring rate is low, but behaves badly as the censoring rate increases, therefore, alternative procedures are also derived based on the empirical likelihood methods developed under a semi-parametric density ratio model defined in the last

Chapter. The performance of these procedures is also evaluated through Monte-Carlo simulations.

In practice, it is also often of interest to compare the magnitude of the interactions between a treatment and two different baseline covariates. For example, in a tumor biomarker study based on the data from a cancer clinical trial, researchers often wants to know whether a new biomarker provides a larger discrimination of treatment effects than an existing biomarker as seen in a study based on data from NCIC CTG MA. 5 trial (O'Malley et al., 2009), where the comparison of interactions between the treatment and two biomarkers (TOP2A and HER2) was of interest. This comparison, together with the cost and technical considerations, will help to decide whether a new biomarker is more useful in practice.

Statistical procedures which can be used for this comparison may be difficult to derive under the Cox model framework through the hazard ratio. In this Chapter, I will also derive some statistical procedures for the comparison of two non-parametric measures of the interaction between a treatment and two covariates and evaluate their performances by Monte-Carlo simulations.

The rest of this Chapter is organized as follows. In Section 7.2, I will present the estimates for the non-parametric interactions based on Patel-Hoel's definition. Procedures for testing the hypothesis related to the presence of the interaction is proposed based on the confidence intervals for the interaction parameter. Section 7.3 derives estimates for the difference of interactions as well as their associated confidence intervals. Procedures for the testing of the null hypothesis that no difference between two interactions are also derived. A summary of results from simulations evaluating and comparing the proposed procedures is provided in Section 7.4. Section 7.5 applies

the proposed approaches to a real data set from the clinical trial. Some conclusions and discussions are presented in Section 7.6 and this Chapter is completed with the proof of main results in Section 7.7.

7.2 Non-parametric Assessment of the Treatment-Covariate Interaction

Consider a two-arm clinical trial, where T_1 and T_2 are two independent random variables for the true time to event in the first and second treatment group, with continuous distribution function $F_1(t)$ and $F_2(t)$, respectively. Suppose that subjects in these two groups can be divided into two subgroups based on a baseline covariate A. Denote two levels of the covariate A as respectively 0 and 1. In this context, a non-parametric measure of the interaction between the treatment and the covariate A can be defined based on the definition of Patel and Hoel (1973):

$$\begin{aligned}\mu_A &= \Pr(T_1 \leq T_2|A=1) - \Pr(T_1 \leq T_2|A=0) \\ &= \int F_1(x|A=1)dF_2(x|A=1) - \int F_1(x|A=0)dF_2(x|A=0),\end{aligned}$$

where $F_j(x|A=a)$ is the distribution function of T_j given $A=a$, $a=0,1,\ j=1,2$. It is easy to see that $\Pr(T_1 \leq T_2|A=a)$ is a measure of the treatment effect in patient subgroup with $A=a$, and therefore, μ_A reflects the difference in the treatment effects between two subgroups defined by covariate A.

Let C_j denote the censoring time in group j with a continuous distribution function $G_j(t)$, $j=1,2$. Due to censoring, T_j can not be observed completely. Instead, what

we can observe is (X_j, δ_j), where

$$\begin{cases} X_j &= \min(T_j, C_j) \\ \delta_j &= I(X_j = T_j) \end{cases} \quad j = 1, 2.$$

Let $(X_{ji}, \delta_{ji}), i = 1, 2, \cdots, n_j$ be independent replications of (X_j, δ_j), which represents the observation from patient in the j-th group. In the following, I will derive statistical inference procedure for μ_A based on the data $\{(X_{ji},\ \delta_{ji}),\ j = 1, 2,\ i = 1, 2, \cdots, n_j\}$.

7.2.1 Inference procedure based on Efron's estimator

Based on the discussion in the last Chapter and data $\{(X_{ji},\ \delta_{ji}),\ j = 1, 2,\ i = 1, 2, \cdots, n_j\}$, a straightforward estimate for μ_A could be:

$$\hat{\mu}_{Ae} = \int \hat{F}_1^{KM}(x|A=1) d\hat{F}_2^{KM}(x|A=1) - \int \hat{F}_1^{KM}(x|A=0) d\hat{F}_2^{KM}(x|A=0), \quad (7.1)$$

where $\hat{F}_j^{KM}(x|A = a)$ is the Kaplan-Meier estimate (Kaplan and Meier, 1958) of $F_j(x|A = a)$ for $j = 1, 2,\ a = 0, 1$. It can be shown that

$$\frac{\sqrt{n}(\hat{\mu}_{Ae} - \mu_A)}{\sqrt{\sigma_{Ae}^2}} \rightarrow_d N(0, 1), \quad (7.2)$$

where the definition of σ_{Ae}^2 and the detailed proof of (7.2) are shown in Section 7.7. However, as the analytic variance σ_{Ae}^2 is difficult to estimate, in practice, we suggest to estimate σ_{Ae}^2 by the jackknife method (Shao and Tu, 1995) and construct the confidence interval for μ_A as:

$$I_{Ae} = \left(\hat{\mu}_{Ae} - \frac{z_{1-\alpha/2}\hat{\sigma}_{Ae}}{\sqrt{n}}, \hat{\mu}_{Ae} + \frac{z_{1-\alpha/2}\hat{\sigma}_{Ae}}{\sqrt{n}} \right),$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of a standard normal distribution and $\hat{\sigma}_{Ae}^2$ is the jackknife estimate of σ_{Ae}^2 defined as

$$\hat{\sigma}_{Ae}^2 = \sum_{k=1}^{n} ((\hat{\mu}_{Ae}^{-k} - \sum_{k=1}^{n} \hat{\mu}_{Ae}^{-k}/n)^2)(n-1)/n,$$

with $\hat{\mu}_{Ae}^{-k}$ the estimate of μ_A based on the sample with the k-th observation left out and equation (7.1).

If we are interested in testing the existence of the interaction between the treatment and covariate A, or mathematically, testing the following hypothesis:

$$\text{Test 1}: \ H_{10}: \ \mu_A = 0 \qquad \text{v.s.} \qquad H_{11}: \ \mu_A \neq 0,$$

we can reject H_{10} at α level if zero does not lie in I_{Ae}. The p-value of this test procedure can be calculated as $2\left[1-\Phi\left(\left|\frac{\sqrt{n}\hat{\mu}_{Ae}}{\hat{\sigma}_{Ae}}\right|\right)\right]$, where Φ is the distribution function of a standard normal variable.

7.2.2 Inference Procedures based on Density Ratio Model and Empirical Likelihood

Since it was shown in the last Chapter that procedures based on the Efron's estimator did not perform well when the censoring rate is high, we can also use both conditional and weighted empirical likelihood methods to derive inference procedures for μ_A under the following flexible semi-parametric density ratio model,

$$\begin{aligned} T_1|A=a &\sim f_1(x|A=a), \\ T_2|A=a &\sim f_2(x|A=a) = \exp(\beta_{0a}+\beta_{1a}h(x,r))f_1(x|A=a), \end{aligned} \tag{7.3}$$

where $a = 0, 1$, β_{0a} and β_{1a} are unknown parameters, $f_1(x|A=a)$ is an unknown density function given $A=a$, and $h(x,r)$ is a Box-Cox transformation (Box and Cox, 1964) of x defined as:

$$h(x,r) = \begin{cases} \dfrac{x^r-1}{r} & \text{when } r \neq 0 \\ \log x & \text{when } r = 0 \end{cases}, \tag{7.4}$$

with an unknown parameter r. Specifically, if we denote the conditional and weighted empirical likelihood estimates of $F_j(x|A = a)$ as respectively $\hat{F}_{jc}(x|A = a)$ and $\hat{F}_{jw}(x|A = a)$, $j = 1, 2,\ a = 0, 1$, then the conditional and weighted empirical likelihood estimates of μ_A can be denoted as $\hat{\mu}_{Ac}$ and $\hat{\mu}_{Aw}$, respectively, where

$$\hat{\mu}_{Au} = \int \hat{F}_{1u}(x|A = 1)d\hat{F}_{2u}(x|A = 1) - \int \hat{F}_{1u}(x|A = 0)d\hat{F}_{2u}(x|A = 0),\ u = c, w,$$

and the confidence intervals I_{Au} of μ_A based on $\hat{\mu}_{Au}$, $u = c, w$, can be defined as:

$$I_{Au} = \left(\hat{\mu}_{Au} - \frac{z_{1-\alpha/2}\hat{\sigma}_{Au}}{\sqrt{n}}, \hat{\mu}_{Au} + \frac{z_{1-\alpha/2}\hat{\sigma}_{Au}}{\sqrt{n}}\right),\ u = c, w,$$

where $\hat{\sigma}_{Au}$ is the bootstrap variance estimate obtained by following the same procedure outlined in Chapter 6. Therefore, based on these estimates, H_{10} can be rejected at α level if the above confidence interval does not cover zero and, similarly, p-value of the test is computed as $2\left[1 - \Phi\left(\left|\frac{\sqrt{n}\hat{\mu}_{Ae}}{\hat{\sigma}_{Ae}}\right|\right)\right]$, for $u = c, w$.

7.3 Comparison of Treatment-Covariate Interactions

Suppose there are two covariates (A and B) and we are interested in the comparison of interactions between the treatment and these two covariates based on the following non-parametric measurement of the interaction between the treatment and a covariate K defined previously:

$$\begin{aligned}\mu_K &= \Pr(T_1 \leq T_2|K = 1) - \Pr(T_1 \leq T_2|K = 0) \\ &= \int F_1(x|K = 1)dF_2(x|K = 1) - \int F_1(x|K = 0)dF_2(x|K = 0),\end{aligned}$$

where $K = A$ or B. Let $D = \mu_A - \mu_B$. In the following, I will present the estimation and confidence interval procedures for D and also the test procedure for the

hypothesis:

$$\text{Test 2}: \ H_{20}: \ D := \mu_A - \mu_B = 0 \qquad \text{v.s.} \qquad H_{21}: \ D \neq 0.$$

7.3.1 Procedures Based on Efron's Estimator

Since $F_j(x|A=a)$ can be considered as a mixture distribution:

$$F_j(x|A=a) = \sum_{b=0}^{1} p_{ab} F_j(x|A=a, B=b),$$

where $p_{ab} = \Pr(B=b|A=a)$, a more efficient estimate for $F_j(x|A=a)$ when we have the information about two baseline covariates would be

$$\tilde{F}_{je}(x|A=a) = \hat{p}_{a1}\hat{F}_j^{KM}(x|A=a, B=1) + \hat{p}_{a0}\hat{F}_j^{KM}(x|A=a, B=0), \tag{7.5}$$

where $\hat{F}_j^{KM}(x|A=a, B=b)$ is the Kaplan-Meier estimate of $F_j(x|A=a, B=b)$ and

$$\hat{p}_{ab} = \frac{\sum_{j=1}^{2}\sum_{i=1}^{n_j} I(A_{ji}=a, B_{ji}=b)}{\sum_{j=1}^{2}\sum_{i=1}^{n_j} I(A_{ji}=a)}. \tag{7.6}$$

Similarly, $F_j(x|B=b)$ can be considered as a mixture distribution:

$$F_j(x|B=b) = \sum_{a=0}^{1} q_{ab} F_j(x|A=a, B=b),$$

with $q_{ab} = \Pr(B=b|A=a)$, thus can be estimated by

$$\tilde{F}_{je}(x|B=b) = \hat{q}_{1b}\hat{F}_j^{KM}(x|A=1, B=b) + \hat{q}_{0b}\hat{F}_j^{KM}(x|A=0, B=b), \tag{7.7}$$

where $\hat{F}_j^{KM}(x|A=a, B=b)$ is the Kaplan-Meier estimate of $F_j(x|A=a, B=b)$ and

$$\hat{q}_{ab} = \frac{\sum_{j=1}^{2}\sum_{i=1}^{n_j} I(A_{ji}=a, B_{ji}=b)}{\sum_{j=1}^{2}\sum_{i=1}^{n_j} I(B_{ji}=b)}. \tag{7.8}$$

Therefore D can be estimated by

$$\begin{aligned}\tilde{D}_e &= \tilde{\mu}_{Ae} - \tilde{\mu}_{Be} \\ &= \left(\int \tilde{F}_{1e}(x|A=1)d\tilde{F}_{2e}(x|A=1) - \int \tilde{F}_{1e}(x|A=0)d\tilde{F}_{2e}(x|A=0)\right) - \\ &\quad \left(\int \tilde{F}_{1e}(x|B=1)d\tilde{F}_{2e}(x|B=1) - \int \tilde{F}_{1e}(x|B=0)d\tilde{F}_{2e}(x|B=0)\right).\end{aligned}$$

We can also show in Section 7.7 that

$$\frac{\sqrt{n}(\tilde{D}_e - D)}{\sqrt{\sigma^2}} \to_d N(0,1), \tag{7.9}$$

where σ^2 is defined in the proof. In practice, the confidence interval for D can be constructed as:

$$I_e = \left(\tilde{D}_e - \frac{z_{1-\alpha/2}\hat{\sigma}}{\sqrt{n}}, \tilde{D}_e + \frac{z_{1-\alpha/2}\hat{\sigma}}{\sqrt{n}}\right),$$

where $\hat{\sigma}^2$ is the jackknife estimate for σ^2. Based on this interval, we may reject H_{20} at α level if zero is not included in I_e.

7.3.2 Procedures Based on Empirical Likelihood and Density Ratio Model

If we assume the following semi-parametric density ratio model:

$$\begin{aligned}T_1|A=a\ B=b &\sim f_1(x|A=a,B=b), \\ T_2|A=a\ B=b &\sim f_2(x|A=a,B=b) \\ &= \exp(\beta_{0ab} + \beta_{1ab}h(x,r))f_1(x|A=a,B=b),\end{aligned} \tag{7.10}$$

where $a=0,1,\ b=0,1$, β_{0ab} and β_{1ab} are unknown parameters, $f_1(x|A=a,B=b)$ is an unknown density function given $A=a,B=b$, and $h(x,r)$ is defined in (7.4), then similarly, according to formula (6.7), (6.8), (6.9) and (6.10) in Chapter 6, we can get respectively the conditional and weighted empirical likelihood estimates $\hat{F}_{jc}(x|A=$

$a, B = b)$ and $\hat{F}_{jw}(x|A = a, B = b)$ for $F_j(x|A = a, B = b)$. Therefore, estimates $\tilde{F}_{jc}(x|A = a)$ and $\tilde{F}_{jw}(x|B = b)$ similar to (7.5) and (7.7) can be derived, leading to the following two semi-parametric estimates of D:

$$\begin{aligned}\tilde{D}_u &= \tilde{\mu}_{Au} - \tilde{\mu}_{Bu} \\ &= \int \tilde{F}_{1u}(x|A=1)d\tilde{F}_{2u}(x|A=1) - \int \tilde{F}_{1u}(x|A=0)d\tilde{F}_{2u}(x|A=0) - \\ &\quad \int \tilde{F}_{1u}(x|B=1)d\tilde{F}_{2u}(x|B=1) - \int \tilde{F}_{1u}(x|B=0)d\tilde{F}_{2u}(x|B=0),\end{aligned}$$

where $u = c$ or w, denoting the conditional likelihood estimate and weighted empirical likelihood estimate, respectively. Again, the bootstrap method can be applied to derive the confidence intervals I_c and I_w based on the semi-parametric density ratio model and test procedures for H_{20} can be defined based on these confidence intervals.

7.4 Simulation Studies

Simulation studies were conducted to evaluate the performance of the procedures proposed above. Table 7.1 summarizes different simulation configurations with different true survival times in the subgroups, which lead to different true interaction parameters as shown in Tables 7.2 and 7.3. The first 4 configurations only involve a single covariate A, while configurations 5 to 9 considered two covariates A and B. For example, in configuration 1, the true survival time in the first treatment group with $A = 1$ is assumed to be log-normal distributed with a scale parameter 2.6 and shape parameter 1, while in the second treatment group with $A = 1$ to be log-normal distributed with a scale parameter 3 and shape parameter 1. The censoring time was chosen to be a uniform variable over 60 to 66 time units and assumed to be the same for all the simulations. This censoring time produces various censoring rates as seen

in Tables 7.2 and 7.3. In all the following simulations, the number of replications was set at 3000.

Table 7.2, in which different simulation numbers indicates different scenarios listed in the first part of Table 7.1, presents a comparison on the performance of three confidence intervals I_{Ae}, I_{Ac} and I_{Aw} given that the assumption (7.3) is true and when there is only one covariate involved. The sample size is fixed at 200, with equal number in each subgroup. From Table 7.2, we can see that when the censoring rate is low, I_{Ae} has the highest coverage probability; while when the censoring rate is high, the confidence intervals based on the empirical likelihood and density ratio model performed better, with the weighted empirical likelihood confidence interval I_{Aw} the best. Besides, all these three methods yield comparable lengths, though in almost all the cases, I_{Ac} has the shortest length, while I_{Aw} has the longest.

Next, we compared the performance of I_e, I_c and I_w, where there are two covariates A and B. The results are shown in Table 7.3. The simulation number in Table 7.3 corresponds with different scenarios listed in the second part of Table 7.1. The sample size was fixed at 320, with equal number in each subgroup. Conclusions similar to Table 7.2 can be drawn: the semi-parametric confidence intervals I_c and I_w perform better than the non-parametric candidate I_e when the censoring rate is high.

7.5 Application to Data Set from a Clinical Trial

We also applied the proposed methods to a data set from a randomized clinical trial conducted by NCIC Clinical Trial Group. This trial was designed to compare two chemotherapy regimens, Cyclophosphamide-Methotrexate-Fluorouracil (CMF) versus Cyclophosphamide-Epirubicin-Fluorouracil (CEF), in women with early stage

Table 7.1: Simulation Numbers and Distribution Assumptions for Tables 7.2 and 7.3

Simulation	A=1		A=0	
Number	j=1	j=2	j=1	j=2
1	LN(2.6,1)	LN(3,1)	LN(3,1)	LN(2.6,1)
2	LN(3,1)	LN(3.3,1)	LN(3.3,1)	LN(3,1)
3	LN(3.3,1)	LN(3.6,1)	LN(3.6,1)	LN(3.3,1)
4	LN(3.55,1)	LN(3.9,1)	LN(3.9,1)	LN(3.55,1)

Simulation	A=1 B=1		A=1 B=0		A=0 B=1		A=0 B=0	
Number	j=1	j=2	j=1	j=2	j=1	j=2	j=1	j=2
5	LN(2.5,1)	LN(2.8,1)	LN(2.5,1)	LN(2.8,1)	LN(2.5,1)	LN(2.5,1)	LN(2.8,1)	LN(2.8,1)
6	LN(2.8,1)	LN(3.2,1)	LN(2.8,1)	LN(3.2,1)	LN(2.8,1)	LN(2.8,1)	LN(3.2,1)	LN(3.2,1)
7	LN(3.2,1)	LN(3.4,1)	LN(3.2,1)	LN(3.4,1)	LN(3.2,1)	LN(3.2,1)	LN(3.4,1)	LN(3.4,1)
8	LN(3.4,1)	LN(3.6,1)	LN(3.4,1)	LN(3.6,1)	LN(3.4,1)	LN(3.4,1)	LN(3.6,1)	LN(3.6,1)
9	LN(3.6,1)	LN(3.8,1)	LN(3.6,1)	LN(3.8,1)	LN(3.6,1)	LN(3.6,1)	LN(3.8,1)	LN(3.8,1)

Table 7.2: Comparison of Estimated Coverage Probabilities and Interval Lengths among I_{Ae} I_{Ac} and I_{Aw}

TRUE level	Simulation Number	TRUE μ	Censoring Rate	Coverage Probability			Interval Length		
				I_{Ae}	I_{Ac}	I_{Aw}	I_{Ae}	I_{Ac}	I_{Aw}
0.99	1	0.223	25%	0.983	0.977	0.973	0.517	0.419	0.494
	2	0.168	35%	0.976	0.973	0.970	0.624	0.457	0.614
	3	0.168	45%	0.943	0.961	0.962	0.759	0.499	0.890
	4	0.195	55%	0.854	0.896	0.969	0.916	0.561	0.993
0.95	1	0.223	25%	0.940	0.933	0.917	0.393	0.338	0.376
	2	0.168	35%	0.935	0.916	0.910	0.474	0.371	0.467
	3	0.168	45%	0.899	0.917	0.916	0.577	0.406	0.686
	4	0.195	55%	0.743	0.825	0.902	0.697	0.451	0.801
0.90	1	0.223	25%	0.888	0.872	0.867	0.330	0.291	0.315
	2	0.168	35%	0.873	0.872	0.855	0.398	0.314	0.390
	3	0.168	45%	0.802	0.857	0.843	0.485	0.344	0.576
	4	0.195	55%	0.715	0.753	0.847	0.584	0.381	0.788

Table 7.3: Comparison of Estimated Coverage Probabilities and Interval Lengths among I_e I_c and I_w

TRUE level	Simulation Number	TRUE Difference	Censoring Rate	Coverage Probability			Interval Length		
				I_e	I_c	I_w	I_e	I_c	I_w
0.99	5	0.168	15%	0.986	0.996	0.981	0.773	0.523	0.766
	6	0.223	26%	0.980	0.987	0.973	0.867	0.638	0.792
	7	0.112	36%	0.971	0.981	0.967	0.878	0.692	0.873
	8	0.112	46%	0.941	0.964	0.971	0.891	0.714	1.014
	9	0.112	58%	0.863	0.877	0.973	0.969	0.780	1.255
0.95	5	0.168	15%	0.955	0.959	0.941	0.580	0.429	0.434
	6	0.223	26%	0.942	0.952	0.935	0.654	0.489	0.531
	7	0.112	36%	0.929	0.943	0.931	0.668	0.567	0.677
	8	0.112	46%	0.893	0.914	0.927	0.678	0.602	0.816
	9	0.112	58%	0.817	0.840	0.913	0.741	0.682	0.971
0.90	5	0.168	15%	0.903	0.911	0.882	0.487	0.370	0.385
	6	0.223	26%	0.887	0.897	0.875	0.549	0.414	0.490
	7	0.112	36%	0.879	0.874	0.859	0.561	0.453	0.583
	8	0.112	46%	0.824	0.858	0.843	0.584	0.498	0.636
	9	0.112	58%	0.742	0.789	0.844	0.628	0.521	0.743

breast cancer. Covariates of interest are the expression levels of two genes (HER2 and TOP2A). 494 patients with expression data for both genes were included in the analysis with average censoring rate of 61.3% on overall survival.

Results based on univariate Cox models reveal that the hazard ratio (CEF over CMF) in HER2 over-expressed group was 0.654 with a 95% confidence interval 0.396 to 1.078, while in HER2 normal group was 1.184 with a 95% confidence interval 0.838 to 1.672 and a p-value of 0.0554 from the treatment-HER2 interaction test. For gene TOP2A, the hazard ratio was 0.423 with a 95% confidence interval from 0.218 to 0.821 in patients with TOP2A alteration, and 1.245 with a 95% confidence interval from 0.880 to 1.761 in TOP2A normal patients. Interaction test between the treatment and gene TOP2A, on the other hand, was highly significant with a p-value 0.0047 from the analysis based on the Cox model.

Figures 7.1-7.4 present the Kaplan-Meier plots for different subgroups. From Figures 7.2 and 7.4 we can see that the proportional hazard assumption may not hold in those subgroups since survival curves for the CEF and CMF groups crossed.

Tables 7.4 and 7.5 summarized the estimates as well as confidence intervals of μ_A and D with p-values for test 1 and test 2 based on respectively the Efron's estimator (Efron), conditional empirical likelihood (CEL) and weighted empirical likelihood (WEL) procedures. Significant interaction between the treatment and HER2 at 0.05 level was only found by the conditional empirical likelihood method, while all except weighted empirical likelihood method showed a significant interaction between the treatment and TOP2A. None of the method showed a significant difference in the interaction between the treatment and these two genes, which implies these two genes may be equally effective in discriminating treatment effects. But sample size may be

Figure 7.1: Kaplan-Meier Curves in HER2 Over-expressed Patients

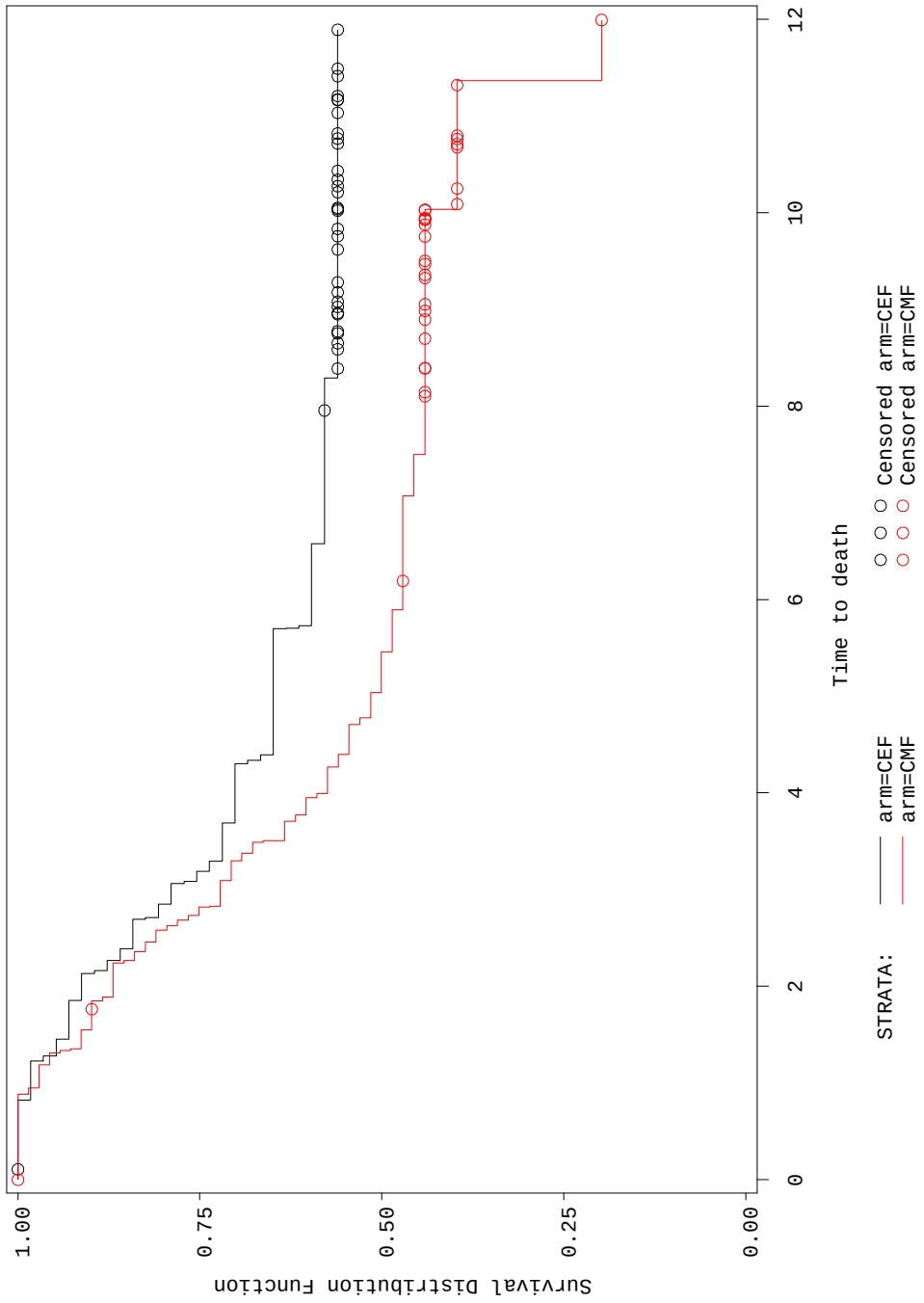


Figure 7.2: Kaplan-Meier Curves in HER2 Normal Patients

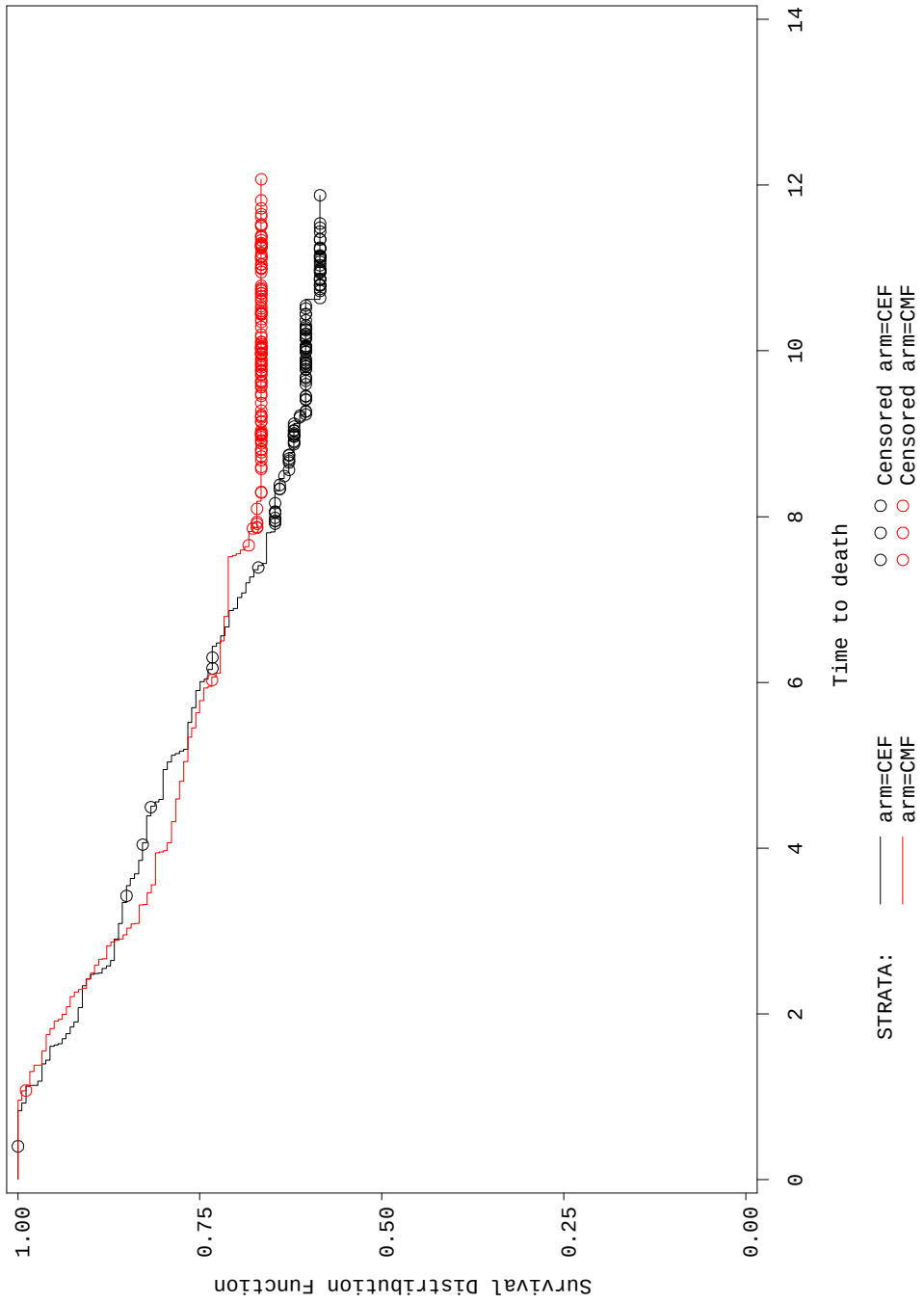


Figure 7.3: Kaplan-Meier Curves in TOP2A Alteration Patients

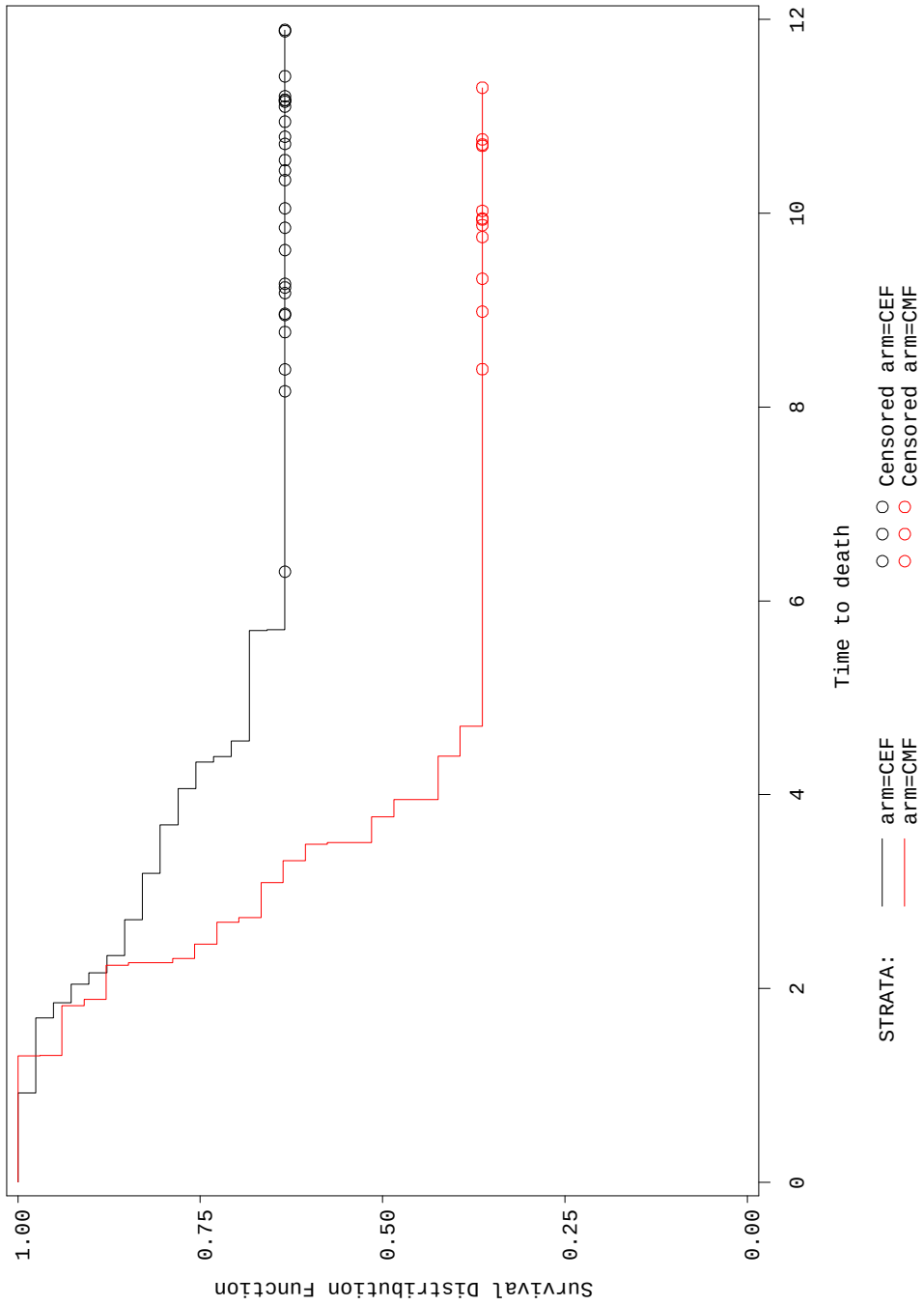


Figure 7.4: Kaplan-Meier Curves in TOP2A Normal Patients

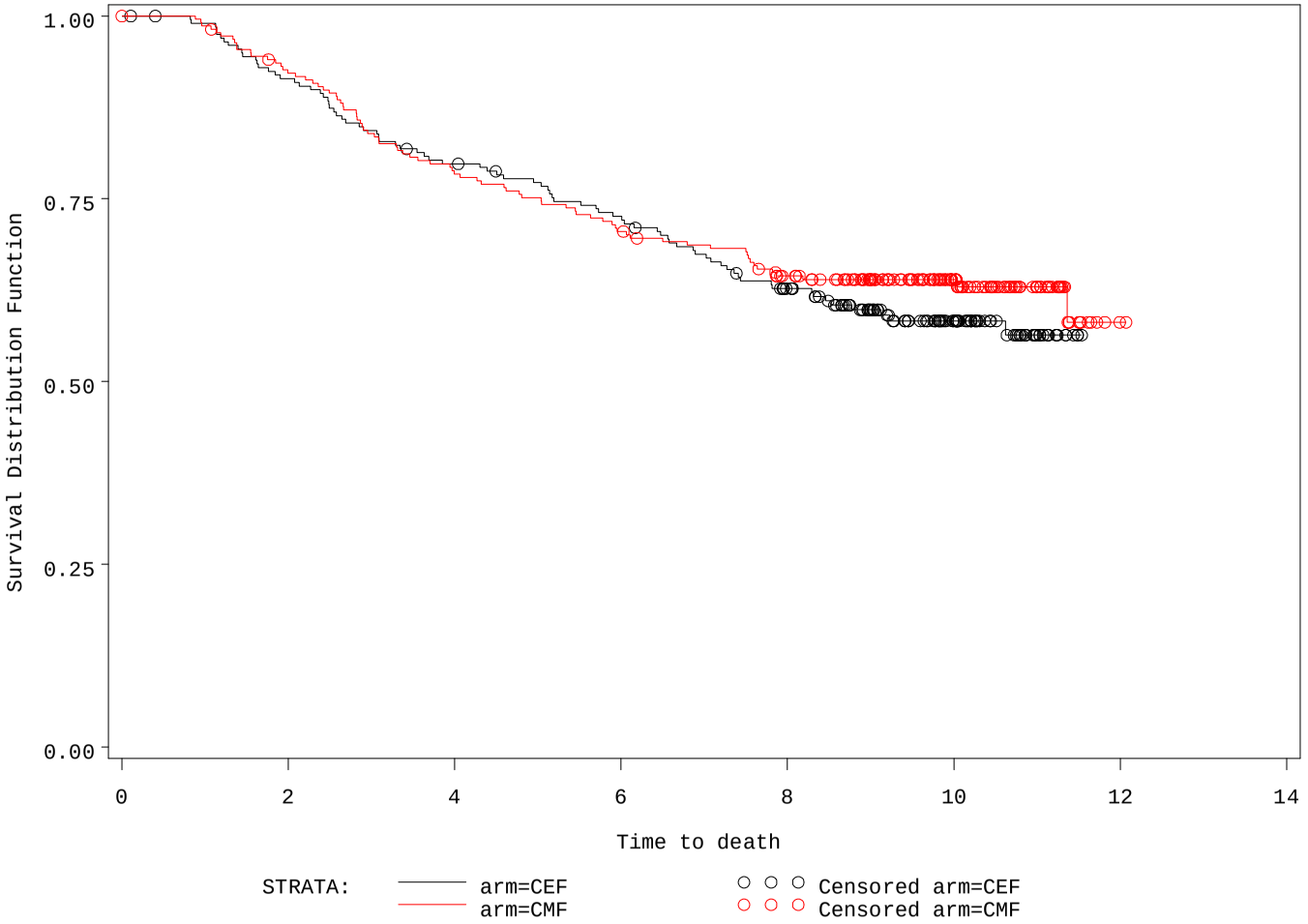


Table 7.4: Comparison of Estimation and 95% Confidence Interval among Three Different Approaches

	Treatment-HER2 interaction	Treatment-TOP2A interaction	Difference of interaction
Efron	0.161 (-0.058,0.380)	0.262 (0.036,0.487)	0.149 (-0.159,0.457)
CEL	0.193 (0.007,0.379)	0.313 (0.127,0.499)	0.179 (-0.050,0.408)
WEL	0.175 (-0.105,0.455)	0.274 (-0.045,0.593)	0.144 (-0.185,0.473)

Table 7.5: Comparison of P-values among Four Different Approaches

	Test 1		Test 2
	Treatment-HER2 interaction	Treatment-TOP2A interaction	Difference of interaction
Cox model	0.0554	0.0047	NA
Efron	0.15	0.023	0.342
CEL	0.041	0.001	0.126
WEL	0.222	0.093	0.391

too small to reach a definite conclusion on this. Meta-analysis with large sample size may be required to draw a definite conclusion.

We can also see from Tables 7.4 and 7.5 that the conditional empirical likelihood based approach produces most liberal results, while the weighted empirical likelihood method always yields most conservative results and Efron method in between, which confirms the results from the simulation study since the censoring rate was quite high in this study.

7.6 Conclusions and Discussions

In this Chapter, a non-parametric estimation procedure based on the Eforn's estimator and associated confidence intervals are derived to quantify the interaction between the treatment and a binary covariate with potentially censored data observed. Under a semi-parametric density ratio model, the conditional and weighted empirical likelihood estimates for this interaction are also obtained. In addition, procedures for the comparison of interactions between the treatment and two covariates are also derived based on both the Efron's estimator and the semi-parametric density ratio model. The proposed approaches are evaluated and compared by Monte-Carlo simulations and applied to a real data set from a cancer clinical trial.

In clinical trials, information on the covariates may be missing for some patients. For the procedures presented in this Chapter, only patients with complete data on the covariate of interest can be included in the analysis. This would lead to a data set which may not balance treatment groups for some other important covariates. The development of procedures by imputing missing covariates or adjusting for potential imbalance through multiple regression or density ratio models is an interesting topic for further investigations.

The procedure developed in this Chapter can be easily generalized to situations where there are multiple treatment groups or baseline covariates and also where the baseline covariates of interest have more than two categories. But it can not be applied to settings where the covariates were measured on a continuous scale. Development of non-parametric procedures to test the interaction between the treatment and a continuous covariate is also of interest. If dichotomization of a continuous covariate

may be desired, the non-parametric measure of interaction presented in this Chapter may be used to find the best cut-off following a similar approach of Contal and O'Quigley (1999) based on the log-rank test.

7.7 Proof of (7.2) and (7.9)

Lemma 7.1. *Let $\hat{G}^{KM}_{jab}$ be the Kaplan-Meier estimate for G_{jab}, which is the censoring distribution in the j-th treatment group given $A = a$ and $B = b$, $j = 1, 2, a = 0, 1, b = 0, 1$. For any distribution function ϕ, denote $\bar{\phi} = 1 - \phi$, $c_\phi = \inf\{x : \ \phi(x) > 0\}$, and $d_\phi = \sup\{x : \ \phi(x) < 1\}$. Assume τ^1_{jab}, τ^2_{jab} are two numbers such that*

$$c_{F_{jab}} < \tau^1_{jab} < \tau^2_{jab} < d_{H_{jab}},$$

where $F_{jab}(t) = F_j(t|A = a, B = b)$ and $\bar{H}_{jab}(t) = \bar{F}_{jab}(t)\bar{G}_{jab}(t)$ for $j = 1, 2, \ a = 0, 1, \ b = 0, 1$. Suppose t is restricted in the interval $[\tau^1_{jab}, \tau^2_{jab}]$, then we have that $\hat{F}^{KM}_{jab}$, the Kaplan-Meier estimate of F_{jab}, is a consistent estimator of F_{jab}, as $n \to \infty$. Besides, assume n_{jab} the number of patients in j-th treatment group with $A = a$ and $B = b$, satisfies that $n_{jab}/n \to \rho_{jab} > 0$, as $n \to \infty$. Then $\sqrt{n}(\hat{F}^{KM}_{jab}(t) - F_{jab}(t)), j = 1, 2, \ a = 0, 1, \ b = 0, 1$, are asymptotically independent centered Gaussian processes with the covariance kernel:

$$\Gamma_{jab}(s, t) = \frac{\bar{F}_{jab}(s)\bar{F}_{jab}(t)}{\rho_{jab}} \int_{-\infty}^{s \wedge t} \frac{dF_{jab}(z)}{\bar{F}_{jab}(z)\bar{H}_{jab}(z)}.$$

Proof. Directly from (7.22) in Efron (1967). □

Proof of (7.2):

Proof. In the proof of (7.2), since only one covariate is considered, we can suppress subscript b and let $F_{ja}(t) = F_j(t|A = a)$ and $H_{ja}(t) = 1 - \bar{F}_{ja}(t)\bar{G}_{ja}(t)$. It is trivial

to show that $\hat{\mu}_{Ae}$ is an asymptotic unbias estimate for μ_A. In the following, we prove that $\hat{\mu}_{Ae}$ is asymptotically normally distributed.

It is easy to see

$$\begin{aligned}\sqrt{n}(\hat{\mu}_{Ae}-\mu_A) &= \left(\sqrt{n}\int \hat{F}_{11}^{KM}(x)d\hat{F}_{21}^{KM}(x)-\sqrt{n}\int F_{11}(x)dF_{21}(x)\right)\\ &\quad -\left(\sqrt{n}\int \hat{F}_{10}^{KM}(x)d\hat{F}_{20}^{KM}(x)-\sqrt{n}\int F_{10}(x)dF_{20}(x)\right) \qquad (7.11)\\ &:= \hat{I}_1-\hat{I}_2,\end{aligned}$$

and

$$\begin{aligned}\hat{I}_1 &= \sqrt{n}\int \hat{F}_{11}^{KM}(x)d\hat{F}_{21}^{KM}(x)-\sqrt{n}\int F_{11}(x)dF_{21}(x)\\ &= \sqrt{n}\left(\int(\hat{F}_{11}^{KM}(x)-F_{11}(x))dF_{21}(x)\right)\\ &\quad -\sqrt{n}\left(\int(\hat{F}_{21}^{KM}(x)-F_{21}(x))dF_{11}(x)\right) \qquad (7.12)\\ &\quad +\sqrt{n}\int\left[\hat{F}_{11}^{KM}(x)-F_{11}(x)\right]d\left[\hat{F}_{21}^{KM}(x)-F_{21}(x)\right].\end{aligned}$$

By Lemma 7.1, the last term in (7.12) is negligible. Besides, denoting

$$\begin{aligned}\hat{I}_{11} &= \sqrt{n}\left(\int(\hat{F}_{11}^{KM}(x)-F_{11}(x))dF_{21}(x)\right),\\ \hat{I}_{21} &= \sqrt{n}\left(\int(\hat{F}_{21}^{KM}(x)-F_{21}(x))dF_{11}(x)\right), \qquad (7.13)\end{aligned}$$

and by Lemma 7.1 again, we have $\hat{I}_{11}$ and $\hat{I}_{21}$ are asymptotically independent normal

random variables with variance σ_{11}^2 and σ_{21}^2, respectively, where

$$\begin{aligned}
\sigma_{11}^2 &= \frac{1}{\rho_{11}} \int\int \Gamma_{11}(s,t) dF_{21}(s) dF_{21}(t) \\
&= \frac{2}{\rho_{11}} \int\int_s^\infty \int_{-\infty}^s \frac{\bar{F}_{11}(s)\bar{F}_{11}(t)}{\bar{F}_{11}(z)\bar{H}_{11}(z)} dF_{11}(z) dF_{21}(t) dF_{21}(s), \\
\sigma_{21}^2 &= \frac{1}{\rho_{21}} \int\int \Gamma_{21}(s,t) dF_{11}(s) dF_{11}(t) \\
&= \frac{2}{\rho_{21}} \int\int_s^\infty \int_{-\infty}^s \frac{\bar{F}_{21}(s)\bar{F}_{21}(t)}{\bar{F}_{21}(z)\bar{H}_{21}(z)} dF_{21}(z) dF_{11}(t) dF_{11}(s).
\end{aligned} \tag{7.14}$$

Thus, (7.12)-(7.14) together yield,

$$\hat{I}_1 \to N(0, \sigma_{11}^2 + \sigma_{21}^2) \tag{7.15}$$

Similarly, if we denote

$$\begin{aligned}
\sigma_{10}^2 &= \frac{1}{\rho_{10}} \int\int \Gamma_{10}(s,t) dF_{20}(s) dF_{20}(t) \\
&= \frac{2}{\rho_{10}} \int\int_s^\infty \int_{-\infty}^s \frac{\bar{F}_{10}(s)\bar{F}_{10}(t)}{\bar{F}_{10}(z)\bar{H}_{10}(z)} dF_{10}(z) dF_{20}(t) dF_{20}(s) \\
\sigma_{20}^2 &= \frac{1}{\rho_{20}} \int\int \Gamma_{20}(s,t) dF_{10}(s) dF_{10}(t) \\
&= \frac{2}{\rho_{20}} \int\int_s^\infty \int_{-\infty}^s \frac{\bar{F}_{20}(s)\bar{F}_{20}(t)}{\bar{F}_{20}(z)\bar{H}_{20}(z)} dF_{20}(z) dF_{10}(t) dF_{10}(s)
\end{aligned} \tag{7.16}$$

then it is easily seen that

$$\hat{I}_2 \to N(0, \sigma_{10}^2 + \sigma_{20}^2). \tag{7.17}$$

Therefore, from (7.11), (7.16) and (7.17), we have

$$\sqrt{n}(\hat{\mu}_{Ae} - \mu_A) \to N(0, \sigma_A^2),$$

where $\sigma_A^2 = \sigma_{11}^2 + \sigma_{21}^2 + \sigma_{10}^2 + \sigma_{20}^2$. □

Proof of (7.9):

Proof. Since

$$\begin{aligned}
\tilde{\mu}_{Ae} &= \int \tilde{F}_{1e}(x|A=1)d\tilde{F}_{2e}(x|A=1) - \int \tilde{F}_{1e}(x|A=0)d\tilde{F}_{2e}(x|A=0) \\
&\approx \int \left(\sum_b p_{1b}\hat{F}^{KM}_{11b}(t)\right) d\left(\sum_b p_{1b}\hat{F}^{KM}_{21b}(t)\right) \\
&\quad - \int \left(\sum_b p_{0b}\hat{F}^{KM}_{10b}(t)\right) d\left(\sum_b p_{0b}\hat{F}^{KM}_{20b}(t)\right) \\
&= \sum_b \sum_{b'} p_{1b}p_{1b'} \int \hat{F}^{KM}_{11b}(t)d\hat{F}^{KM}_{21b'}(t) - \sum_b \sum_{b'} p_{0b}p_{0b'} \int \hat{F}^{KM}_{10b}(t)d\hat{F}^{KM}_{20b'}(t),
\end{aligned}$$

if we denote

$$\xi_c = \begin{cases} 1 & c=1 \\ -1 & c=0 \end{cases},$$

then we have

$$\tilde{\mu}_{Ae} \approx \sum_a \sum_b \sum_{b'} \xi_a p_{ab}p_{ab'} \int \hat{F}^{KM}_{1ab}(t)d\hat{F}^{KM}_{2ab'}(t).$$

By the same technique used in (7.12), we have

$$\begin{aligned}
\sqrt{n}(\tilde{\mu}_{Ae} - \mu_A) &\approx \sqrt{n}\sum_a \sum_b \sum_{b'} \xi_a p_{ab}p_{ab'} \left(\int \hat{F}^{KM}_{1ab}(t)d\hat{F}^{KM}_{2ab'}(t) - \int F^{KM}_{1ab}(t)dF^{KM}_{2ab'}(t) \right) \\
&\approx \sum_a \sum_b \sum_{b'} \xi_a p_{ab}p_{ab'} \left[\int \sqrt{n}(\hat{F}^{KM}_{1ab}(t) - F_{1ab}(t))dF_{2ab'}(t) \right. \\
&\quad \left. - \int \sqrt{n}(\hat{F}^{KM}_{2ab'}(t) - F_{2ab'}(t))dF_{1ab}(t) \right].
\end{aligned} \tag{7.18}$$

Similarly,

$$\begin{aligned}
\sqrt{n}(\tilde{\mu}_{Be} - \mu_B) &\approx \sum_b \sum_a \sum_{a'} \xi_b q_{ab}p_{a'b} \left[\int \sqrt{n}(\hat{F}^{KM}_{1ab}(t) - F_{1ab}(t))dF_{2a'b}(t) \right. \\
&\quad \left. - \int \sqrt{n}(\hat{F}^{KM}_{2a'b}(t) - F_{2a'b}(t))dF_{1ab}(t) \right].
\end{aligned} \tag{7.19}$$

Applying Lemma 7.1 again we have $\sqrt{n}(\tilde{\mu}_{Ae} - \mu_A)$ and $\sqrt{n}(\tilde{\mu}_{Be} - \mu_B)$ are both asymptotically centered normal variables. This leads to the asymptotical normality of $\sqrt{n}\tilde{D}_e = \sqrt{n}(\tilde{\mu}_{Ae} - \tilde{\mu}_{Be})$. It remains to compute the variance of $\sqrt{n}\tilde{D}$.

By (7.18) and (7.19), we can get

$$\begin{aligned}
\mathrm{Var}(\sqrt{n}(\tilde{\mu}_{Ae} - \mu_A)) &= \sum_a \sum_b \sum_{b'} p_{ab}^2 p_{ab'}^2 \left[\int \int \Gamma_{1ab}(s,t) dF_{2ab'}(s) dF_{2ab'}(t) \right. \\
&\quad \left. + \int \int \Gamma_{2ab'}(s,t) dF_{1ab}(s) dF_{1ab}(t) \right], \\
\mathrm{Var}(\sqrt{n}(\tilde{\mu}_{Be} - \mu_B)) &= \sum_b \sum_a \sum_{a'} q_{ab}^2 q_{a'b}^2 \left[\int \int \Gamma_{1ab}(s,t) dF_{2a'b}(s) dF_{2a'b}(t) \right. \\
&\quad \left. + \int \int \Gamma_{2a'b}(s,t) dF_{1ab}(s) dF_{1ab}(t) \right].
\end{aligned} \tag{7.20}$$

In addition, we have,

$$
\begin{aligned}
& \mathrm{Cov}(\sqrt{n}(\tilde{\mu}_{Ae}-\mu_A),\sqrt{n}(\tilde{\mu}_{Be}-\mu_B)) \\
= \;& \mathrm{Cov}\Bigg\{\sum_a\sum_b\sum_{b'}\xi_a p_{ab}p_{ab'}\Bigg[\int\sqrt{n}(\hat{F}^{KM}_{1ab}(t)-F_{1ab}(t))dF_{2ab'}(t) \\
& -\int\sqrt{n}(\hat{F}^{KM}_{2ab'}(t)-F_{2ab'}(t))dF_{1ab}(t)\Bigg], \\
& \sum_\beta\sum_\alpha\sum_{\alpha'}\xi_\beta q_{\alpha\beta}p_{\alpha'\beta}\Bigg[\int\sqrt{n}(\hat{F}^{KM}_{1\alpha\beta}(t)-F_{1\alpha\beta}(t))dF_{2\alpha'\beta}(t) \\
& -\int\sqrt{n}(\hat{F}^{KM}_{2\alpha'\beta}(t)-F_{2\alpha'\beta}(t))dF_{1\alpha\beta}(t)\Bigg]\Bigg\} \\
= \;& \sum_a\sum_b\sum_{b'}\sum_\beta\sum_\alpha\sum_{\alpha'}\xi_a\xi_\beta p_{ab}p_{ab'}q_{\alpha\beta}q_{\alpha'\beta}\int\int\mathrm{Cov}[\sqrt{n}(\hat{F}^{KM}_{1ab}(t)-F_{1ab}(t)), \\
& \sqrt{n}(\hat{F}^{KM}_{1\alpha\beta}(t)-F_{1\alpha\beta}(t))]dF_{2ab'}(t)dF_{2\alpha'\beta}(t) \\
& +\sum_a\sum_b\sum_{b'}\sum_\beta\sum_\alpha\sum_{\alpha'}\xi_a\xi_\beta p_{ab}p_{ab'}q_{\alpha\beta}q_{\alpha'\beta}\int\int\mathrm{Cov}[\sqrt{n}(\hat{F}^{KM}_{2ab'}(t)-F_{2ab'}(t)), \\
& \sqrt{n}(\hat{F}^{KM}_{2\alpha'\beta}(t)-F_{2\alpha'\beta}(t))]dF_{1ab}(t)dF_{1\alpha\beta}(t) \\
& -\sum_a\sum_b\sum_{b'}\sum_\beta\sum_\alpha\sum_{\alpha'}\xi_a\xi_\beta p_{ab}p_{ab'}q_{\alpha\beta}q_{\alpha'\beta}\int\int\mathrm{Cov}[\sqrt{n}(\hat{F}^{KM}_{1ab}(t)-F_{1ab}(t)), \\
& \sqrt{n}(\hat{F}^{KM}_{2\alpha'\beta}(t)-F_{2\alpha'\beta}(t))]dF_{2ab'}(t)dF_{1\alpha\beta}(t) \\
& -\sum_a\sum_b\sum_{b'}\sum_\beta\sum_\alpha\sum_{\alpha'}\xi_a\xi_\beta p_{ab}p_{ab'}q_{\alpha\beta}q_{\alpha'\beta}\int\int\mathrm{Cov}[\sqrt{n}(\hat{F}^{KM}_{2ab'}(t)-F_{2ab'}(t)), \\
& \sqrt{n}(\hat{F}^{KM}_{1\alpha\beta}(t)-F_{1\alpha\beta}(t))]dF_{1ab}(t)dF_{2\alpha'\beta}(t).
\end{aligned}
\tag{7.21}
$$

Since $\sqrt{n}(\hat{F}^{KM}_{1..}(t)-F_{1..}(t))$ and $\sqrt{n}(\hat{F}^{KM}_{2..}(t)-F_{2..}(t))$ are independent, it is easy to

see the last two terms in (7.21) are 0. It follows from independency that

$$\begin{aligned}
&\mathrm{Cov}(\sqrt{n}(\tilde{\mu}_{Ae}-\mu_A),\sqrt{n}(\tilde{\mu}_{Be}-\mu_B)) \\
=\ & \sum_a\sum_b\sum_{b'}\sum_{\alpha'}\xi_a\xi_b p_{ab}p_{ab'}q_{ab}q_{\alpha' b}\int\int \Gamma_{1ab}(s,t)dF_{2ab'}(s)dF_{2\alpha' b}(t) \\
&+\sum_a\sum_{b'}\sum_b\sum_{\alpha}\xi_a\xi_{b'} p_{ab}p_{ab'}q_{\alpha b'}q_{ab'}\int\int \Gamma_{2ab'}(s,t)dF_{1ab}(s)dF_{1ab'}(t) \qquad (7.22)\\
=\ & \sum_a\sum_b\sum_{b'}\sum_{\alpha}\xi_a\xi_b p_{ab}p_{ab'}q_{ab}q_{\alpha b}\int\int \Gamma_{1ab}(s,t)dF_{2ab'}(s)dF_{2\alpha b}(t) \\
&+\sum_a\sum_b\sum_{b'}\sum_{\alpha}\xi_a\xi_b p_{ab}p_{ab'}q_{ab}q_{\alpha b}\int\int \Gamma_{2ab}(s,t)dF_{1ab'}(s)dF_{1ab}(t).
\end{aligned}$$

To finish our proof, we can define

$$\sigma^2=\mathrm{Var}(\sqrt{n}(\tilde{\mu}_{Ae}-\mu_A))+\mathrm{Var}(\sqrt{n}(\tilde{\mu}_{Be}-\mu_B))-2\mathrm{Cov}(\sqrt{n}(\tilde{\mu}_{Ae}-\mu_A),\sqrt{n}(\tilde{\mu}_{Be}-\mu_B)),$$

with three terms defined in (7.20) and (7.22). □

Chapter 8

Summary and Future Work

In this thesis, I first reviewed current literatures on the measures of treatment effect and their statistical procedures in cancer clinical trials as well as on the applications of the empirical likelihood method for survival data and statistical inferences on a two sample semi-parametric density ratio model, two of the major tools in the thesis. I then developed new statistical inference procedures for two important measures of the treatment effect, the time-dependent hazard ratio and the probabilistic index, and applied the probabilistic index to assess the interaction between the treatment and covariates. Specifically,

In Chapter 5, with the hazard rate function estimated by a non-parametric kernel method with a under-smoothing bandwidth parameter, an empirical likelihood confidence interval for the hazard ratio function is proposed and shown by simulations to outperform the existing inferences methods based on the normal approximation (Tu, 2007) in terms of the coverage probability.

In Chapter 6, assuming that the density functions of the survival time in two treatment groups are linked by a density ratio model, the conditional and weighted

empirical likelihood method are applied to estimate the probabilistic index with the confidence intervals constructed by the bootstrap re-sampling method. Simulation studies have shown that the conditional empirical likelihood based confidence interval performs better when the censoring rate is low, while the weighted empirical likelihood method have the advantages when the censoring rate is high.

In Chapter 7, both approaches based on the Efron's estimator and the semi-parametric density ratio model are applied to estimate and construct confidence intervals for a non-parametric measure of the treatment-covariate interaction. The approaches are also extended to compare the strength of interactions between the treatment and two covariates. It is shown in the simulation studies that the approach based on the Efron's estimator is satisfactory when the censoring rate is low. When the censoring is modest, the conditional likelihood method could be used. The weighed empirical likelihood method yields the smallest type I error, but also with the lowest power.

All the approaches proposed in this thesis are applied to real data sets from cancer clinical trials, which demonstrate they are feasible in the practical applications.

I have listed the potential future research topics at the end of Chapter 5-7, which included development of simultaneous confidence bands for the hazard ratio function as well as statistical procedures which can take into account of additional covariates in the evaluation of the treatment effect based on the probabilistic index and assessment the heterogeneity of treatment effects.

Bibliography

L. Acion, J.J. Peterson, S. Temple, and S. Arndt. Probabilistic index: an intuitive non-parametric approach to measuring the size of treatment effects. *Statistics in Medicine*, 25:591–602, 2006.

J.A. Anderson. Multivariate logistic compounds. *Biometrika*, 66:17–26, 1979.

Y.C. Bassiakos, X. Meng, and S. Lo. A general estimator of the treatment effect when the data are heavily censored. *Biometrika*, 78:741–748, 1991.

P.J. Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimates. *Annals of Statistics*, 1:1071–1095, 1973.

G.E.P. Box and D.R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26:211–252, 1964.

L.C. Brumback, M.S. Pepe, and T.A. Alonzo. Using the roc curve for gauging treatment effect in clinical trials. *Statistics in Medicine*, 25:575–590, 2006.

M. Buyse. Letter to the editor: Reformulating the hazard ratio to enhance communication with clinical investigators. *Clinical Trials*, 5:641–642, 2008.

K. F. Cheng and C. K. Chu. Semiparametric density estimation under a two-sample density ratio model. *Bernoulli*, 10:583–604, 2004.

M. Cheng, P. Hall, and D. Tu. Confidence bands for hazard rates under random censorship. *Biometrika*, 93:357–366, 2006.

C. Contal and J. O'Quigley. An application of changepoint methods in studying the effect of age on survival in breast cancer. *Computational Statistics and Data Analysis*, 30:253–270, 1999.

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

T.F. Cox and G. Ferry. Robust logistic discrimination. *Biometrika*, 78:841–849, 1991.

S. Csorgo and L. Horvath. The rate of strong uniform consistency for the product-limit estimator. *Probability Theory and Related Fields*, 62:411–426, 1983.

B. Efron. The two-sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 4:831–853, 1967.

E.C. Fieller. Some problems in interval estimation. *Journal of Royal Statistical Society, Series B*, 16:175–185, 1954.

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

K. Fokianos. Merging information for semiparametric density estimation. *Journal of Royal Statistical Society, Series B*, 66:941–958, 2004.

K. Fokianos and I. Kaimi. On the effect of misspecifying the density ratio model. *Annals of the Institute for Statistical Mathematics*, 58:475–497, 2006.

K. Fokianos and J.F. Troendle. Inference for the relative treatment effect with the density ratio model. *Statistical Modeling*, 7:155–173, 2007.

E.A. Gehan. A generalized wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika*, 52:203–223, 1965.

P.B. Gilbert, L.J. Wei, M.R. Kosorok, and J.D. Clemens. Simultaneous inferences on the contrast of two hazard functions with censored observations. *Biometrics*, 58: 773–780, 2002.

P. Hall and A.B. Owen. Empirical likelihood confidence bands in density estimation. *Journal of Computational and Graphical Statistics*, 2:273–289, 1993.

F.E. Harrell, R.M. Califf, D.B. Pryor, K.L. Lee, and R.A. Rosati. Evaluating the yield of medical tests. *The Journal of the American Medical Association*, 247:2543–2546, 1982.

S. Jiang and D. Tu. Empirical likelihood confidence intervals for ratio of hazard rates under right censorship. *Statistics and Its Interface*, 3:455–464, 2010.

E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

C.S. Karapetis, S. Khambata-Ford, D.J. Jonker, C.J. O'Callaghan, D. Tu, N.C. Tebbutt, R.J. Simes, H. Chalchal, J. Shapiro, S. Robitaille, T. Price, L. Shepherd, H-J. Au, C. Langer, M.J. Moore, and J.R. Zalcberg. K-ras mutations and benefit from

cetuximab in advanced colorectal cancer. *New England Journal of Medicine*, 359: 1757–1765, 2008.

R. Kay and S. Little. Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, 74:495–501, 1987.

B. Kedem, E. Kim, A. Voulgaraki, and B.I. Graubard. Two-dimensional semiparametric density ratio modeling of testicular germ cell data. *Statistics in Medicine*, 28:2147–2159, 2009.

A. Kezioua and S. Leoni-Aubin. On empirical likelihood for semiparametric two-sample density ratio models. *Journal of Statistical Planning and Inference*, 138: 915–928, 2008.

J.A. Koziol and Z. Jia. The concordance index C and Mann-Whitney parameter $Pr(X > Y)$ with randomly censored data. *Biometrical Journal*, 51:467–474, 2009.

M.N. Levine, K.I. Pritchard, V.H.C. Bramwell, L.E. Shepherd, D. Tu, and N. Paul. Randomized trial comparing Cyclophosphamide, Epirubicin, and Fluorouracil with Cyclophosphamide, Methotrexate, and Fluorouracil in premenopausal women with node-positive breast cancer: Update of National Cancer Institute of Canada Clinical Trials Group Trial MA5. *Journal of Clinical Oncology*, 23:5166–5170, 2005.

G. Li. On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statistics and Probability Letters*, 25:95–104, 1995.

S.H. Lo, Y.P. Mack, and J.L. Wang. Density and hazard rate estimation for censored data via strong representation of the kaplan-meier estimator. *Probability Theory and Related Fields*, 80:461–473, 1989.

R.A. Maller and S. Zhou. Estimating the proportion of immunes in a censored sample. *Biometrika*, 79:731–739, 1992.

I.W. McKeaguea and Y. Zhao. Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statistics and Probability Letters*, 60:405–415, 2002.

B.K. Moser and M.H. McCann. Reformulating the hazard ratio to enhance communication with clinical investigators. *Clinical Trials*, 5:248–252, 2008.

H. Muss, D. Tu, J. Ingle, S. Martino, N. Robert, J. Pater, T. Whelan, M.J. Palmer, M. Piccart, L.E. Shepherd, K.I. Pritchard, Z. He, and P.E. Goss. Efficacy, toxicity and quality of life in older women with early breast cancer treated with letrozole or placebo after five years of tamoxifen: NCIC CTG intergroup trial MA. 17. *Journal of Clinical Oncology*, 26:1956–1964, 2008.

F.P. O'Malley, S. Chia, D. Tu, L.E. Shepherd, M.N. Levine, V.H. Bramwell, and K.I. Andrulis, I.L. Pritchard. Topoisomerase II alpha and responsiveness of breast cancer to adjuvant chemotherapy. *Journal of National Cancer Institute*, 101:644–650, 2009.

A.B. Owen. *Empirical likelihood*. Chapman and Hall, 2001.

A.B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75:237–249, 1988.

K.M. Patel and D.G. Hoel. A nonparametric test for interaction in factorial experiments. *Journal of the American Statistical Association*, 68:615–620, 1973.

J. Qin. Inferences for case-control and semiparametric two-sample density ratio model. *Biometrika*, 85:619–630, 1998.

J. Qin and B. Zhang. Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika*, 90:585–596, 2003.

J. Qin and B. Zhang. A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 84:609–618, 1997.

H. Ramlau-Hansen. Smoothing counting process intensities by means of kernel functions. *Annals of Statistics*, 11:453–466, 1983.

J. Ren. Weighted empirical likelihood ratio confidence intervals for the mean with censored data. *Annals of the Institute of Statistical Mathematics*, 53:498–516, 2001.

J. Ren. Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *The Annals of Statistics*, 36:147–166, 2008.

M. Schemper. Non-parametric analysis of treatment-covariate interaction in the presence of censoring. *Statistics in Medicine*, 7:1257–1266, 1988.

W.F. Scott. The estimation of the hazard ratio in clinical trials and in meta-analysis. *The Mathematical Scientist*, 25:54–58, 2000.

J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer, 1995.

J. Shen and S. He. Empirical likelihood for the difference of two survival functions under right censorship. *Statistics and Probability Letters*, 76:169–181, 2006.

J. Shen and S. He. Empirical likelihood confidence intervals for hazard and density functions under right censorship. *Annals of the Institute of Statistical Mathematics*, 60:575–589, 2008.

Y. Shen, J. Qin, and J.P. Costantino. Inference of tamoxifen's effects on prevention of breast cancer from a randomized controlled trial. *Journal of American Statistics Association*, 102:1235–1244, 2007.

D.M. Stablein, W.H. Carter, and J.W. Novak. Analysis of survival data with non-proportional hazard functions. *Controlled Clinical Trials*, 2:149–159, 1981.

D.R. Thomas and G.L. Grunkemeier. Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70: 865–871, 1975.

D. Tu. Under-smoothed kernel confidence intervals for the hazard ratio based on censored data. *Biometrical Journal*, 49:474–483, 2007.

D. Tu and J. Pater. Some issues related to applications of coxs models in cancer clinical trials. *Controlled Clinical Trials*, 20:84S, 1999.

D. Tu, K. Shalay, and J. Pater. Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. *Drug Information Journal*, 34: 511–523, 2000.

Q. Wang and P. Lai. Empirical likelihood calibration estimation for the median treatment difference in observational studies. *Computational Statistics and Data Analysis*, 55:1596–1609, 2011.

X. Xiang. Law of the logarithm for density and hazard rate estimation for censored data. *Journal of Multivariate Analysis*, 49:278–286, 1994.

B. Zhang. Quantile estimation under a two-sample semi-parametric model. *Bernoulli*, 6:491–511, 2000.

Y. Zhao and M Zhao. Empirical likelihood for the contrast of two hazard functions with right censoring. *Statistics and Probability Letters*, 81:392–401, 2011.

W. Zhou. Statistical inference for $Pr(X < Y)$. *Statistics in Medicine*, 27:257–279, 2008.

Y. Zhou and H. Liang. Empirical-likelihood-based semiparametric inference for the treatment effect in the two-sample problem with censoring. *Biometrika*, 92:271–282, 2005.