UNIVERSITI PUTRA MALAYSIA

PARAMETRIC AND NONPARAMETRIC INFERENCE FOR PARTLY INTERVAL-CENSORED FAILURE TIME DATA

AZZAH MOHAMMAD ALHARPY

FS 2013 53

PARAMETRIC AND NONPARAMETRIC INFERENCE FOR PARTLY INTERVAL-CENSORED FAILURE TIME DATA

AZZAH MOHAMMAD ALHARPY

DOCTOR OF PHILOSOPHY
UNIVERSITI PUTRA MALAYSIA

2013

PARAMETRIC AND NONPARAMETRIC INFERENCE FOR PARTLY INTERVAL-CENSORED FAILURE TIME DATA

By

AZZAH MOHAMMAD ALHARPY

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

September 2013

COPYRIGHT

DEDICATIONS

To

My late father

Who has supported me all the way, May Allah rest his soul in heaven

My lovely mother

For her unlimited love, care and support

My brothers and sisters

For their great encouragement and support

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

PARAMETRIC AND NONPARAMETRIC INFERENCE FOR PARTLY INTERVAL-CENSORED FAILURE TIME DATA

By

AZZAH MOHAMMAD ALHARPY

September 2013

Chair: Noor Akma Ibrahim, PhD

Faculty: Faculty of Science

Survival analysis is used in many fields for analysis of data, particularly in medical and biological science. In this context the event of interest is often death, the onset of disease or the disappearance of disease's symptoms. The time to event is called failure time, and this failure time may be observed exactly and recorded or may occurred between two inspection times. Data that include both exact failure data and interval-censored data is called partly interval-censored data. This phenomenon often happens in clinical trials and health studies that are followed by periodic follow-ups. Comparison of survival functions is one of the main objectives in survival studies. Thus, this thesis focuses on the aspect of inferential comparison problem for survival functions in the existence of partly interval-censored failure time data. The research is divided into two parts, parametric and nonparametric inferences.

The parametric maximum likelihood estimator, and a score test and likelihood

ratio test for this kind of failure time data are constructed under Weibull distributions by using direct approach (without imputation) and indirect approach (with multiple imputation technique).

The nonparametric maximum likelihood estimator and the development of nonparametric test approach for comparison of survival function of two samples or more in the existence of partly interval-censored failure time data are constructed where the Turnbull self-consistency equation is modified and then subsequently used in the multiple imputation technique.

The behavior of parametric and nonparametric maximum likelihood estimators, and the development of parametric and nonparametric tests approach for comparison of survival function of two samples in the existence of this type of censored data are also studied under the non-proportional hazard by using Piecewise exponential distribution.

Simulation studies are carried out to assess the performance of the method and approach that have been developed. The simulation results indicate that the developed tests statistics work well and the good points of a certain method depend on a specific situation. A modified secondary data set from breast cancer study has been used to illustrate the proposed tests.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENTAKBIRAN BERPARAMETER DAN TAK BERPARAMETER BAGI DATA MASA KEGAGALAN TERTAPIS-SELANG SEBAHAGIAN

Oleh

AZZAH MOHAMMAD ALHARPY

September 2013

Pengerusi: Noor Akma Ibrahim, PhD

Fakulti: Fakulti Sains

Analisis mandirian digunakan dalam pelbagai bidang untuk menganalisis data terutamanya dalam bidang sains perubatan dan biologi. Dalam konteks ini peristiwa yang sering menjadi tumpuan adalah kematian, permulaan penyakit atau kehilangan gejala penyakit. Masa sehingga peristiwa ini berlaku dikenali sebagai masa kegagalan dan masa kegagalan ini boleh dicerap dengan tepat dan direkodkan atau mungkin boleh berlaku diantara dua pemeriksaan. Data yang merangkumi kedua-dua jenis data kegagalan yang tepat dan juga tertapis selang dipanggil data tertapis-selang sebahagian. Fenomena ini sering berlaku dalam ujian klinikal dan kajian kesihatan yang disusuli dengan susulan berkala. Perbandingan diantara fungsi mandirian merupakan objektif utama dalam kajian mandirian. Yang demikian tumpuan tesis ini adalah untuk melihat dari aspek pentakbiran terhadap masalah perbandingan fungsi mandirian dengan kehadiran data masa kegagalan tertapis-selang sebahagian. Penyelidikan ini terbahagi kepada dua bahagian iaitu pentakbiran berparameter dan tak berparameter.

Penganggar kebolehjadian maksimum berparameter, ujian skor dan ujian nisbah kebolehjadian bagi data masa kegagalan jenis ini dibina berdasarkan taburan Weibull dengan menggunakan pendekatan secara langsung (tanpa imputasi) dan pendekatan secara tak langsung (teknik imputasi berganda).

Penganggar kebolehjadian maksimum tak berparameter dan pengembangan pendekatan ujian tak berparameter untuk perbandingan fungsi mandirian dua sampel atau lebih dengan kehadiran data masa kegagalan tertapis-selang sebahagian dibina dengan mengubahsuai persamaan kosistenan-kendiri Turnbull, seterusnya digunakan dalam teknik imputasi berganda.

Tingkah laku penganggar kebolehjadian maksimum berparameter dan tak berparameter dan perkembangan pendekatan ujian berparameter dan tak berparameter bagi membandingkan fungsi mandirian dua sampel dengan kehadiran data tertapis jenis ini dikaji juga dibawah model bahaya tak berkadaran menggunakan taburan eksponen cebis demi cebis.

Kajian simulasi telah dijalankan untuk menilai keupayaan kaedah dan pendekatan yang telah dibangunkan. Keputusan simulasi menunjukkan yang ujian statistik yang dibangunkan berfungsi dengan baik dan kelebihan sesuatu kaedah bergantung kepada situasi yang tertentu. Satu set data sekunder kanser payudara diubahsuai dan digunakan untuk mengilustrasi ujian yang telah dibangunkan.

ACKNOWLEDGMENTS

First and foremost, I am grateful to ALLAH (s. w. t.) for giving me the strength and patience to complete this thesis.

My deepest gratitude goes to my father Mohammad. To my great sadness, my father passed away on 15^{th} May 2009 due to his illness. He was a very supportive and encouraging person, who came with me to Malaysia and helped me immensely to start my study, which I hope will have significant impact on my future career. I hope he is proud of this thesis.

As far as academic goes, I would like to express my deep appreciation and sincere thanks to my supervisor, Prof. Dr. Noor Akma Ibrahim for her invaluable guidance, wise counsels and her continuous assistance. Her undiminished support throughout the many challenging stages of this work undoubtedly made this project possible.

My appreciation and sincere thanks also to Assoc. Prof. Dr. Jayanthi Arasan and Dr. Isa Daud as members of my supervisory committee for their helpful comments and cooperation.

To my family, I am outmost grateful to my beloved mother for her unlimited love, care, support and encouragement, without my mother and her prayers I would not be at this stage of my career.

My special thanks go to my brothers (Ali, Atiah, Hulayyil, Fahad, Abdullah) and my sisters for their sacrifices and care. Also my thanks go to all my relatives and

friends for their encouragement.

My final thanks go to Taibah University, Al Madinah Al Munawwarah, Saudi Arabia, for offering me the opportunity to complete my studies in Malaysia.

I certify that a Thesis Examination Committee has met on 20 September 2013 to conduct the final examination of Azzah Mohammad Alharpy on her thesis entitled "Parametric and Nonparametric Inference for Partly Interval-Censored Failure Time Data" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Habshah binti Midi, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Mohd Rizam bin Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Yong Zulina Zubairi, PhD
Associate Professor
University of Malaya
Malaysia
(External Examiner)

Emmanuel Lesaffre, PhD
Professor
University Rotterdam
Holland
(External Examiner)

NORITAH OMAR, PhD
Associate Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 20 November 2013

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy.

The members of the Supervisory Committee were as follows:

Noor Akma Ibrahim, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Isa Bin Daud, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

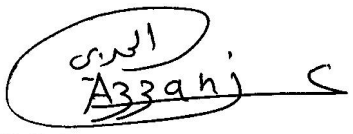
Jayanthi Arasan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.



AZZAH MOHAMMAD ALHARPY

Date: 20 September 2013

TABLE OF CONTENTS

		Page
ABSTRACT		ii
ABSTRAK		iv
ACKNOWLEDGMENTS		vi
APPROVAL		viii
DECLARATION		x
LIST OF TABLES		xiv
LIST OF FIGURES		xviii
LIST OF ABBREVIATIONS		xix
CHAPTER		
1	**INTRODUCTION**	1
	1.1 Basic Formulation in Survival Analysis	2
	1.2 Types of Interval-Censored Data	4
	1.3 Independent Interval Censoring	6
	1.4 Proportional and Non-proportional Hazards Model	7
	1.5 Weibull Distribution	9
	1.6 Multiple Imputation	10
	1.7 Problem Statement	13
	1.8 Objectives	14
	1.9 Outline of Thesis	15
2	**LITERATURE REVIEW**	17
	2.1 General Review	17
	2.2 Parametric Test for Interval-Censored Data	19
	2.3 Turnbull's Algorithm	22
	2.4 Log-rank Test for Exact Data	25
	2.5 Generalized Log-Rank Test for Interval-Censored Failure Time Data	26
	2.6 Summary	31
3	**PARAMETRIC TEST FOR PARTLY INTERVAL-CENSORED DATA**	33
	3.1 Introduction	33
	3.2 Partly Interval-Censored Data	33
	3.3 Analysis Using the Weibull Model	35
	3.4 PMLE for Partly Interval-Censored Data	36
	3.5 Generalized Score Test for Partly Interval-Censored Data	37
	3.5.1 Generalized Score Test Without Multiple Imputation Technique	38

3.5.2 Generalized Score Test Via Multiple Imputation Technique 39
3.6 Generalized Likelihood Ratio Test for Partly Interval-Censored Data 41
3.6.1 Generalized Likelihood Ratio Test Without Multiple Imputation Technique 41
3.6.2 Generalized Likelihood Ratio Test Via Multiple Imputation Technique 42
3.7 Simulation Studies 42
3.7.1 Simulation A 43
3.7.2 Results and Discussion A 45
3.7.3 Simulation B 57
3.7.4 Results and Discussion B 57
3.8 An Application 64
3.9 Conclusions 66

4 **GENERALIZED LOG-RANK TEST FOR PARTLY INTERVAL-CENSORED DATA BY USING MULTIPLE IMPUTATION TECHNIQUE** 68
4.1 Introduction 68
4.2 Generalized Log-Rank Test 68
4.2.1 Assumptions 69
4.2.2 NPMLE for Partly Interval-Censored Data 69
4.2.3 A Method Based on Multiple Imputation 72
4.3 Simulation Studies 74
4.3.1 Simulation A 74
4.3.2 Results and Discussion A 75
4.3.3 Simulation B 82
4.3.4 Results and Discussion B 83
4.4 Comparison of the Generalized Log-Rank Test and Huang's Test under Party Interval-Censored Data 87
4.5 An Application 91
4.6 Concluding Remarks 92

5 **TEST PROCEDURES FOR PARTLY INTERVAL-CENSORED DATA UNDER NON-PROPORTIONAL HAZARDS MODEL** 94
5.1 Introduction 94
5.2 Piecewise Exponential Model 94
5.3 Parametric Test 96
5.4 Nonparametric Test 103
5.5 Simulation Studies 104
5.5.1 Simulation A 104
5.5.2 Results and Discussion A 105
5.5.3 Simulation B 115
5.5.4 Results and Discussion B 116
5.6 An application 123
5.7 Concluding Remarks 123

6 **GENERAL CONCLUSION AND RECOMMENDATION FOR FUTURE RESEARCH** 125
6.1 Introduction 125
6.2 Summary of Results 125
6.3 Contribution 128
6.4 Recommendations for Future Research 128

BIBLIOGRAPHY 130
APPENDICES 136
BIODATA OF STUDENT 354
LIST OF PUBLICATIONS 355

LIST OF TABLES

Table		Page
3.1	Simulation results for Weibull distribution with parameters (1.5,6) under a direct approach with 50% exact observations based on 500 replications and different sample sizes.	46
3.2	Simulation results for Weibull distribution with parameters (0.6,3) under a direct approach with 50% exact observations based on 500 replications and different sample sizes.	47
3.3	Simulation results for Weibull distribution with parameters (1,5) under a direct approach with 50% exact observations based on 500 replications and different sample sizes.	48
3.4	Simulation results for Weibull distribution with parameters (1.5,6) under an indirect approach with 50% exact observations based on 500 replications and different sample sizes.	50
3.5	Simulation results for Weibull distribution with parameters (0.6,3) under an indirect approach with 50% exact observations based on 10 imputations, 500 replications and different sample sizes.	51
3.6	Simulation results for Weibull distribution with parameters (1,5) under an indirect approach with 50% exact observations based on 10 imputations, 500 replications and different sample sizes.	52
3.7	Simulation results of relative efficiency to compare between two estimators (parametric with direct and parametric with indirect) under Weibull distribution with parameters (1.5,6) with different percentage of exact observations based on 10 imputations, 500 replications, 100 sample size and different $\acute{q}$.	54
3.8	Simulation results of relative efficiency to compare between two estimators (parametric with direct and parametric with indirect) under Weibull distribution with parameters (0.6,3) with different percentages of exact observations based on 10 imputations, 500 replications, 100 sample size and different $\acute{q}$.	55
3.9	Simulation results of relative efficiency to compare between two estimators (parametric with direct and parametric with indirect) under Weibull distribution with parameters (1,5) with different percentages of exact observations based on 10 imputations, 500 replications, 100 sample size and different $\acute{q}$.	56

3.10 Estimated power and size of test with 30% exact observations based on simulation data with 10 multiple imputation and 1000 replications at the significance level 0.05 59

3.11 Estimated power and size of test with 50% exact observations based on simulation data with 10 multiple imputations and 1000 replications at the significance level 0.05 62

3.12 Time to cosmetic deterioration (in months) in breast cancer patients with two treatments. 65

4.1 Bias, Var and MSE of nonparametric estimate of survival function using Weibull distribution with parameters (1.5,6) and 50% exact observations based on 500 replications and different sample sizes. 76

4.2 Bias, Var and MSE of nonparametric estimate of survival function using Weibull distribution with parameters (0.6,3) and 50% exact observations based on 500 replications and different sample sizes. 77

4.3 Bias, Var and MSE of nonparametric estimate of survival function using Weibull distribution with parameters (1,5) and 50% exact observations based on 500 replications and different sample sizes. 78

4.4 Simulation results of relative efficiency to compare between two estimates (parametric with indirect and nonparametric) of survival function under Weibull distribution with parameters (1.5,6) 79

4.5 Simulation results of relative efficiency to compare between two estimates (parametric with indirect and nonparametric) of survival function under Weibull distribution with parameters (0.6,3) 80

4.6 Simulation results of relative efficiency to compare between two estimates (parametric with indirect and nonparametric) of survival function under Weibull distribution with parameters (1,5) 81

4.7 Estimated power and size of test with 30% exact observations based on simulation data with 10 multiple imputations and 10,000 replications at the significance level 0.05 (0.01) 84

4.8 Estimated power and size of test with 50% exact observations based on simulation data with 10 multiple imputations and 10,000 replications at the significance level 0.05 (0.01) 85

4.9 Estimated power and size of test with 30% exact observations based on simulation data with 10 multiple imputations and 10,000 replications at the significance level 0.05. The values in the parentheses represent Huang's test. 89

4.10 Estimated power and size of test with 50% exact observations based on simulation data with 10 multiple imputation and 10,000 replications at the significance level 0.05. The values in the parentheses represent Huang's test. 90

5.1 Bias, Var and MSE of parametric estimate of survival function for piecewise exponential distribution with one cutpoint with 50% exact observations based on 500 replications and different sample sizes. 106

5.2 Bias, Var and MSE of parametric estimate of survival function for piecewise exponential distribution with two cutpoints with 50% exact observations based on 500 replications and different sample sizes. 107

5.3 Bias, Var and MSE of nonparametric estimate of survival function for piecewise exponential distribution with one cut point with 50% exact observations based on 500 replications and different sample sizes. 109

5.4 Bias, Var and MSE of nonparametric estimate of survival function for piecewise exponential distribution with two cutpoints with 50% exact observations based on 500 replications and different sample sizes. 110

5.5 MSE of parametric estimate of survival function for piecewise exponential distribution with one cutpoint with different percentages of exact observations based on 10 imputations, 500 replications, 100 sample size and different values of $\acute{q}$. 111

5.6 MSE of parametric estimate of survival function for piecewise exponential distribution with two cutpoints with different percentages of exact observations based on 10 imputations, 500 replications, 100 sample size and different values of $\acute{q}$. 112

5.7 MSE of nonparametric estimate of survival function for piecewise exponential distribution with one cutpoint with different percentages of exact observations based on 10 imputations, 500 replications, 100 sample size and different values of $\acute{q}$. 113

5.8 MSE of nonparametric estimate of survival function for piecewise exponential distribution with two cutpoints with different percentages of exact observations based on 10 imputations, 500 replications, 100 sample size and different values of $\acute{q}$. 114

5.9 Estimated power and size of test with 30% exact observations based on simulation data with 10 multiple imputations and 1000 replications at the significance level 0.05 (0.01) 118

5.10 Estimated power and size of test with 50% exact observations based on simulation data with 10 multiple imputations and 1000 replications at the significance level 0.05 (0.01) 120

5.11 Estimated power and size of test with 30% exact observations based on simulation data with 10 multiple imputations and $10,000$ replications at the significance level 0.05 (0.01) 121

5.12 Estimated power and size of test with 50% exact observations based on simulation data with 10 multiple imputations and $10,000$ replications at the significance level 0.05 (0.01) 122

LIST OF FIGURES

Figure		Page
1.1	Density Functions for Different Weibull Distribution.	10
1.2	Hazard Functions for Different Weibull Distribution.	11
3.1	PMLE of Survival Function of Time to Cosmetic Deterioration.	66
4.1	Q-Q Plot for Generalized Log-Rank Test.	86
4.2	NPMLE of Survival Function of Time to Cosmetic Deterioration.	92
5.1	Different Configurations of Hazards.	117
5.2	PMLE of Survival Function of Time to Cosmetic Deterioration under Piecewise Exponential Destribution.	124

LIST OF ABBREVIATIONS

LRT	Likelihood Ratio Test
LRTD	Likelihood Ratio Test with Direct Approach
LRTI	Likelihood Ratio Test with Indirect Approach
MSE	Mean Square Erore
NPE	Nonparametric Estimate
NPMLE	Nonparametric Maximum Likelihood Estimator
PE	Parametric Estimate
PMLE	Parametric Maximum Likelihood Estimator
RE	Relative Efficiency
ST	Score Test
STD	Score Test with Direct Approach)
STI	Score Test with Indirect Approach)
W.I	With Imputation
W.O	Without Imputation

CHAPTER 1

INTRODUCTION

Survival analysis is a statistical method important in the analysis of life time, particularly in medical and biological sciences. The outcome variable of interest is time recorded for an event's occurence. In the context of medical and biological studies, the event of interest is often death, the onset of a disease or the disappearance of a disease's symptoms. The time to event of interest is called either survival time or failure time and the probability that the subject survives beyond a specified time is calculated by a basic formula called "survival function".

One of the complications which arises in survival analysis is the presence of censored data. Censoring occurs when the information about the failure time of some subjects is incomplete. Different circumstances can produce many types of censored data, including right-censored, left-censored and interval-censored data. Interval-censored data arises when the event of interest can not be immediately observed and it is only known to have appeared through a random interval of time.

Survival analysis with interval-censored data has been developed over the past three decades and written research is extensive. For example, Peto and Peto (1972) discussed partly interval-censored data whereby they treated their exact data as an interval-censored data, separating exact observation times by very short intervals. In an article by Peto (1973), the procedure was only to estimate the distribution function when data are interval-censored.

Turnbull (1976) described a scheme of censored failure time data and derived the self-consistency equation for computing the maximum likelihood estimator of survival function. Huang and Wellner (1995) proved the asymptotic normality of a

class of linear functions of the nonparametric maximum likelihood estimator of a distribution function with case I interval-censored data.

Interval-censored data are divided into many subcategories, including case I interval-censored data, case II interval-censored data, doubly interval-censored, mixed interval-censored and partly interval-censored. Partly interval-censored data is an important subcategory of interval-censored data. It arises in medical and health studies, which entail periodic follow-ups. Partly interval-censored data originates from an event of interest that is observed directly for some subjects, but for remaining subjects, the exact time of the event is unknown, except that it falls within a specific time interval.

1.1 Basic Formulation in Survival Analysis

The basic quantity employed to describe failure time phenomena is the survival function, the probability of an individual's survival beyond time t. It is defined as

$$S(t) = P(T > t) \tag{1.1}$$

where T is a non-negative random variable denoting the failure time.
If T is a continuous random variable, the survival function is the complement of a cumulative distribution function, that is,

$$S(t) = 1 - F(t) \tag{1.2}$$

where $F(t) = P(T \leq t)$. Furthermore, the survival function is the integral of the probability density function, that is,

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du \tag{1.3}$$

thus

$$f(t) = -\frac{d\,S(t)}{dt} \quad . \tag{1.4}$$

When T is a discrete random variable a different technique is required. Discrete random variables in survival analysis created due to rounding off measurements, collection of failure times into intervals, or when lifetime refers to an integral number of units. Let T be the discrete random variable taking values, where $0 = t_0 < t_1 < t_2 < \cdots$, with probability mass function $p(t_j) = P(T = t_j)$ $j = 1, 2, \cdots,$. Therefore, the survival function of T is

$$S(t) = P(T > t) = \sum_{t_j > t} p(t_j) \quad . \tag{1.5}$$

Additionally, the hazard function and the cumulative hazard function of T are also fundamental in survival analysis. Hazard function is also known as the conditional failure rate in reliability and can be defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | \, T \geq t) \quad . \tag{1.6}$$

If T is a continuous random variable, then

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \ln[S(t)]}{dt} \quad . \tag{1.7}$$

A related function is the cumulative hazard function $H(t)$, defined by

$$H(t) = \int_o^t h(u)\, du = -\ln[S(t)] \quad . \tag{1.8}$$

It is easy to see that

$$S(t) = \exp[-H(t)] = \exp[-\int_0^t h(u)\, du] \quad . \tag{1.9}$$

When T is a discrete random variable, the hazard function is given by

$$h(t_j) = P(T = t_j | \, T \geq t_j) = \frac{p(t_j)}{S(t_{j-1})} \quad j = 1, 2, \ldots . \tag{1.10}$$

And the cumulative hazard function $H(t)$ is defined as

$$H(t) = \sum_{t_j \leq t} h(t_j) \quad . \tag{1.11}$$

Note that, $S(t)$, $h(t)$ or $H(t)$ uniquely determines the distribution of T.

1.2 Types of Interval-Censored Data

Interval-censored data is one of the obstacles which arises in survival analysis. There are several different types of interval-censored data. They are as follows:

1. Case I interval-censored data, also called current status data, arises when each individual is subjected to observation only at a single follow-up time, and thus, the event of interest (failure) is only observed either to have or have not occurred before the observation time. That is, the failure time of interest is either left- or right-censored data respectively (Keiding, 1991; Groeneboom and Wellner, 1992; Koul and Yi, 2006).

   Case I interval-censored data usually occur in tumorigenicity tests. In these tests, the tumor start time of animals is commonly of prime interest but not observable. Instead, tumor status is commonly known at death (either natural death or euthanization for scientific study). Thus, the tumor start time is expected only to be less or greater than the death time. There are many authors who discussed the current status data arising from survival studies such as Huang and Wellner (1995); Huang (1996), Rossini and Tsi-

atis (1996), Lin et al. (1998), Shen (2000), Ghosh (2001), Martinussen and Scheike (2002), Xue et al. (2004) and Sun (2006).

2. Case II interval-censored data, also called general interval-censored data, is defined as data which refers to a situation when the event of interest cannot be immediately observed and is only known to have appeared through a random interval of time. Left (right) censoring is a special case of interval censoring in which the left (right) end point is $0(\infty)$.

   Case II interval-censored data arises in several of medical and health studies. For example, in a study which compares time to cosmetic deterioration of breasts for breast cancer patients treated with radiotherapy and radiotherapy plus chemotherapy, patients were examined at each clinical visit for breast retraction. The breast retraction is only known to take place between two clinical visits or right censored at the end of the study. The objective of the study is to compare the patients who received adjuvant chemotherapy to those who did not and to decide whether chemotherapy affects the rate of deterioration of the cosmetic state (Finkelstein, 1986; Pan, 2000; Lim and Sun, 2003; Huang et al., 2008).

3. Doubly interval-censored data refers to the survival time of interest, which is the elapsed time between two related events called the initial and the end events, and the observations on the occurrences of both events could be interval-censoring (Gruttola and Lagakos, 1989; Sun, 1997, 2001, 2004).
   If X is the time to the initial event in which $X \in (L_X, R_X]$ and Y is the time to the end event in which $Y \in (L_Y, R_Y]$, then the random variable $T = Y - X$ corresponds to the survival time of interest. Doubly interval-censored data is reduced to right-censored or interval-censored data if the occurence of the initial event was observed exactly (Deng et al., 2009). This kind of data

often arise in many fields such as biometry studies and reliability research. The articles that addressed the doubly interval-censored data arising from survival studies include Kim et al. (1993), Gomez and Lagakos (1994), Li and Yu (1997), Gomez and Call (1999), Fang and Sun (2001), Sun et al. (2004), Sun (2006) and Zhang et al. (2009).

4. Mixed interval-censored data refers to the survival time of interest, which is observed either to belong to an interval, or to be in right-censoring (Zhao and Sun, 2004).

5. Partly interval-censored data arise when the exact failure times are observed of some subjects, but for the remaining subjects, the failure time of interest is not observable, but is only known to be bracketed between two examination times (Huang, 1999). An example of this type of partly interval-censored data is presented by the Framingham Heart Disease study. In this study, times of the first occurrence of the subcategory angina pectoris in coronary heart disease patients are of interest. For some patients, the event time is recorded precisely, but for the remaining patients, time is recorded only between two clinical examinations (see Feinleib et al., 1975; Odell et al., 1992). More details about this kind of data is presented in Chapter 3.

1.3 Independent Interval Censoring

Independent interval censoring is a factual assumption in survival analysis. Independent interval censoring is the condition whereby the method that generates the censoring is independent of the subject's failure time distribution. For instance, T is failure time of interest and L and R are the two observed values such that $T \in (L, R]$. Then the independent censoring process for interval-censored data can

be expressed by

$$P(L < T \leq R | L = l, R = r) = P(l < T \leq r) \quad .$$

. This means the joint survival function of the two observed values L and R is free from any parameters contributory in the survival function of T. More importantly, it should be noted that the independent interval censoring is non-informative interval censoring while the opposite is not always true (Betensky, 2000; Oller et al., 2004; Sun, 2006).

1.4 Proportional and Non-proportional Hazards Model

Proportional hazards model is a common semi-parametric regression model used for analyzing survival data, proposed by Cox (1972) to examine the effect of predictor variable on survival time. Proportional hazards model has extensively been used in medical testing analysis and reliability engineering.

Proportional hazards model is usually written in terms of the hazard model, (see Kleinbaum, 1996).

$$h(t, \mathbf{X}) = h_0(t)\, e^{\sum_{i=1}^{p} \beta_i X_i} \tag{1.12}$$

where $h_0(t)$ is called baseline hazard and $e^{\sum_{i=1}^{p} \beta_i X_i}$ is called the exponential. If the baseline hazard is specified, the model is called parametric model, but if the baseline hazard is unspecified, the model is called nonparametric model.

The proportional hazards model (Cox and Oakes, 1984) is expressed by

$$h_1(t, \mathbf{X}) = \psi(\mathbf{X})\, h_0(t) \tag{1.13}$$

in which the explanatory vector $\mathbf{X}$ is constant over time for any subject and $h_0(t)$ is the hazard for the subject under the standard conditions, $\mathbf{X} = \mathbf{0}$ and we require $\psi(\mathbf{0}) = 1$. Then, the survival function under the proportional hazards model is given by

$$S_1(t, \mathbf{X}) = [S_0(t)]^{\psi(\mathbf{X})} \quad . \tag{1.14}$$

Here $S_0(t)$ is a survival function corresponding to the hazard function $h_0(t)$. If $\psi(\mathbf{X}) = 1$, there is no difference between the two survival curves, and if $\psi(\mathbf{X}) > 1$, then the subject with survival function 1 has a lower survival rate. Correspondingly, if $\psi(\mathbf{X}) < 1$, then the subject with survival function 1 has a higher survival rate.

Comparatively, in Cox's non-proportional hazards model, the hazard ratio is represented as a step function of time. The hazard is given by

$$h(t, x) = h_0(t)\, exp\,(\sum_{i=1}^{p}(\beta_i + \gamma_{ji})x_i) \tag{1.15}$$

where $j = (1, \cdots, r)$, $\gamma_{1i} = 0$. In non-proportional hazards model, the hazard ratio will be constant within each of the r pre-specified time intervals but change between the intervals. That means, the hazard ratio equals to $exp\,(\beta_i)$ in the first interval and $exp(\beta_i + \gamma_{ji})$ in the subsequent intervals for $j = 2, \cdots, r$ (Basar, 2006).

In this thesis we assume the parametric model is without covariate. This means the failure time satisfies a specific distribution . In the case of proportional hazards model we presume that the failure times follow Weibull distribution and the hazard ratio between two distributions is

$$\frac{h_2(t)}{h_1(t)} = exp(\beta), \quad \text{for all} \;\; t \geq 0.$$

This kind of model is applied in Chapter 3.

In the case of non-proportional hazards model we suppose that the failure times follow a piecewise exponential distribution and the hazard ratio between two distributions with two intervals is

$$\frac{h_2(t)}{h_1(t)} = exp\,(\beta_1)\; I_{(t \leq t_0)} + exp\,(\beta_2)\; I_{(t > t_0)}.$$

Where t_0 represents the cut point. If $\beta_1 = 0$ and $\beta_2 \neq 0$, then we have late hazard difference, and if $\beta_1 \neq 0$ and $\beta_2 = 0$, then we have early hazard difference. If $\beta_2 = -\beta_1$, then we have cross hazard. This kind of model is applied in Chapter 5.

1.5 Weibull Distribution

The Weibull distribution is a commonly used distribution for studying lifetime models, biological and medical sciences and reliability. Let T be a random variable following the Weibull distribution with shape parameter a and scale parameter b. The Weibull distribution density function (Raymond, 1977) is given by

$$f(t) = \frac{a}{b}(\frac{t}{b})^{a-1}\, exp\,[-(\frac{t}{b})^a] \quad t > 0, a > 0, b > 0 \tag{1.16}$$

Figure 1.1 shows the density functions for different Weibull distribution. The cumulative Weibull distribution function and survival function are given by

$$F(t) = 1 - exp\,[-(\frac{t}{b})^a] \quad t > 0, a > 0, b > 0 \tag{1.17}$$

and

$$S(t) = exp\,[-(\frac{t}{b})^a] \quad t > 0, a > 0, b > 0 \tag{1.18}$$

respectively.

The hazard function of Weibull distribution is given by

$$h(t) = \frac{a}{b}\,(\frac{t}{b})^{a-1} \quad t > 0, a > 0, b > 0 \tag{1.19}$$

Figure 1.2 shows the hazard functions for different Weibull distribution. It is easy to see that the Weibull distribution reduces to the exponential distribution and has constant hazard over time if the shape parameter $a = 1$. The hazard function increases over time if $a > 1$. The hazard function decreases over time if $a < 1$.

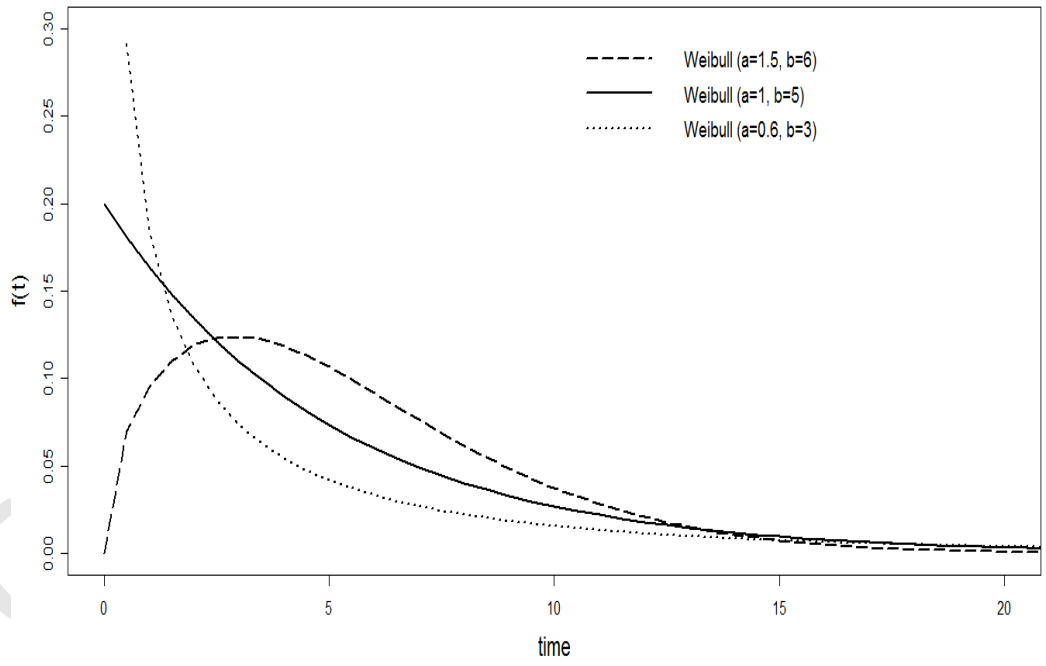


Figure 1.1: Density Functions for Different Weibull Distribution.

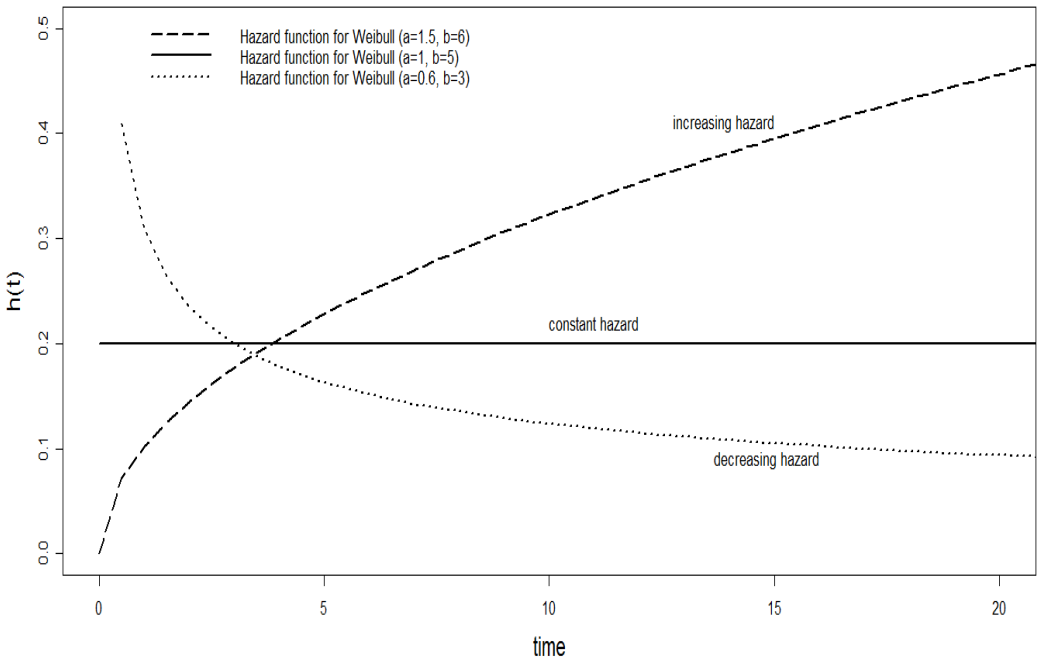


Figure 1.2: Hazard Functions for Different Weibull Distribution.

1.6 Multiple Imputation

Imputation method is one of common tools to solve missing values. Missing values arise in many fields of study. For example, in medical and health studies, it is required for the participants to undergo periodic follow-ups for the examination of characteristic related to condition of interest. The missing values permanently occur for several reasons. For handling the missing values there are various imputation methods such as single imputation, multiple imputation and others.

Single imputation is often applied because it is intuitively attractive. In single imputation, we fill in each missing value by predicted value. The obvious shortcoming in single imputation is that we replace the missing values by a single values and then treat it as if it were a true values. As a result, single imputation ignores uncertainty and always underestimates the variance. Multiple imputation rectifies

this shortcoming, by taking into account both within imputation uncertainty and between imputation uncertainty. In this thesis the focus will be on multiple imputation.

Multiple imputation technique was developed by Rubin (1987) as a general technique for handling data sets with missing values. Application of the technique requires three steps: imputation, analysis and pooling.

Imputation step: Impute (fill in) the missing value several times, creating many augmented data sets.

Analysis step: Analyze each augmented data set separately.

Pooling step: Integrate the analysis results into a final result.

More formally, suppose that we are interested in estimating an unknown parameter vector θ. Let $\hat{\theta}$ be an estimator of θ when the data set is complete and $\hat{\Sigma}$ is a covariance estimator associated with $\hat{\theta}$. Now, we use multiple imputation to obtain multiple sets (Y sets) of complete data from the given incomplete data set. Then, we will obtain Y estimates, $\left(\hat{\theta}_y, \hat{\Sigma}_y\right)$, $y = 1, \cdots, Y$.

Rubin's multiple imputation estimators are given by

$$\hat{\theta}_* = \sum_{y=1}^{Y} \frac{\hat{\theta}_y}{Y} \quad ,$$

where $\hat{\theta}_*$ is the estimate of θ based on the Y imputation, and

$$\hat{\Sigma}_* = \frac{1}{Y}\sum_{y=1}^{Y} \hat{\Sigma}_y + \left(1 + \frac{1}{Y}\right) \sum_{y=1}^{Y} \frac{\left(\hat{\theta}_y - \hat{\theta}_*\right)\left(\hat{\theta}_y - \hat{\theta}_*\right)^T}{Y - 1}$$

is the estimate of the covariance, where the first term is the average within the

imputation covariance associated with the estimate, and the second term is the between imputation covariance of the estimate multiplying by $\left(1+\frac{1}{Y}\right)$, where the factor $\left(1+\frac{1}{Y}\right)$ is an adjustment for using a finite number of imputations.

The multiple imputation method has been discussed in research analyzing interval-censored failure time data. This method is used to reduce the interval-censored data to right-censored data, which can be handled by using specified methods for right-censored data (for an example see Pan, 2000; Chen and Sun, 2010). This method is also used to reduce the interval-censored data to exact data, which can be handled by using specified methods for exact data (for an example see Huang et al., 2008). There are many authors who employed the imputation methods for interval-censored data such as Dorey et al. (1993), Satten et al. (1998), Betensky and Finkelstein (1999), Bebchuk and Betensky (2000) and Pan (2001).

1.7 Problem Statement

Many statistical approaches have been developed to solve the problems that arise in the survival analysis. One of these problems is comparison of survival distribution for two samples or more when data are incomplete. In most previously published research the comparison problem of survival distribution for two samples or more has been solved for cases with right-censored data and interval-censored data. Multiple imputation technique is one method that has been used to solve the comparison problem when the data are right-censored or interval-censored. In contrast, not many research has been considered for partly interval-censored data and it is still ongoing but so far limited. For example, Kim (2003) used the proportional hazard model for regression analysis of partly interval-censored data. Elfaki et al. (2012) studied the parametric Cox's proportional hazard model for partly interval-censored data. Both the articles discuss a parametric comparison

of survival function by fitting proportional hazard model regression. Zhao et al. (2008) discussed nonparametric test for partly interval-censored data and evaluated their proposed test under proportional hazard model using simulation studies.

It should be noted that the previous published researches were focused on parametric and nonparametric tests for proportional hazard model only and to our knowledge, there is no any application of multiple imputation to partly interval-censored data so far. This thesis is motivated by the idea that existing comparison problem for survival distribution of two samples or more in the presence of partly interval-censored data is needed to examine the impact of imputation on the results.

In this thesis, the focus will be on the problem of comparing survival distributions of two samples or more in the presence of partly interval-censored failure time data by using the multiple imputation technique under proportional and non-proportional hazards model. This research will be divided into two parts, the first part will be devoted to the parametric test for partly interval-censored data via multiple imputation. We will construct a parametric maximum likelihood estimator using Weibull distribution with partly interval-censored data in order to carry out the test under the proportional hazards model. Also, we will construct a parametric maximum likelihood estimator for piecewise exponential distribution with partly interval-censored data in order to implement the test under the non-proportional hazards model.

The second part of the research will be devoted to the nonparametric test for partly interval-censored data via multiple imputation. We will modify the self-consistency algorithm in terms of partly interval-censored data and we will construct the non-

parametric test in the presence of partly interval-censored data and analyze the generalized log-rank test via multiple imputation. We will then study the proposed test under proportional and non-proportional hazards model.

1.8 Objectives

The aim of this research is to establish parametric and nonparametric tests for comparing survival functions of two samples or more in the presence of partly interval-censored data via multiple imputation technique.
The main objectives of this study are as follows:

1. To establish the parametric and the nonparametric estimators of survival function in the presence of partly interval-censored data.

2. To develop a parametric and a nonparametric tests in the presence of partly interval-censored data via multiple imputation technique to address the comparison problem for two samples or more under proportional hazards model using Weibull distribution.

3. To develop a parametric and a nonparametric tests in the presence of partly interval-censored data via multiple imputation technique to address the comparison problem for a two-sample under non-proportional hazards model using piecewise exponential distribution.

4. To conduct a simulation study to assess the properties of the survival function estimates and to investigate the performance of the proposed tests under the proportional hazards model and non-proportional hazards model.

5. Application of the developed models to a modified real data set.

1.9 Outline of Thesis

This thesis is organized into six chapters. Chapter 2 provides a review of current related literature such as the research conducted on parametric and nonparametric tests which compare two samples or more in the presence of interval-censored failure time data.

Chapter 3 starts with the description for partly interval-censored data. Then, the parametric maximum likelihood estimator (PMLE) for survival function is constructed under Weibull distribution. Following that, parametric tests such as the score test and the likelihood ratio test are performed in the presence of the partly interval-censored data using two approaches, direct (without imputation) and indirect (with imputation). Conclusions are then recorded based on the simulation data and a modified secondary data set from breast cancer.

Chapter 4 studies a nonparametric test with partly interval-censored data via multiple imputation. Turnbull's self-consistency algorithm is modified in terms of partly interval-censored data to construct the nonparametric maximum likelihood estimator (NPMLE) for survival function. Then, the generalized log-rank test is carried out in the presence of partly interval-censored data using multiple imputation technique. Following that, a comparison is made between a generalized log-rank test and Huang's test under partly interval-censored data. Finally based on the simulation data and a modified secondary data set from breast cancer the conclusions are drawn.

Chapter 5 concentrates on the parametric and nonparametric tests under the non-proportional hazards model with partly interval-censored failure time data via multiple imputation technique. The PMLE for survival function is constructed under

piecewise exponential distribution in the presence of the partly interval-censored data. Then, the parametric tests such as the score test and the likelihood ratio test are implemented in the presence of partly interval-censored data using multiple imputation technique. Following that, a generalized log-rank test with partly interval-censored data is applied under non-proportional hazards model. Additionally, the simulation data and a modified secondary data set from breast cancer are used. Conclusions from this chapter are drawn.

Finally, Chapter 6 summarizes the study and suggests some recommendations for further research.

It should be mentioned that all simulation studies were implemented by using the R programming language.

BIBLIOGRAPHY

Andersen, P. K. and Ronn, B. B. 1995. A nonparametric test for comparing two samples where all observations are either left- or right-censored. . *Biometrics* 51: 323–329.

Basar, E. 2006. Non-Proportional Hazards with Application to Kidney Transplant Data. *Commun. Fac. Sci. Univ. Ank. Series A1* 55 (2): 55–63.

Bebchuk, J. D. and Betensky, R. A. 2000. Multiple imputation for simple estimation of the hazard function based on interval censored data. *Statistics in Medicine* 19: 405–419.

Betensky, R. A. 2000. On nonidentifiability and noninformative censoring for curent status data. *Biometrika* 87: 218–221.

Betensky, R. A. and Finkelstein, D. M. 1999. An extension of Kendall's coefficient of concordance to bivariate interval censored data. *Statistics in Medicine* 18: 3101–3109.

Chen, L. and Sun, J. 2010. A multiple imputation approach to the analysis of interval-censored failure time data with the additive hazards model. *Computational Statistics and Data Analysis* 54: 1109–1116.

Cox, D. R. 1972. Regression models and life tables (with discussion). . *Journal of the Royal Statistical Society: Series B* 34 (5): 187–220.

Cox, D. R. and Oakes, D. 1984. *Analysis of Survival Data*. London: Chapman and Hall.

Dempster, A. P., Laird, N. M. and Rubin, D. B. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1): 1–38.

Deng, D., Fang, H. and Sun, J. 2009. Nonparametric estimation for doubly censored failure time data. *Journal of Nonparametric Statistics* 21 (7): 801–814.

Dinse, G. E. 1994. A comparison of tumor incidence analyses applicable in single-sacrifice animal experiments. . *Statistics in Medicine* 13: 689–708.

Dorey, F. J., Little, R. J. A. and Schenker, N. 1993. Multiple imputation for threshold-crossing data with interval censoring. . *Statistics in Medicine* 12: 1589–1603.

Elfaki, F. A. M., Azram, M. and Usman, M. 2012. Parametric Cox's model for partly interval-censored data with application to AIDS studies. . *International Journal of Applied Physics and Mathematics* 2 (5): 352–354.

Fang, H. and Sun, J. 2001. Consistency of nonparametric maximum likelihood estimation of a distribution function based on doubly interval-censored failure time data. *Statistics and Probability Letters* 55: 311–318.

Fang, H., Sun, J. and Lee, M.-L. T. 2002. Nonparametric survival comparison for interval-censored continuous data. . *Statistica Sinica* 12: 1073–1083.

Fay, M. P. 1996. Rank invariant tests for interval-censored data under the grouped continuous model . *Biometrics* 52: 811–822.

Fay, M. P. 1999. Comparing several score tests for interval-censored data. *Statistics in Medicine* 18 (3): 273–285.

Fay, M. P. and Shih, J. H. 1998. Permutation tests using estimated distribution functions. . *Journal of the American Statistical Association* 93: 387–396.

Feinleib, M., Kannel, W. B. and Garrison, R. J. 1975. The Framingham Offspring Study: Design and preliminary data. *Preventive Medicine* 4: 518.

Finkelstein, D. M. 1986. A proportional hazards model for interval-censored failure time data. *Biometrics* 42 (4): 845–854.

Fleming, T. R. and Harrington, D. P. 1991. *Counting Process and Survival Analysis*. New York: John Wiley.

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

Ghosh, D. 2001. Efficiency considerations in the additive hazards model with current status data. . *Statistica Neerlandica* 55: 367–376.

Gomez, G. and Call, M. L. 1999. Nonparametric estimation with doubly censored data. *Journal of Applied Statistics* 26: 45–58.

Gomez, G. and Lagakos, S. W. 1994. Estimation of the infection time and latency distribution of AIDS with doubly censored data. *Biometrics* 50: 204–212.

Groeneboom, P. and Wellner, J. A. 1992. Information Bounds and Nonparametric Maximum Likelihood Estimation. *DMV Seminar* 19. Birkhauser Verlag,Basel.

Gruttola, V. D. and Lagakos, S. W. 1989. Analysis of Doubly-Censored Survival Data, with Application to AIDS. *Biometrics* 45 (1): 1–11.

Huang, J. 1996. Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics* 24: 540–568.

Huang, J. 1999. Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica* 9: 501–519.

Huang, J., Lee, C. and Yu, Q. 2008. A generalized log-rank test for interval-censored failure time data via multiple imputation. *Statistics in Medicine* 27: 3217–3226.

Huang, J. and Wellner, J. A. 1995. Asymptotic normality of the NPMLE of linear functionals for interval-censored data, case 1. *Statistica Neerlandica* 49 (2): 153–163.

Jongbloed, G. 1998. The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics* 7: 310–321.

Kalbfleisch, J. D. and Prentice, R. L. 2002. *The Statistical Analysis of Failure Time Data*. 2nd edn. New York: Wiley.

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

Kim, J. 2003. Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society, Series B* 65: 489–502.

Kim, J., Kang, R. and Nam, C. M. 2006. Logrank-type tests for comparing survival curves with interval-censored data. *Computational Statistics and Data Analysis* 50: 3165–3178.

Kim, M. Y., De Gruttola, V. and Lagakos, S. W. 1993. Analyzing doubly censored data with covariates, with application to AIDS. . *Biometrics* 49: 13–22.

Klein, J. B. and Moeschbereger, M. L. 1997. *Survival Analysis Techniques for Censored and Truncated Data*. 2nd edn. New York: Springer.

Kleinbaum, D. G. 1996. *Survival Analysis: A Self-Learning Text*. 1st edn. New York: Springer.

Koul, H. L. and Yi, T. 2006. Goodness-of-fit testing in interval censoring case 1. *Statistics and Probability Letters* 76: 709–718.

Li, L. and Yu, Q. 1997. Self-consistent estimators of Survival functions with doubly-censored data. . *Communications in Statistics- Theory and Methods* 26 (11): 2609–2621.

Lim, H. J. and Sun, J. 2003. Nonparametric test for interval-censored failure time data. *Biometrical Journal* 45 (3): 263–276.

Lin, D. Y., Oakes, D. and Ying, Z. 1998. Additive hazards regression with current status data. . *Biometrika* 85: 289–298.

Martinussen, T. and Scheike, T. H. 2002. Efficient estimation in additive hazards regression with current status data. . *Biometrika* 89: 649–658.

Odell, P. M., Anderson, K. M. and D'agostino, R. B. 1992. Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. *Biometrics* 48: 951–959.

Oller, R., Gomez, G. and Calle, M. L. 2004. Interval censoring: model characterizations for the validity of the simlifide likelihood. *Canadian Journal of Statistics* 32: 315–326.

Pan, W. 1999. A comparison of some two-sample tests with interval-censored data. . *Journal of Nonparametric Statistics* 12: 133–146.

Pan, W. 2000. A two-sample test with interval-censored data via multiple imputation. *Statistic in Medicine* 19 (1): 1–11.

Pan, W. 2001. A multiple imputation approach to regression analysis for doubly censored data with application to AIDS studies. *Biometrics* 57: 1245–1250.

Peto, R. 1973. Experimental survival curves for interval-censored data. *Applied Statistics* 22: 86–91.

Peto, R. and Peto, J. 1972. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistic Society, Series A* 135 (2): 185–207.

Petroni, G. R. and Wolfe, R. A. 1994. A two sample test for stochastic ordering with interval-censored data. *Biometrics* 50: 77–87.

Raymond, K. W. W. 1977. Weibull Distribution, Iterative Likelihood Techniques and Hydrometeorological Data. *Journal of Applied Meteorology* 16: 1360–1364.

Rossini, A. J. 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.

Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Satten, G. A., Datta, S. and Williamson, J. M. 1998. Inference based on imputed failure times for the proportional hazards model with interval-censored data. . *Journal of the American Statistical Association* 93: 318–327.

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.

Shen, X. 2000. Linear regression with current status data. . *Journal of the American Statistical Association* 95: 842–852.

Sun, J. 1996. A non-parametric test for interval-censored failure time data with application to AIDS studies. *Statistics in Medicine* 15 (13): 1378–1395.

Sun, J. 1997. Self-consistency estimation of distributions based on truncated and doubly censored data with application to AIDS cohort studies. *Lifetime Data Analysis* 3: 305–313.

Sun, J. 2001. Nonparametric test for doubly interval-censored failure time data. *Lifetime Data Analysis* 7: 363–375.

Sun, J. 2004. Statistical analysis of doubly interval-censored failure time data. *Handbook of Statistics: Survival Analysis* 23: 105–122.

Sun, J. 2006. *The statistical analysis of interval-censored failure time data*. 2nd edn. New York: Springer.

Sun, J. and Kalbfleisch, J. D. 1993. The analysis of current status data on point processes. . *Journal of the American Statistical Association* 88: 1449–1454.

Sun, J. and Kalbfleisch, J. D. 1996. Nonparametric tests of tumor prevalence data. . *Biometrics* 52: 726–731.

Sun, J., Liao, Q. and Pagano, M. 1999. Regression analysis of doubly censored failure time data with applications to AIDS studies. . *Biometrics* 55: 909–914.

Sun, J., Lim, H. J. and Zhao, X. 2004. An independence test for doubly censored failure time data. *Biometrical Journal* 46: 503–511.

Sun, J., Zhao, Q. and Zhao, X. 2005. Generalized log-rank tests for interval-censored failure time data. *Scandinavian Journal of Statistics* 32: 49–57.

Tang, M. X., Tsai, W. Y., Marder, K. and Mayeux, R. 1995. Linear rank testsfor doubly censored data. *Statistics in Medicine* 14: 2555–2563.

Turnbull, B. W. 1976. The empirical distribution function with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistic Society, Series B* 38 (3): 290–295.

Wellner, J. A. and Zhan, Y. 1997. A hybird algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association* 92: 945–959.

Xue, H., Lam, K. F. and Li, G. 2004. Sieve maximum likelihood estimation for semiparametric regression models with current status data. . *Journal of the American Statistical Association* 99: 346–356.

Zhang, B. y., liu, W. and Zhan, Y. 2001. A nonparametric two-sample test of the failure function with interval censoring case 2. *Biometrika* 88 (3): 677–686.

Zhang, W., Zhang, Y., Chaloner, K. and Stapleton, J. T. 2009. Imputation methods for doubly censored HIV data. . *Journal of Statistical Computation and Simulation* 79 (10): 1245–1257.

Zhang, Y., Liu, W. and Wu, H. 2003. A simple nonparametric two-sample test for the distribution function of event time with interval censored data. . *Journal of Nonparametric Statistics* 16: 643–652.

Zhao, Q. and Sun, J. 2004. Generalized log-rank test for mixed interval-censored failure time data. *Statistics in Medicine* 23 (10): 1621–1629.

Zhao, X., Zhao, Q., Sun, J. and Kim, J. S. 2008. Generalized log-rank tests for partly interval-censored failure time data. *Biometrical Journal* 50 (3): 375–385.