NONPARAMETRIC AND SEMIPARAMETRIC METHODS FOR INTERVAL-CENSORED FAILURE TIME DATA

A Dissertation

Presented to

the Faculty of the Graduate School

University of Missouri–Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

CHAO ZHU

Dr. (Tony) Jianguo Sun, Dissertation Supervisor

AUGUST, 2006

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

NONPARAMETRIC AND SEMIPARAMETRIC METHODS FOR INTERVAL-CENSORED FAILURE TIME DATA

Presented by Chao Zhu

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of acceptance.

Dr. (Tony) Jianguo Sun

Dr. Nancy Flournoy

Dr. Tim Wright

Dr. Athanasios Micheas

Dr. Hong He

To my parents

ACKNOWLEDGEMENTS

I am greatly indebted to my dissertation supervisor, Dr. Tony Sun, for his kindly providing guidance throughout the development of this dissertation. His direction and comments have been of the greatest help at all times. This work could not have been done without him.

I am also very grateful for having an exceptional doctoral committee and wish to thank Drs. Nancy Flournoy, Tim Wright, Athanasios Micheas and Hong He for their continual support and constructive comments.

I extend many thanks to all persons in the Department of Statistics at the University of Missouri-Columbia for their helps during my graduate education, especially Zhigang Zhang, Qiang Zhao, Lianming Wang, Liuquan Sun and Xingwei Tong, for their useful discussions.

Finally, I would like to thank my fiancee, Luyan Dai, for her love, patience and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NONPARAMETRIC AND SEMIPARAMETRIC METHODS FOR INTERVAL-CENSORED FAILURE TIME DATA

Chao Zhu

Dr. (Tony) Jianguo Sun, Dissertation Supervisor

ABSTRACT

Interval-censored failure time data commonly arise in follow-up studies such as clinical trials and epidemiology studies. For their analysis, what interests researcher most includes comparisons of survival functions for different groups and regression analysis. This dissertation, which consists of three parts, consider these problems on two types of interval-censored data by using nonparametric and semiparametric methods.

In Chapter 2, we discuss a goodness-of-fit test for checking the proportional odds (PO) model with interval-censored data. The PO model has a feature that allows the ratio of two hazard functions to be monotonic and converge to one. Hence, it provides an important tool for modeling the situation where hazard functions are nonproportional. We derive a procedure for testing the PO model, which is a generalization of Dauxois and Kirmani (2003) for right-censored data. Simulation studies suggest that the proposed test works well and we apply the test to a real dataset from an AIDS cohort study.

Chapters 3 considers nonparametric comparison of survival functions. For this, several test procedures have been proposed for interval-censored failure time data in which distributions of censoring intervals are identical among different treatment groups. Sometimes these distributions may not be the same and depend on treatments. A class of test statistics is proposed for situations where the distributions may be different for subjects in different treatment groups. The asymptotic normality of the test statistics is established and the test procedure is evaluated by simulations, which suggest that it works well. An illustrative example is provided.

Chapter 4 discusses semiparametric regression analysis of two-sample current status

data. For their regression analysis, One limitation of commonly used models is that they cannot be used to situations where survival functions cross. We consider a class of two-sample models that include these commonly used models as special cases and especially, are appropriate for crossing survival functions. Some estimating equation-based approaches are presented and the proposed estimates of regression parameters are shown to be consistent and asymptotically normally distributed. The method is evaluated using simulation studies and applied to a set of current status data arising from a tumorgenicity experiment.

CHAPTER 1

INTRODUCTION

1.1 Basic Quantities in Survival Analysis

Survival analysis, or time-to-event data analysis is used predominately in biomedical science where the interest is in observing time to death either of patients or of laboratory animals. It has also been used widely in social sciences where interest is on analyzing time to events such as job change, marriage, birth of children and so forth. The engineering science has also contributed to the development of survival analysis which is called "reliability analysis" or "failure time analysis" in this field, where the main focus is on modeling the time of machines or electronic components to break down. The data arising from these fields are usually referred to as survival data, time-to-event data, or failure time data. Note that the failure time, usually denoted by T, is a nonnegative random variable.

The survival function of T is defined as $S(t) = P(T \geq t) = 1 - F(t)$, where $F(t)$ is the cumulative distribution function (CDF). $S(t)$ is the probability that an individual experiences the event no earlier than time t. In survival analysis, the survival function of a failure time is preferred over the cumulative distribution function because it is more intuitive and easier to communicate with people in applied fields where survival data occur such as medical sciences.

In addition to the survival function, the hazard function and the cumulative hazard

function of T are also commonly used in modeling T because of their conveniences. When T is continuous, the hazard function of T is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t) = \frac{f(t)}{S(t)} = -[\frac{d}{dt}\{S(t)\}]/S(t),$$

where $f(t) = dF(t)/dt$ is the density function of T. Note that $\lambda(t)$ is the instantaneous failure rate at time t given that an individual survives up to time t^-. The cumulative hazard function is defined as

$$\Lambda(t) = \int_0^t \lambda(u)du.$$

It is easy to see that

$$S(t) = \exp[-\Lambda(t)] = \exp[-\int_0^t \lambda(u)du].$$

Thus, $S(t)$, $\lambda(t)$, or $\Lambda(t)$ uniquely determines the distribution of T.

If T is a discrete random variable taking values $0 = t_0 < t_1 < t_2 \cdots$, the hazard function is defined as

$$\lambda(t_j) = P(T = t_j | T \geq t_j) = \frac{f(t_j)}{S(t_j)}, j = 1, 2, \cdots$$

where $S(t_0) = 1$ and $f(t_j) = S(t_j) - S(t_{j+1}), j = 1, 2, \cdots$. The cumulative hazard function is defined as

$$\Lambda(t) = \sum_{t_j \leq t} \lambda(t_j).$$

1.2 Typical Censoring Mechanisms and Examples

Imagine that you are a researcher in a hospital for studying the effectiveness of a new treatment for a generally terminal disease. The major variable of interest could be the number of days (failure time T) that the patient with the disease survives. In principle,

if everyone dies, one could use the standard parametric and nonparametric statistics for describing the average survival and for comparing the new treatment with traditional treatments. However, at the end of the study there may be patients who survive over the entire study period, in particular among those patients who entered the hospital (and the research project) late in the study. Also there may be other patients with whom we lose contact. Surely, one would not want to exclude all of these patients from the study by declaring them to be missing (since most of them are "survivors" and, therefore, they reflect on the success of the new treatment method). These observations, which contain only partial information, are called censored observations (e.g., patient A survived at least 4 months before he moved away and we lost contact. The term censoring was first used by Hald, 1949).

Above is an example of right censoring. It is the most commonly encountered censoring mechanism in many fields such as clinical trials, environmental science, insurance, and manufacturing. Two common types of right censoring are *Type I* and *Type II censoring*. The Type I right censoring means that there is a fixed censoring time C and the exact failure time X of an individual is known if and only if X is less than or equal to C. If X is greater than C, his or her event time is censored at C. The data from this type of experiments can be conveniently represented by pairs of random variables (T, δ), where δ indicates whether the survival time is observed ($\delta = 1$) or censored ($\delta = 0$) and T is equal to X if the survival time is observed and C if it is censored, i.e., $T = min(X, C)$.

Type II right censoring means that a study continues until r failures occur, where r is a predetermined integer ($r < n$). Experiments involving Type II censoring are often used in testing of equipment life. Here, all items are put on test at the same time and the test

is terminated when r of the n items have failed. Such an experiment may save time and money because it could take a very long time for all items to fail. Also the statistical treatment of Type II censored data is simpler in some sense because the data consist of the r smallest survival times in a random sample of n survival times and the theory of order statistics is directly applicable.

A failure time T associated with a specific individual in a study is considered to be *left censored* if it is less than a censoring time C_l, that is, the event of interest has already occurred for the individual before that person enters the study at time C_l. For such individuals, we know that they have experienced the event some time before time C_l, but their exact event time is unknown. For example, on a survey questionnaire, the investigator wonders when the individual first used marijuana. A subject is then left censored if he/she admits that he/she has used it before but cannot recall when the first time was.

Interval censoring is another type of censoring mechanism. There exist two types of interval-censored data, *case I* and *II interval-censored data* (Groeneboom and Wellner, 1992; Sun, 2005). The former, which is also often referred to as current status data, means that each subject is observed only once and thus the failure event of interest is observed only to have occurred before the observation time or not yet. In other words, the failure time of interest T is either left- or right-censored. Case I interval-censored data commonly occur in, for example, tumorigenicity experiments. In these experiments, the tumor onset time of animals is usually of main interest but not observable. Instead, only tumor status is usually known at death (either natural death or being sacrificed). Thus, the tumor onset time is known only to be less or greater than the death time.

Case II interval-censored data mean that the failure time T is known only to belong to an interval, say $[L, R]$. They reduce to case I interval-censored data if the interval includes either 0 or infinity. This type of data arises in many medical and health studies that entail periodic follow-ups. In this situation, an individual due for scheduled observations for a clinically observed change in disease status may miss some observations and may return with a changed status, thus contributing an interval-censored time of the occurrence of the change. Another example arises in the acquired immune deficiency syndrome (AIDS) studies that concern the human immunodeficiency virus (HIV) infection and the AIDS incubation time (the time from HIV infection to AIDS diagnosis). In this case if a subject is HIV positive at the beginning of the study, his or her HIV infection time is usually determined by a retrospective study of the subject's history. Thus only an interval given by the last HIV negative test and the first HIV positive test is known for the HIV infection time.

Another way to represent a case II interval-censored observation is to use $\{ U, V, \delta_1 = I(T \leq U), \delta_2 = I(U < T \leq V), \delta_3 = 1 - \delta_1 - \delta_2 \}$ assuming that each subject is observed twice, where U and V are two random variables satisfying $U \leq V$ with probability 1. This formulation is convenient and often used, for example, in a theoretical investigation of an inference procedure. Both representations give rise to the same likelihood function. Note that although (U, V) representation seems natural, it is not common to have interval-censored data collected or given in these formats in practice. However, it is much easier and more natural to impose assmptions such as independence with T on them than on (L, R) representation, which is often needed for derivation of the asymptotic properties of inference procedures. For data given in (U, V) representation, one can easily obtain the

corresponding data with (L, R) representation. More discussion on this is given in later chapters.

1.3 Parametric and Semiparametric Models in Survival Analysis

In this section, we review some commonly used parametric and semiparametric models in survival analysis.

1.3.1 Parametric models

Parametric models (for the failure time T) naturally smooth the data by "borrowing" information from adjacent points. With growing computing power and existing statistical programming languages, it is relatively simple to work with exact likelihood for interval-censored data with a variety of parametric models. Among other distributions, the exponential, Weibull and log-logistic distributions are mostly used in practice. We shall briefly introduce the latter two distributions in the following.

Suppose that T is continuous. By Weibull distribution, we mean that T has density function

$$f(t) = \alpha\,\eta\,t^{\alpha-1}\,\exp\{-\eta\,t^{\alpha}\},$$

where $\alpha > 0$ and $\eta > 0$. Thus the survival function of T is

$$S(t) = \exp\{-\eta\,t^{\alpha}\}$$

and the hazard function is

$$\lambda(t) = \alpha\,\eta\,t^{\alpha-1}.$$

The corresponding cumulative hazard function is

$$\Lambda(t) = \int_0^t \alpha\,\eta\,t^{\alpha-1}dt = \eta\,t^{\alpha}.$$

Note that the Weibull distribution is flexible enough to accommodate increasing ($\alpha > 1$), decreasing ($\alpha < 1$), or constant ($\alpha = 1$) hazard rates. When $\alpha = 1$, the Weibull distribution reduces to the exponential distribution with $\lambda(t) = \eta$.

A failure time T is said to follow the log-logistic distribution if its logarithm, $Y = \ln(T)$, follows the logistic distribution, a distribution closely resembling the normal distribution. Its survival function and hazard rate may be written as

$$S(t) = \frac{1}{1+\eta\,t^{\alpha}}$$

and

$$\lambda(t) = \frac{\alpha\,\eta\,t^{\alpha-1}}{1+\eta\,t^{\alpha}}.$$

The numerator of the hazard function is the same as the Weibull hazard, but the denominator allows the hazard to possess the following characteristics: monotone decreasing for $\alpha \leq 1$, and for $\alpha > 1$, the hazard rate increases initially to a maximum at time $[(\alpha-1)/\eta]^{1/\alpha}$ and then decreases to zero as time approaches infinity.

1.3.2 The proportional hazards model

Although the analysis of interval-censored data based on parametric models can be simple and efficient if the model is correctly specified, they are not widely used since the model choice is hard to determine in many situations. Instead, one usually looks for semiparametric or nonparametric methods. For the former, a common used model is the proportional hazards (PH) model, also referred to as the Cox model (Cox, 1972). It

specifies the hazard function of a continuous survival time T to have the form

$$\lambda(t|Z) = \lambda_0(t) \exp\{\beta' Z\}$$

given covariates Z which may depend on time, where β is a $p \times 1$ vector of unknown regression parameters and $\lambda_0(t)$, the baseline hazard, is an unknown and unspecified function. Note that the proportionality comes from the fact that, for example, if we look at two individuals with covariate values Z and Z^*, the ratio of their hazard functions is

$$\frac{\lambda(t|Z)}{\lambda(t|Z^*)} = \frac{\lambda_0(t) \exp[\sum_{k=1}^{p} \beta_k Z_k]}{\lambda_0(t) \exp[\sum_{k=1}^{p} \beta_k Z_k^*]} = \exp\left[\sum_{k=1}^{p} \beta_k (Z_k - Z_k^*)\right],$$

which is a constant, where $\beta = (\beta_1, ..., \beta_p)$. Usually β is of the main interest and can be estimated independently by the partial likelihood approach (Cox, 1975) when right-censored data are observed. This appealing property of the PH model, together with its great flexibility, has made it one of the most popular models in survival analysis during the past three decades.

Diamond et al. (1986) were the first to use the PH model on case I interval censored data. Their methods, however, require estimation of the baseline hazard $\lambda_0(t)$. Huang (1996a) gave a systematic treatment of the proportional hazards model under case I interval censoring. He showed that, under certain regularity conditions, $\hat{\beta}_n$, the maximum likelihood estimator (MLE) of β, is consistent with an $n^{1/2}$ convergence rate and has an asymptotic normal distribution with the limiting variance given by the inverse of the Fisher information of β. However, $\hat{\Lambda}_n$, the MLE of the cumulative hazard function, is only consistent with an $n^{1/3}$ convergence rate, and its asymptotic distribution is unknown. Finkelstein (1986) proposed to use the Newton-Raphson algorithm to compute the MLE of the regression parameter β and the baseline cumulative hazard function for

case II interval censoring. Satten (1996) proposed a marginal likelihood method to fit the proportional hazards model to case II interval-censored data, and Pan (2000) applied a multiple imputation approach for comparing two treatments. Betensky et al. (2002) proposed a local likelihood method mainly for estimating the baseline hazard function, while Cai and Betensky (2003) introduced piecewise linear penalized spline for the same purpose.

1.3.3 The proportional odds model

An important alternative to the PH model is the proportional odds (PO) model, which assumes that

$$\log\{F(t|Z)/S(t|Z)\} = h(t) + \beta' Z,$$

where $F(t|Z)$ and $S(t|Z)$ denote the distribution function and the survival function of T given Z, respectively, and $h(t)$ is a baseline monotone-increasing function, also referred to as the baseline log odds. The original PO model was developed by McCullagh (1980) for analyzing ordinal data. Although this model is not as commonly used as the PH model when censored data are observed, partly due to the lack of a widely accepted estimation procedure for regression parameter β, it does provide certain flexibility that the PH model can not. For instance, the initial effect of treatment, or the differences between stages of the disease at diagnosis, may diminish with time and the hazard functions of different groups of patients should become more similar. In this case, two hazard functions from different treatment groups are not proportional, but changing with time. Thus the assumption of the PH model, which requires a constant ratio for two hazard functions, is then violated. One of the earliest applications of the PO model on interval censoring was

given by Dinse and Lagakos (1983). Rossini and Tsiatis (1996) also discussed the fitting of this model to case I interval censoring with approximating the baseline log odds by step functions, thus obtaining consistent and asymptotic normal estimators for β. Huang and Rossini (1997) and Rabinowitz et al. (2000) considered the sieve estimation and the approximated score function methods, respectively. For asymptotic properties and computation of the MLEs of β and $h(t)$, see Huang and Wellner (1996).

1.4 Nonparametric Survival Analysis

In addition to the semiparametric models discussed in the previous section, nonparametric methods for the analysis of survival data have also attracted much attention. Similar to the semiparametric methods, nonparametric methods do not require the knowledge of the underlying distribution of the failure time T. Hence it provides a flexible way to deal with the data in many practical situations. In this section, we review some classical problems that can be addressed by using nonparametric methods.

1.4.1 Nonparametric Estimation of a Survival Function

One of the basic research problems in survival analysis is the estimation of a survival function, for which numerous methods have been proposed under different censoring mechanisms. For right-censored data, consider a survival study that consists of n independent subjects. Let $S(t)$ denote the true survival function and $t_0 = 0 < t_1 < t_2 < \ldots < t_{k+1} = \infty$ the observed failure times. Define

$d_j =$ the number of failures at t_j,

$r_j =$ the number of subjects at risk at t_j^-,

$c_j =$ the number of subjects censored in $[t_j, t_{j+1})$,

$t_{j1}, \ldots, t_{jc_j}$ =censored survival times in $[t_j, t_{j+1})$ $\quad j = 0, 1, \ldots, k$.

The likelihood function is then proportional to

$$L = \prod_{j=0}^{k} \left\{ [S(t_j) - S(t_j+)]^{d_j} \prod_{r=1}^{c_j} S(t_{jr}+) \right\}$$

and the nonparametric maximum likelihood estimator (NPMLE) of $S(t)$ is given by Kaplan-Meier estimator

$$\hat{S}(t) = \prod_{j|t_j<t} \frac{r_j - d_j}{r_j}$$

(Kaplan and Meier, 1958).

A closely related estimator of a survival function is given by $\tilde{S}(t) = \exp\{-\tilde{\Lambda}(t)\}$, where $\tilde{\Lambda}(t)$ is the Nelson-Aalen estimator of the cumulative hazard function and has the form

$$\tilde{\Lambda}(t) = \sum_{j:t_{(j)} \leq t} \frac{d_j}{r_j}$$

(Nelson, 1972; Aalen, 1978).

For case I interval-censored data or current status data, suppose that F denotes the CDF of the survival time of interest. Then the NPMLE of F can be shown to be equal to the isotonic regression of $\{d_1/n_1, ..., d_m/n_m\}$ with weights $\{n_1, ..., n_m\}$, where $d_j = \sum_{i \in S_j} I(T_i \leq s_j)$, $n_j = |S_j|$ and S_j denotes the set of subjects who are observed at s_j, $j = 1, ..., m$. Thus by using the max-min formula for an isotonic regression (Barlow et al., 1972), the NPMLE of F can be written as

$$\hat{F}_n(s_j) = \max_{u \leq j} \min_{v \geq j} (\sum_{l=u}^{v} d_j / \sum_{l=u}^{v} n_j).$$

It can be shown that the above $\hat{F}_n$ is consistent. Furthermore, as $n \to \infty$ and at fixed time point t_0, $\hat{F}_n(t_0)$ has a limiting, non-normal distribution at $n^{1/3}$ or $(n \log n)^{1/3}$

convergence rate depending on if the probability of observing $T = t_0$ is zero or away from zero. Note that this is different than the usual $n^{1/2}$-covergence rate. However, the integral of $\hat{F}_n$ and its linear functionals can be shown to have asymptotic normal distribution with $n^{1/2}$-convergence (Huang et al., 1995 and Geskus, 1999). Anderson et al.(1995) utilized this property and constructed a nonparametric test procedure based on the asymptotic normality for comparing two survival functions with case I interval-censored data.

For case II interval-censored data, suppose that observed data can be represented by $\{I_i\}_{i=1}^n$, where $I_i = [L_i, R_i)$ is the interval observed to contain the unobserved survival time associated with the ith subject. If $L_i = 0$, we have a left-censored observation and if $R_i = \infty$, we have a right-censored observation. Let $\{s_j\}_{j=0}^{m+1}$ denote the unique ordered elements of $\{0, \{L_i\}_{i=1}^n, \{R_i\}_{i=1}^n, \infty\}$, α_{ij} be the indicator of the event $[s_{j-1}, s_j) \subseteq I_i$ and $p_j = F(s_j) - F(s_{j-1})$. Then the likelihood function of $p = (p_1, \ldots, p_{m+1})'$ is proportional to

$$L(p) = \prod_{i=1}^{n} \{F(R_i) - F(L_i)\} = \prod_{i=1}^{n} (\sum_{j=1}^{m+1} \alpha_{ij} p_j)$$

and the problem of finding the nonparametric maximum likelihood estimator of F becomes that of maximizing $L(p)$ with respect to p subject to $\sum_{j=1}^{m+1} p_j = 1$ and $p_j \geq 0 (j = 1, \ldots, m+1)$

To maximize $L(p)$ with respect to p, a simple and common way is to use the self-consistency algorithm proposed by Turnbull (1976). In this case, the estimator of p_j can be easily obtained by using the following equation on the p_j:

$$p_j = \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_{ij} p_j}{\sum_{l=1}^{m+1} \alpha_{il} p_l} , \qquad for\ j = 1, \ldots, r\ .$$

Note that the above self-consistency algorithm can be seen as a special case of the EM algorithm. Although it is easy to implement, it has been known to have a slow convergence rate. Alternatively, Groeneboom and Wellner (1992) developed a convex minorant algorithm, which converges faster than the self-consistency algorithm. However, both algorithms are iterative and in fact, there is no closed form for the NPMLE of F.

1.4.2 Comparisons of Survival Functions

The comparison of survival functions is a major goal of many survival studies such as clinic trials. There usually exist two general approaches for the comparison. One approach is to use semiparametric regression techniques, and the other is to use nonparametric test procedures. In the first approach, treatment indicators are included in regression models as covariates. Then certain types of tests, such as the score test, can be developed to test whether or not the corresponding regression coefficients are zero. In the second approach, distribution free procedures are developed to compare survival functions. Most such procedures use the ranks of failure times instead of the actual failure times, and they assume that censoring time distributions are the same across treatment groups.

In the case of nonparametric comparisons for right-censored data, the log-rank test (Mantel 1966), a generalization of the Savage test (1956), is the most commonly used procedure. It can be shown that the log-rank test statistic is actually the same as the score statistic from the partial likelihood under the PH model. In other words, the log-rank test is the locally most powerful test. Other nonparametric test procedures include: the weighted log-rank tests (Gehan, 1965; Breslow, 1970; Peto and Peto, 1972; Harrington and Fleming, 1982) and the procedures based on the differences between weighted Kaplan-

Meier estimates (Pepe and Fleming, 1989). For the first class, different weights can be used to adjust the sensitivity of the tests to the difference between hazard functions over time. However, the test procedures in this class could have low power if the hazard functions cross. In the second class, the tests may not be sensitive to the hazard differences because they are based on the differences of estimated survival functions. Obviously, such tests would not be efficient if the survival functions cross.

Several nonparametric test procedures have been developed to compare failure time distributions for interval-censored data. In addition to the test derived by Anderson et al. (1995) as above, Sun (1996) developed a log-rank type test, which is a counterpart of the log-rank test used for right-censored data, and Pan (2000) proposed a two-sample test using a multiple imputation approach. Petroni and Wolfe (1994) considered procedures based on the differences between the estimated survival functions. Lim and Sun (2003) investigated three classes of procedures based on the differences between the estimated survival functions, estimated hazard functions, and estimated cumulative hazard functions, respectively, using different distance measures. Most existing procedures, such as those proposed by Sun (1996) and Petroni and Wolfe (1994), can be viewed as special cases of their approaches. However, for most nonparametric test procedures, the methods are ad-hoc and the asymptotic properties of the test statistics are unknown. Also, they do not reduce to the log-rank test, the locally most powerful test, in the case of right-censored data.

1.5 Outline

The remaining part of this dissertation is organized as follows. Chapter 2 discusses

the goodness-of-fit test of the proportional odds model with interval-censored data. As mentioned in Section 1.3, the PH model is the most commonly used model for regression problems. However, this model has been found to be inappropriate for some data sets due to the fact that hazard functions from different treatment groups are not proportional, but changing with time. In contrast, the PO model has a feature that allows the ratio of two hazard functions to be monotonic and converge to one. Hence, it provides an important tool for modeling the situation above. Unfortunately, there are no methods available for checking the PO model with interval-censored data. Corresponding to this, a procedure for testing the PO model is derived and its performance is evaluated by a simulation study. In addition, the proposed test procedure is applied to a data set from an AIDS cohort study.

In Chapter 3, we consider the nonparametric comparison of two survival functions in the presence of unequal censoring. Most existing methods assume that the distribution of observation times for two samples are identical. However, there exist cases that the observation times may depend on the treatments (covariates). A comparison not accounting for differences in observation times could seriously overestimate or underestimate the treatment difference. A new test procedure is thus established. Simulation studies are conducted to compare the proposed test with two other procedures. Finally, an application from an AIDS cohort study is provided for illustration.

Chapter 4 considers semiparametric regression analysis of two-sample current status data. In practice, there exist situations when the data provide evidence of crossing hazard functions. For example, a treatment could be effective in the long run but may have certain adverse effects during the early stage. In this situation, the hazard functions may

cross. The commonly used semiparametric models mentioned above do not accommodate such a crossing phenomenon. In this chapter, we describe a two-sample semiparametric model that can accommodate crossing survival functions. The parameters in this model are two summary parameters that represent the short-term and long-term hazard ratios respectively. The model includes the proportional hazards model and the proportional odds model as special cases. Simulation studies show that the estimators perform well. In addition, a real dataset from a carcinogenicity experiment is provided for illustration purpose.

This dissertation concludes with Chapter 5, which discusses several directions for future research.

CHAPTER 2

TESTING THE PROPORTIONAL ODDS MODEL FOR INTERVAL-CENSORED DATA

2.1 Introduction

Consider a survival study that involves two independent survival variables T_1 and T_2 with continuous distributions $F_1(t)$ and $F_2(t)$. The proportional odds (PO) model postulates that

$$\frac{1 - F_2(t)}{F_2(t)} = e^{\beta} \frac{1 - F_1(t)}{F_1(t)} ,$$

where β is a constant. Define $\phi_i(t) = (1 - F_i(t))/F_i(t)$, $i = 1, 2$. Then the PO model can be rewritten as $\phi_2(t) = \alpha \, \phi_1(t)$, where $\alpha = e^{\beta}$. That is, the odds of the survival between the two samples are proportional to each other. Let $\lambda_i(t)$ denote the hazard function corresponding to $F_i(t)$. Under the PO model, we have that

$$\frac{\lambda_2(t)}{\lambda_1(t)} = \frac{1}{1 + (\alpha - 1)\{1 - F_1(t)\}} ,$$

which is a monotonic function and converges to 1 as $t \to \infty$.

The PO model is attractive in many situations. This is especially the case when the ratio of the two hazards are not proportional, but changing with time. One of such example is that treatment effect diminishes along with time. Many authors have discussed inference about the PO model (Dabrowska and Doksum, 1988; Huang and Rossini, 1997; Rossini and Tsiatis, 1996; Murphy, Rosssini and van der Vaart, 1997; Shen, 1998). In particular,

Dauxois and Kirmani (2003) developed a procedure for testing the PO model for two sample right-censored failure time data. For inference based on interval-censored failure time data, Huang and Rossini (1997) and Rossini and Tsiatis (1996) proposed some sieve estimation approaches. Huang and Wellner (1997) and Rabinowitz et al. (2000) considered the same problem and studied the full likelihood approach and an approximate conditional likelihood approach, respectively.

Let $w_1(t)$ and $w_2(t)$ be two positive known weight functions such that the ratio w_1/w_2 is an increasing function. Define $\theta_{ij} = \int_{\tau_1}^{\tau_2} w_i(t)\,\phi_j(t)\,dt$, where τ_1 and τ_2 are prespecified constants such that $\tau_1 < \tau_2$ with $F_i(\tau_1) > 0$ and $F_i(\tau_2) < 1$, $i, j = 1, 2$. To test the PO model or the hypothesis $H_0 : \phi_2(t) = \alpha\,\phi_1(t)$ for all $t > 0$ and some $\alpha > 0$ against $H_1 :$ $\phi_2(t)$ and $\phi_1(t)$ are not proportional, Dauxois and Kirmani (2003) proposed to use the statistic

$$q(w_1, w_2) = \theta_{11}\,\theta_{22} - \theta_{12}\,\theta_{21} \tag{2.1}$$

with replacing $1 - F_i(t)$ by their Kaplan-Meier estimators. In the following, we generalize the above test procedure to the interval-censored failure time data situation.

Interval-censored data have become common as described in Chapter 1. However, there does not seem to exist a procedure to test the PO model for interval-censored data. Note that in this case, the Kaplan-Meier estimator does not exist anymore and also due to the significant difference between right-censoring and interval-censoring, the theory developed in Dauxois and Kirmani (2003) cannot be directly generalized to interval-censored data. For example, the Kaplan-Meier estimator has a $\sqrt{n}$ convergence rate, but the nonparametric maximum likelihood estimator of a survival function for interval-censored data may only have a $n^{1/3}$ convergence rate (Groeneboom, 1996). Fortunately,

Geskus and Groeneboom (1999) showed that under some conditions, the linear functional of the nonparametric maximum likelihood estimator from interval-censored data still has the usual $\sqrt{n}$ convergence rate.

In the following, we first discuss in Section 2.2 the generalization of the test procedure given in Dauxois and Kirmani (2003) to two sample interval-censored data situations and the related asymptotic theory is established. Note that although the idea behind the generalized test is straightforward, the implementation and the derivation of asymptotic properties of the test are not trivial due to complex structure of interval-censored data. Section 2.3 considers the testing of the hypothesis H_0 against H_1 using the theory given in Section 2.2 and two implementation procedures are presented. In Section 2.4, the test procedure given in the previous sections is generalized to situations where there exists a categorical covariate or K different populations with $K \geq 2$. Simulation results for assessing the performance of the proposed method are reported in Section 2.5 and Section 2.6 applies the method to a set of interval-censored data arising from an AIDS study. The chapter concludes with some remarks in Section 2.7.

2.2 Asymptotic theory for two- sample interval-censored data

In this section, we first consider situations where only two sample interval-censored data are available for T_1 and T_2 defined above. By this, we mean that T_i is not observable except for knowing that it belongs to some interval given by

$$\{U_i, V_i, \Delta_{i1} = I(T_i \leq U_i), \Delta_{i2} = I(U_i < T_i \leq V_i)\},$$

where $U_i \leq V_i$ are random monitoring times for T_i and $I(\cdot)$ is the indicator function,

$i = 1, 2$. In the following, we assume that T_i is independent of (U_i, V_i) and the observed data are

$$\{U_{ij}, V_{ij}, \Delta_{i1}^{(j)}, \Delta_{i2}^{(j)}\ , i = 1, 2, j = 1, ..., n_i\}\,,$$

where $\{U_{ij}, V_{ij}, \Delta_{i1}^{(j)}, \Delta_{i2}^{(j)}\}$ are i.i.d. replicates of $(U_i, V_i, \Delta_{i1}, \Delta_{i2})$.

Now consider the testing of the hypothesis H_0. Let $\hat{F}_i(t)$ denote the nonparametric maximum likelihood estimator of $F_i(t)$ based on the interval-censored data

$$\{U_{ij}, V_{ij}, \Delta_{i1}^{(j)}, \Delta_{i2}^{(j)}\,; j = 1, ..., n_i\}$$

and define

$$\hat{\phi}_i(t) = \frac{1 - \hat{F}_i(t)}{\hat{F}_i(t)}\ , \ i = 1, 2.$$

Motivated by the statistic given in equation (2.1), we propose to base the test on the statistic

$$Q_n(w_1, w_2) \ = \ \left(\frac{n_1 n_2}{n}\right)^{1/2} \ \left[\hat{\theta}_{11}\hat{\theta}_{22} - \hat{\theta}_{12}\hat{\theta}_{21}\right]\ , \tag{2.2}$$

where $n \ = \ n_1 \ + \ n_2$ and

$$\hat{\theta}_{ru} \ = \ \int_{\tau_1}^{\tau_2} w_r(t)\hat{\phi}_u(t)dt\ , \ r = 1, 2, u = 1, 2\,.$$

It is easy to see that if the hypothesis H_0 is true, $Q_n(w_1, w_2)$ should be close to zero. Thus H_0 should be rejected in favor of the hypothesis H_1 if $\mid Q_n(w_1, w_2)\mid$ is too large.

To employ the statistic Q_n, we need to establish its asymptotic distribution under the hypothesis H_0. For this end, let $G_i(u, v, \delta_{i1}, \delta_{i2})$ denote the distribution function of $(U_i, V_i, \Delta_{i1}, \Delta_{i2})$ and $h_i(u, v)$ the density function of (U_i, V_i) with the marginal density functions h_{i1} and h_{i2} for U_i and V_i, respectively. Define

$$C_1(t) = \frac{\theta_{12}w_2(t) - \theta_{22}w_1(t)}{F_1(t)^2}$$

and

$$C_2(t) = \frac{\theta_{21} w_1(t) - \theta_{11} w_2(t)}{F_2(t)^2}\,.$$

Also let Ψ_{F_i} denote the solution to the following Fredholm integral equation

$$\Psi_{F_i}(t) = d_i(t)\left\{ C_i(t) - \int_{\tau_1}^{\tau_2} \frac{\Psi_{F_i}(t) - \Psi_{F_i}(s)}{|F_i(t) - F_i(s)|} h_i^*(t,s) ds \right\},\ i = 1, 2, \tag{2.3}$$

where $d_i(t) = F_i(t)(1-F_i(t))/[h_{i1}(t)(1-F_i(t)+h_{i2}(t)F_i(t)]$ and $h_i^*(t,s) = h_i(t,s)+h_i(s,t)$. Also define

$$\Phi_i(u, v, \delta_{i1}, \delta_{i2}) \;=\; -\delta_{i1}\frac{\Psi_{F_i}(u)}{F_i(u)} - \delta_{i2}\frac{\Psi_{F_i}(v) - \Psi_{F_i}(u)}{F_i(v) - F_i(u)} + (1 - \delta_{i1} - \delta_{i2})\frac{\Psi_{F_i}(v)}{1 - F_i(v)}\,.$$

Assume that the regularity conditions (A)-(D) of Fang, Sun and Lee (2002) hold about the random monitoring times (U_i, V_i) $(i = 1, 2)$. Then the asymptotic normality of $Q_n(w_1, w_2)$ is given in the following theorem.

Theorem 2.1. *Assume that the weight functions $w_i(t)$ $(i = 1, 2)$ have bounded derivatives on $[\tau_1, \tau_2]$ and $n_1/n \to \rho$ $(0 < \rho < 1)$ as $n \to \infty$. Then under the above conditions and H_0, $Q_n(w_1, w_2)$ has an asymptotic normal distribution with mean zero and variance*

$$\sigma^2 = (1-\rho)\int \Phi_1^2(u, v, \delta_{11}, \delta_{12}) dG_1(u, v, \delta_{11}, \delta_{12}) + \rho \int \Phi_2^2(u, v, \delta_{21}, \delta_{22}) dG_2(u, v, \delta_{21}, \delta_{22}). \tag{2.4}$$

The proof of this theorem is sketched in the Appendix. In the next section, we describe the use of the results given above for testing H_0.

2.3 Two-sample test procedure

To test the hypothesis H_0 by using the statistic Q_n, we present two implementation approaches based on the above theorem. One is to directly apply Q_n by deriving a consistent estimate of the asymptotic variance σ^2 and the other is to employ a simple bootstrap procedure.

First we consider estimation of σ^2. For this, note that $\hat{F}_i$ $(i = 1, 2)$ only has mass at observation times and according to Theorem 3.5 of Groeneboom (1996), $\Psi_{\hat{F}_i}$ is absolutely continuous with respect to $\hat{F}_i$ and a step function with jumps at the time points where $\hat{F}_i$ jumps. Let $0 < t_1^{(i)} < ... < t_{m_i}^{(i)} < \infty$ denote the time points at which $\hat{F}_i$ has jumps and $z_j^{(i)} = \hat{F}_i(t_j^{(i)})$, $i = 1, 2$, $j = 1, ..., m_i$. Also let $\hat{H}_i$, $\hat{H}_{i1}$ and $\hat{H}_{i2}$ denote the empirical distributions of (U_i, V_i), U_i and V_i, respectively. Define

$$\Delta_j(h_{il}) = \int_{t_j^{(i)}}^{t_{j+1}^{(i)}} h_{il}(t)dt \approx \int_{t_j^{(i)}}^{t_{j+1}^{(i)}} d\hat{H}_{il}(t)\,,\ l = 1, 2,$$

$$\Delta_{jk}(h_i) = \int_{u=t_j^{(i)}}^{t_{j+1}^{(i)}} \int_{v=t_k^{(i)}}^{t_{k+1}^{(i)}} h_i(u, v)dudv \approx \int_{u=t_j^{(i)}}^{t_{j+1}^{(i)}} \int_{v=t_k^{(i)}}^{t_{k+1}^{(i)}} d\hat{H}_i(u, v),$$

$$d_j^{(i)} = \frac{z_j^{(i)}(1 - z_j^{(i)})}{\Delta_j(h_{i1})(1 - z_j^{(i)}) + \Delta_j(h_{i2})z_j^{(i)}},$$

$$\Delta_j(C_i) = \int_{t_j^{(i)}}^{t_{j+1}^{(i)}} dC_i(t)dt \approx \int_{t_j^{(i)}}^{t_{j+1}^{(i)}} d\hat{C}_i(t)$$

$j, k = 1, ..., m_i$, $i = 1, 2$, where

$$\hat{C}_1(t) = \frac{\hat{\theta}_{12}w_2(t) - \hat{\theta}_{22}w_1(t)}{\hat{F}_1(t)^2}\ ,\ \hat{C}_2(t) = \frac{\hat{\theta}_{21}w_1(t) - \hat{\theta}_{11}w_2(t)}{\hat{F}_2(t)^2}\,.$$

Let $y_j^{(i)} = \Psi_{\hat{F}_i}(t_j^{(i)})$. Then it can be shown that the vector $y^{(i)} = (y_1^{(i)}, ..., y_{m_i}^{(i)})'$ $(i = 1, 2)$ is the unique solution to the following set of linear equations

$$y_j^{(i)} \left\{ (d_j^{(i)})^{-1} + \sum_{k<j} \frac{\Delta_{kj}(h_i)}{z_j^{(i)} - z_k^{(i)}} + \sum_{k>j} \frac{\Delta_{jk}(h_i)}{z_k^{(i)} - z_j^{(i)}} \right\} = \Delta_j(C_i) + \sum_{k<j} \frac{\Delta_{kj}(h_i)}{z_j^{(i)} - z_k^{(i)}} y_k^{(i)} + \sum_{k>j} \frac{\Delta_{jk}(h_i)}{z_k^{(i)} - z_j^{(i)}} y_k^{(i)}$$

for $j = 1, ..., m_i$ (Theorem 3.1 of Geskus and Groeneboom, 1999).

For each i, define

$$\hat{\Phi}_i(u,v,\delta_{i1},\delta_{i2}) = -\delta_{i1}\frac{\Psi_{\hat{F}_i}(u)}{\hat{F}_i(u)} - \delta_{i2}\frac{\Psi_{\hat{F}_i}(v)-\Psi_{\hat{F}_i}(u)}{\hat{F}_i(v)-\hat{F}_i(u)} + (1-\delta_{i1}-\delta_{i2})\frac{\Psi_{\hat{F}_i}(v)}{1-\hat{F}_i(v)}\,.$$

It follows from the uniform consistency of $\hat{H}_i$, $\hat{H}_{i1}$, $\hat{H}_{i2}$ and $\hat{F}_i$ that $\hat{\Phi}_i(u,v,\delta_{i1},\delta_{i2})$ is a uniformly consistent estimator of $\Phi_i(u,v,\delta_{i1},\delta_{i2})$. This naturally yields a consistent estimator of σ^2 given by

$$\hat{\sigma}^2 = \frac{n_2}{n}\int \hat{\Phi}_1^2(u,v,\delta_{11},\delta_{12})d\hat{G}_1(u,v,\delta_{11},\delta_{12}) + \frac{n_1}{n}\int \hat{\Phi}_2^2(u,v,\delta_{21},\delta_{22})d\hat{G}_2(u,v,\delta_{21},\delta_{22})\,, \tag{2.5}$$

where $\hat{G}_i$ is the empirical estimator of G_i. Hence the test of the hypothesis H_0 can be carried out by using the statistic $Q_n(w_1,w_2)/\hat{\sigma}$ based on the standard normal distribution.

Note that the above estimator $\hat{\sigma}^2$ is very technically involved due to the complexity of the estimator $\Psi_{\hat{F}_i}$. Thus the above procedure could be complicated and demanding in computation, especially when the number of jumps of $\hat{F}_i$ is not small. For this, we suggest the following simple bootstrap procedure.

Let M be a prespecified integer and $Q_n^{(0)}$ denote the observed value of the test statistic Q_n. For each i $(=1,2)$ and l $(1 \le l \le M)$, draw a simple random sample

$$D_i^{(l)} = \{\, U_{ij}^{(l)}, V_{ij}^{(l)}, \Delta_{i1}^{(jl)}, \Delta_{i2}^{(jl)},\ j = 1,...,n_i \,\}$$

with replacement from the observed data on T_i. Let $Q_n^{(l)}$ denote the value of statistic Q_n calculated based on the generated data set $\{\, D_1^{(l)}, D_2^{(l)} \,\}$. It follows from the theorem given in the previous section that under H_0 and when n is large, the bootstrap samples $\{\, Q_n^{(l)}\,;\, l = 1,...,M \,\}$ follow a normal distribution. The variance of $Q_n^{(0)}$ can then be estimated by the sample variance, say $\hat{\sigma}_b^2$, of the $Q_n^{(l)}$'s and the hypothesis H_0 can be tested by using the statistic $Q_n^* = Q_n^{(0)}/\hat{\sigma}_b$ based on the standard normal distribution.

Similar bootstrap procedures have been used by Fang, Sun and Lee (2002) and Monaco, Cai and Grizzle (2005) among others.

To implement the above test procedure, one needs to determine the maximum likelihood estimator $\hat{F}_i$ of F_i. For this, several procedures are available (Gentleman and Geyer, 1994; Sun, 2004) and perhaps the simplest procedure, which will be used below, is the self-consistency algorithm given in Turnbull (1976). Also one needs to choose τ_1 and τ_2 and it is easy to see that to test H_0 against H_1 over all possible t, one should select them to make the interval $[\tau_1, \tau_2]$ as large as possible. Dauxois and Kirmani (2003) suggested to choose them such that

$$0 < \hat{F}_i(\tau_1) < 10^{-3} \ , \ 0 < 1 - \hat{F}_i(\tau_2) < 10^{-3} \, ,$$

$i = 1, 2$. Another choice that one has to make is the selection of weight functions w_1 and w_2 and different weight functions give different test statistics. It is apparent that these weight functions set up the measurement scales for the null hypothesis. If the null hypothesis is true, the test statistic Q_n should be close to zero no matter what scales are used and otherwise, Q_n is away from zero. For the following numerical studies, we consider several choices including the natural and simple functions $w_1(t) = 1$ and $w_2(t) = 1/(1+t)$.

It should be noted that in practice, interval-censored data may be given in the form $\{(L_{ij}, R_{ij}] \, , \, i = 1, 2, j = 1, ..., n_i\}$, where $(L_{ij}, R_{ij}]$ is the interval within which the failure time of the jth subject from the ith group is observed to occur. This form is commonly used in practice, while the form used above is more convenient and usually used for the situation where the asymptotic property of an approach for interval-censored failure time data is of interest. There is no difference between the two forms in terms of implementation of the test procedure proposed here and other inference procedures (Huang and Wellner,

1997).

2.4 K- sample test procedure

Now we consider situations where study subjects come from K different populations. Let T_{ij} denote the survival variable of interest from subject j in population i with the cumulative distribution $F_i(t)$, $j = 1, ..., n_i$, $i = 1, ..., K$. As before, define

$$\phi_i(t) = \frac{1 - F_i(t)}{F_i(t)} \ , \ \ i = 1, ..., K$$

and suppose that one is interested in testing the null hypothesis H_0' versus the alternative hypothesis H_1', where

$H_0' : \ \phi_i(t) = \alpha_i \, \phi_1(t)$ for all $t > 0$ and some constants $\alpha_i > 0, \ i = 2, ..., K$,

H_1' : some $\phi_i(t)$ and $\phi_1(t)$ are not proportional.

Furthermore, suppose that for the T_{ij}'s, only interval-censored data are available and have the form

$$\{ \, U_{ij}, V_{ij}, \Delta_{i1}^{(j)}, \Delta_{i2}^{(j)} \, ; \, j = 1, ..., n_i, i = 1, ..., K \, \}$$

as before, where $\Delta_{i1}^{(j)} = I(T_{ij} \leq U_{ij})$, $\Delta_{i2}^{(j)} = I(U_{ij} < T_{ij} \leq V_{ij})$. In the following, it is assumed that T_{ij} is independent of (U_{ij}, V_{ij}).

As before, let $\hat{F}_i(t)$ denote the nonparametric maximum likelihood estimator of $F_i(t)$ based on the interval-censored data $\{ \, U_{ij}, V_{ij}, \Delta_{i1}^{(j)}, \Delta_{i2}^{(j)} \, ; \, j = 1, ..., n_i \, \}$ and define

$$\hat{\phi}_i(t) = \frac{1 - \hat{F}_i(t)}{\hat{F}_i(t)} \ , \ \ \hat{\theta}_{rk} = \int_{\tau_1}^{\tau_2} w_r(t) \, \hat{\phi}_k(t) \, dt \, ,$$

$i, \, r, \, k = 1, ..., K$. In the above, τ_1 and τ_2 are defined as in the previous sections and the $w_r(t)$'s are some positive known weight functions such that w_1/w_r is an increasing

function. To test H_0' versus H_1', as in Section 2.2, we propose to use the statistic $\boldsymbol{Q}_n(\boldsymbol{w}) = (Q_n^{(2)}(w_1, w_2), ..., Q_n^{(K)}(w_1, w_K))'$, where $\boldsymbol{w} = (w_1, ..., w_K)'$ and

$$Q_n^{(i)}(w_1, w_i) = \left(\frac{n_1 n_i}{n}\right)^{1/2} \left(\hat{\theta}_{11}\hat{\theta}_{ii} - \hat{\theta}_{1i}\hat{\theta}_{i1}\right)$$

with $n = \sum_{i=1}^{K} n_i$. It is apparent that if H_0' is true, all $Q_n^{(i)}(w_1, w_i)$ should be close to zero.

For the null asymptotic distribution of $\boldsymbol{Q}_n(\boldsymbol{w})$, following the notation used in the previous sections, let $G_i(u, v, \delta_{i1}, \delta_{i2})$ denote the distribution function of $(U_{i1}, V_{i1}, \Delta_{i1}^{(1)}, \Delta_{i2}^{(1)})$ and $h_i(u, v)$ the density function of (U_{i1}, V_{i1}) with the marginal density functions h_{i1} and h_{i2} for U_{i1} and V_{i1}, respectively, $i = 1, ..., K$. Define

$$C_{i1}(t) = \frac{\theta_{1i} w_i(t) - \theta_{ii} w_1(t)}{F_1(t)^2} \;, \quad C_{i2}(t) = \frac{\theta_{i1} w_1(t) - \theta_{11} w_i(t)}{F_i(t)^2} \,.$$

Also let Ψ_{i1} and Ψ_{i2} denote the solutions to the Fredholm integral equations

$$\Psi_{i1}(t) = d_1(t) \left\{ C_{i1}(t) - \int_{\tau_1}^{\tau_2} \frac{\Psi_{i1}(t) - \Psi_{i1}(s)}{|F_1(t) - F_1(s)|} h_1^*(t, s) ds \right\}$$

and

$$\Psi_{i2}(t) = d_i(t) \left\{ C_{i2}(t) - \int_{\tau_1}^{\tau_2} \frac{\Psi_{i2}(t) - \Psi_{i2}(s)}{|F_i(t) - F_i(s)|} h_i^*(t, s) ds \right\} ,$$

respectively, where $h_i^*(t, s) = h_i(t, s) + h_i(s, t)$ and

$$d_i(t) = \frac{F_i(t)(1 - F_i(t))}{h_{i1}(t)(1 - F_i(t)) + h_{i2}(t) F_i(t)} \,.$$

Also define

$$\Phi_{i1}(u, v, \delta_{11}, \delta_{12}) = -\delta_{11} \frac{\Psi_{i1}(u)}{F_1(u)} - \delta_{12} \frac{\Psi_{i1}(v) - \Psi_{i1}(u)}{F_1(v) - F_1(u)} + (1 - \delta_{11} - \delta_{12}) \frac{\Psi_{i1}(v)}{1 - F_1(v)}$$

and

$$\Phi_{i2}(u, v, \delta_{i1}, \delta_{i2}) = -\delta_{i1} \frac{\Psi_{i2}(u)}{F_i(u)} - \delta_{i2} \frac{\Psi_{i2}(v) - \Psi_{i2}(u)}{F_i(v) - F_i(u)} + (1 - \delta_{i1} - \delta_{i2}) \frac{\Psi_{i2}(v)}{1 - F_i(v)} \,,$$

and assume that the regularity conditions (A)-(D) of Fang, Sun and Lee (2002) hold about the random monitoring times (U_{ij}, V_{ij}) $(i = 1, ..., K, j = 1, ..., n_i)$. Then one can generalize the theorem 2.1 as follows.

Theorem 2.2 *Assume that the weight functions* $w_i(t)$ $(i = 1, ..., K)$ *have bounded derivatives on* $[\tau_1, \tau_2]$ *and* $n_i/n \to \rho_i$ $(0 < \rho_i < 1)$ *as* $n \to \infty$. *Then under the above conditions and* H_0, $\boldsymbol{Q}_n(\boldsymbol{w})$ *converges in distribution to a normal random vector with mean zero and a covariance matrix* $\Sigma = (\sigma_{ij})$, $i, j = 2, ..., K$, *where*

$$\sigma_{ii} = \rho_i \int \Phi_{i1}^2(u, v, \delta_{11}, \delta_{12}) dG_1(u, v, \delta_{11}, \delta_{12}) + \rho_1 \int \Phi_{i2}^2(u, v, \delta_{i1}, \delta_{i2}) dG_i(u, v, \delta_{i1}, \delta_{i2}),$$

and for $i \neq j$,

$$\sigma_{ij} = (\rho_i \rho_j)^{1/2} \int \Phi_{i1}^2(u, v, \delta_{11}, \delta_{12}) dG_1(u, v, \delta_{11}, \delta_{12}).$$

The proof of the theorem given above is similar to that of the theorem 2.1 and thus omitted. Based on this theorem, one can carry out the test of H_0' using the statistic $\boldsymbol{Q}(\boldsymbol{w})\, \hat{\Sigma}^{-1}\, \boldsymbol{Q}'(\boldsymbol{w})$, where $\hat{\Sigma}$ is a consistent estimator of Σ. For the implementation, as in Section 3, one can easily develop a simple bootstrap procedure similar to that described for the two sample situation.

2.5 Numerical Studies

This section reports some results obtained from simulation studies conducted for assessing the performance of the proposed approach for testing the PO model. In the study, we focused on the two sample situation and to generate T_1 and T_2, the log-logistic

distributions $F_1(t) = F(1, \beta_1)$ and $F_2(t) = F(\alpha, \beta_2)$ were used, where

$$F(\alpha, \beta) = \frac{(t/\alpha)^\beta}{1 + (t/\alpha)^\beta}$$

and α and β are constants. This gives $\phi_2(t)/\phi_1(t) = \alpha^{\beta_2} t^{\beta_1 - \beta_2}$ and H_0 and H_1 correspond to $\beta_1 = \beta_2$ and $\beta_1 \neq \beta_2$, respectively.

For censoring intervals, we mimicked interval-censored data commonly arising from periodic follow-up studies and first generated a right-censoring time C_{ij} from the uniform distribution $U(0, A)$, where A is a positive constant chosen to control the percentage of right-censored observations. Given C_{ij}, if T_{ij}, the above generated failure time for the jth subject from group i, is greater than C_{ij}, we defined $L_{ij} = C_{ij}$ and $R_{ij} = \infty$. That is, T_{ij} was right-censored. If $T_{ij} \leq C_{ij}$, we defined $L_{ij} = \max(0, T_{ij} - a_1)$ and $R_{ij} = \min(T_{ij} + a_2, C_{ij})$, where a_1 and a_2 are random numbers generated independently from the uniform distribution $U(0, B)$. Here B is a positive constant controlling the length of censoring intervals. For the variance estimation of Q_n, the simple bootstrap procedure was used and in the study, we took τ_1 and τ_2 to be the smallest and largest possible values.

Table 1 presents the estimated size and power of the test procedure based on Q_n for testing H_0 at the significance level of 5% based on 1000 replications, $M = 500$, $n_1 = 100$, $n_2 = 150$, $w_1(t) = 1$ and $w_2(t) = 1/(1 + t)$. Here we considered the situations with $\alpha = 2$, $\beta_1 = 1$, 1.5, 2, or 3, $\beta_2 = 1$, and the percentage of right-censored observations being 10%, 20% or 30%. The top half of the table is for the case where $B = 0.5$ and the bottom half is for the case where $B = 1$. The results suggest that the test procedure seems to have right size and reasonable power. As expected, the power decreases when the length of censoring intervals increases.

To investigate the dependence of the power of the proposed test procedure on the

sample size and weight function, we also performed simulations using different sample sizes and weight functions. For example, under the same set-up as in Table 1 but with $n_1 = 200$ and $n_2 = 300$, we obtained powers of 0.610, 0.467 and 0.403 for the situations with $\beta_1 = 1.5$, $B = 1$, and 10%, 20% and 30% right-censorings, respectively. For the exact same situation but with $n_1 = 100$, $n_2 = 150$, $w_1(t) = t$ and $w_2(t) = 1/(1+t)$, the test gave powers of 0.398, 0.284 and 0.231, respectively. These results indicate that as expected, the power of the test procedure increases as the sample size increases and could depend on the selection of weight functions.

To evaluate the normal approximation given in Theorem, we studied the normal quantile plots of the standardized test statistics. Figure 1 displays such plots for $B = 0.5$ and 1 with $\alpha = 2$, $\beta_1 = \beta_2 = 1$ under 10%, 20% and 30% right censoring percentages, respectively. They suggested that the normal approximation works well.

2.6 An Application

To illustrate the proposed methodology, consider the data presented in Tables 2 & 3, which are reproduced from DeGruttola and Lagakos (1989). The data arose from a cohort study on hemophiliacs that consists of 262 persons with hemophilia treated since 1978. All patients were at the risk of being infected by HIV due to contaminated blood that they received for their hemophilia. By the end of study, 197 subjects were confirmed to be infected with HIV and among these infected subjects, 25 were found infected at their first tests for the infection. Since the determination of HIV infection was based on periodic blood test results, only interval-censored data were obtained for the infection times. One objective of the study was to investigate the relationship between their HIV infection rate

and the amount of blood that they received.

For this, in the original study, the patients were classified into two groups as the lightly and heavily treated groups. In the former group (157), the patients received less than 1000 $\mu g/kg$ in each year and in the other group (105), the patients received at least 1000 $\mu g/kg$ of blood for at least one year between 1982 and 1985. In the study, the observed time intervals for the HIV infection were measured in 6-month intervals.

Define T_1 and T_2 to be the times to HIV infection for the patients in the lightly and heavily treated groups, respectively. To examine the appropriateness of the PO model for the data set and the relationship of the distribution functions of T_1 and T_2, we applied the test procedure proposed in the previous sections and obtained $Q_n^{(0)} = 9.5547$ with the estimated standard error being 9.3053. In the procedure, we used $w_1(t) = 1$ and $w_2(t) = 1/(1+t)$. This yielded a p-value of 0.305 for testing H_0 against H_1 and suggests that the PO model seems to be appropriate for the data. By using $w_1(t) = t$ and $w_2(t) = 1/(1+t)$, we obtained a p-value of 0.384 and the same conclusion. For the results, we took $\tau_1 = 6$ and $\tau_2 = 17$, the smallest and largest possible time points for T_2.

To further investigate the fit of the PO model to the problem, we obtained the estimators of the separate log odds ratio functions, $\log \hat{\phi}_i(t)$, corresponding to the two groups and they are presented in Figure 2. Note that if H_0 is true, the two curves should be roughly parallel to each other. Figure 2 again suggests that the PO model seems to fit the data well.

2.7 Concluding Remarks

This chapter proposed a goodness-of-fit test procedure for the PO model for interval-

censored failure time data. The analysis of interval-censored data has recently attracted a lot of attention and several models including the PO model have been investigated for their regression analysis. However, there seems to exist little research in the literature on the development of formal approaches that can be used for model checking. One reason is that the censoring mechanism involved in interval-censored data is much more difficult to deal with than that in right-censored data in addition to less information given by the interval-censored data. This can be seen from the problem considered here. The simulation results suggested that the procedure given here seems to perform reasonably well for practical situations.

Although the focus here is on K sample situations, the test procedure proposed in the previous sections can be applied to situations with categorical covariates. However, it does not seem to be straightforward to generalize the idea used here to continuous covariate situations, for which some different test procedures need to be developed for testing the PO model. Another important question that was not fully discussed in the previous sections is the selection of optimal weight functions for a given situation. As usual, this is a very difficult question (Sengupta, Bhattacharjee and Rajeev, 1998) and the existence of interval-censoring makes it even more challenging. Of course, one may first need to ask the existence of such weight functions, for which we have no clear answer. One other direction for future research is the asymptotic validity of the simple bootstrap procedure described in Section 2.3. Note that as mentioned before, several authors used similar procedures (Fang, Sun and Lee, 2002; Monaco, Cai and Grizzle, 2005), but no theoretical justification was given. Although the simulation study indicates that it works well, it would be helpful and desirable to provide some justifications.

CHAPTER 3

A NONPARAMETRIC TEST FOR INTERVAL-CENSORED DATA WITH UNEQUAL CENSORING

3.1 Introduction

As discussed in Chapter 1, one of the primary objectives in clinical trials and epidemiological studies is to compare survival functions. In this case, one usually prefers to apply nonparametric methods due to the lack of knowledge about the underlying distributions of the failure time of interest. In this chapter, we consider such nonparametric comparison problems when only interval-censored failure time data are available. For survival comparison based on interval-censored data, a few test procedures have been proposed (Finkelstein, 1986; Self and Grossman, 1986; Fay, 1996; Pan, 2000; Petroni and Wolfe, 1994; Zhang et al., 2001, 2003; Zhao and Sun, 2004). However, most of them assume that censoring intervals or observation times for all subjects have the same distribution function, which obviously may not be true in practice. A failure to take into account this difference in distributions could seriously overestimate or underestimate the treatment difference. One exception is given by Sun (1999), who considered survival comparison based on case I interval-censored data when the distributions of observation times differ among different treatment groups.

In the following, we discuss the same problem as that in Sun (1999) for case II interval-censored data. Specifically, we consider the two-sample survival comparison problem and a

class of test statistics is presented in Section 3.2 that allow the distributions of observation times to be different between two treatment groups. The statistics are constructed based on linear functionals of estimated survival functions and are generalizations of those used in Zhang et al. (2001). The asymptotic normality of the test statistic is established. Monte Carlo simulation studies are performed to evaluate the finite sample properties of the proposed approach in Section 3.3 and Section 3.4 applies it to an AIDS cohort study. Some concluding remarks are given in Section 3.5.

3.2 Statistical Methods

Consider a survival study that consists of n independent subjects randomly assigned to one of two treatments. For subject i, let T_i denote the failure time of interest and assume that only an interval-censored observation on it is available. Specifically, suppose that the observed information includes two random variables U_i and V_i with $U_i \leq V_i$ and the indicator variables $\delta_{1\,i} = I(T_i \leq U_i)$, $\delta_{2\,i} = I(U_i < T_i \leq V_i)$ and $\delta_{3\,i} = 1 - \delta_{1i} - \delta_{2i}$, where I is the indicator function. It will be assumed that U_i and V_i are independent of T_i. The variables $\delta_{1\,i}$, $\delta_{2\,i}$ and $\delta_{3\,i}$ indicate whether the survival event of interest for subject i has occurred before U_i, within the interval $(U_i, V_i]$, or after V_i. We assume that the failure time and the observation times are independent.

Define $N_i(t) = I(T_i \leq t)$, a counting process indicating if the survival event of interest has occurred by time t, and let z_i be 0-1 treatment indicator, $i = 1, ..., n$. Also let $F_l(t)$ denote the failure time distribution function for subjects with $z_i = l$, $l = 0, 1$. Then the observed data consist of $\{\, (U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, z_i)\,;\, i = 1, ..., n\,\}$ or $\{(U_i, V_i, N_i(U_i), N_i(V_i), z_i)\,;\ \ i = 1, ..., n\,\}$ and the goal is to test the hypothesis H_0 :

$F_0(t) = F_1(t)$.

To construct a test statistic for H_0, let $H_1^{(l)}(u)$, $H_2^{(l)}(v)$ and $H^{(l)}(u,v)$ denote marginal and joint distribution functions of the U_i's and V_i's for subjects with $z_i = l$, respectively, $l = 0, 1$. Assume that the support of F_0 and F_1 is given by a finite interval $[0, \tau]$. Motivated by the weighted Kaplan-Meier test statistics for right-censored data (Fleming and Harrington, 1991) and the statistics given in Zhang et al. (2001), we consider the functional

$$g(F) = \int\int_{0 \le u \le v \le \tau} \{F(u)\eta(u) + F(v)\eta(v)\}\, dH^{(1)}(u,v)\,, \tag{3.1}$$

where $\eta(u)$ is a known bounded weight function. Let $\hat{F}_0$ and $\hat{F}_1$ denote the estimates of F_0 and F_1, respectively. Then a natural test statistic for H_0 is given by

$$Q = n^{1/2}\left\{g(\hat{F}_0) - g(\hat{F}_1)\right\}$$

for given $\eta(u)$. It is apparent that under H_0, Q should be around zero.

For estimation of F_0 and F_1, note that we can divide the observed data into two sets of current status data given below:

$$\{(U_i, N_i(U_i), z_i)\,;\ i = 1, ..., n\}\ ,\ \{(V_i, N_i(V_i), z_i)\,;\ i = 1, ..., n\}\,.$$

One way to estimate F_0 and F_1 is to combine these two data sets together, but treat them as independent samples. Then we have a single larger set of current status data and can define $\hat{F}_l$ to be the maximum likelihood estimator based on this larger data set from subjects with $z_i = l$, $l = 0, 1$. The same idea was used by Zhang et al. (2001) among others and one advantage of this approach is that $\hat{F}_0$ and $\hat{F}_1$ have closed forms. More comments on this are given below.

To test H_0 using statistic Q, we need to derive the null asymptotic distribution of Q. To this end, let $h_1^{(l)}(u)$ and $h_2^{(l)}(v)$ denote the marginal density functions of the $U_i's$ and $V_i's$ for subjects with $z_i = l$, respectively, $l = 0, 1$. It will be assumed that these functions are positive and satisfy

$$\frac{h_1^{(1)}(\cdot)}{h_1^{(0)}(\cdot)} = \frac{h_2^{(1)}(\cdot)}{h_2^{(0)}(\cdot)} = R(\cdot)\,. \tag{3.2}$$

Let $\xi = \eta \cdot R$. Then we have the following result.

Theorem 3.1. *Suppose that $\eta \circ F^{-1}$ and $\xi \circ F^{-1}$ are bounded Lipschitz functions and $n_0/n \to p$ $(0 < p < 1)$ as $n \to \infty$, where $n_0 = \sum_{i=1}^{n}(1 - z_i)$. Then under H_0 and $n \to \infty$,*

$$Q \to N(0\,,\, \frac{A_0}{p} + \frac{A_1}{1-p})$$

in distribution, where

$$A_0 = \int_0^\tau F_0(u)(1 - F_0(u))\xi^2(u)dH_1^{(0)}(u) + \int_0^\tau F_0(v)(1 - F_0(v))\xi^2(v)dH_2^{(0)}(v)$$
$$+2\int\int_{0\le u\le v\le\tau} F_0(u)(1 - F_0(v))\xi(u)\xi(v)dH^{(0)}(u, v)$$

and

$$A_1 = \int_0^\tau F_1(u)(1 - F_1(u))\eta^2(u)dH_1^{(1)}(u) + \int_0^\tau F_1(v)(1 - F_1(v))\eta^2(v)dH_2^{(1)}(v)$$
$$+2\int\int_{0\le u\le v\le\tau} F_1(u)(1 - F_1(v))\eta(u)\eta(v)dH^{(1)}(u, v)\,.$$

The proof of the above theorem is sketched in the Appendix. Condition (3.2) means that the ratio of the density functions between the first observation times across the two groups is the same as that between the second observation times across the two groups. In other words, the mechanism or reasons behind the difference between the observation

times in the two groups are the same for the first and second observation times, which is the case for many medical studies. One situation in which condition (3.2) holds is of course when $H^{(0)}(u,v) = H^{(1)}(u,v)$. That is, the observation times have the same distribution and in this case, $R(u) = 1$. In the following, we will consider situations where $H^{(0)}(u,v) \neq H^{(1)}(u,v)$ but condition (3.2) still holds.

Using the above theorem, for large n, one can test H_0 by the statistic

$$T_\eta = \frac{n^{1/2}\int_0^\tau \{\,[\hat{F}_0(u) - \hat{F}_1(u)]\,\eta(u)\,d\hat{H}_1^{(1)}(u) + [\hat{F}_0(v) - \hat{F}_1(v)]\,\eta(v)\,d\hat{H}_2^{(1)}(v)\}}{(\,n_0 n^{-1}\,\hat{A}_0 + n_1 n^{-1}\,\hat{A}_1)^{1/2}}$$

based on the standard normal distribution, where $n_1 = n - n_0$,

$$\hat{A}_0 = \int_0^\tau \hat{F}_0(u)(1-\hat{F}_0(u))\hat{\xi}^2(u)d\hat{H}_1^{(0)}(u) + \int_0^\tau \hat{F}_0(v)(1-\hat{F}_0(v))\hat{\xi}^2(v)d\hat{H}_2^{(0)}(u)$$
$$+2\int\int_{0\le u\le v\le\tau} \hat{F}_0(u)(1-\hat{F}_0(v))\hat{\xi}(u)\hat{\xi}(v)d\hat{H}^{(0)}(u,v)$$

and

$$\hat{A}_1 = \int_0^\tau \hat{F}_1(u)(1-\hat{F}_1(u))\eta^2(u)d\hat{H}_1^{(1)}(u) + \int_0^\tau \hat{F}_1(v)(1-\hat{F}_1(v))\eta^2(v)d\hat{H}_2^{(1)}(v)$$
$$+2\int\int_{0\le u\le v\le\tau} \hat{F}_1(u)(1-\hat{F}_1(v))\eta(u)\eta(v)d\hat{H}^{(1)}(u,v)\,.$$

In the above, $\hat{H}_1^{(l)}$, $\hat{H}_2^{(l)}$ and $\hat{H}^{(l)}$ denote the empirical distributions of U_i's, V_i's and (U_i, V_i)'s for subjects with $z_i = l$, respectively, $l = 0, 1$, and $\hat{\xi}(\cdot)$ is an estimate of $\xi(\cdot)$.

In the application of the above test procedure, different η gives different test statistics and it is apparent that the simplest one is $\eta(u) = 1$. For estimation of $\xi(u) = \eta(u)\,R(u)$, a simple approach, which is used in the following numerical studies, is to replace $h_1^{(1)}$ and $h_1^{(0)}$ in (3.2) with their empirical estimates for given η. Another approach is to use smooth estimates of $h_1^{(1)}$ and $h_1^{(0)}$ such as kernel estimates in estimation of R.

3.3 Numerical Studies

Monte Carlo simulation studies were conducted to investigate the performance of the proposed test procedure. In these studies, it was assumed that there are two treatment groups and they have the same number of subjects. The failure time T_i was generated from Weibull distributions with the shape and scale parameters α_1 and β for group 1 and α_2 and $\theta\,\beta$ for group 2, respectively. For observation times, for $l = 0, 1$, we first independently generated U_i and W_i from Gamma(p_l, λ_l) and Gamma(q, λ_l), respectively, where

$$\lambda_l = \lambda \left[\frac{\Gamma(p_l + q)}{\Gamma(p_l)} \right]^{1/q}$$

with p_l, q and λ being some constants. Then we took $V_i = U_i + W_i$, which follows Gamma$(p_l + q, \lambda_l)$. This gives

$$\frac{h_2^{(l)}(t)}{h_1^{(l)}(t)} = \frac{\lambda_l^{p_l+q} t^{p_l+q-1} e^{-\lambda_l t}/\Gamma(p_l+q)}{\lambda_l^{p_l}\, t^{p_l-1} e^{-\lambda_l t}/\Gamma(p_l)} = \lambda_l^q\, t^q \cdot \frac{\Gamma(p_l)}{\Gamma(p_l+q)} = (\lambda t)^q\,.$$

and thus condition (3.2) holds. For the results reported below, we took $\eta(\cdot) = 1$, $p_1 = 0.2$, $p_2 = 0.4$, $q = 3$ and $\lambda = 0.8$.

Tables 4 & 5 present the estimated size and power at the significance level 0.05 of the proposed test procedure (NPTU) based on 1000 sets of simulated data with $n_1 = n_2 = 50$ or 100, $\beta = 0.5$ or 1, $\theta = 1$, 1.5, 2 or 3, and α_1 and α_2 taking value 0.5, 1 or 1.5. Here the three different values of α_1 and α_2 give decreasing, constant and increasing hazard rates, respectively. Note that under the model used here, the null hypothesis H_0 is equivalent to $\alpha_1 = \alpha_2$ and $\theta = 1$. For comparison, by assuming that the underlying true model is known, we also calculated the estimated size and power of the parametric likelihood ratio test (PLRT) for H_0 and included them in Tables 4 & 5. In addition, Table 4 gives

the estimated size of the test procedure (NPT) given in Zhang et al. (2001) by assuming that the distributions of observation times are the same between the two groups. That is, $H^{(0)}(u, v) = H^{(1)}(u, v)$.

It can be seen from Table 4 that the proposed test has reasonable size and power. Especially, its size and power are quite close to those of the parametric likelihood ratio test for most situations, which is optimal for the situations considered here. As expected, both size and power become better when the sample size increases. On the other hand, the test that ignores the difference between the distributions of observation times does not seem to have the proper size.

We also investigated the approximation of the standard normal distribution of the test statistic T_η under H_0 for sample sizes of 50 and 100, respectively. The plots indicate that the approximation works well (Figures 3 and 4).

3.4 An Application

In this section, we apply the proposed method to the AIDS cohort study discussed in Chapter 2. One objective of the study was to compare the HIV infection rates between the two groups.

To apply the proposed approach to test if survival functions of the time to HIV infection between the two treatment groups are identical, we first check if the distributions of censoring intervals are the same. For this, we obtained empirical estimates of the joint distributions $H^{(0)}(u, v)$ and $H^{(1)}(u, v)$ based on subjects within each treatment separately and display them in Figures 5 & 6. It seems from the figures that the two distributions are quite different and this suggests that the proposed test procedure should be used. The

application of the proposed test gave $T_{\eta=1} = 3.603$, yielding a p-value of < 0.001. The result indicates that the patients in the two different groups had significantly different risk to become HIV infected. To confirm this, Figure 7 presents the nonparametric estimators $\hat{F}_0$ and $\hat{F}_1$ used in the test statistic of the distribution functions of time to HIV infection for patients in the two groups. It seems to be consistent with the result given by the test procedure.

3.5 Concluding Remarks

In the preceding sections, a class of test statistics was proposed for two-sample survival comparison based on interval-censored failure time data. The key advantage of the approach over existing test procedures is that it allows different distributions of censoring intervals or observation times between two treatment groups, which often occurs in practice. Failure to take into account such differences in treatment comparison can either underestimate or overestimate treatment difference (Sun, 1999). The simulation results suggest that the presented approach works reasonably well for practical situations.

In constructing the test statistics Q, instead of using functional $g(F)$, an alternative is to apply different functionals such as

$$\int\int_{0\le u\le v\le \tau} \{F(u)\eta(u) + F(v)\eta(v)\}\, dH^{(0)}(u,v)\,.$$

A test procedure can be similarly developed by using the functional given above. In the development of the test procedure given above, another modification that one may apply is to use the maximum likelihood estimators of F_0 and F_1 based on observed interval-censored data instead of $\hat{F}_0$ and $\hat{F}_1$. As commented before, one disadvantage of this

approach is that the maximum likelihood estimators do not have closed forms, which would make the implementation of the test procedure much harder. Also the derivation of asymptotic distribution of the resulting test statistics is not easy to obtain (Zhang et al., 2001). Of course, one advantage of such test procedures is that they may be more efficient. However, the efficiency gain may not be significant based on the simulated results given in Section 3.3.

This chapter discussed the situation where the distributions of censoring intervals or observation times may differ between two treatment groups, but the observation times are independent of the survival time of interest. A more complicated situation that may occur in practice is that the observation times and the survival time of interest are correlated. In this case, for treatment comparison based on interval-censored data, a different test procedure would be needed that can take into account the correlation.

CHAPTER 4

SEMIPARAMETRIC REGRESSION ANALYSIS OF TWO-SAMPLE CURRENT STATUS DATA

4.1 Introduction

As commented in Chapter 1, current status data or case I interval-censored data arise in many fields including animal carcinogenicity experiments, demographical studies, econometrics, and epidemiological studies where the variable of interest is the time to occurrence of a certain event. By current status data, we mean that each subject is observed at only one time point and no information is available on subjects between their entry times and observation time points. That is, for each subject, one only knows whether the event of interest has occurred before the observation time and the occurrence time is either left- or right-censored. For example, in carcinogenicity experiments, animals are usually examined only at death or sacrifice time for evidence of a malignancy. In these situations, the time to tumor onset is of interest, but not directly observable. Instead, one knows only the age at death or sacrifice and whether or not the tumor is present at that time.

For the two-sample semiparametric modeling, the proportional hazards model is perhaps the most widely used model and under this model, the hazard ratio for the two groups is a constant. Sometimes the constant hazard ratio may be in question and in this case, one can use the proportional odds model, which allows the time-dependent hazard ratio. One shortcoming of these models is that they do not apply if the two hazard or

survival functions cross and this can happen in, for example, a medical study where a treatment may be effective in long run but can have certain adverse effects during the early stage. In this situation, the hazard functions or survival functions may cross and one needs different models rather than these discussed above. Corresponding to this, in this chapter, we consider a class of two-sample semiparametric models that are sometimes referred to as short-term and long-term hazard ratio models (Yang and Prentice, 2005) when only current status data are available. One feature of these models is that they can accommodate crossing survival functions. Also they include the proportional hazards model and the proportional odds model as special cases and thus can be used for model checking.

The remainder of this chapter is organized as follows. Section 4.2 introduces the two-sample hazard ratio model along with some notation and the assumptions used throughout the chapter and in Section 4.3, an estimating equation-based procedure is presented for estimation of regression parameters. Furthermore, the asymptotic properties of the proposed estimate are established along with a simple bootstrap procedure for covariance estimation of the proposed estimate. In Section 4.4, we report some results from a simulation study conducted for evaluating the proposed estimation procedure and they suggest that the approach works well for practical situations. Section 4.5 provides an illustrative example arising from a carcinogenicity experiment and some concluding remarks are given in Section 4.6.

4.2 Two-sample Hazard Ratio Model

Consider a survival study that consists of two groups, control and treatment groups.

Suppose that the underlying failure times are absolutely continuous and let $\lambda_C(t)$ and $\lambda_T(t)$ denote the hazard functions of the failure times corresponding to the two groups. Define $S_C(t) = \exp\{-\int_0^t \lambda_C(s)ds\}$, the survival function for subjects in the control group, $\tau_0 = \sup\{t : S_c(t) > 0\}$,

$$\theta_1\,\theta_2 = \lim_{t\to 0} \frac{\lambda_T(t)}{\lambda_C(t)} \;,\;\; \theta_2 = \lim_{t\to \tau_0} \frac{\lambda_T(t)}{\lambda_C(t)}\,.$$

The short-term and long-term two sample hazard ratio model postulates that

$$\lambda_T(t) = \frac{\theta_1\,\theta_2}{\theta_1 + (1-\theta_1)S_C(t)}\,\lambda_C(t)\,,\;\; (t \le \tau_0)\,. \tag{4.1}$$

It is easy to see that $\theta_1\,\theta_2$ and θ_2 represent the short-term and long-term hazard ratios, respectively, while θ_1 denotes the ratio of the short-term ratio and long-term ratio. Also the ratio of the two hazard functions $\lambda_T(t)$ and $\lambda_C(t)$ is not constant as the proportional hazards model, and it is monotonically increasing if $\theta_1 < 1$ and monotonically decreasing if $\theta_1 > 1$. If $\theta_1 = 1$, meaning that the short-term and long-term effects are the same, model (4.1) gives the proportional hazards model, and if $\theta_2 = 1$, we have the proportional odds model.

Let

$$R(t) = \frac{1 - S_C(t)}{S_C(t)}\,,\, (t < \tau_0)\,. \tag{4.2}$$

Then the survival functions S_C and S_T for the two groups have the forms

$$S_C(t) = \frac{1}{1 + R(t)} \;,\;\; S_T(t) = \frac{1}{\{1 + \theta_1\,R(t)\}^{\theta_2}}\,. \tag{4.3}$$

It can be easily shown that S_C and S_T cross if either $\theta_1\,\theta_2 < 1$ and $\theta_2 > 1$, or $\theta_1\,\theta_2 > 1$ and $\theta_2 < 1$.

In the following, we assume that only current status data are available. Specifically, for subject i, suppose that it is observed only once at time C_i and the observed information

for the failure time of interest T_i is given by $\delta_i = I(T_i \leq C_i)$, indicating if the survival event of interest has occurred before or at C_i, $i = 1, ..., n$. Define $Z_i = 1$ if subject i is in the treatment group and 0 otherwise. Then the observed data are

$$\{ C_i , \delta_i , Z_i ; i = 1, ..., n \}$$

and the survival function for subject i can be written as

$$S_i(t; \beta) = \{ 1 + \exp(Z_i \beta_1) R(t) \}^{-\exp(Z_i \beta_2)} \tag{4.4}$$

from (4.3), where $\beta = (\beta_1, \beta_2)'$, which will be referred to as regression parameters, with $\beta_1 = \ln(\theta_1)$ and $\beta_2 = \ln(\theta_2)$. Let $n_1 = \sum_{i=1}^{n} Z_i$, the number of subjects in the treatment group. In the following, we assume that $\lim_{n \to \infty} n_1/n = \rho$ with $0 < \rho < 1$ and that given Z_i, T_i and C_i are independent.

4.3 Estimation of Regression Parameters

In this section, we consider estimation of regression parameters β. For this, suppose that C_i is a positive, continuous variable with the hazard function $\lambda_0(t)$ and the cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$. For each i, define $N_i(t) = \delta_i I(C_i \leq t)$, which is a counting process with the intensity process

$$Y_i(t) \, p_i(t) \, d\Lambda_0(t)$$

(Lin et al., 1998), where $Y_i(t) = I(C_i \geq t)$ and $p_i(t) = S_i(t, \beta)$, $i = 1, ..., n$. Also define the counting process $N_i^C(t) = I(C_i \leq t)$, whose intensity process is given by $Y_i(t) \, d\Lambda_0(t)$, $i = 1, ..., n$. These yield martingales

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \, p_i(s) \, d\Lambda_0(s)$$

and

$$M_i^C(t) \,=\, N_i^C(t) \,-\, \int_0^t Y_i(s)\, d\Lambda_0(s)$$

with respect to the σ-filtrations $\mathcal{F}_t \,=\, \sigma\{N_i(s), Y_i(s), Z_i(s) \,:\, s \le t, i \,=\, 1, \cdots, n\}$ and $\mathcal{F}_t^C \,=\, \sigma\{N_i^C(s), Y_i(s), Z_i(s) \,:\, s \le t, i \,=\, 1, \cdots, n\}$, respectively.

Let $\{m_i(t)\}_{i=1}^n$ be some independently and identically distributed (i.i.d.) random processes that may be functions of the observed data and unknown parameters. Define

$$S^{(j)}(\beta, t) \,=\, \frac{1}{n} \sum_{i=1}^{n} Y_i(t)\, m_i(t)\, q_i^{\otimes j}(t)\,,$$

where $q_i(t) \,=\, \partial m_i/\partial\beta$, $j \,=\, 0, 1$. For estimation of β, we first assume that $R(t)$ and $\Lambda_0(t)$ are known. Then motivated by the partial score function under the proportional hazards model, one can use the estimating function

$$U(\beta, R, \Lambda_0) \,=\, \sum_{i=1}^{n} \int_0^\infty \left\{ q_i(t) \,-\, \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\} dM_i(t)$$

and define an estimate of β as the solution to $U(\beta, R, \Lambda_0) \,=\, 0$. In this approach, of course, we need to estimate $R(t)$ and $\Lambda_0(t)$. Also one needs to choose the $m_i(t)$'s such that the estimation function $n^{-1/2}\, U(\beta, R, \Lambda_0)$ with R and Λ_0 replaced by their estimates has an asymptotic normal distribution with mean zero, which leads to the unbiasedness of the resulting estimate.

For estimation of $R(t)$, one can first use the data from the subjects in the control group to obtain the nonparametric maximum likelihood estimator of S_C, which has the $n^{1/3}$ convergence rate (Groeneboom and Wellner, 1992). Then one can estimate $R(t)$ using (4.2). In the following, we assume that there exists a uniformly consistent estimator $\hat{R}(t)$ such that

$$|\, \hat{R}(t) \,-\, R(t)\,| \,=\, O_p(n^{-1/3})$$

uniformly for $t \leq \tau_0$. For $\Lambda_0(t)$, it can be easily estimated by the Nelson-Aalen estimator given by

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i^C(s)}{\sum_{j=1}^n Y_j(s)}$$

based on the observed data on the C_i's. These lead to an estimate of $M_i(t)$ given by

$$d\hat{M}_i(t) = dN_i(t) - Y_i(t)\,\hat{p}_i(t)\,d\hat{\Lambda}_0(t)\,.$$

Given the estimates defined above, it is natural to estimate β by $\hat{\beta}$ defined as the solution to $\hat{U}(\beta) = 0$, where

$$\hat{U}(\beta) = U(\beta, \hat{R}, \hat{\Lambda}_0) = \sum_{i=1}^{n} \int_0^\infty \left\{ \hat{q}_i(t) - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} \right\} d\hat{M}_i(t)$$

and $\hat{q}_i(t)$ and $\hat{S}^{(j)}(\beta,t)$ denote $q_i(t)$ and $S^{(j)}(\beta,t)$ with $R(t)$ and/or $\Lambda_0(t)$ replaced by their estimates given above. To obtain $\hat{\beta}$, we need to specify $m_i(t)$ and for this, we propose to use

$$m_i(t) = \frac{\partial p_i(t)}{\partial R}\bigg|_{R=R(t)} = -\,p_i(t)\,\frac{\exp\{Z_i(\beta_1 + \beta_2)\}}{1 + \exp(Z_i\beta_1)\,R(t)}\,.$$

This gives

$$q_i(t) = q_i(t;\beta,R) = -\,\frac{Z_i p_i(t)\exp\{Z_i(\beta_1+\beta_2)\}}{1+\exp(Z_i\beta_1)\,R(t)}$$

$$\times \begin{pmatrix} [1-\exp\{Z_i(\beta_1+\beta_2)\}R(t)]/\{1+\exp(Z_i\beta_1)R(t)\} \\ 1-\exp(Z_i\beta_2)\log\{1+\exp(Z_i\beta_1)R(t)\} \end{pmatrix}.$$

In the Appendix, we will show that $n^{-1/2}\,\hat{U}(\beta)$ can be written as a sum of n i.i.d. zero-mean random vectors. It follows that $\hat{\beta}$ is consistent and its distribution can be approximated by a normal distribution.

For estimation of the covariance matrix of $\hat{\beta}$, it can be seen from the Appendix that it is possible to derive a consistent estimate but it would be quite complicated. Instead, we

propose to use the following simple bootstrap procedure. Let M be a prespecified integer. For each l ($1 \leq l \leq M$), draw a simple random sample of size n denoted by

$$D^{(l)} = (C_i^{(l)}, \delta_i^{(l)}, Z_i^{(l)} ; i = 1, \cdots, n)$$

from the observed data with replacement. Let $\hat{\beta}^{(l)}$ denote the estimate of β defined above based on the data set $D^{(l)}$. Then for large n, $(\hat{\beta}^{(l)})_{l=1}^{M}$ follow the same distribution as $\hat{\beta}$ and thus the covariance matrix of $\hat{\beta}$ can be estimated by the sample covariance matrix of the $\hat{\beta}^{(l)}$'s.

4.4 Numeric Studies

Some numeric studies were conducted to investigate the performance of the proposed estimation procedure for practical situations with the focus on the bias of the proposed estimates of regression parameters, the bootstrap procedure and the normal approximation. In the study, we generated the failure time T_i based on model (4.4) with $\lambda_C(t) = 1$ or $S_C(t) = e^{-t}$ and different values of regression parameters. In particular, we considered $\beta = (0, 0)'$ (no group difference), $(0, 1)'$ (the proportional hazards model), $(1, 0)'$ (the proportional odds model), $(1, 1)'$, $(-1, -1)'$, $(2, -1)'$, and $(-2, 1)'$. The last two choices represent the situations in which the group or treatment effect is initially negative (positive) but gradually becomes positive (negative). Also in this last two situations, the two corresponding survival functions cross. Figures 8-11 display survival functions for cases $(0, 0)'$, $(0, 1)'$, $(-1, -1)'$, and $(2, -1)'$. The observation times C_i's were assumed to follow the uniform distribution $U(0, 4)$.

Table 6 summarizes the simulation results based on 1000 replications, $M = 300$, and

$n_1 = n_2 = 100$ or 200. In particular, it gives the estimated bias (Bias) defined as the means of regression parameter estimates minus their true values, the sample standard errors (SSE) of the estimates of regression parameters, and the means of the bootstrap sample standard errors (BSE). These results suggest that the proposed estimates seem to be unbiased and the bootstrap procedure seems to give reasonable variance estimates. When n increases from 100 to 200, as expected, there is a substantial reduction in the biases as well as the estimated standard errors.

To assess the normal approximation to the distribution of $\hat{\beta}$, we studied the quantile plots of the standardized $\hat{\beta}$ against the standard normal distribution for the various cases considered in Table 6. Figure 12 displays two of these plots that correspond to $\beta = (2, -1)'$ and suggest that the normal approximation works well.

4.5 An Application

In this section, we apply the methodology proposed in the previous sections to a set of current status data arising from a carcinogenicity experiment. This study was originally reported by Hoel and Walburg (1972) and concerns lung tumors on 144 male RFM mice. The experiment involves two groups, conventional environment (CE) and germfree environment (GE), and the observation times C_i's are the death or sacrifice times of the animals. At each death or sacrifice time point, the presence or absence of lung tumors was examined and the lung tumors were found in 27 out of the 96 mice assigned to the CE compared with 35 out of the 48 mice assigned to the GE. Since lung tumors are generally regarded as nonlethal, it is reasonable to assume that the tumor onset times T_i's are independent of the death or sacrifice times C_i's.

To fit model (4.1) to the data, define $Z_i = 0$ for the mice in the CE group and 1 otherwise. The application of the estimation procedure gives $\hat{\beta} = (0.092, 0.478)'$ with the 95% confidence intervals (-0.273, 0.457) and (0.195, 0.761) for β_1 and β_2, respectively, based on $M = 500$ bootstrap samples. These give $\hat{\theta}_1 = 1.095$ and $\hat{\theta}_2 = 1.613$ and their 95% confidence intervals are, respectively, (0.730, 1.460) and (1.330, 1.896). These results indicate that there exists a consistent group or treatment effect throughout the whole study and that maybe one can fit the data to the proportional hazards model. This latter conclusion is reinforced by Figure 13, which shows the estimated survival functions for the two groups given using the proposed inference procedure and by fitting the proportional hazards model, respectively.

4.6 Conclusion and Discussion

This chapter discussed the analysis of two-sample current status data that commonly occur in many studies. For the analysis, a class of short-term and long-term hazard ratio models are described and the inference procedure was proposed with the focus on estimation of short-term and long-term effect parameters. The asymptotic properties of the proposed parameter estimates were established and the simulation study suggests that these estimates work well for practical situations. A major advantage of the models considered here is that they include some commonly used models as special cases and allow crossing survival functions. In particular, as shown in Section 4.4, the methodology can be used as a model-checking procedure for the proportional hazards model and other models that are included in model (4.1) as special cases.

There exist several directions for future research. One is that we only considered the

two-sample situation and it would be useful to generalize the proposed methodology to regression analysis of general current status data. Another question of interest is the validity of the bootstrap variance estimation procedure. Although the simulation results indicate that it works reasonably well for practical situations, no rigorous proof to its validity is available yet.

CHAPTER 5

FUTURE RESEARCH

There exist many open questions in the analysis of interval-censored data. In this chapter, we shall discuss several potential directions for future investigation that closely related to the questions investigated in the previous chapters.

5.1 Testing the Proportional Hazards Model for Interval-censored Data

Regression diagnostics are used to check, for example, the goodness of fit and assumptions about regression models. For right-censored data, some residual based methods such as Cox-Snell residuals are available for assessing the fit of the PH model. Also some graphical techniques such as score residuals and arjas plots are available for checking the assumptions of the PH model. However, there seems to have few existing methods for testing the PH model with interval-censored data. Recently, Yuen et al. (2005) presented a goodness-of-fit test based on the leveraged bootstrap to check the adequacy of the PH model for current status data. A generalization of their method to case II interval-censored data is a possible direction for future research.

5.2 Nonparametric Tests for Comparing Survival Functions for Interval-censored Data in the Presence of Dependent Censoring

A key advantage of the test proposed in Chapter 3 is that it allows different distributions

of censoring intervals or observations times between two treatment groups. However, it still assumes that the failure time T is independent to the observation times within treatment group as most researchers did. It would be useful if we can release this assumption and derive a different test such as a log-rank type test (DiRienzo, 2003) based on interval-censored data for a more general situation. That is, compare survival functions when the censoring intervals depend on failure times and treatment groups, for which there seems to be no established method. In addition, we have also assumed that covariates do not exist. To deal with covariates, some regression models and related inference procedures would be needed.

5.3 Efficient Estimation for the Short-term and Long-term Hazards Ratios

As discussed in Chapter 4, the estimation procedure for short- and long-term hazard ratios performs reasonably well. However, it may not be the most efficient way. It would be useful if we can derive efficient estimators by using, for example, full likelihood approaches. Also, based on the estimators, one can derive a hypothesis test of identical survival functions. Due to the advantage of the model, one can expect to obtain more powers while testing for the crossed survival functions as compared to traditional tests such as the log-rank type of tests. In addition, recall that the model we proposed includes the proportional hazards model and the proportional odds model as special cases. Therefore, hypothesis tests for regression model checking can be derived based on asymptotic normality properties on parameters.

Bibliography

[1] Aalen, O. O. (1978), "Nonparametric inference for a family of counting process", *The Annals of Statistics* **6**, 701–726.

[2] Anderson, P. K. & Rønn, B. B. (1995), "A nonparametric test for comparing two samples where all observations are either left- or right-censored", *Biometrics* **51**, 323–329.

[3] Betensky, R. A., Lindsey, J.C., Ryan, L. M. and Wand, M. P. (2002), "A local likelihood proportional hazards model for interval censored data", *Statistics in Medicine* **21**, 263–275.

[4] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972), *Statistical Inference under Order Restrictions*, New York: Wiley.

[5] Cai, T. and Betensky, R. A. (2003), "Hazard regression for interval-censored data with penalized spline", *Biometrics* **59**, 570–579.

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

[7] Cox, D. R. (1975), "Partial likelihood", *Biometrika* **62**, 269–276.

[8] Dauxois, J. Y. and Kirmani, S. N. U. A. (2003), "Testing the proportional odds model under random censoring", *Biometrika* **90**, 913–922.

[9] De Gruttola, V. G. and Lagakos, S. W. (1989), "Analysis of doubly-censored survival data, with application to AIDS", *Biometrics* **45**, 1–11.

[10] Diamond, I. D., McDonald, J. W. and Shah, I. H. (1986), "Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan", *Demography* **23**, 607–620.

[11] Dinse, G. E. and Lagakos, S. W. (1983), "Regression analysis of tumor prevalence data", *Applied Statistics* **32**, 236–248.

[12] DiRienzo, A. G. (2003), "Nonparametric comparison of two survival-time distributions in the presence of dependent censoring", *Biometrics* **59**, 497–504.

[13] Efron, B. (1981), "Censoring data and the bootstrap", *Journal of the American Statistical Association* **76**, 312–319.

[14] Fang, H., Sun, J., and Lee, M-L T. (2002), "Nonparametric survival comparisons for interval-censored continuous data", *Statist. Sinica* **12**, 1073–1083.

[15] Finkelstein, D. M. (1986), "A proportional hazards model for interval-censored failure time data", *Biometrics* **42**, 845–854.

[16] Finkelstein, D. M., Goggins, W. B. and Schoenfeld, D. A. (2002), "Analysis of failure time data with dependent interval censoring", *Biometrics* **58**, 298–304.

[17] Finkelstein, D. M. and Wolfe, R. A. (1985), "A semiparametric model for regression analysis of interval-censored failure time data", *Biometrics* **41**, 933–945.

[18] Fleming, T. R. and Harrington, D. P. (1991), "Counting process and survival analysis", John Wiley: New York.

[19] Fay, M. P. (1996), "Rank invariant tests for interval censored data under the grouped continuous model", *Biometrics* **52**, 811–822.

[20] Gentleman, R. and Geyer, C. J. (1994), "Maximum likelihood for interval censored data: Consistency and computation", *Biometrika* **81**, 618–623.

[21] Geskus, R. and Groeneboom, P. (1999), "Asymptotically optimal estimation of smooth functionals for interval censoring, case 2", *Ann. Statist.* **27**, 627–674.

[22] Groeneboom, P. (1996).] "Lectures on inverse problems", In *Lecture Notes in Math.*, 1648. Springer-Verlag, Berlin.

[23] Groeneboom, P. and Weller, J. A. (1992), "Information bounds and nonparametric maximum likelihood estimation", DMV seminar, Band **19**, Birkhauser, New York.

[24] Harrington, D. P. and Fleming, T. R. (1982), "A class of rank test procedures for censored survival data", *Biometrika* **69**, 553–566.

[25] Hoel, D. G. and Walburg, H. E. (1972), "Statistical analysis of survival experiments", *J. Natn. Cancer Inst.* **49**, 361–372.

[26] Huang, J., Rossini, A. J. and Wellner, J. A. (1996), "Efficient estimation for proportional hazards model with "case 2" interval censoring", *Technical Report* 251, Department of Statistics and Actuarial Science, University of Iowa.

[27] Huang, J. and Rossini, A. J. (1997), "Sieve estimation for the proportional-odds failure-time regression model with interval censoring", *J. Amer. Statist. Assoc.* **92**, 960–967.

[28] Huang, J. and Wellner, J. A. (1995), "Asymptotic normality of the NPMLE of linear functionals for interval censored data, Case I", *Statistica Neerlandica* **49**, 153–163.

[29] Kalbfleisch, J. D. and Prentice, R. L. (2002), *The statistical analysis of failure time data*, Second Edition, John Wiley: New York.

[30] Kaplan, E. L. and Meier, P. (1958), "Nonparametric estimation from incompete observation", *Journal of the American Statistical Association* **53**, 457–481.

[31] Keiding, N. (1991), "Age-specific incidence and prevalence: a statistical perspective (with discussion)", *Journal of Royal Statistical Society A* **154**, 371–412.

[32] Lin, D. Y., Oakes, D. and Ying, Z. (1998), "Additive hazards regression with current status data", *Biometrika* **85**, 289–298.

[33] McCullagh, P. (1980), "Regression models for ordinal data (with discussion)", *Journal of Royal Statistical Society B* **42**, 109–142.

[34] Murphy, S. A., Rossini, A. J., and van der Vaart, A. W. (1997), "Maximum likelihood estimation in the proportional odds model", *J. Amer. Statist. Assoc.* **92**, 968–976.

[35] Monaco, J., Cai, J. and Grizzle (2005), "Bootstrap analysis of multivariate failure time data", *Statist. Med.* **24**, 3387–3400.

[36] Pan, W. (2000), "A two-sample test with interval censored data via multiple imputation", *Statistics in Medicine* **19**, 1–12.

[37] Petroni, G. R. and Wolfe, R. A. (1994), "A two sample test for stochastic ordering with interval-censored Data", *Biometrika* **50**, 77–87.

[38] Rabinowitz D., Betensky, R. A. and Tsiatis, A. A. (2000), "Using conditional logistic regression to fit proportional odds models to interval censored data", *Biometrics* **56**, 511–518.

[39] Rossini, A. and Tsiatis, A. A. (1996), "A semiparametric proportional odds regression model for the analysis of current status data", *Journal of the American Statistical Association* **91**, 713–721.

[40] Satten (1996), "Rank-based inference in the proportional hazards model for interval censored data", *Biometrika* **83**, 355–370.

[41] Self, S. G. and Grossman, E. A. (1986), "Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers", *Biometrics* **42**, 521–530.

[42] Sengupta, D., Bhattacharjee, A. and Rajeev, B. (1998), "Testing for the proportionality of hazards in two samples against the increasing cumulative hazard ratio alternative", *Scandinavian Journal of Statistics* **25**, 637–647.

[43] Shen, X. (1998), "Proportional odds regression and sieve maximum likelihood estimation", *Biometrika* **85**, 165–177.

[44] Sun, J. (1996), "A nonparametric test for interval-censored failure time data with application to AIDS studies", *Statistic in Medicine* **15**, 1387–1395.

[45] Sun, J. (1999), "A nonparametric test for current status data with unequal Censoring", *J. Roy. Statist. Soc. Ser. B* **61**, 243–250.

[46] Sun, J. (2005), "Interval censoring", *Encyclopedia of Biostatistics*, Second Edition, John Wiley & Sons Ltd., 2603–2609.

[47] Turnbull, B. W. (1976), "The empirical distribution function with arbitrarily grouped censored and truncated data", *JRSS* B **38**, 290–295.

[48] Yang, S. and Prentice, R. (2005), "Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data", *Biometrika* **92**, 1–17.

[49] Yeun, K., Shi, J. and Zhu, L. "A goodness-of-fit test for the proportional hazards model with current status data", *2005 Joint Meeting of CSPS/IMS*, Beijing, China.

[50] Zhang, Y., Liu, W. and Wu, W. (2003), "A simple nonparametric two-sample test for the distribution function of event time with interval censored data", *J. Nonparametr. Stat.* **15**, 643–652.

[51] Zhang, Y., Liu, W. and Zhan, Y. (2001), "A nonparametric two-sample test of the failure function with interval censoring case 2", *Biometrika* **88**, 677–686.

[52] Zhao, Q. and Sun, J. (2004), "Generalized log-rank test for mixed interval-censored failure time data", *Statistics in Medicine* **23**, 1621–1629.

APPENDIX

Proof of Theorem 2.1

Proof. Note that for each pair $(i,\ j)\ (=1,2)$, we have

$$\hat{\theta}_{ij}-\theta_{ij}\ =\ -\int_{\tau_1}^{\tau_2}\frac{w_i(t)}{F_j^2(t)}(\hat{F}_j(t)-F_j(t))dt+\int_{\tau_1}^{\tau_2}\frac{w_i(t)}{\hat{F}_j(t)F_j^2(t)}(\hat{F}_j(t)-F_j(t))^2dt\,. \qquad (A1)$$

Hence it follows from Corollary 4.3 of Groeneboom (1996) that

$$|\hat{\theta}_{ij}-\theta_{ij}|=O_p(n^{-1/2})\,. \qquad (A.2)$$

Also note that under H_0, $q(w_1,w_2)\ =\ 0$. Then by (A.1) and (A.2) we have

$$Q_n(w_1,w_2)=\Big(\frac{n_1n_2}{n}\Big)^{1/2}\Big[\theta_{22}(\hat{\theta}_{11}-\theta_{11})+\theta_{11}(\hat{\theta}_{22}-\theta_{22})-\theta_{21}(\hat{\theta}_{12}-\theta_{12})-\theta_{12}(\hat{\theta}_{21}-\theta_{21})\Big]+o_p(1)$$

$$=\Big(\frac{n_1n_2}{n}\Big)^{1/2}\left[\int_{\tau_1}^{\tau_2}C_1(t)(\hat{F}_1(t)-F_1(t))dt+\int_{\tau_1}^{\tau_2}C_2(t)(\hat{F}_2(t)-F_2(t))dt\right]+o_p(1)\,. \quad (A.3)$$

For the terms at the right-hand side, by following the proof of the theorem of Fang, Sun and Lee (2002), it can be shown that

$$\int_{\tau_1}^{\tau_2}C_i(t)(\hat{F}_i(t)-F_i(t))dt=-\int\Phi_i(u,v,\delta_{i1},\delta_{i2})d\Big[\hat{G}_i(u,v,\delta_{i1},\delta_{i2})-G_i(u,v,\delta_{i1},\delta_{i2})\Big]. \qquad (A.4)$$

Thus it follows from $E\Phi_i(U_i,V_i,\Delta_{i1},\Delta_{i2})\ =\ 0$, $n_1/n\ \to\ \rho$, (A.3), (A.4) and the central limit theorem that $Q_n(w_1,w_2)$ is asymptotically normal with mean zero and variance σ^2. □

Proof of Theorem 3.1

Proof. To prove the asymptotic normality of Q, first note that under H_0, we can rewritten it as

$$Q \;=\; \left(\frac{n}{n_0}\right)^{1/2} Q_0 \;-\; \left(\frac{n}{n_1}\right)^{1/2} Q_1 \,,$$

where

$$Q_0 \;=\; n_0^{1/2} \left\{ g(\hat{F}_0) \;-\; g(F_0) \right\}$$

and

$$Q_1 \;=\; n_1^{1/2} \left\{ g(\hat{F}_1) \;-\; g(F_1) \right\}.$$

Thus it is sufficient to show that Q_0 and Q_1 converge in distribution to independent normal random variables with mean zero and variances A_0 and A_1, respectively.

Define $S_l \;=\; \{i : \, z_i = l\}$, $l \;=\; 0, 1$. For Q_1, following the proof of Theorem 1 of Zhang et al. (2001), it can be easily shown that we have

$$Q_1 \;=\; U_1 \;+\; o_p(1) \,,$$

where

$$U_1 \;=\; n_1^{-1/2} \sum_{i \in S_1} \left\{ \left[\delta_{1i} \;-\; F_1(u_i) \right] \eta(u_i) \;+\; \left[\delta_{1i} \;+\; \delta_{2i} \;-\; F_1(v_i) \right] \eta(v_i) \right\} \,,$$

which clearly has an asymptotic normal distribution with mean zero and variance A_1.

For Q_0, under condition (3.2), we have

$$\begin{aligned}
Q_0 &= n_0^{1/2} \int_0^\tau \left\{ [\hat{F}_0(u) - F_0(u)]\eta(u) dH_1^{(1)}(u) + [\hat{F}_0(v) - F_0(v)]\eta(v) dH_2^{(1)}(v) \right\} \\
&= n_0^{1/2} \int_0^\tau \left\{ [\hat{F}_0(u) - F_0(u)]\eta(u) \frac{h_1^{(1)}(u)}{h_1^{(0)}(u)} dH_1^{(0)}(u) + [\hat{F}_0(v) - F_0(v)]\eta(v) \frac{h_2^{(1)}(v)}{h_2^{(0)}(v)} dH_2^{(0)}(v) \right\} \\
&= n_0^{1/2} \int\int_{0 \le u \le v \le \tau} \left\{ [\hat{F}_0(u) - F_0(u)]\eta(u) R(u) dH_1^{(0)}(u) + [\hat{F}_0(v) - F_0(v)]\eta(v) R(v) dH_2^{(0)}(v) \right\} \\
&= n_0^{1/2} \int\int_{0 \le u \le v \le \tau} \left\{ [\hat{F}_0(u) - F_0(u)]\xi(u) + [\hat{F}_0(v) - F_0(v)]\xi(v) \right\} dH^{(0)}(u, v) \,.
\end{aligned}$$

Then as Q_1, we have that

$$Q_0 = U_0 + o_p(1) \,,$$

where

$$U_0 = n_0^{-1/2} \sum_{i \in S_0} \{ [\delta_{1i} - F_0(u_i)]\xi(u_i) + [\delta_{1i} + \delta_{2i} - F_0(v_i)]\xi(v_i) \} \,,$$

which obviously has an asymptotic normal distribution with mean zero and variance A_0.

It is apparent that U_0 and U_1 are independent and this completes the proof. □

Asymptotic Normality of $n^{-1/2}\,\hat{U}(\beta)$ in Section 4.3

Let $M_i(t)$ and $\hat{M}_i(t)$ and other notation be defined as in the previous sections. To see the asymptotic normality of $n^{-1/2}\,\hat{U}(\beta)$, note that it can be rewritten as

$$\begin{aligned}
\frac{1}{\sqrt{n}}\,\hat{U}(\beta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty \left(\hat{q}_i(t) - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right)\left(dN_i(t) - Y_i(t)\hat{p}_i(t)d\hat{\Lambda}_0(t)\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty \left(\hat{q}_i(t) - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right) dM_i(t)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty \left(\hat{q}_i(t) - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right) Y_i(t)\{p_i(t)-\hat{p}_i(t)\}d\hat{\Lambda}_0(t)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty \left(\hat{q}_i(t) - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right) Y_i(t)\hat{p}_i(t)\left(d\Lambda_0(t) - d\hat{\Lambda}_0(t)\right).
\end{aligned}$$

To show that $n^{-1/2}\,\hat{U}(\beta)$ can be written as a sum of n i.i.d. zero-mean vectors and has an asymptotic normal distribution, it is sufficient to prove that each of the three terms in the right side of the above equation has the same property or converges to zero in probability. For the first term, it is easy to see that it is equal to $n^{-1/2}\,U(\beta,R,\Lambda_0) + o_p(1)$ as $n \to \infty$ based on the fact that the M_i's are i.i.d. martingales. For the second term, using the Taylor series expansion to $p_i(t) - \hat{p}_i(t)$ at $R = \hat{R}(t)$, we have

$$p_i(t) - \hat{p}_i(t) = \hat{m}_i(t)\,\{R(t) - \hat{R}(t)\} + D^2p_i(R^*)\,\{R(t) - \hat{R}(t)\}^2\,,$$

where D^2p_i denotes the second derivative of p_i with respect to R and R^* is some fixed value between $R(t)$ and $\hat{R}(t)$. It then follows that the second term can be written as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty \left(\hat{q}_i(t) - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right) Y_i(t)\hat{m}_i(t)\{R(t) - \hat{R}(t)\}\,d\Lambda_0(t)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\infty \left(\hat{q}_i(t) - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right) Y_i(t)D^2p_i(t)(R^*)\{R(t) - \hat{R}(t)\}^2\,d\Lambda_0(t)\,.
\end{aligned}$$

In the above equation, it is apparent that the first part is equal to zero and the second part converges to zero in probability.

Finally for the third term, note that

$$\hat{\Lambda}_0(t) \,-\, \Lambda_0(t) \,=\, \sum_{k=1}^{n} \int_0^t \frac{dM_k^C(s)}{\sum_{j=1}^{n} Y_j(s)}\,.$$

This leads to

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\infty} \left(\hat{q}_i(t) \,-\, \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} \right) Y_i(t) \hat{p}_i(t) \left(d\Lambda_0(t) \,-\, d\hat{\Lambda}_0(t) \right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\infty} \left(\hat{S}^{(3)}(\beta,t) \,-\, \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\, \hat{S}^{(2)}(\beta,t) \right) \left(\frac{1}{n} \sum_{j=1}^{n} Y_j(t) \right)^{-1} dM_k^C(t)\,,
\end{aligned}$$

where $\hat{S}^{(3)}(\beta,t)$ and $\hat{S}^{(4)}(\beta,t)$ are defined as

$$S^{(j)}(\beta,t) \,=\, \frac{1}{n} \sum_{i=1}^{n} Y_i(t)\, p_i(t)\, q_i^{\otimes(j-2)}(t) \;\; (j \,=\, 3,4)$$

with R replaced by its estimate. It is easy to see that the above summation can be written as a sum of i.i.d. zero-mean vectors.

Table 1: Estimated Empirical Sizes and Powers of the Test Procedure

Right-censoring	β_1			
percentages	1	1.5	2	3
		$B = 0.5$		
$q = 10\%$	0.052	0.523	0.675	0.714
$q = 20\%$	0.048	0.423	0.632	0.691
$q = 30\%$	0.050	0.320	0.554	0.622
		$B = 1$		
$q = 10\%$	0.050	0.305	0.519	0.556
$q = 20\%$	0.052	0.246	0.454	0.514
$q = 30\%$	0.043	0.191	0.357	0.473

Table 2: Interval-censored HIV Infection Data

X_L	X_R		X_L	X_R		X_L	X_R		X_L	X_R
					Heavily treated					
15	∞	(2)	16	∞	(3)	17	∞	(3)		
10	11		1	16		12	13		13	15
14	16		12	14		14	15		13	16
14	15		13	15		9	12		14	15
1	11		12	14		11	12		15	16
15	16		1	13		10	11		5	7
5	7		15	15		14	15		12	13
12	13		1	14		14	15		10	11
10	11		8	10		15	16		9	10
10	12		1	14		1	15		1	13
14	15		3	15		12	13		14	15
9	10		14	15		15	16		1	15
1	14		11	13		10	11		1	7
9	12		1	11		12	13		13	14
10	15		13	15		1	12		7	10
1	15		9	12		7	15		14	16
11	13		11	13		11	13		1	6
8	15		10	11		12	13		7	9
12	13		9	13		13	14		9	12
3	14		10	11		14	15		7	9
12	13		13	14		1	7		3	7
10	11		13	15		10	12		5	7
9	11		1	10		9	13		5	8
10	11		13	15		1	7		10	12
10	12		8	10		9	12		10	12
10	14									

Observations for 262 hemophilia patients by amount of blood received. Numbers in parentheses denote multiplicities.

Table 3: Interval-censored HIV Infection Data (continued)

X_L	X_R	X_L	X_R		X_L	X_R		X_L	X_R	
Lightly treated										
1	∞	15	∞	(19)	16	∞	(31)	17	∞	(10)
18	∞	10	15		12	14		1	15	
1	15	1	15		10	12		1	16	
15	16	3	10		8	15		8	13	
1	12	13	14		5	11		14	16	
1	11	9	14		8	16		11	12	
1	17	1	18		1	15		11	16	
8	12	9	13		1	15		13	14	
9	14	1	5		1	16		12	15	
9	12	13	15		4	11		1	16	
1	15	14	15		1	12		14	15	
1	14	6	13		13	14		15	16	
7	12	12	14		12	14		1	13	
12	13	13	15		15	16		1	15	
13	15	8	16		10	12		14	15	
11	15	13	15		3	16		6	8	
15	16	11	14		13	14		12	14	
7	10	1	12		1	15		12	13	
1	15	10	16		11	14		1	14	
12	13	9	14		12	14		11	12	
1	11	1	16		12	13		14	15	
1	15	15	16		11	12		13	13	
13	14	10	12		6	12		1	12	
1	3	11	14		1	5		10	11	
7	13	12	13		6	13		11	14	

Table 4: Empirical Sizes of the Proposed Test Procedure (nominal size=5%)

Parameters				$n_1 = n_2 = 50$			$n_1 = n_2 = 100$		
β	θ	α_1	α_2	NPTU	PLRT	NPT	NPTU	PLRT	NPT
0.5	1.0	0.5	0.5	3.8	4.5	8.6	4.4	4.5	8.1
0.5	1.0	1.0	1.0	5.3	5.0	8.5	5.2	5.1	7.6
0.5	1.0	1.5	1.5	7.4	5.8	9.8	6.2	5.4	9.0
1.0	1.0	0.5	0.5	4.8	5.4	7.3	4.5	5.4	7.1
1.0	1.0	1.0	1.0	3.6	5.0	7.9	4.0	5.2	8.0
1.0	1.0	1.5	1.5	4.2	6.0	7.6	5.2	5.4	7.3

Table 5: Empirical Powers of the Proposed Test Procedure

Parameters				$n_1 = n_2 = 50$		$n_1 = n_2 = 100$	
β	θ	α_1	α_2	NPTU	PLRT	NPTU	PLRT
0.5	1.5	0.5	0.5	10.3	10.6	15.0	18.6
0.5	1.5	1.0	1.0	17.4	22.4	28.6	41.4
0.5	1.5	1.5	1.5	27.5	35.3	42.8	53.5
0.5	2.0	0.5	0.5	22.5	26.6	42.4	43.9
0.5	2.0	1.0	1.0	48.1	60.4	77.2	88.0
0.5	2.0	1.5	1.5	61.8	75.1	90.8	95.7
0.5	3.0	0.5	0.5	48.9	54.2	83.6	88.2
0.5	3.0	1.0	1.0	89.4	94.3	100.0	100.0
0.5	3.0	1.5	1.5	97.2	100.0	100.0	100.0
1.0	1.5	0.5	0.5	9.2	10.6	12.6	17.1
1.0	1.5	1.0	1.0	23.0	27.5	38.0	49.8
1.0	1.5	1.5	1.5	35.4	40.3	63.7	70.4
1.0	2.0	0.5	0.5	21.4	24.6	40.1	44.4
1.0	2.0	1.0	1.0	55.3	67.2	84.0	92.3
1.0	2.0	1.5	1.5	84.6	91.2	97.6	100.0
1.0	3.0	0.5	0.5	44.8	51.4	81.1	87.6
1.0	3.0	1.0	1.0	94.0	95.6	100.0	100.0
1.0	3.0	1.5	1.5	99.8	100.0	100.0	100.0
1.0	2.0	0.5	1.0	59.7	73.0	86.8	96.4
1.0	2.0	1.5	0.5	35.8	99.3	55.3	100.0

Table 6: Summary Statistics for the Numerical Studies

		Bias		SSE		BSE	
β	n	β_1	β_2	β_1	β_2	β_1	β_2
(0, 0)	100	0.0042	-0.0023	0.0184	0.0212	0.0214	0.0189
	200	0.0019	-0.0011	0.0097	0.0111	0.0109	0.0092
(0, 1)	100	-0.0198	0.0107	0.0201	0.0840	0.0184	0.0739
	200	-0.0035	0.0041	0.0043	0.0392	0.0034	0.0350
(1, 0)	100	-0.0064	-0.0109	0.0461	0.2815	0.0403	0.2911
	200	-0.0023	-0.0011	0.0298	0.1884	0.0240	0.1947
(1, 1)	100	0.0214	0.0609	0.1207	0.4148	0.1188	0.3944
	200	0.0153	0.0411	0.1004	0.3061	0.0987	0.2901
(-1, -1)	100	0.0036	0.0044	0.0170	0.0207	0.0183	0.0211
	200	0.0024	0.0021	0.0118	0.0143	0.0130	0.0155
(2, -1)	100	-0.0357	-0.0238	0.2143	0.3034	0.1928	0.3011
	200	-0.0161	-0.0166	0.1415	0.2097	0.1235	0.2021
(-2, 1)	100	0.0055	-0.0524	0.1573	0.4477	0.1628	0.4011
	200	0.0031	-0.0266	0.1015	0.3314	0.1135	0.3021

Table 7: Lung Tumor Data on RFM Male Mice

Necropsy Finding	Individual ages at death (days)*											
	A. Conventional mice (96)											
Lung	381	477	485	515	539	563	*565*	582	*603*	616	624	650
tumor	651	656	659	672	679	698	*702*	709	723	731	775	779
	795	811	839									
No Lung	45	198	215	217	257	262	266	371	431	447	454	459
Tumor	475	479	484	500	502	503	505	508	516	531	541	553
	556	570	572	575	577	585	588	594	600	601	608	614
	616	632	632	638	642	642	642	644	644	647	647	653
	659	660	662	663	667	667	673	673	677	689	693	718
	720	721	728	760	762	773	777	815	886			
	B. Germfree mice (48)											
Lung	546	609	692	692	710	752	773	781	782	789	808	810
Tumor	814	842	846	851	871	873	876	888	888	890	894	*896*
	911	913	914	914	916	921	921	926	936	945	1008	
No Lung	412	524	647	648	695	785	814	817	851	880	913	942
Tumor	986											

* Italicized ages represent mice dying of lung tumors.

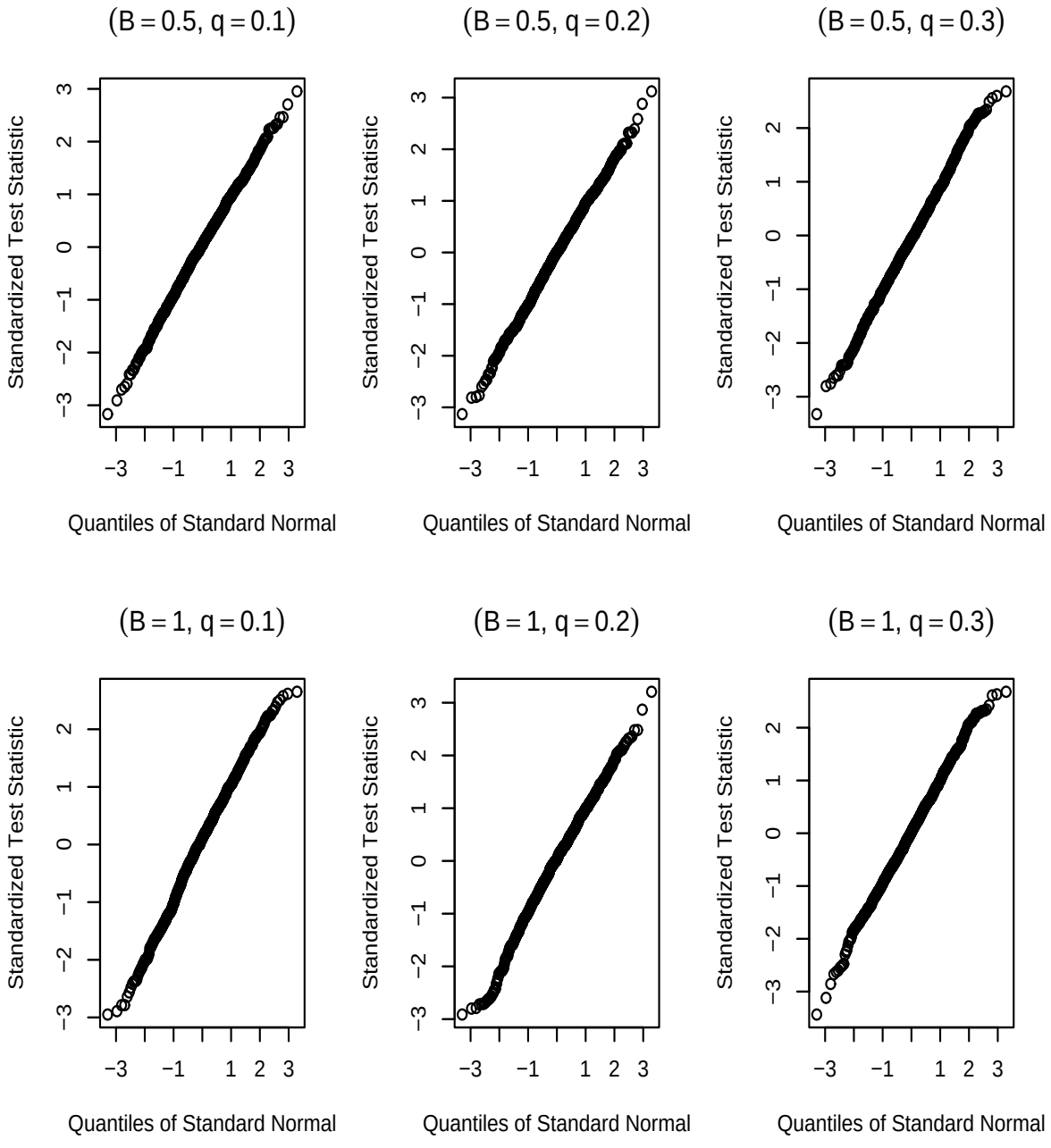


Figure 1: Normal Quantile Plots

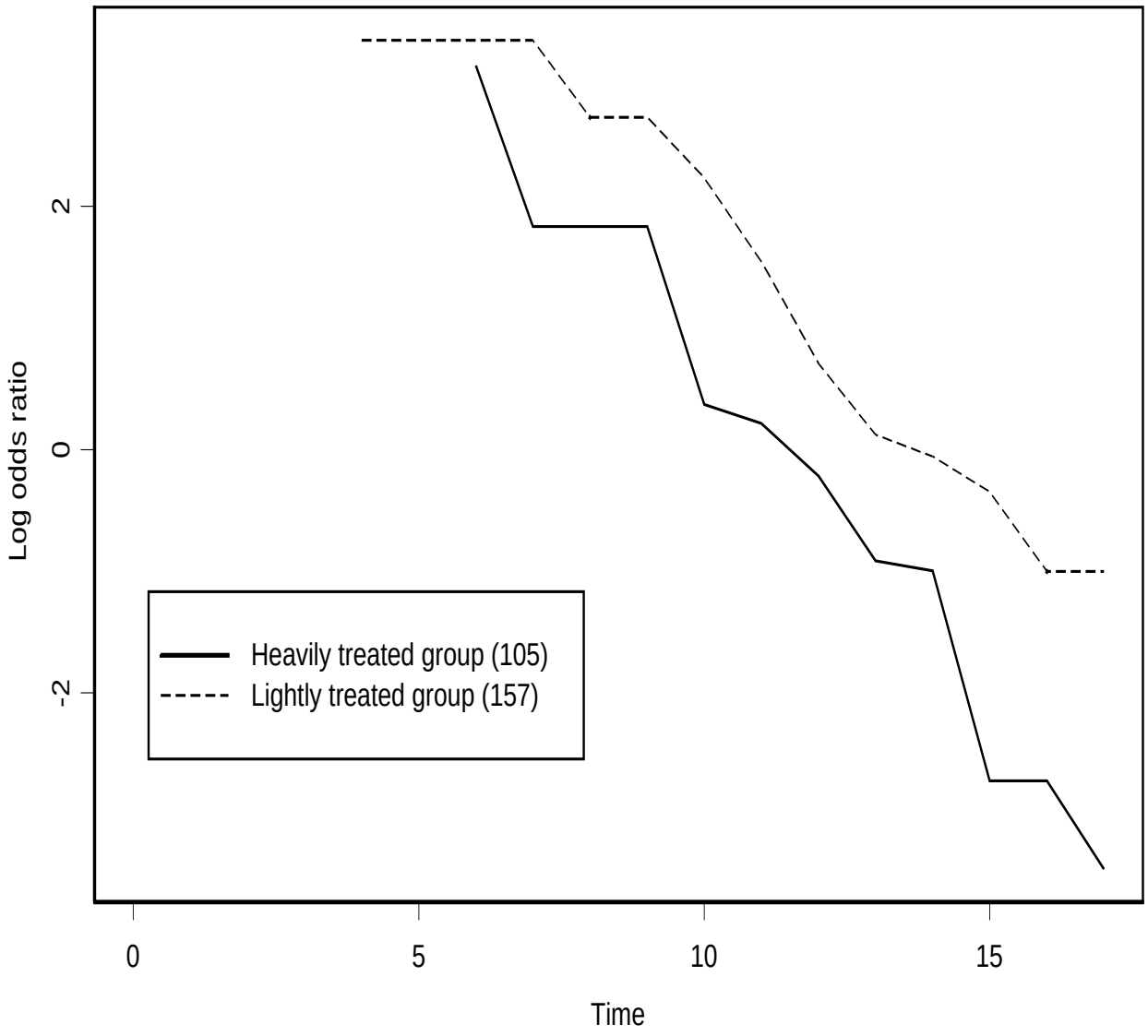


Figure 2: Estimated Log Odds Ratio for Two Groups

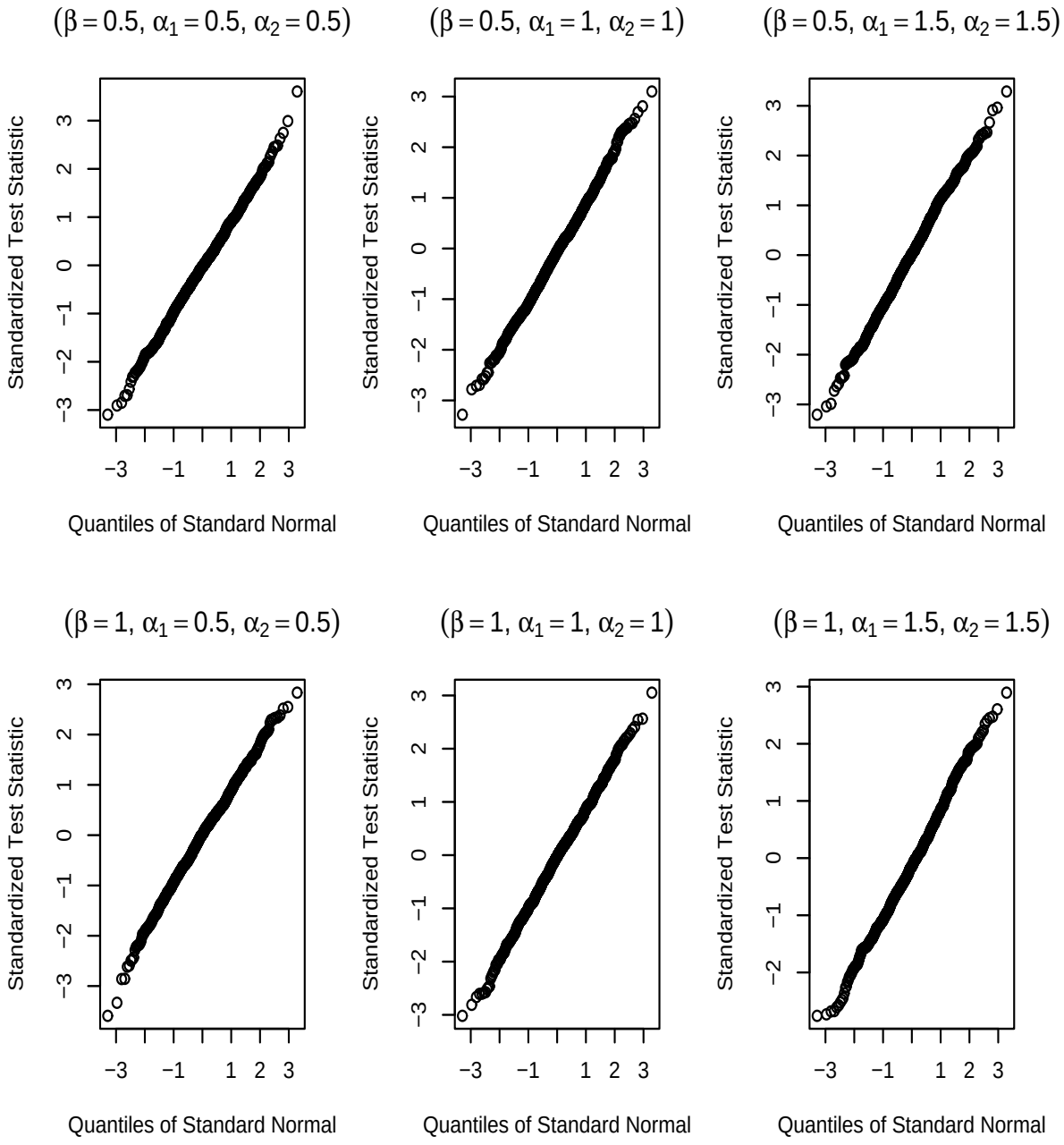


Figure 3: Normal Quantile Plots for $n = 50$

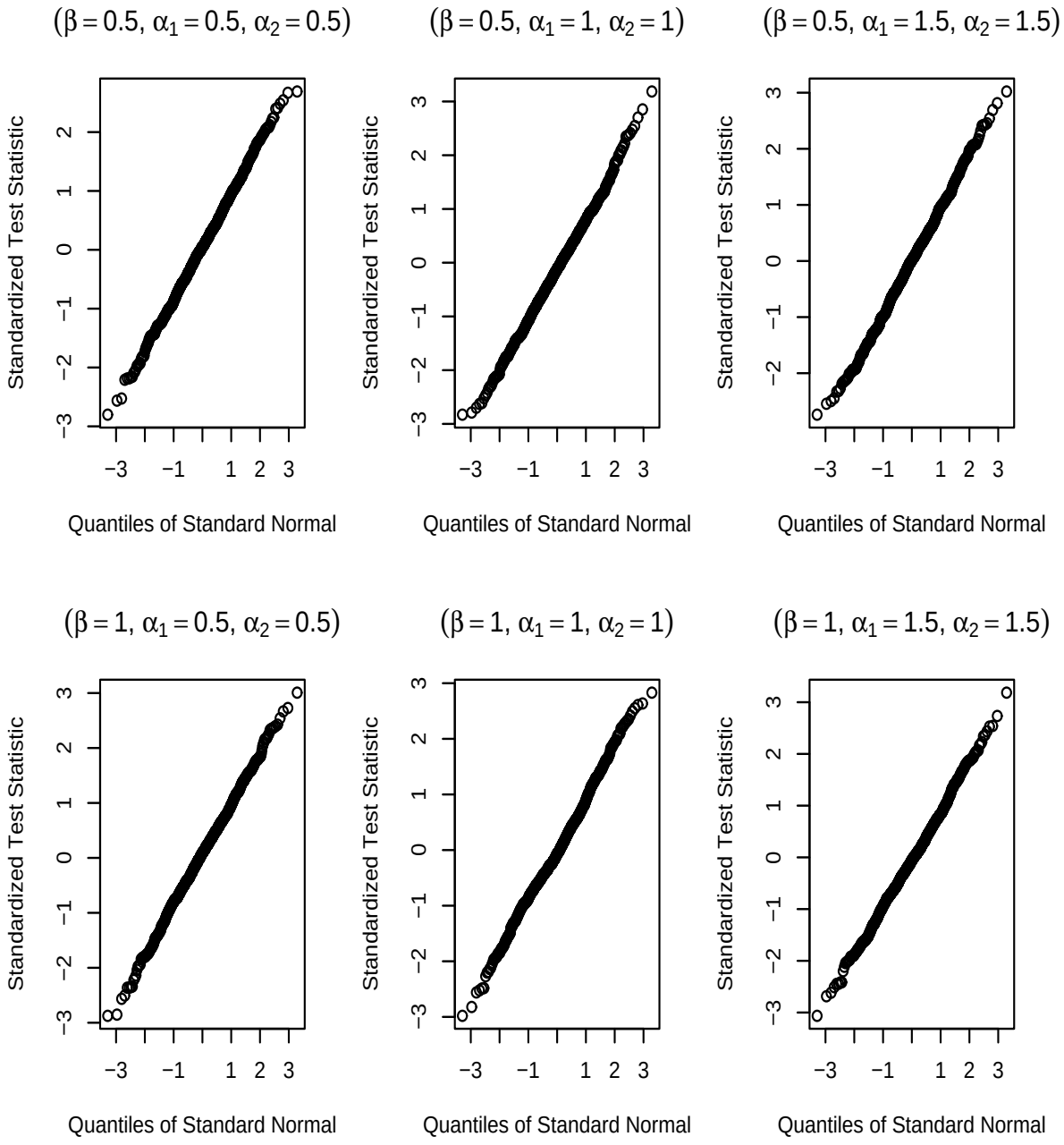


Figure 4: Normal Quantile Plots for $n = 100$

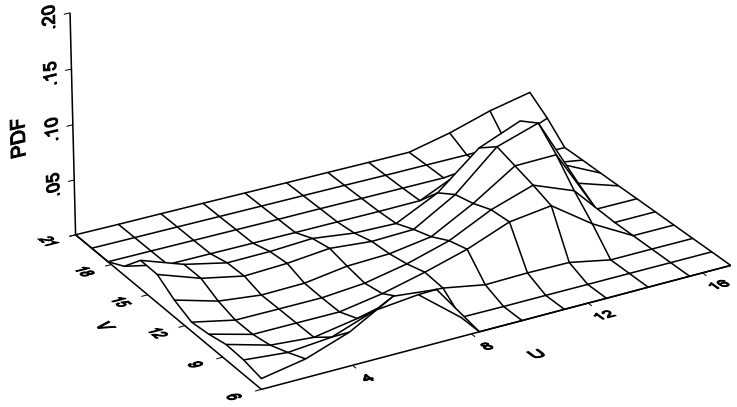


Figure 5: Joint Empirical Distributions of Observation Times (Heavily Treated Group)

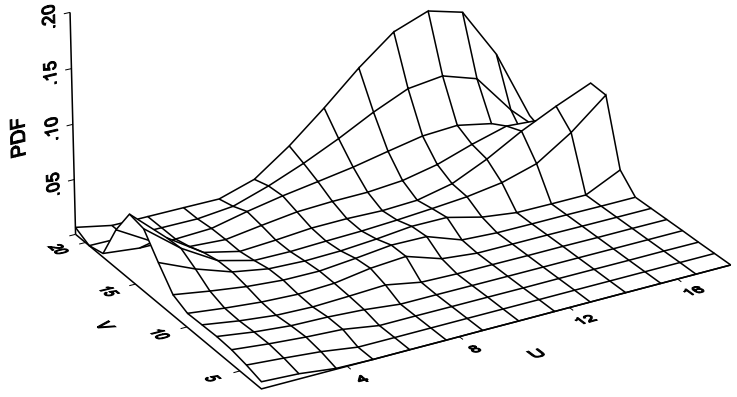


Figure 6: Joint Empirical Distributions of Observation Times (Lightly Treated Group)

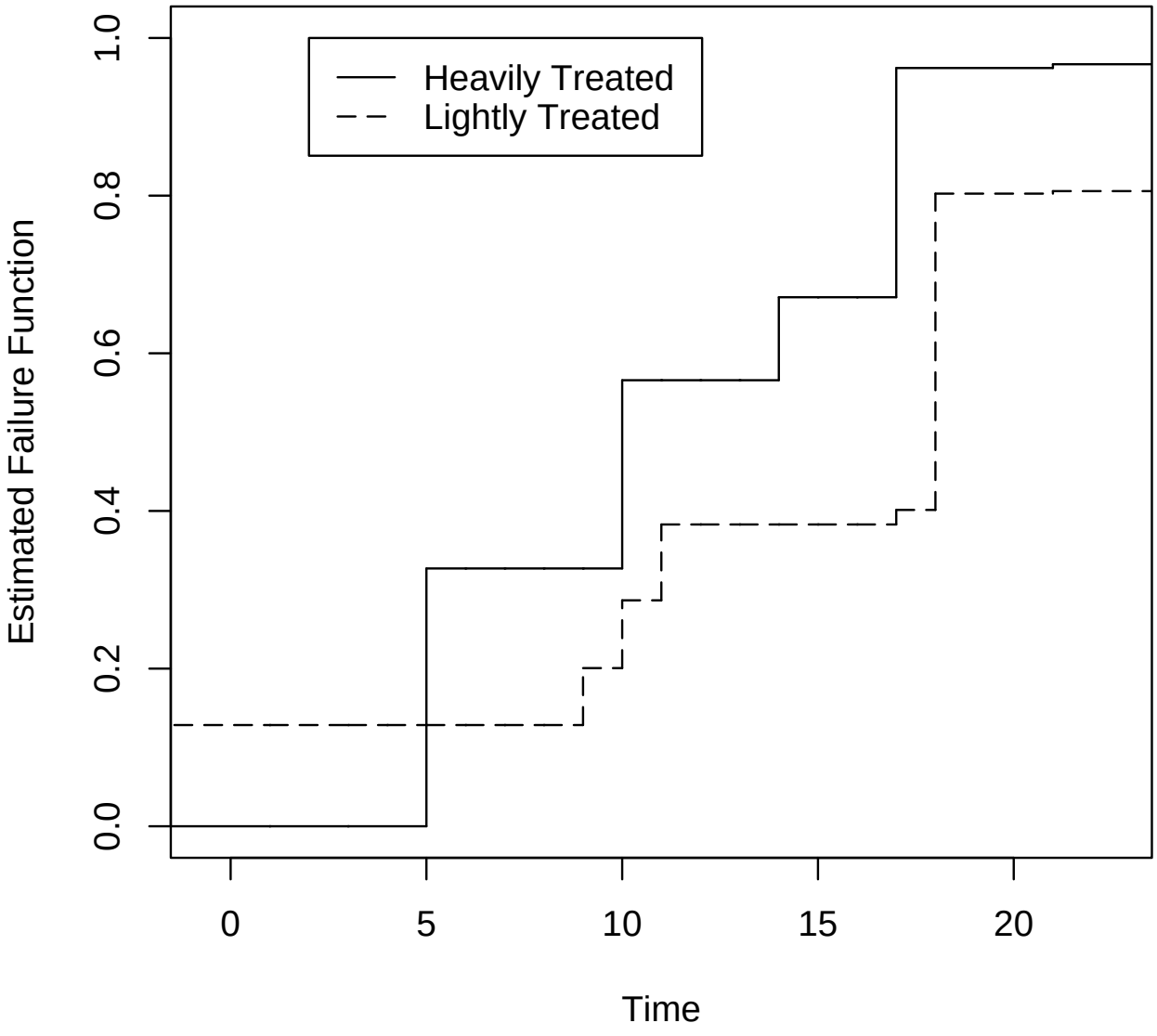


Figure 7: Nonparametric Estimators of the Distribution Functions of Time to HIV Infection for the AIDS Cohort Study

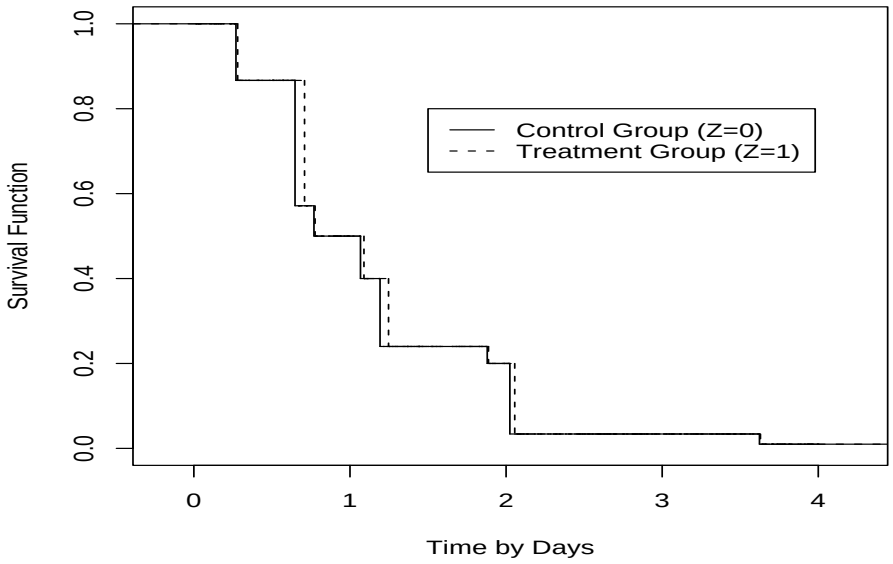


Figure 8: Survival Functions for $\beta_1 = 0$, $\beta_2 = 0$

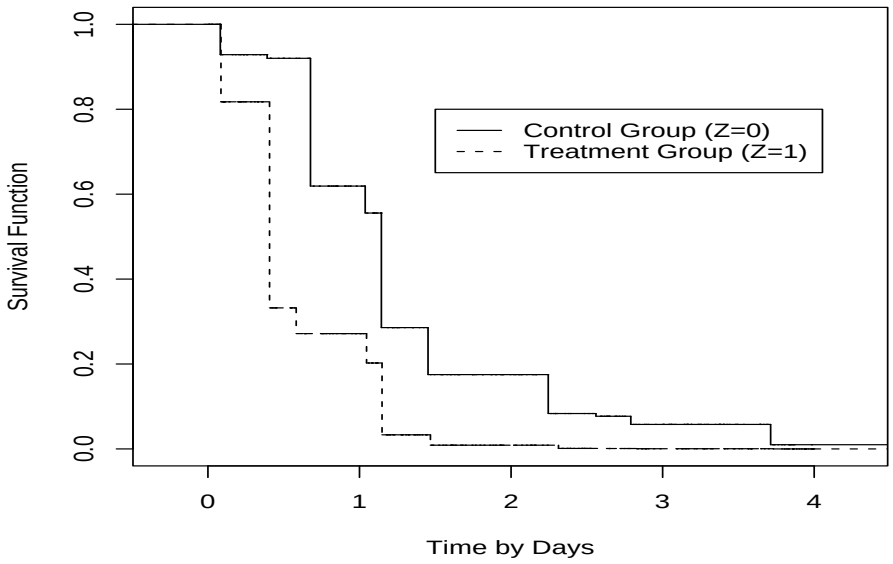


Figure 9: Survival Functions for $\beta_1 = 0$, $\beta_2 = 1$

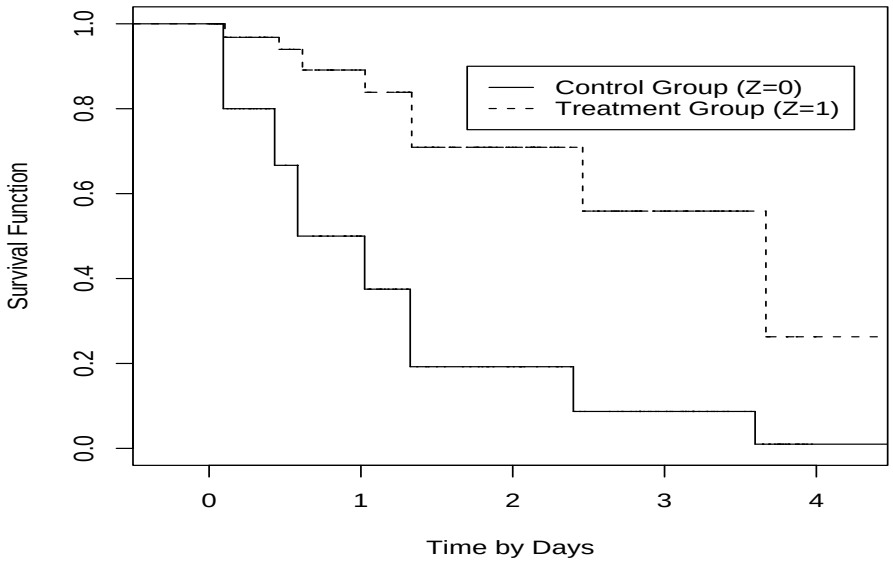


Figure 10: Survival Functions for $\beta_1 = -1$, $\beta_2 = -1$

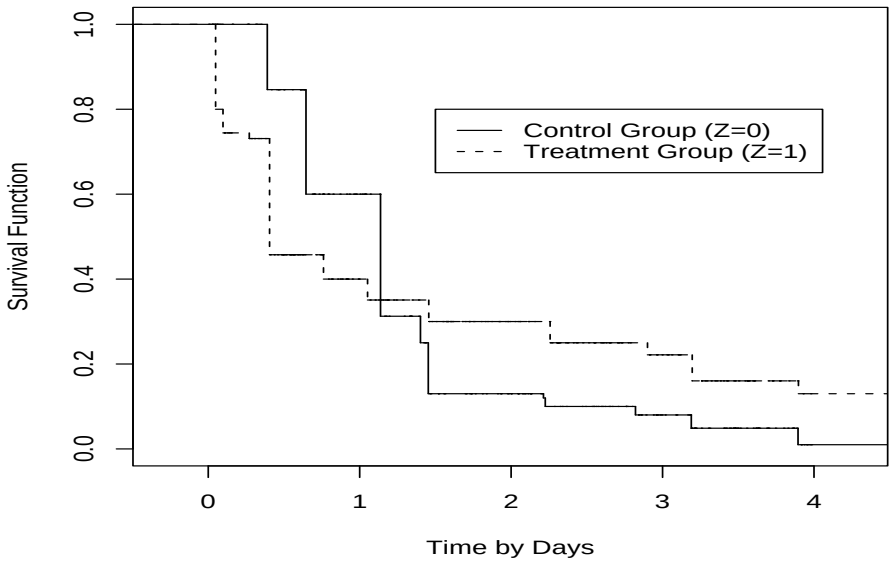
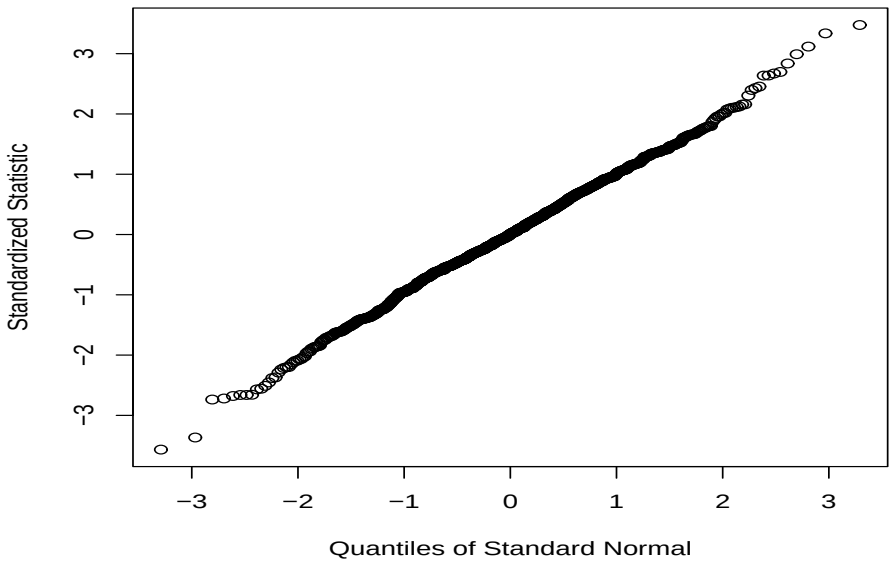


Figure 11: Survival Functions for $\beta_1 = 2$, $\beta_2 = -1$



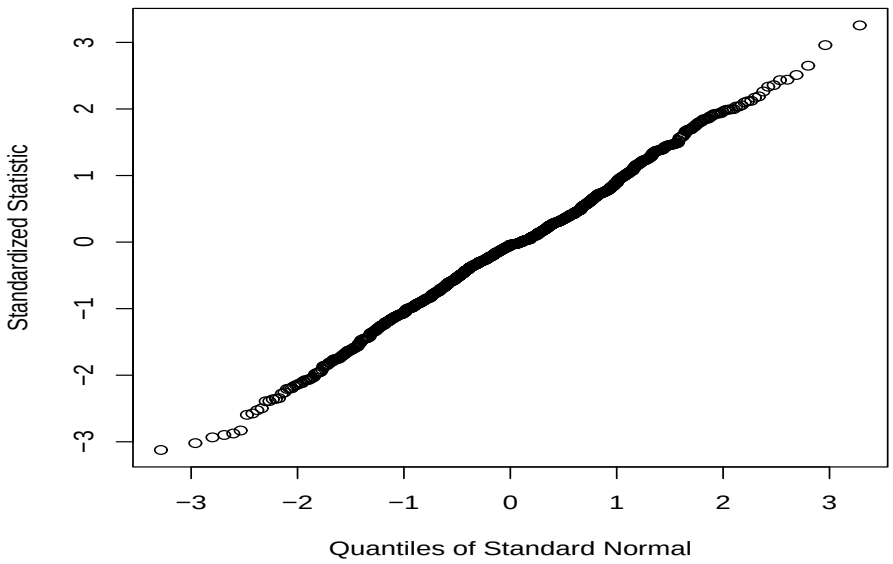


Figure 12: Normal Quantile Plots for $\beta_1 = 2$ (top) and $\beta_2 = -1$ (bottom)

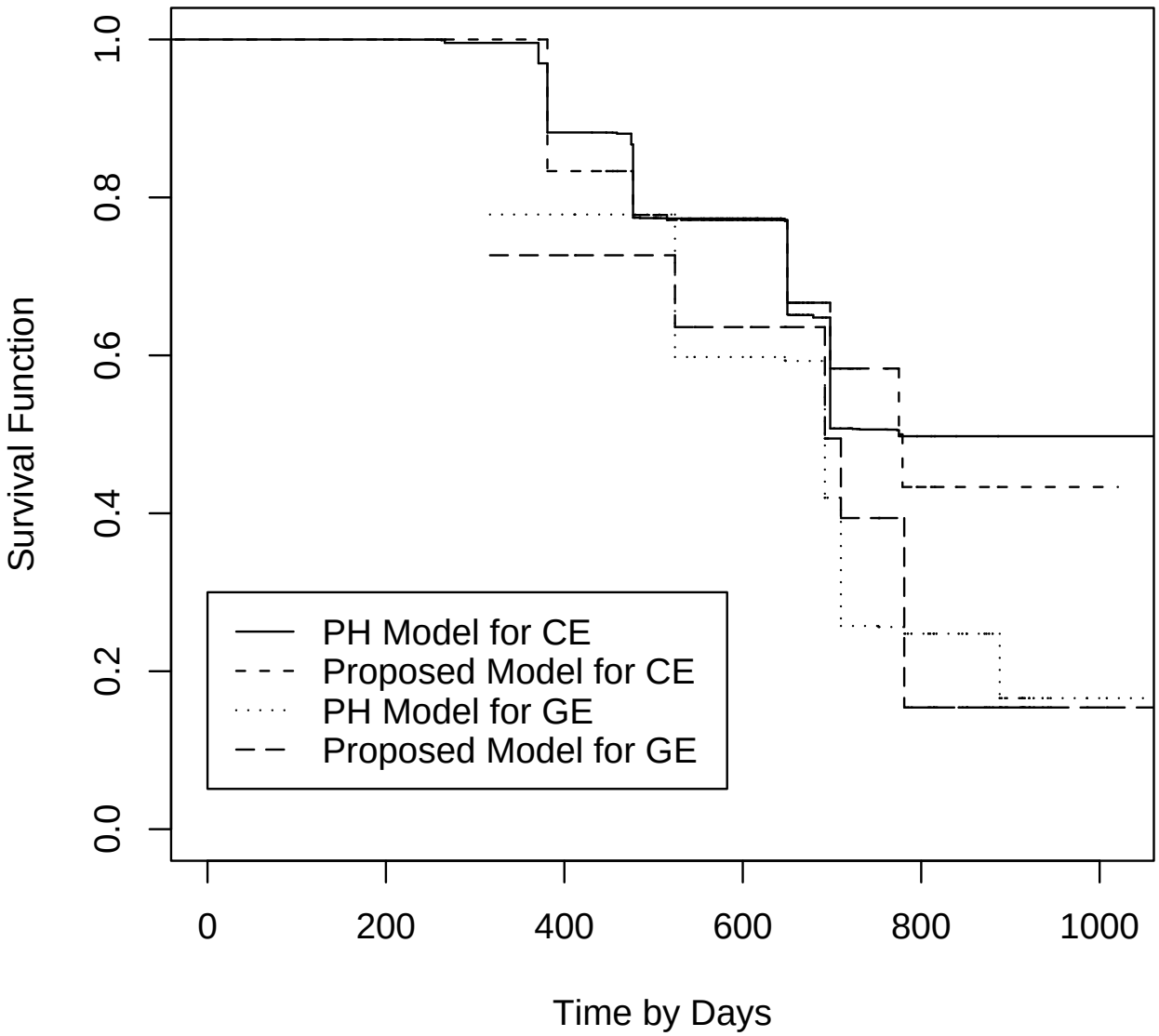


Figure 13: Estimates of Survival Functions of Time to Lung Tumor Onset

VITA

Chao Zhu was born on August 2, 1979, in Shanghai, P.R. China. After attending public schools in Anhui Province, he received the following degrees: B.S. in Information Management and Information Systems from the University of Science and Technology of China (2001); Master of Arts (2003) and Ph.D. (2006) in Statistics from the University of Missouri-Columbia. In July 2006, he will join the Bristol-Myers Squibb Company as a Research Biostatistician in Wallingford, Connecticut.