INFERENCES ON MEDIAN FAILURE TIME FOR CENSORED SURVIVAL DATA

by

Shaowu Tang

B.S. in Applied Mathematics, Wuhan University, China, 1993

M.S. in Applied Mathematics, Wuhan University, China, 1996

Ph.D in Mathematics, Jacobs University Bremen, Germany, 2005

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Shaowu Tang

It was defended on

July 22, 2010

and approved by

Jong-Hyeon Jeong, Ph.D, Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Joseph P. Costantino, Dr.PH, Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Maria Mori Brooks, Ph.D, Associate Professor, Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh

Dissertation Director: Jong-Hyeon Jeong, Ph.D, Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

INFERENCES ON MEDIAN FAILURE TIME FOR CENSORED SURVIVAL DATA

Shaowu Tang, MS

University of Pittsburgh, 2010

In this thesis two approaches of inferences on median failure times are developed to compare the difference of median failure times between two groups of censored survival data.

The first one is to generalize the Mood's median test - which is designed to deal with complete data - to censored survival data. To this end, two groups of censored survival data are pooled and then the estimated pooled median failure time is obtained from the product-limit estimator. A score is assigned for each observation to indicate the probability whether it survives after the pooled median failure time or not and for each group the scores are summed to summarize the number of observations whose survival time is larger than or equal to pooled median survival time, which results in a 2×2 contingency table with non-integer entries. Four 2×2 contingency tables with integer entries are then derived and a corrected test statistic is defined as the weighted sum of the statistics from the four 2×2 contingency tables which is shown to be approximately distributed as χ^2 with 1 degree of freedom for large samples.

The second approach is proposed to construct a 95% confidence interval for the difference of median failure times between two groups of censored survival distributions. Since the median failure time is approximately normally distributed for large samples, the estimated median failure times for each group are obtained by product-limit method and their standard

errors are computed through bootstrap samples from the original data. Theory of construction for 95% confidence interval for the difference of median failure times is investigated for the standard normal distributions and it can be used for general normal distributions by translation and rescaling.

Extensive numerical studies are carried out to test the appropriateness of the two approaches and the results show that the approaches developed in the paper are easy to implement and the results are promising, compared to the results from published papers. The proposed methods will facilitate more accurate analysis of survival data under censoring, which are commonly collected from clinical studies that influence public health.

TABLE OF CONTENTS

LIST OF TABLES

PREFACE

I am grateful to Professor Jong-Hyeon Jeong for his positive guidance throughout my research. His stimulating comments and arguments have been a constant source of inspiration. I am also in great debt to Professor Joseph P. Costantino for his valuable guidance and support in selecting appropriate courses during my master program.

1.0 INTRODUCTION

1.1 MOTIVATION

Since the median is less sensitive to the outliers, in survival analysis, the median failure time is a natural and popular quantitative measure for comparing the treatment effects in randomized clinical trials and gives rise to many researchers' interest. In early 1980's one-sample confidence interval procedures for the median failure time have been studied by many investigators ([2], [4] and [3]), and much effort has been also put on two-sample or $k-$sample median test for censored data since then ([5], [8], [6] and [11]).

Two-sample inference procedures for median failure times introduced in [5] and [8] are based on one-sample confidence intervals in which the null hypothesis of equal population medians will be rejected when two confidence intervals are disjoint. In those papers the asymptotic properties of the quantile intervals have been investigated and the procedures were applied to data from a colorectal cancer clinical trial to compare four treatments. The results showed that the procedures have the same Pitman efficiency as Mood's median test. However, one drawback in these procedures is that these methods involve the underlying density function for non-shift models, and in general it is difficult to estimate the density function well for censored data. Therefore there is difficulty in applying these methods for real problems in practice.

In order to avoid the estimation of density function, Su and Wei proposed a simple and nonparametric inference procedure for comparing two median failure times in [6], where a quantity with parameters of interest and a nuisance parameter are defined and the quantity is minimized over the nuisance parameter to get a test statistic which is shown to be approximately χ^2 distributed with 1 degree of freedom, based on the theory in [7]. The method proposed in this paper can be applied to inference on ratio or difference of median failure times.

In order to avoid the estimation of the density function for non-shift model shown in [2] and the minimization procedure presented in [6], in this thesis two approaches of inferences on median failure times are developed. The first approach is proposed to construct a 95% confidence interval for the difference of median failure distributions between two groups of censored survival times. Since the median failure time is approximately normally distributed for large samples, the estimated median failure times for each group are obtained by the product-limit method and their standard errors are computed via bootstrap samples from the original data. Theory of constructing 95% confidence interval for the difference in median failure times is investigated for the standard normal distributions and it can be used for general normal distributions by translation and rescaling.

The second one is to generalize the Mood's median test - which is designed to deal with complete data - to censored survival data. To this end, two groups of censored survival data are pooled and then the estimated pooled median failure time is obtained from the product-limit method. A score is assigned for each observation to indicate the probability whether it survives after the pooled median survival failure time or not and for each group the scores are summed to summarize the number of observations whose survival time is larger than or equal to pooled median survival time, which results in a 2×2 table with non-integer entries.

Four 2×2 contingency tables with integer entries are then derived and a test statistic is formed as the weighted sum of the statistics from the four 2×2 contingency tables which is shown to be approximately distributed as χ^2 with 1 degree of freedom for large samples.

Extensive numerical studies are carried out to test the validity of the two approaches and the results show that the approaches developed in the thesis are easy to implement and the results are promising, compared to the results from [6] and [11].

1.2 PRELIMINARIES FOR SURVIVAL ANALYSIS

In this section a brief introduction is given to the fundamentals of survival analysis which can be found in many well-written books (for example [10]).

Let X be the time to event which is a nonnegative random variable from a homogeneous population, and C be independent censoring variable from an arbitrary censoring distribution H. Then the survival function - the probability of an individual surviving beyond time x - is defined as

$$S(x) = \Pr(X > x). \tag{1.2.1}$$

From the fact that $f(x) = F'(x) = (1 - S(x))'$, the probability density function $f(x)$ is

$$f(x) = -\frac{dS(x)}{dx}. \tag{1.2.2}$$

Denoting the median time of X by θ, then it holds

$$\theta = \min\{x : S(x) \leq \frac{1}{2}\}. \tag{1.2.3}$$

In clinical trials, time-to-event data are often censored, i.e., the failure time of an observation is only partially known. Under right censoring, the exact failure time X is observed if and only if X is less than or equal to the censoring time C. Therefore, the right censored

data can be represented by pairs of random variables (T, δ), where δ indicates whether X corresponds to an event ($\delta = 1$) or is censored ($\delta = 0$), and T is equal to X for $\delta = 1$ or to C for $\delta = 0$, i.e., $T = \min(X, C)$. In this thesis only right-censored data are under consideration.

The standard estimator of the survival function, known as Kaplan-Meier or Product-Limit estimator, is defined as

$$\hat{S}(t) = \begin{cases} 1, & \text{if } t < t_1 \\ \Pi_{t_i \leq t}[1 - \frac{d_i}{Y_i}], & \text{if } t_1 \leq t, \end{cases} \tag{1.2.4}$$

where one assumes that the events occur at D distinct times $t_1 < t_2 < \cdots < t_D$, and that at time t_i there are d_i events. Furthermore, Y_i is the number of individuals who are at risk at time t_i, i.e., Y_i is the count of the number of individuals who are alive at t_i or experience the event of interest at t_i.

The variance of the product-limit estimator is estimated by Greenwood's formula

$$\hat{V}[\hat{S}(t)] = \hat{S}^2(t) \sum_{t_i \leq t} \frac{d_i}{Y_i(Y_i - d_i)}. \tag{1.2.5}$$

Similarly denoting the estimated median failure time by $\hat{\theta}$, then $\hat{\theta}$ can be obtained by

$$\hat{\theta} = \min\{t : \hat{S}(t) \leq \frac{1}{2}\}. \tag{1.2.6}$$

Formulae (1.2.4) and (1.2.6) will be used in the simulations in the following chapters.

2.0 TWO-SAMPLE CONFIDENCE INTERVAL FOR THE DIFFERENCE IN MEDIAN FAILURE TIMES

This chapter contains 4 sections. The first section reviews the method and simulation results proposed in [6]. In the second section the theoretical results are presented for constructing a 95% confidence interval for the difference of median failure times. Extensive numerical simulations are carried out in the third section and the results are compared to those in [6]. In the last section the method derived in section 2 is applied to National Surgical Adjuvant Breast and Bowl Project (NSABP) B-04 data and the result is compared to that in [11].

2.1 METHOD PRESENTED IN [SU93]

In this section the methodology and results presented in [6] will be briefly reviewed.

Let $S_i(\cdot)$ be the survival function for group $i (i = 1, 2)$. Let θ_{i0} and $\hat{S}_i(\cdot)$ be the corresponding median and Kaplan-Meier estimates. Then the estimates $\hat{\theta}_i$ for θ_{i0} can be obtained by solving

$$\hat{\theta}_i = \min\{t : \hat{S}_i(t) \leq \frac{1}{2}\} \text{ for } i = 1, 2 \tag{2.1.1}$$

respectively.

To make inference about $\tau_0 = g(\theta_{10}, \theta_{20})$ for some given function g, consider the quantity

$$W(\tau_0, \theta_1) = \frac{(\hat{S}_1(\theta_1) - 0.5)^2}{\sigma_1^2(\hat{\theta}_1)} + \frac{(\hat{S}_2(h(\tau_0, \theta_1)) - 0.5)^2}{\sigma_2^2(\hat{\theta}_1)}, \tag{2.1.2}$$

where θ_1 is a nuisance parameter, $\sigma_i^2(t)$ is the usual Greenwood's formula for the variance of $\hat{S}_i(t)$, and $\theta_{20} = h(\tau_0, \theta_{10})$. Note that one can choose $g(\cdot)$ such that τ_0 is the ratio θ_{20}/θ_{10} or the difference $\theta_{20} - \theta_{10}$. By defining $G(\tau_0)$ by

$$G(\tau_0) = \min_{\theta_1} W(\tau_0, \theta_1), \tag{2.1.3}$$

it was shown in [6] that

$$G(\tau_0) \sim \chi_1^2. \tag{2.1.4}$$

Table 1: Empirical coverage probabilities of CIs for $\tau_0 = \theta_{20}/\theta_{10} = 1$ for $S_1(t) = S_2(t) = \exp(-t)$

		Censoring Proportions					
n_i	Nominal level	.0,.0	.1,.1	.25,.25	.4,.4	.1,.25	.1,.4
30	.95	.973	.970	.967	.990	.983	.984
	.90	.941	.934	.929	.962	.950	.940
	.85	.894	.889	.883	.928	.917	.902
	.80	.839	.834	.831	.896	.879	.857
50	.95	.976	.964	.970	.977	.972	.981
	.90	.940	.920	.926	.950	.934	.943
	.85	.890	.870	.881	.909	.885	.915
	.80	.850	.838	.824	.855	.840	.873
100	.95	.967	.959	.974	.967	.968	.957
	.90	.918	.909	.936	.922	.928	.923
	.85	.877	.875	.886	.879	.892	.880
	.80	.822	.833	.841	.848	.849	.833

Table 2: Empirical coverage probabilities of CIs for $\tau_0 = \theta_{20}/\theta_{10} = 1$ for $S_1(t) = \exp(-t), S_2(t) = 1 - \Phi(\log(1.44t))$

		Censoring Proportions					
n_i	Nominal level	.0,.0	.1,.1	.25,.25	.4,.4	.1,.25	.1,.4
30	.95	.972	.979	.985	.985	.976	.973
	.90	.944	.949	.953	.954	.936	.941
	.85	.898	.898	.918	.919	.886	.894
	.80	.858	.845	.874	.885	.838	.855
50	.95	.975	.962	.969	.978	.971	.971
	.90	.930	.923	.937	.938	.926	.931
	.85	.885	.878	.898	.998	.878	.886
	.80	.846	.834	.858	.850	.836	.838
100	.95	.959	.964	.972	.974	.967	.970
	.90	.925	.934	.926	.941	.930	.933
	.85	.883	.881	.884	.900	.885	.888
	.80	.834	.836	.835	.848	.845	.845

In [6] extensive simulation was carried out and the results were reported. The results showed the empirical coverage probabilities of $100(1-\alpha)\%$ confidence interval I for θ_{20}/θ_{10} with various survival distributions S_1 and S_2, censoring distributions and confidence coefficients. The censoring variables are generated from $U(0, c_i)$ with various c_i's, which are determined by some pre-specified censoring proportions for $i = 1, 2$. The results are summarized in Tables 1 and 2. Each entry in Tables 1 and 2 is generated independently from 1000 Monte Carlo simulations. For example, the empirical coverage probability 0.97 implies that among 1000 intervals I only 30 intervals do not cover the true $\tau_0 = 1$. The results in Table 1 and Table 2 showed that the method proposed in [6] is conservative.

2.2 CONSTRUCTION OF THE 95% CI FOR THE DIFFERENCE IN MEDIAN FAILURE TIMES

In this section we will show how to construct the 95% confidence interval for the difference in median failure times from one-sample confidence intervals. It was shown in [1] that the asymptotic distribution of $\hat{\theta}_i$ is normal. Therefore, in this section we will derive the theoretical results based on the normality assumption. Since it holds

$$\frac{\hat{\theta}_i - \theta_{i0}}{\text{s.e.}(\hat{\theta}_i)} \sim \text{AN}(0,1) \text{ for } i = 1, 2. \tag{2.2.1}$$

Without loss of generality, we start from standard normal distribution. For given $X, Y \sim N(0,1)$ and $\alpha > 0$, the problem can be formulated by writing L_X, U_X, L_Y and U_Y such that

$$\Pr(L_X \leq X \leq U_X) = 1 - \alpha, \tag{2.2.2}$$

$$\Pr(L_Y \leq Y \leq U_Y) = 1 - \alpha, \tag{2.2.3}$$

implies

$$\Pr(L_X - U_Y \leq X - Y \leq U_X - L_Y) = 1 - \alpha. \tag{2.2.4}$$

Recall that $X, Y \sim N(0,1)$ implies that $\frac{X-Y}{\sqrt{2}} \sim N(0,1)$. Therefore equation (2.2.4) is equivalent to

$$\Pr(\frac{L_X - U_Y}{\sqrt{2}} \leq \frac{X - Y}{\sqrt{2}} \leq \frac{U_X - L_Y}{\sqrt{2}}) = 1 - \alpha. \tag{2.2.5}$$

In the sequel we will consider the following two cases.

2.2.1 Case 1: confidence intervals of X, Y are symmetric

Since in practice the confidence intervals are in general symmetric about the mean, let us firstly consider the case

$$L_X = -U_X = Z_{\frac{\alpha}{2}}, \text{ and } L_Y = -U_Y = Z_{\frac{\alpha}{2}}. \tag{2.2.6}$$

Note that in this case, it holds

$$\frac{L_X - U_Y}{\sqrt{2}} = \sqrt{2}Z_{\frac{\alpha}{2}} = -\frac{U_X - L_Y}{\sqrt{2}}, \tag{2.2.7}$$

and thus it holds

$$\Pr(\sqrt{2}Z_{\frac{\alpha}{2}} \leq \frac{X - Y}{\sqrt{2}} \leq \sqrt{2}Z_{1-\frac{\alpha}{2}}) > 1 - \alpha, \tag{2.2.8}$$

since $[Z_{\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}] \subset [\sqrt{2}Z_{\frac{\alpha}{2}}, \sqrt{2}Z_{1-\frac{\alpha}{2}}]$. Therefore in this case it is too conservative. If one chooses

$$L_X = -U_X = Z_{\frac{\alpha'}{2}}, \text{ and } L_Y = -U_Y = Z_{\frac{\alpha'}{2}} \tag{2.2.9}$$

for some α' and still wants (2.2.4) holds, then it should satisfy

$$\frac{L_X - U_Y}{\sqrt{2}} = \sqrt{2}Z_{\frac{\alpha'}{2}} = Z_{\frac{\alpha}{2}} \tag{2.2.10}$$

which implies that

$$\alpha' = 2\Phi(\frac{\sqrt{2}}{2}Z_{\frac{\alpha}{2}}) = 2\Phi(\frac{\sqrt{2}}{2}\Phi^{-1}(\frac{\alpha}{2})), \tag{2.2.11}$$

where $\Phi(\cdot)$ is the cdf of standard normal distribution. Since $\frac{\sqrt{2}}{2}Z_{\frac{\alpha}{2}} > Z_{\frac{\alpha}{2}}$, it holds $\alpha' > \alpha$.

2.2.2 Case 2: confidence intervals of X, Y are nonsymmetric

In section 2.2.1 we have shown that if the confidence intervals of X and Y are symmetric, then (2.2.4) doesn't hold. In fact, the confidence interval of $X - Y$ is conservative, i.e., the confidence interval of $X - Y$ is larger than $100(1 - \alpha)\%$. To get the exact $100(1 - \alpha)\%$ confidence interval of $X - Y$, one needs to increase the significant level α of X and Y from α to $\alpha' = 2\Phi(\frac{\sqrt{2}}{2}\Phi^{-1}(\frac{\alpha}{2}))$.

Now in this section we loose the assumption of symmetry of confidence intervals of X and Y, i.e., assume $L_X = Z_{\alpha_1}$ and $L_Y = Z_{\alpha_2}$ for some $0 < \alpha_1, \alpha_2 < \alpha$. Therefore (2.2.4) is equivalent to

$$\Pr(\frac{Z_{\alpha_1} + Z_{\alpha-\alpha_2}}{\sqrt{2}} \leq \frac{X - Y}{\sqrt{2}} \leq -\frac{Z_{\alpha-\alpha_1} + Z_{\alpha_2}}{\sqrt{2}}) = 1 - \alpha. \tag{2.2.12}$$

Note that in (2.2.12) there are two unknown parameters α_1 and α_2 for one equation. To solve this equation, one can assume one more condition holds:

$$\frac{Z_{\alpha_1} + Z_{\alpha-\alpha_2}}{\sqrt{2}} = \frac{Z_{\alpha-\alpha_1} + Z_{\alpha_2}}{\sqrt{2}} = Z_{\frac{\alpha}{2}}, \tag{2.2.13}$$

i.e., we assume that the confidence interval of $X - Y$ is symmetric about the mean. Next lemma shows the relationship between α_1 and α_2:

Lemma 2.2.1. *It holds*

$$Z_{\alpha_1} + Z_{\alpha-\alpha_2} = Z_{\alpha-\alpha_1} + Z_{\alpha_2} \text{ iff } \alpha_1 = \alpha_2. \tag{2.2.14}$$

Proof. $\Leftarrow$: It is trivial.

$\Rightarrow$: Suppose α_1 and α_2 are one solution of $Z_{\alpha_1} + Z_{\alpha-\alpha_2} = Z_{\alpha-\alpha_1} + Z_{\alpha_2}$. Now we want to show that $\alpha_1 = \alpha_2$. Without loss of generality, we assume that α_2 is fixed and define

$$g(\alpha_1) := Z_{\alpha_1} - Z_{\alpha-\alpha_1} + [Z_{\alpha-\alpha_2} - Z_{\alpha_2}]. \tag{2.2.15}$$

Then it holds

$$\frac{dZ_{\alpha_1}}{d\alpha_1} = \frac{d\Phi^{-1}(\alpha_1)}{d\alpha_1} = \frac{1}{f(Z_{\alpha_1})} \tag{2.2.16}$$

and

$$\frac{dZ_{\alpha-\alpha_1}}{d\alpha_1} = \frac{d\Phi^{-1}(\alpha-\alpha_1)}{d\alpha_1} = -\frac{1}{f(Z_{\alpha-\alpha_1})}. \tag{2.2.17}$$

Therefore it holds

$$\frac{dg(\alpha_1)}{d\alpha_1} = \frac{1}{f(Z_{\alpha_1})f(Z_{\alpha-\alpha_1})}[f(Z_{\alpha-\alpha_1}) - f(Z_{\alpha_1})]. \tag{2.2.18}$$

Since $\alpha_1 \neq \frac{\alpha}{2}$, then α_1 satisfies either $\alpha_1 \in (0, \frac{\alpha}{2})$ or $\alpha_1 \in (\frac{\alpha}{2}, \alpha)$. In either case one finds that $\frac{dg(\alpha_1)}{d\alpha_1}$ is strictly monotone, which implies that $g(\alpha_1) = 0$ has at most one solution, given α_2 constant. Since $\alpha_1 = \alpha_2$ is one solution of $g(\alpha_1) = 0$, it is also the only solution of $g(\alpha_1) = 0$, which finishes the proof of the lemma.

□

Lemma 2.2.1 implies that we only need to check if there exists $0 < \alpha_1 < \alpha$ such that

$$\frac{Z_{\alpha-\alpha_1} + Z_{\alpha_1}}{\sqrt{2}} = Z_{\frac{\alpha}{2}}. \tag{2.2.19}$$

Note that since α_1 and $\alpha - \alpha_1$ are symmetric about $\frac{\alpha}{2}$, without loss of generality we assume that $\alpha_1 \in (\frac{\alpha}{2}, \alpha)$. The following theorem shows that there doesn't exist such α_1 that equation (2.2.4) holds.

Theorem 2.2.2. *There is no α_1 with $\frac{\alpha}{2} < \alpha_1 < \alpha$ such that (2.2.4) holds.*

Proof. Note that if $\alpha_1 \in (\frac{\alpha}{2}, \alpha)$, then $\alpha - \alpha_1 \in (0, \frac{\alpha}{2})$, and therefore it holds

$$\alpha - \alpha_1 < \frac{\alpha}{2} < \alpha_1 < \alpha \Rightarrow Z_{\alpha-\alpha_1} < Z_{\alpha_1} \Rightarrow f(Z_{\alpha-\alpha_1}) < f(Z_{\alpha_1}). \tag{2.2.20}$$

Similarly let us define

$$h(\alpha_1) := Z_{\alpha-\alpha_1} + Z_{\alpha_1} - \sqrt{2}Z_{\frac{\alpha}{2}}. \tag{2.2.21}$$

Thus

$$\frac{dh(\alpha_1)}{d\alpha_1} = \frac{1}{f(Z_{\alpha_1})f(Z_{\alpha-\alpha_1})}[f(Z_{\alpha-\alpha_1}) - f(Z_{\alpha_1})] < 0, \tag{2.2.22}$$

which implies that $h(\alpha_1)$ is strictly decreasing in $(\frac{\alpha}{2}, \alpha)$.

Note that

$$h(\frac{\alpha}{2}) = (2-\sqrt{2})Z_{\frac{\alpha}{2}} < 0 \Rightarrow h(\alpha_1) < 0 \text{ for all } \alpha_1 \in (\frac{\alpha}{2}, \alpha), \tag{2.2.23}$$

which ends the proof. □

Theorem 2.2.2 shows that under the assumption of symmetry of confidence interval for $X - Y$, there is no solution. If we replace the α in equations (2.2.2) and (2.2.3) by some α', then under the assumption of symmetry of confidence interval for $X - Y$, one finds that

Theorem 2.2.3. *If α' satisfies that*

$$\alpha' > 2\Phi(\frac{\sqrt{2}}{2}Z_{\frac{\alpha}{2}}) = 2\Phi(\frac{\sqrt{2}}{2}\Phi^{-1}(\frac{\alpha}{2})), \tag{2.2.24}$$

then there exists some $0 < \alpha_1 = \alpha_2 < \alpha'$ such that

$$\frac{Z_{\alpha'-\alpha_1} + Z_{\alpha_1}}{\sqrt{2}} = Z_{\frac{\alpha}{2}}. \tag{2.2.25}$$

Proof. Without loss of generality we assume that $\alpha_1 \in (\frac{\alpha'}{2}, \alpha')$ and define

$$\tilde{h}(\alpha_1) := Z_{\alpha'-\alpha_1} + Z_{\alpha_1} - \sqrt{2}Z_{\frac{\alpha}{2}}. \tag{2.2.26}$$

Then again $\tilde{h}(\alpha_1)$ is strictly decreasing in $(\frac{\alpha'}{2}, \alpha')$.

Note that

$$\tilde{h}(\alpha' - \varepsilon) = Z_\varepsilon + Z_{\alpha'-\varepsilon} - \sqrt{2}Z_{\frac{\alpha}{2}} < 0 \tag{2.2.27}$$

holds for small enough $\varepsilon > 0$. Furthermore, it holds

$$\tilde{h}(\frac{\alpha'}{2}) = 2Z_{\frac{\alpha'}{2}} - \sqrt{2}Z_{\frac{\alpha}{2}} > 0 \tag{2.2.28}$$

if given

$$\alpha' > 2\Phi(\frac{\sqrt{2}}{2}Z_{\frac{\alpha}{2}}) = 2\Phi(\frac{\sqrt{2}}{2}\Phi^{-1}(\frac{\alpha}{2})). \tag{2.2.29}$$

Since $\tilde{h}(\alpha_1)$ is strictly decreasing in $(\frac{\alpha'}{2}, \alpha')$ and it holds

$$\tilde{h}(\alpha'-) \cdot \tilde{h}(\frac{\alpha'}{2}) < 0, \tag{2.2.30}$$

there exist one $\alpha_1 \in (\frac{\alpha'}{2}, \alpha')$ such that equation (2.2.25) holds. Note that since given α', $\tilde{h}(\alpha_1) = 0$ is a nonlinear equation, which implies one can obtain the root of $\tilde{h}(\alpha_1) = 0$ by bisection method. □

2.3 SIMULATION STUDIES

For simplicity in this section we will only consider Case 1 discussed in Section 2.2.1. Assume that α' is defined by (2.2.11), then for $i = 1, 2$, it holds

$$\Pr(Z_{\frac{\alpha'}{2}} < \frac{\hat{\theta}_i - \theta_{i0}}{\text{s.e.}(\hat{\theta}_i)} < Z_{1-\frac{\alpha'}{2}}) = 1 - \alpha', \tag{2.3.1}$$

which is equivalent to

$$\Pr[\hat{\theta}_i - Z_{1-\frac{\alpha'}{2}} \cdot \text{s.e.}(\hat{\theta}_i) < \theta_{i0} < \hat{\theta}_i + Z_{1-\frac{\alpha'}{2}} \cdot \text{s.e.}(\hat{\theta}_i)] = 1 - \alpha'. \tag{2.3.2}$$

By defining

$$L_i = \hat{\theta}_i - Z_{1-\frac{\alpha'}{2}} \cdot \text{s.e.}(\hat{\theta}_i) \text{ and } U_i = \hat{\theta}_i + Z_{1-\frac{\alpha'}{2}} \cdot \text{s.e.}(\hat{\theta}_i) \tag{2.3.3}$$

it holds

$$\Pr(L_2 - U_1 < \theta_{20} - \theta_{10} < U_2 - L_1) = 1 - \alpha, \tag{2.3.4}$$

i.e., the 95% confidence interval for $\theta_{20} - \theta_{10}$ is

$$(L_2 - U_1, U_2 - L_1) = ((\hat{\theta}_2 - \hat{\theta}_1) - Z_{1-\frac{\alpha'}{2}}[\text{s.e.}(\hat{\theta}_1) + \text{s.e.}(\hat{\theta}_2)], (\hat{\theta}_2 - \hat{\theta}_1) + Z_{1-\frac{\alpha'}{2}}[\text{s.e.}(\hat{\theta}_1) + \text{s.e.}(\hat{\theta}_2)]). \tag{2.3.5}$$

From (2.3.5) in order to construct the 95% confidence interval for $\theta_{20} - \theta_{10}$, one needs to compute $\hat{\theta}_i$ and s.e.$(\hat{\theta}_i)$ for $i = 1, 2$. Recall that $\hat{\theta}_i$ are estimated median failure times and can be obtained by (2.1.1). In this thesis we plan to approximate s.e.$(\hat{\theta}_i)$ in terms of bootstrap techniques. Assume that (T_i, δ_i) are two groups of censored survival data for $i = 1, 2$. For each i, one randomly generate D bootstrap samples $\{(T_i^j, \delta_i^j)\}_{j=1}^D$ from the original censored survival data (T_i, δ_i) with replacement at the same length. Therefore for each group one can obtain a set of estimated median survival times $\{\hat{\theta}_i^j\}_{j=1}^D$ by (1.2.4) and (2.1.1). Then one can estimate s.e.$(\hat{\theta}_i)$ by

$$\text{s.e.}(\hat{\theta}_i) \approx \text{sd}(\{\hat{\theta}_i^j\}_{j=1}^D) \text{ for } i = 1, 2. \tag{2.3.6}$$

In general one can get a good estimate for s.e.$(\hat{\theta}_i)$ provided $50 \leq D \leq 200$. We repeated the simulations presented in Section 2.1 with $D = 50$ and the results are shown in Tables 3 and 4.

Table 3: Empirical coverage probabilities of CIs $\theta_{20} - \theta_{10} = 0$ for $S_1(t) = S_2(t) = \exp(-t)$

			Mean	Censoring	Proportions		
n_i	α	.43,.43	.28,.28	.1,.1	.01,.01	.1,.28	.1,.43
30	0.05	0.043	0.045	0.048	0.052	0.055	0.046
	0.1	0.09	0.081	0.099	0.097	0.112	0.089
	0.15	0.12	0.115	0.134	0.141	0.125	0.117
	0.2	0.165	0.167	0.18	0.186	0.188	0.177
50	0.05	0.038	0.044	0.050	0.058	0.051	0.053
	0.1	0.079	0.079	0.096	0.087	0.102	0.095
	0.15	0.121	0.134	0.149	0.147	0.158	0.127
	0.2	0.165	0.199	0.194	0.189	0.207	0.167
100	0.05	0.042	0.044	0.064	0.057	0.051	0.048
	0.1	0.098	0.105	0.103	0.111	0.101	0.088
	0.15	0.140	0.121	0.164	0.154	0.143	0.131
	0.2	0.169	0.176	0.205	0.200	0.201	0.179

Table 4: $\theta_{20} - \theta_{10} = 0$ and $S_1(t) = \exp(-t), S_2(t) = 1 - \Phi(\log(1.44t))$

			Mean	Censoring	Proportions		
30	0.05	0.037	0.038	0.050	0.055	0.049	0.044
	0.1	0.085	0.092	0.098	0.097	0.098	0.091
	0.15	0.111	0.125	0.150	0.144	0.119	0.127
	0.2	0.169	0.163	0.202	0.203	0.173	0.178
50	0.05	0.040	0.046	0.046	0.054	0.049	0.047
	0.1	0.077	0.098	0.111	0.109	0.090	0.091
	0.15	0.118	0.129	0.133	0.134	0.130	0.145
	0.2	0.176	0.171	0.188	0.192	0.181	0.178
100	0.05	0.044	0.048	0.045	0.056	0.055	0.051
	0.1	0.104	0.094	0.100	0.106	0.104	0.099
	0.15	0.157	0.138	0.145	0.153	0.143	0.145
	0.2	0.176	0.172	0.195	0.200	0.193	0.176

Compared to the results in Tables 1 and 2 in Section 2.1, one finds that the method proposed in this section is less conservative and behaves much better for heavily censored data. Note that in our simulations we choose such c_i's that the censoring percentages are similar to those in [6] but we can't produce the data with the same censoring percentages.

2.4 SIMULATION FOR NSABP B-04 DATA

In this section we will apply the method proposed in **Section 2.2** and **Section 2.3** and compare the result to [11].

As mentioned in [11], the NSABP B-04 study was designed to compare radical mastectomy with a less extensive surgery (total mastectomy) with or without radiation therapy. A total of 1079 women with clinically negative axillary nodes went through radical mastectomy, total mastectomy without axillary dissection but with post-operative irradiation, or total mastectomy plus axillary dissection if their nodes became positive. A total of 586 women with clinically positive axillary nodes experienced either radical mastectomy or total mastectomy without axillary dissection but with post-operative irradiation. About 90% of all patients were either followed for at least 25 years or were known to have died. The proportion of patients still alive is less than 30% among node-negative patients and less than 20% among node-positive patients. At time $t = t_0 = 0$, the censoring proportion in node-negative patients is about 27%, while the censoring proportion in node-positive patients is about 16%.

The report presented in [11] showed that at time $t = 0$ it holds

$$\hat{\theta}_1 = 12.46, \hat{\theta}_2 = 6.87, \tau_0 = \frac{\hat{\theta}_2}{\hat{\theta}_1} = 0.55 \text{ with } 95\% \text{ CI} = (0.49, 0.63), \tag{2.4.1}$$

which implies that the median lifetimes are significantly different between node-negative and node-positive breast cancer patients at the 5% significant level at $t = t_0 = 0$.

The R-output is

```
Call:
survdiff(formula = Surv(time, event) ~ group, data = data)

           N Observed Expected (O-E)^2/E (O-E)^2/V
group=0 1079      792      912      15.8      54.8
group=1  586      491      371      38.7      54.8

 Chisq= 54.8  on 1 degrees of freedom, p= 1.34e-13
```

which show the overall lifetime distributions are significantly different between node-negative and node-positive breast cancer patients at the 5% significant level.

We have already shown that in 95% CI of $\theta_{20} - \theta_{10}$ it holds

$$L_2 - U_1 < \theta_{20} - \theta_{10} < U_2 - L_1 \text{ and } L_i < \theta_{i0} < U_i. \tag{2.4.2}$$

Therefore it is easy to get the corresponding interval of the ratio θ_{10}/θ_{20} as

$$\frac{L_2}{U_1} < \frac{\theta_{20}}{\theta_{10}} < \frac{U_2}{L_1}. \tag{2.4.3}$$

We repeated the simulation for $D = 50, 100, 200, 500, 1000, 5000, 10000$. The results are summarized in the following table:

Table 5: Simulation of 95% CI for NSABP B-04

D	s.e.$(\hat{\theta}_1)$	s.e.$(\hat{\theta}_2)$	95% CI for $\theta_{10} - \theta_{20}$	CI for θ_{20}/θ_{10}
50	0.6326	0.2956	$(4.3054, 6.8780)$	$(0.4842, 0.6283)$
100	0.6770	0.3082	$(4.2263, 6.9570)$	$(0.4807, 0.6331)$
200	0.5913	0.2937	$(4.3651, 6.8182)$	$(0.4865, 0.6250)$
500	0.6810	0.3040	$(4.2254, 6.9568)$	$(0.4809, 0.6329)$
1000	0.6611	0.3173	$(4.2356, 6.9477)$	$(0.4805, 0.6330)$
5000	0.6460	0.3080	$(4.2695, 6.9138)$	$(0.4823, 0.6308)$
10000	0.6556	0.3098	$(4.2536, 6.9296)$	$(0.4812, 0.6317)$

The results show that the method is quite robust for bootstrap sample size D and the confidence interval of θ_{20}/θ_{10} derived from the 95% CI for $\theta_{10} - \theta_{20}$ by (2.4.3) is consistent to that reported in [11], which is $(0.49, 0.63)$.

3.0 GENERALIZED MOOD'S MEDIAN TEST FOR CENSORED SURVIVAL DATA

This chapter contains 4 sections too. The first section reviews the Mood's median tests for complete data while in the second section the generalized Mood's median tests for censored survival data is presened. Extensive numerical simulations are carried out in the third section and the results are compared to those in [6] and the method derived in section 2 is applied to National Surgical Adjuvant Breast and Bowl Project (**NSABP**) data B-04 again and the result is compared to that in [11]. In the last section we draw a conclusion for our investigation.

3.1 MOOD'S MEDIAN TEST FOR COMPLETE DATA

The Mood's median test is a useful and general nonparametric test for comparing two or more independent samples which assumes that observations are independent both within and between samples and the distributions of the populations the samples were drawn from all have the same shape. Similar to the sign test, the median test is very robust against outliers, and fairly robust against differences in the shapes of the distributions. Since the Mood's median doesn't take into account the difference between each observation and the pooled median, it has poor power for normally distributed data, even worse power for short-tailed distributions, but good relative power for heavy-tailed (outlier-rich) distributions.

Assume that $\{X_1, X_2, \cdots, X_m\}$ and $\{Y_1, Y_2, \cdots, Y_n\}$ are two independent samples and assume that θ_{pool} is the median for pooled data. Then for each sample one can count the number of elements which are greater than θ_{pool}. The results can be summarized into the following 2×2 contingency table:

	Sample X	Sample Y	Total
$> \theta_{pool}$	n_{11}	n_{21}	n_1
$\leq \theta_{pool}$	n_{12}	n_{22}	n_2
Total	m_1	m_2	N

where $N = n + m, m_1 = m$ and $m_2 = n$. For each n_{ij} one can define the expected number μ_{ij} which are defined as

$$\mu_{ij} = \frac{m_i n_j}{N}, \text{ where } i = 1, 2; j = 1, 2.$$

Then the test statistic X^2 satisfies

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \sim \chi^2_{(i-1)\times(j-1)} = \chi^2_1. \tag{3.1.1}$$

The Chi-squared approximation improves as $\{\mu_{ij}\}$ increase, and $\{\mu_{ij} \geq 5\}$ is usually sufficient for a good approximation, otherwise Fisher's exact test can be used to test the null hypothesis.

3.2 GENERALIZED MOOD'S MEDIAN TEST FOR CENSORED SURVIVAL DATA

In last section we recall the two-sample Mood's median test for complete data. In this section we plan to generalize it to incomplete data, i.e., survival data which are observed or right-censored.

Suppose m_i independent observations are drawn from the ith population, $i = 1, 2$. Let T_{ij} be the jth observed survival data from the ith population with $1 \leq j \leq m_i,\ i = 1, 2$. For each T_{ij} one also observes δ_{ij} which indicates whether T_{ij} is censored or not.

For each sample one can compute the Kaplan-Meier estimate $\hat{S}_i(t) = 1 - \hat{F}_i(t)$. Furthermore, one can obtain the pooled median $\hat{\theta}_{pool}$ from the pooled data, i.e.,

$$\hat{\theta}_{pool} = \min\{t : \hat{S}_{pool}(t) \leq \frac{1}{2}\}, \tag{3.2.1}$$

where $\hat{S}_{pool}(t)$ is the Kaplan-Meier estimate of survival time $S(t)$ for the pooled data.

Now for each sample one can estimate n_{i1} by

$$\hat{n}_{i1} = \sum_{k=1}^{m_i} \Pr(X_{ik} > \hat{\theta}_{pool}), \tag{3.2.2}$$

where X_{ik} is the true event time of kth observation in ith population, and $\Pr(X_{ik} > \hat{\theta}_{pool})$ estimates the probability that the true failure time of the kth observation is beyond $\hat{\theta}_{pool}$. Therefore, it holds

$$\Pr(X_{ik} > \hat{\theta}_{pool}) = \begin{cases} 1, & \text{if } T_{ik} > \hat{\theta}_{pool} \text{ and } \delta_{ik} = 1; \\ 1, & \text{if } T_{ik} \geq \hat{\theta}_{pool} \text{ and } \delta_{ik} = 0; \\ 0, & \text{if } T_{ik} \leq \hat{\theta}_{pool} \text{ and } \delta_{ik} = 1; \\ \hat{q}_{ik}, & \text{if } T_{ik} < \hat{\theta}_{pool} \text{ and } \delta_{ik} = 0. \end{cases} \tag{3.2.3}$$

Obviously it holds

$$\hat{q}_{ik} = \Pr(X_{ik} > \hat{\theta}_{pool} | T_{ik} < \hat{\theta}_{pool}, \delta_{ik} = 0) = \Pr(X_{ik} > \hat{\theta}_{pool} | X_{ik} > T_{ik}) = \frac{\Pr(X_{ik} > \hat{\theta}_{pool})}{\Pr(X_{ik} > T_{ik})} = \frac{\hat{S}_i(\hat{\theta}_{pool})}{\hat{S}_i(T_{ik})} \tag{3.2.4}$$

for $i = 1, 2$ and $k = 1, \cdots, m_i$. Now one can define

$$\hat{n}_{i2} = \sum_{k=1}^{m_i} \Pr(X_{ik} \leq \hat{\theta}_{pool}) = m_i - \hat{n}_{i1}. \tag{3.2.5}$$

Table 6: 2×2 table generated from generalized Mood's median test

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	$\hat{n}_{11}$	$\hat{n}_{21}$	$\hat{n}_1$
$\leq \hat{\theta}_{pool}$	$\hat{n}_{12}$	$\hat{n}_{22}$	$\hat{n}_2$
Total	m_1	m_2	N

Therefore one gets a 2×2 table as in Table 7.

Note that although $\hat{n}_{i1} + \hat{n}_{i2} = m_i$ holds for $i = 1, 2$, in general $\hat{n}_{i1}, \hat{n}_{i2}$ and $\hat{n}_i$ are not integers. To generate a 2×2 contingency table similar to which in Section 3.1, let us define

$$\tilde{n}_{i1} = \lfloor \hat{n}_{i1} \rfloor := \inf\{m \in \mathbf{N} : m \leq \hat{n}_{i1}\} \text{ and } \tilde{n}_{i2} = m_i - \tilde{n}_{i1}, \tag{3.2.6}$$

where $\mathbf{N}$ is the set of positive integers. Obviously if $\tilde{n}_{i1}$ happens to be an integer, then it holds $\tilde{n}_{i1} = \hat{n}_{i1}$ and $\tilde{n}_{i2} = \hat{n}_{i2}$. Otherwise it holds

$$\tilde{n}_{i1} < \hat{n}_{i1} < \tilde{n}_{i1} + 1 \text{ and } \tilde{n}_{i2} - 1 < \hat{n}_{i2} < \tilde{n}_{i2}. \tag{3.2.7}$$

Now we can define four 2×2 contingency tables:

1.

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	$\tilde{n}_{11}$	$\tilde{n}_{21}$	n_1^1
$\leq \hat{\theta}_{pool}$	$\tilde{n}_{12}$	$\tilde{n}_{22}$	n_2^1
Total	m_1	m_2	N

2.

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	$\tilde{n}_{11} + 1$	$\tilde{n}_{21}$	n_1^2
$\leq \hat{\theta}_{pool}$	$\tilde{n}_{12} - 1$	$\tilde{n}_{22}$	n_2^2
Total	m_1	m_2	N

3.

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	$\tilde{n}_{11}$	$\tilde{n}_{21}+1$	n_1^3
$\leq \hat{\theta}_{pool}$	$\tilde{n}_{12}$	$\tilde{n}_{22}-1$	n_2^3
Total	m_1	m_2	N

4.

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	$\tilde{n}_{11}+1$	$\tilde{n}_{21}+1$	n_1^4
$\leq \hat{\theta}_{pool}$	$\tilde{n}_{12}-1$	$\tilde{n}_{22}-1$	n_2^4
Total	m_1	m_2	N

Therefore one can obtain 4 statistics of X_i^2 for 4 tables respectively. Next we will define the test statistic for censored data by choosing some weights to the 4 statistics X_i^2. To this end, let us define

$$\hat{n}_{11} = \tilde{n}_{11} + \lambda \text{ and } \hat{n}_{21} = \tilde{n}_{21} + \eta. \tag{3.2.8}$$

Note that in each table since the column margins are fixed, so there are 2 degree of freedom to choose the entries off four in each table. Therefore, we will compare the first two entries in the first row of each table to $\hat{n}_{11}$ and $\hat{n}_{21}$ and weight X_i^2 according to their differences, i.e., we can define the weights as

$$\begin{aligned} \varpi_1 &= (1-\lambda)(1-\eta), \\ \varpi_2 &= \lambda(1-\eta), \\ \varpi_3 &= \lambda\eta, \\ \varpi_4 &= \eta(1-\lambda). \end{aligned} \tag{3.2.9}$$

Then the corrected test statistic can be defined as

$$X^2 := \sum_{i=1}^{4} \varpi_i X_i^2. \tag{3.2.10}$$

Since X^2 is a correction for $\{X_i^2\}_{i=1}^4$ and $X_i^2 \sim \chi_1^2$, we still approximate X^2 by χ_1^2 distribution.

Another correction can be done for p-values, i.e., assuming already having the p-values p_i corresponding to 4 cases (X_i^2) respectively, one can define the corrected p-value p as

$$p := \sum_{i=1}^{4} \varpi_i p_i. \tag{3.2.11}$$

Equ. (3.2.11) is very useful especially for small samples. When the assumption of $X_i^2 \sim \chi_1^2$ may not hold, one still can compute p_i by Fisher's exact test and define p by Equ. (3.2.11). In this paper we use Equ. (3.2.10), given $n_i \geq 30$.

3.3 SIMULATION STUDIES

Firstly we simulate the method for the data taken from Weibull distribution with $\kappa = 2$ and $\rho = 0.05$ with sample size $n = 100, 200, 500$. To this end we generate 2 sets of random numbers from uniform distribution $U(0, 1)$ and transform one into Weibull distributed random numbers and then divide them randomly into 2 groups by utilizing the other set of random numbers, for example, one can recode the second set into set with entries $0/1$ according to the fact that if each entry is larger than or equal to 0.5 or not. Therefore we have a set of pairs $\{(X_i, g_i)\}$ where X_i are Weibull distributed random random numbers and $g_i = 0/1$ is the group indication. Now for some carefully selected c one generates another set of random numbers $\{C_i\}$ from $U(0, c)$. Now define

$$T = \min\{X, C\} \text{ and } \delta = \begin{cases} 1, & \text{if } X <= C, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3.1}$$

Therefore one obtains a triple $\{T, \delta, g\}$ where T is the observed failure time, δ is the censoring indicator and g is the group indicator. Now method proposed in Section 3.2 can be used on T, δ, g.

Table 7: Test for Weibull distribution with $\kappa = 2$ and $\rho = 0.05$ for $\alpha = 0.05$

c	n=100	100	200	200	500	500
	Censoring	Significant	Censoring	Significant	Censoring	Significant
30	0.53	0.01	0.55	0.014	0.582	0.012
40	0.36	0.018	0.445	0.018	0.46	0.017
80	0.26	0.021	0.225	0.037	0.216	0.033
100	0.16	0.033	0.18	0.037	0.164	0.027
500	0.01	0.04	0.02	0.056	0.03	0.05
1000	0.04	0.05	0.01	0.055	0.028	0.052
10000	0	0.058	0	0.055	0.002	0.065

Table 8: Generalized Mood's median test for $S_1(t) = S_2(t) = \exp(-t)$

			Mean	Censoring	Proportions		
n_i	α	.43,.43	.28,.28	.1,.1	.01,.01	.1,.28	.1,.43
30	0.05	0.02	0.043	0.05	0.072	0.05	0.04
	0.1	0.066	0.064	0.083	0.086	0.091	0.084
	0.15	0.105	0.12	0.156	0.21	0.159	0.124
	0.2	0.15	0.184	0.203	0.188	0.192	0.187
50	0.05	0.025	0.035	0.048	0.042	0.055	0.049
	0.1	0.066	0.092	0.085	0.095	0.082	0.078
	0.15	0.103	0.134	0.136	0.125	0.154	0.129
	0.2	0.159	0.173	0.181	0.206	0.178	0.184
100	0.05	0.027	0.034	0.037	0.05	0.04	0.045
	0.1	0.077	0.082	0.097	0.108	0.107	0.081
	0.15	0.11	0.111	0.156	0.132	0.126	0.126
	0.2	0.155	0.177	0.192	0.214	0.184	0.193

Table 9: Generalized Mood's median test for $S_1(t) = \exp(-t), S_2(t) = 1 - \Phi(\log(1.44t))$

			Mean	Censoring	Proportions		
n_i	α	.43,.43	.28,.28	.1,.1	.01,.01	.1,.28	.1,.43
30	0.05	0.032	0.044	0.047	0.070	0.043	0.038
	0.10	0.085	0.093	0.094	0.076	0.093	0.096
	0.15	0.109	0.143	0.161	0.185	0.122	0.138
	0.20	0.163	0.191	0.187	0.187	0.212	0.165
50	0.05	0.030	0.034	0.050	0.046	0.038	0.037
	0.10	0.077	0.076	0.091	0.100	0.091	0.079
	0.15	0.108	0.119	0.151	0.125	0.135	0.126
	0.20	0.157	0.172	0.207	0.197	0.197	0.175
100	0.05	0.035	0.040	0.044	0.051	0.035	0.052
	0.10	0.063	0.074	0.096	0.113	0.098	0.076
	0.15	0.108	0.130	0.140	0.135	0.138	0.114
	0.20	0.168	0.188	0.189	0.180	0.175	0.177

The results in Table 7 show that the method is conservative for heavily censored data and converges to Mood's median test as censoring proportion tends to 0. While comparing the results in Tables 8 and 9 to those in Table 1 and 2, it shows that the generalized Mood's median test is less conservative than the method proposed in [6].

Now let us apply the generalized Mood's median test to NSABP B-04 data. The 2×2 table is given in Table 10.

Table 10: 2×2 table generated by generalized Mood's method for NSABP B-04 data

	Group 0	Group 1	Total
$> \theta_{pool}$	605.3622	222.4423	827.8045
$\leq \theta_{pool}$	473.6378	363.5577	837.1955
Total	1079	586	1665

Obviously λ and η satisfy $\lambda = 0.3622$ and $\eta = 0.4423$. From (3.2.9) the weights are

$$\varpi_1 = 0.3557,\ \varpi_2 = 0.2020,\ \varpi_3 = 0.1602 \text{ and } \varpi_4 = 0.2821. \tag{3.3.2}$$

Similarly, the four 2×2 contingency tables are

1.

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	605	222	827
$\leq \hat{\theta}_{pool}$	474	364	838
Total	1079	586	1665

2.

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	606	222	828
$\leq \hat{\theta}_{pool}$	473	364	837
Total	1079	586	1665

3.

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	605	223	828
$\leq \hat{\theta}_{pool}$	474	363	837
Total	1079	586	1665

4.

	Sample one	Sample two	Total
$> \hat{\theta}_{pool}$	606	223	829
$\leq \hat{\theta}_{pool}$	473	363	836
Total	1079	586	1665

Therefore the 4 corresponding test statistics are

$$X_1^2 = 50.24362,\ x_2^2 = 50.75627\ X_3^2 = 49.30443 \text{ and } X_4^2 = 49.81243 \tag{3.3.3}$$

and from (3.2.10) one has the test statistic $X^2 = 50.01316$ with χ_1^2 p-value 7.835433×10^{-13}, which implies the difference of mean failure times is highly significant, obviously consistent to the result in [11]. Similarly, from Equ. (3.2.11), the corrected p-value is 1.522602×10^{-12}, which is less conservative than 7.835433×10^{-13}, but it is still highly significant.

Note that the value of the test statistic X^2 for NSABP B-04 data is similar to those of $X_i^2, i = 1, 2, 3, 4$, since NSABP B-04 data is a large sample case. Therefore, one can replace X^2 by any X_i^2 and still get the same conclusion. In general this is not true for small sample cases, thus X^2 defined in (3.2.10) is important in this situation. Another important issue is that, although one can define a test statistic X^2 for Table 6 via (3.1.1), no proof has been shown that $X^2 \sim \chi_1^2$ holds for non-integer entries. Furthermore, when one of the expected number μ_{ij} is less than 5, Fisher's exact test should be applied to get the p-value from the hypergeometric distribution, which can't be used for non-integer entries and we recommend to use Equ. (3.2.11) instead of Equ. (3.2.10).

3.4 CONCLUSION

The simulation results presented in Chapters 2 and 3 indicate that the methods we proposed in this thesis are both satisfactory and easy to implement, compared to the results in [6]. When applied to the NSABP B-04 data, the results are consistent to those in [11]. The generalized Mood's median test will converge to the Mood's median test when censoring proportion tends to 0, but it is more conservative than the method we proposed in **Chapter 2**, while the latter is quite robust to the choice of bootstrap sample size D. Already having studied the type I error of the methods, the power analysis will be studied in the future to get an idea how the new methods are.

APPENDIX A

R CODE FOR GENERALIZED MOOD'S MEDIAN TEST FOR NSABP B-04

```
# Load package Survival
rm(list=ls())
library(survival)

# Load the data

data<-read.table("J:\\MS thesis\\simulation\\data.for.tang", head=T,as.is=T)


#Define the number of observations in group 1
n1<- length(data$time[data$group==0])
n2<- length(data$time[data$group==1])



# Estimate of survival time
fit<-survfit(Surv(time,event)~1,data=data)
```

```
# Find the estimated median time
test_median<-min(fit$time[fit$surv<=0.5])
surv_median<-min(fit$surv[fit$surv<=0.5])




# Define variable count which counts the number of entries
# which is larger than the estimate sample median

count<-rep(0,length(data$time))

# if (time1>test_median)&(event1=1) then count=1
count[which(data$time>test_median & data$event==1)]<-1

# if (failure>=test_median)&(censor=0) then count=1
count[which(data$time>=test_median & data$event==0)]<-1

# if (failure<=test_median)&(censor=1) then count=0
# Since the default value of count is 0, one can omit the following command
count[which(data$time<=test_median & data$event==1)]<-0

# if (failure<test_median)&(censor=0) then count=surv_median/survival
aa<- which(data$time<test_median & data$event==0)
nn<-length(aa)
for (i in 1:nn){
count[aa[i]]<-surv_median/summary(fit,time=c(data$time[aa[i]]))[1]
```

```
# Combine count into dataset
data2<-data.frame(cbind(data,count))

# Compute the sum of count in group 1
sum_1<-sum(data2$count[data2$group==1])

# Compute the sum of count in group 0
sum_0<-sum(data2$count[data2$group==0])

# Define the interger part of two sums
a<- floor(sum_0)
b<- floor(sum_1)

# Define the digit part of two sums
lambda<-sum_0-a
eta<- sum_1-b

# Define 4 weights
w1<- (1-lambda)*(1-eta)
w2<- lambda*(1-eta)
w3<- (1-lambda)*eta
w4<- lambda*eta

# Define a function to derive chi-square test for 2X2 table
test_chisq<- function(a,b,c,d){
m1<-a+b
m2<-c+d
```

```
n1<-a+c
n2<-b+d
N<-a+b+c+d
o1<-m1*n1/N
o2<-m1*n2/N
o3<-m2*n1/N
o4<-m2*n2/N
test<-(a-o1)^2/(o1)+(b-o2)^2/(o2)+(c-o3)^2/(o3)+(d-o4)^2/(o4)
}


# Compute 4 test statistics for 4 2X2 table
chi1<-test_chisq(a,b,n1-a,n2-b)
chi2<-test_chisq(a+1,b,n1-a-1,n2-b)
chi3<-test_chisq(a,b+1,n1-a,n2-b-1)
chi4<-test_chisq(a+1,b+1,n1-a-1,n2-b-1)

# Define the test statistic by weighted sum of chi1 to chi4
chi<-w1*chi1+w2*chi2+w3*chi3+w4*chi4

# Compute the p-value
p-value<- dchisq(chi,df=1)
```

APPENDIX B

R CODE FOR THE CONSTRUCTION OF 95% CI FOR NSABP B-04

```
  # Load package Survival
rm(list=ls())
library(survival)

# Load the data

data<-read.table("J:\\MS thesis\\simulation\\data.for.tang", head=T,as.is=T)

# Log-rank test

survdiff(Surv(time,event)~group,data=data)

# Plot of survival time for each group

plot(survfit(Surv(time,event)~ group, data=data), lty=1:2,mark.time=F, ylab="Survival Ra

# Subset two subsets
```

```
data1<- data[data$group==0,c(1,2)]
data2<- data[data$group==1,c(1,2)]

# Define significant level
alpha <- 0.05

# Define significant level for each geoup
alpha1<- 2*pnorm(sqrt(2)/2*qnorm(alpha/2))

# Estimate of survival time for two groups
fit1<-survfit(Surv(time,event)~1,data=data1)
fit2<-survfit(Surv(time,event)~1,data=data2)

# Find the estimated median survival time for two groups
test_median1<-ifelse(min(fit1$surv)<0.5,min(fit1$time[fit1$surv<=0.5]),min(fit1$time))
test_median2<-ifelse(min(fit2$surv)<0.5,min(fit2$time[fit2$surv<=0.5]),min(fit2$time))

# Define number of bootstrap repitition
k <- 50

# Define two vectors containing medians of bootstrap sample for two groups
med1<-rep(0,k)
med2<-rep(0,k)

# Bootstrap iteration
for (j in 1:k){
sim1<-data1[sample(length(data1$time),replace=T),]
sim2<-data2[sample(length(data2$time),replace=T),]
```

```
fit11<-survfit(Surv(time,event)~1,data=sim1)
fit12<-survfit(Surv(time,event)~1,data=sim2)
med1[j]<-ifelse(min(fit11$surv)<0.5,min(fit11$time[fit11$surv<=0.5]),min(fit11$time))
med2[j]<-ifelse(min(fit12$surv)<0.5,min(fit12$time[fit12$surv<=0.5]),min(fit12$time))
}

# Generating lower and upper limit for 95% CI for median1-median2
low1<-test_median1-qnorm(1-alpha1/2)*sd(med1)
low2<-test_median2-qnorm(1-alpha1/2)*sd(med2)
upper1<-test_median1+qnorm(1-alpha1/2)*sd(med1)
upper2<-test_median2+qnorm(1-alpha1/2)*sd(med2)

low<-low1-upper2
upper<- upper1-low2

show(c(low1,upper1))
show(c(low2,upper2))

show(c(sd(med1),sd(med2)))

show(c(low,upper))
```

BIBLIOGRAPHY

[1] N. Reid. *Estimating the median survival time.* Biometrika, **68**:601-608, 1981.

[2] Ron Brookmeyer and John Crowley. *A confidence interval for the median survival time.* Biometrics, **38**:29-41, 1982.

[3] John D. Emerson. *Nonparametric confidence intervals for the median in the presence of right censoring.* Biometrics, **38**:17-27, 1982.

[4] Christopher Jennison. *Repeated confidence intervals for the median survival time.* Biometrika, **33(1)**:45-51, 1984.

[5] Thomas P. Hettmansperger. *Two-sample inference based on one-sample sign statistics.* Appl. Statist., **49**:603-607, 1985.

[6] John Q. Su and L.J. Wei. *Nonparametric estimation for the difference or ratio of median failure times.* Biometrics, **49**:603-607, 1985.

[7] I.V. Basawa and H.L. Koul. *Large-sample statistics based on quadratic dispersion.* International Statistical Review, **56(3)**:199-219, 1988.

[8] Jane-ling Wang and Thomas P. Hettmansperger. *Two-sample inference for median survival times based on one-sample procedures for censored survival data.* Journal of the American Statistical Association, **85(410)**:529-536, 1990.

[9] Geoffrey McLachlan and David Peel. *Finite mixture models.* John Wiley & Sons, Inc., 2000.

[10] John P. Klein and Melvin L. Moeschberger. *Survival analysis: techniques for censored and truncated data.* Springer, 2003.

[11] Jong-Hyeon Jeong, Sin-Ho Jung and Joseph P. Costantino. *Nonparametric inference on median residual life function.* Biometrics, **64**:157-163, 2008.