12-2007

Diagnostics for Choosing between Log-Rank and Wilcoxon Tests

Ruvie Lou Maria Custodio Martinez
Western Michigan University

DIAGNOSTICS FOR CHOOSING BETWEEN LOG-RANK AND WILCOXON TESTS

by

Ruvie Lou Maria Custodio Martinez

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics
Dr. Joshua D. Naranjo, Advisor

Western Michigan University
Kalamazoo, Michigan
December 2007

DIAGNOSTICS FOR CHOOSING BETWEEN LOG-RANK AND WILCOXON TESTS

Ruvie Lou Maria Custodio Martinez, Ph.D.

Western Michigan University, 2007

Two commonly used tests for comparison of survival curves are the generalized Wilcoxon procedure of Gehan(1965) and Breslow(1970) and the Log-rank test proposed by Mantel(1966) and Cox(1972). In applications, the Log-rank test is used after checking for validity of the proportional hazards (PH) assumption, with Wilcoxon being the fallback method when the PH assumption fails.

However, the relative performance of the two procedures depend not just on the PH assumption but also on the pattern of differences between the two curves. We will show that the crucial factor is whether the differences tend to occur early or late in time. We propose diagnostics to measure early-or-late differences between two survival curves. A pretest based on either diagnostic will help the user choose the more efficient test under various patterns of treatment differences.

TABLE OF CONTENTS

Table of Contents – Continued

CHAPTER

Table of Contents – Continued

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivation and Background

The two-sample location problem is a fundamental technique used to determine the possible existence of treatment effect. In survival analysis we often test for whether treatment made a difference in the survival rates of the two groups. The two commonly used nonparametric tests for comparison of two survival distributions are the generalized Wilcoxon procedure of Gehan (1965) and Breslow (1970) and the Log-rank test proposed by Mantel(1966) and Cox(1972). Both tests are based on the ranks of the observations (Lee and Go 1997). The Log-rank test was derived by Cox under the assumption of proportional hazards and has been shown by Peto (1972) to be the locally most powerful rank-invariant test when there is a single parameter of interest and censorship is equal. Lee, Desu and Gehan (1975) have shown that when the hazard ratio is nonconstant the generalized Wilcoxon test can be more powerful than the Log-rank test.

Based on the above information there are situations where each of the two methods is more powerful than the other. In Chapter 2, we perform a power comparison of the two methods under various distributions and treatment effect

structure. In Chapter 3, we investigate properties of diagnostics for choosing between Log-rank test and Wilcoxon test. The investigation shows that it may be more useful to discriminate based not on proportional hazards assumption, but on whether treatment differences occur earlier or later in the range of comparison. In Chapter 4, we propose diagnostics for early and late treatment differences that will help the user choose between the two methods of testing equality of two survival distributions. Simulation results are presented for adaptive tests based on proposed diagnostics in Chapter 5 and Chapter 6.

1.2 Definition of Terms

In summarizing survival data, there are two functions of great interest, the survivor function and the hazard function.

We consider a population of individuals; for each individual we observe either the time to failure or the time to "loss" or "censoring". For the censored individuals we only know that the time to failure is greater than the censoring time.

We denote by T a random variable representing failure time. Let T have,

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \tag{1.1}$$

Let $S(t)$ be the survivor function,

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

and let $h(t)$ be the hazard function

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \tag{1.3}$$

The survivor function, the probability density function and the hazard function are equivalent ways of describing a continuous probability distribution. Given any one of them, we can compute for the other two. The relationship between the $p.d.f$ and the survivor function is,

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \tag{1.4}$$

While the hazard function can be expressed as,

$$h(t) = \frac{f(t)}{S(t)} \tag{1.5}$$

Together, equations (1.4) and (1.5) imply that

$$h(t) = -\frac{d}{dt} \log S(t) \tag{1.6}$$

Integrating both sides of equation (1.6) gives the expression for the survivor function in terms of the hazard function:

$$S(t) = \exp\left\{-\int_0^t h(u)du\right\} \tag{1.7}$$

The cumulative hazard function is given by

$$H(t) = \int_0^t h(u)du \tag{1.8}$$

Consequently we can write $S(t) = \exp[-H(t)]$ from (1.7)

Using equation (1.5), this formula leads to

$$f(t) = h(t) \exp\left\{-\int_0^t h(u)du\right\} \tag{1.9}$$

1.2.1 The Product-Limit Estimation

The standard estimator of the survival function, proposed by Kaplan and Meier (1958), is called the Product-Limit estimator. Let us denote the distinct failure times by

$$t_1 < t_2 < t_3 \ldots < t_k \tag{1.10}$$

The basic computations for the Kaplan-Meier survival curve rely on the computation of conditional survival probabilities. In particular, the probability

$$P[T \geq t_i \mid T \geq t_{i-1}] \tag{1.11}$$

which can be interpreted as the probability of an individual surviving to a specific time, given that the individual survived to the previous time. Another probability is the unconditional probability of survival,

$$P[T \geq t_i] \tag{1.12}$$

which represents the probability of survival to a specific time. Below is the relationship between the unconditional probability and the conditional probability:

$$P[T \geq t_i] = P[T \geq t_i \mid T \geq t_{i-1}]P[T \geq t_{i-1}] \tag{1.13}$$

We can apply this approach to get

$$P[T \geq t_i] = P[T \geq t_i \mid T \geq t_{i-1}]P[T \geq t_{i-1} \mid T \geq t_{i-2}]P[T \geq t_{i-2}] \tag{1.14}$$

and we can continue along these lines to get

$$P[T \geq t_i] = P[T \geq t_i \mid T \geq t_{i-1}]P[T \geq t_{i-1} \mid T \geq t_{i-2}]P[T \geq t_{i-2}] \ldots P[T \geq t_0] \tag{1.15}$$

The last probability represents the probability of an individual surviving at the start of the study, this probability has to be 1. Therefore, the unconditional probability is equal to the cumulative product of the conditional probabilities.

At each time point we observe the following: d_i, the number of deaths or failures at time t_i; c_i, the number of censored observations at time t_i and any between t_i and t_{i-1}; $n_i = n_{i-1} - d_{i-1} - c_{i-1}$, which is the number of individuals at risk entering the interval $(t_{i-1}, t_i]$. We also define n_0 as the total number of individuals in the study, c_0 as the number of censored observations prior to the first death or failure and $d_0 = 0$. The conditional probability of survival is given by

$$P[T \geq t_i \mid T \geq t_{i-1}] = 1 - \frac{d_i}{n_i} \tag{1.16}$$

And the unconditional probability of survival is the cumulative product of the conditional probabilities given by:

$$S(t) = P[T \geq t_i] = \prod_{j=1}^{i} \left(1 - \frac{d_j}{n_j}\right) \tag{1.17}$$

For uncensored data this is the usual sample survivor function; some of the asymptotic properties of (1.17) are given by Kaplan and Meier (1958) and by Efron (1967).

1.2.2 Proportional Hazards Model

D. R. Cox in his 1972 paper introduced two significant innovations. First, he proposed a model that is standardly referred to as the poportional hazards model. Second, he proposed a new estimation method called partial likelihood. The term Cox regression refers to the combination of the model and estimation method (Allison, 1995). In this paper, we will focus on the proportional hazards model more than the estimation method. The Cox proportional hazards model has been widely used in the biomedical field (Leemis, 1995) and recently there has been an increasing interest in its application in reliability engineering.

The Cox proportional hazards model can be written as

$$h_i(t/x_{i1}\ldots x_{ip}) = h_0(t)\exp\{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\}. \tag{1.18}$$

This equation states that the hazard for individual i at time t is the product of two factors:

1. a baseline hazard function $h_0(t)$ that is unspecified
2. a linear function of a set of fixed p covariates $(x_1, \ldots, x_p)$, which is exponentiated.

The function $h_0(t)$ can be considered as the hazard function for an individual whose covariates are all equal to 0.

Taking the logarithm of both sides, we can rewrite the model in (1.18) as

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \tag{1.19}$$

where $\alpha(t) = \log h_0(t)$. If we further specify $\alpha(t) = \alpha$, which is the simplest function that says the hazard is constant over time, we will get the exponential model. If we specify $\alpha(t) = \alpha t$, we will get the Gompertz model. And if we specify $\alpha(t) = \alpha \log(t)$, we will have a Weibull model. Thus, the proportional hazards model is generalization of the Exponential, Weibull and Gompertz model (Allison, 1995). However, the great attraction of the Cox proportional hazards model is that $h_0(t)$ can take any form.

It is called proportional hazards model because the hazard for any individual is a fixed proportion of the hazard for any other individual. To illustrate, take the ratio of the hazards for two individuals i and j, and use (1.18):

$$\frac{h_i(t)}{h_j(t)} = \exp\{\beta_1(x_{i1} - x_{j1}) + \cdots + \beta_p(x_{ip} - x_{jp})\} \tag{1.20}$$

The $h_0(t)$ cancels out of the numerator and denominator. As a result, the ratio of the hazards is constant over time.

To illustrate a two-sample problem, let

$$x_i = \begin{cases} 0 & \text{if individual belongs to Group 1} \\ 1 & \text{if individual belongs to Group 2} \end{cases}$$

Then following (1.18) the hazard functions for Group 1 and Group 2 are $h_1(t)$ and $h_2(t) = h_1(t)\psi$ respectively where $\psi = \exp(\beta)$. Lehmann(1953) showed that in the continuous case the survivor functions are related by $S_2(t) = [S_1(t)]^{\psi}$.

Lemma 1.1 (Lehmann alternative) *Let $S_1(t)$ and $S_2(t)$ be the survival functions of Group 1 and Group 2 respectively. If $h_2(t) = \psi h_1(t)$ then the survivor functions are related by $S_2(t) = [S_1(t)]^{\psi}$.*

Proof.

$$\begin{aligned} h_2(t) &= h_1(t)\psi \\ -\int_0^t h_2(u)du &= -\int_0^t \psi h_1(u)du \\ \exp\left\{-\int_0^t h_2(u)du\right\} &= \exp\left\{-\int_0^t \psi h_1(u)du\right\} \\ \exp\left\{-\int_0^t h_2(u)du\right\} &= \exp\left\{-\psi\int_0^t h_1(u)du\right\} \end{aligned}$$

Using (1.7), S(t) $= \exp\left\{-\int_0^t h(u)du\right\}$ we have,

$$S_2(t) = [S_1(t)]^{\psi} \tag{1.21}$$

The Lehmann alternative (1.21) may be seen as the equivalent to the proportional hazards model in the two-sample problem.

1.2.3 Description of Log-rank Test and Wilcoxon Test

We begin by considering that there are n_1 and n_2 patients who are allocated between two groups. We consider tests of the hypothesis that the patient groups have the same survival distribution against the alternative that the patients in Group 2 have improved survival. Suppose that there are r distinct death times, t_1, t_2, ..., t_r, across the two groups, and at a certain time t_j there are d_{1j} individuals who died in Group 1 and d_{2j} individuals in Group 2, for j=1,2,..., r. For each particular death time we can create the following contingency table

Table 1.1: Number of deaths at the j'th death time in Group 1 and Group 2

Group	**Number of deaths at t_j**	**Number of surviving beyond t_j**	**Number at risk just before t_j**
1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
2	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Total	d_j	$n_j - d_j$	n_j

Here d_{1j} is a random variable with a *hypergeometric* distribution with null mean given by $e_{1j} = \frac{n_{1j}d_j}{n_j}$ and variance $v_{1j} = \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}$. Combining the information from each 2 x 2 table for each t_j gives an overall measure of the deviation of the observed values of d_{1j} from their expected values we have the statistic

$$U_L = \sum_{j=1}^{r}(d_{1j} - e_{1j}). \tag{1.22}$$

This statistic will have a zero mean, since $E(d_{1j}) = e_{1j}$. Since the deaths are independent of each other, the variance of U_L is the sum of the variances of the d_{1j},

$$var(U_L) = \sum_{j=1}^{r}(v_{1j}) = V_L. \tag{1.23}$$

This procedure is similar to the Mantel-Haenszel test and some authors refer to it as the Cox-Mantel test (Mantel and Haenszel, 1959). Thus, the test statistic

$$\frac{U_L^2}{V_L} \tag{1.24}$$

summarize the extent to which the observed survival times in the two groups of data deviate from those expected under the null hypothesis of no group differences. The test statistic will have an asymptotic chi-squared distribution with one degree of freedom under the null hypothesis (Cox, 1972)

The generalized Wilcoxon test of Gehan (1965a,b) is a generalization of Wilcoxon's two sample rank sum test for the case of censored data. The Wilcoxon test is based on the statistic

$$U_W = \sum_{j=1}^{r} n_j(d_{1j} - e_{1j}). \tag{1.25}$$

with variance $V_W = \sum_{j=1}^{r} n_j^2 v_{1j}$, and so the Wilcoxon test statistic is

$$\frac{U_W^2}{V_W} \tag{1.26}$$

which will have an asymptotic chi-squared distribution with one degree of freedom under the null hypothesis (Gehan, 1965). This differ from the Log-rank statistic (1.24) by the weights n_j, the total number of individuals at risk at time t_j. Thus the Wilcoxon statistic gives greater weight to differences occurring near the beginning

of a study, and is less sensitive to events occurring when very few individuals under study remain alive (Taron and Ware, 1977).

Another version of Wilcoxon's two sample rank sum test for complete observations is described by Peto and Peto (1972). The weight used are the Kaplan-Meier estimate of the survival function,

$$U_{PP} = \sum_{j=1}^{r} \tilde{S}(t)(d_{1j} - e_{1j}). \tag{1.27}$$

This test reduces to Gehan's Wilcoxon test when there are no censored observations. (Lee, Desu and Gehan, 1975)

1.2.4 Some Survival Distributions (Weibull, Log-normal and Log-logistic)

Weibull Distribution

The Weibull distribution is a continuous probability distribution characterized by two parameters, γ (shape parameter) and λ (scale parameter), with *probability density function*

$$f(t) = \lambda\gamma(\lambda t)^{\gamma-1} \exp\{-(\lambda t)^{\gamma}\}. \tag{1.28}$$

The mean and variance are $\frac{\Gamma(1+1/\gamma)}{\lambda}$ and $\frac{\Gamma(1+2/\gamma) - [\Gamma(1+1/\gamma)]^2}{\lambda^2}$, respectively, where $\Gamma(\gamma) = \int_0^\infty u^{\gamma-1}e^{-u}du$ is the gamma function. The *survival function* is expressed as

$$S(t) = \exp[-(\lambda t)^{\gamma}] \tag{1.29}$$

and the corresponding *hazard function* as

$$h(t) = \lambda\gamma(\lambda t)^{\gamma-1}. \tag{1.30}$$

When γ=1, the hazard rate remains constant as time increases (this is the exponential case). The hazard rate increases when γ >1 and decreases when γ <1 as t increases. Since Weibull distribution may be used to model increasing, decreasing and constant risk, it has a broader application. The following published papers used the Weibull distribution to model survival times.

1. Elketroussi M, Fan DP. 1991. Time trends of smoking cessation analyzed with six mathematical survival models. *Int. J. Biomed. Comput.* 27:231-44

2. Hendricks JCM, Medley GF, Van Griensven GJP, Coutilinho RA, Heisterkamp SH, et al. 1993. The treatment-free incubation period of AIDS in a cohort of homosexual men. *AIDS* 7:231-39

3. Juckett DA, Rosenberg B. 1993. Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. *Mech. Ageing Dev.* 69:1-31

4. Scott BR, Hahn FF. 1980. A model that leads to the Weibull distribution function to characterize early radiation response probabilities. *Health Phys* 39:521-30

5. Williams AJ, Al-Katib A, Wong GY, Jhanwar SC, Chaganti RSK, et al 1978. Efficient analysis of Weibull survival data from experiments on heterogeneous patient populations. *Biometrics* 34:209-22

Lognormal Distribution

The lognormal is defined as the distribution of a variable whose logarithm follows a normal distribution. Consider T as the random variable whose $\log T$ is normally distributed with mean μ and variance σ^2, then T has a lognormal distribution with parameters μ and σ^2. The *probability density function* is

$$f(t) = \frac{1}{t\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right]. \tag{1.31}$$

The mean and variance are $\exp(\mu + 0.5\sigma^2)$ and $[\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$, respectively. The *survival function* is expressed as

$$S(t) = 1 - \Phi\left[\frac{\ln t - \mu}{\sigma}\right], \tag{1.32}$$

and the corresponding *hazard function* as

$$h(t) = \frac{f(t)}{S(t)}. \tag{1.33}$$

The hazard function increases initially to a maximum and then decreases, almost as the median is passed, toward zero as time approaches infinity (non-monotonic hazards). The hazard function is an inverse-bath tub shape. The following published papers used the Lognormal distribution to model survival times.

1. Ahmed FE, Hattis D, Wolke RE, Steinman D. 1993. Human health risks due to consumption of chemically contaminated fishery products. *Environ. Health Perspect.* 101:297-302

2. Horner RD. 1987. Age at onset of Alzheimer's disease: clue to the relative importance of etiologic factors? *Am. J. Epidemiol.* 126:409-14

3. Larsen RI, McDonnell WF, Horstman DH. 1991. An air quality data analysis system for interrelating effects, standards, and needed source reducations: Part II. A lognormal model relating human lung function decrease to O_3 exposure. *J. Air Waste Manage. Assoc.* 41:455-59

Log-logistic Distribution

Similar to the lognormal distribution, if the variable log T has a logistic distribution, the variable T follows the log-logistic distribution. The log-logistic distribution also has two parameters, $\alpha = \frac{1}{\sigma}$ and $\lambda = \exp\left(-\frac{\mu}{\sigma}\right)$. The *probability density function* is

$$f(t) = \frac{\alpha\lambda t^{\alpha-1}}{[1+\lambda t^{\alpha}]^2}. \tag{1.34}$$

The mean and variance are $\frac{\pi \csc(\pi/\alpha)}{\alpha\lambda^{1/\alpha}}$ if $\alpha > 1$, and $\frac{2\pi \csc(2\pi/\alpha)}{\alpha\lambda^{2/\alpha}} - E[T]^2$, if $\alpha > 2$.The *survival function* is expressed as

$$S(t) = \frac{1}{1+\lambda t^{\alpha}} \tag{1.35}$$

and the corresponding *hazard function* as

$$h(t) = \frac{\alpha\lambda t^{\alpha-1}}{1+\lambda t^{\alpha}} \tag{1.36}$$

When α <1 the hazard rate decreases from infinity toward 0 when α=1, it decreases from λ to 0, and when α >1, it increases from 0 to a maximum and then decreases toward 0. The following published papers used the log-logistic distribution to model survival times.

1. Conkin, J. 2001. A Log Logistic Survival Model Applied to Hypobaric Decompression Sickness. NASA. *http://techreports.larc.nasa.gov/cgi-bib/NTRS*

2. Elketroussi M, Fan DP. 1991. Time trends of smoking cessation analyzed with six mathematical survival models. *Int. J. Biomed. Comput.* 27:231-44

3. Schmidt P, Witte DA. 1988. *Predicting Recidivism Using Survival Models.* NY: Springer-Verlag

Table 1.2: Summary of density, survival and hazard function with mean and variance

Distribution	**Parameters**	**Density, survival and hazard function Mean and Variance**
Weibull	$\lambda,\ \gamma > 0$	$f(t) = \lambda\gamma(\lambda t)^{\gamma-1}\exp\{-(\lambda t)^{\gamma}\}$ $S(t) = \exp[-(\lambda t)^{\gamma}]$ $h(t) = \lambda\gamma(\lambda t)^{\gamma-1}$ $\mu = \frac{\Gamma(1+1/\gamma)}{\lambda}$ $\sigma^2 = \frac{\Gamma(1+2/\gamma) - [\Gamma(1+1/\gamma)]^2}{\lambda^2}$
Lognormal	$\mu,\ \sigma > 0$	$f(t) = \frac{1}{t\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{lnt-\mu}{\sigma}\right)^2\right]$ $S(t) = 1 - \Phi\left[\frac{lnt-\mu}{\sigma}\right]$ $h(t) = \frac{f(t)}{S(t)}$ $\mu = \exp(\mu + 0.5\sigma^2)$ $\sigma^2 = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$
Log-logistic	$\alpha = \frac{1}{\sigma}$ $\lambda = \exp\left(-\frac{\mu}{\sigma}\right)$	$f(t) = \frac{\alpha\lambda t^{\alpha-1}}{[1+\lambda t^{\alpha}]^2}$ $S(t) = \frac{1}{1+\lambda t^{\alpha}}$ $h(t) = \frac{\alpha\lambda t^{\alpha-1}}{1+\lambda t^{\alpha}}$ $\mu = \frac{\pi\csc(\pi/\alpha)}{\alpha\lambda^{1/\alpha}}$ if $\alpha > 1$ $\sigma^2 = \frac{2\pi\csc(2\pi/\alpha)}{\alpha\lambda^{2/\alpha}} - E[T]^2$, if $\alpha > 2$

1.3 Proportional Hazards Model, Lehmann Alternative and the Weibull Distribution

A general model for comparing two groups of survival times is the proportional hazards model. To illustrate a two-sample problem under the proportional hazards model, let

$$x_i = \begin{cases} 0 & \text{if ith individual belongs to Group 1} \\ 1 & \text{if ith individual belongs to Group 2} \end{cases}$$

Then the hazard function for the ith individual is

$$h(t) = h_0(t)\exp(\beta x_i) = \begin{cases} h_0(t) & \text{if ith individual belongs to Group 1} \\ h_0(t)\psi & \text{if ith individual belongs to Group 2} \end{cases}$$

where $\psi = \exp(\beta)$. Consequently, the hazard functions for the two groups are:

$$\begin{aligned} h_1(t) &= h_0(t) \\ h_2(t) &= h_0(t)\psi \end{aligned}$$

So the relationship between the two hazard functions is,

$$h_2(t) = h_1(t)\psi \tag{1.37}$$

From Lemma 1.1, condition (1.37) is equivalent to assuming $S_2(t) = S_1(t)^{\psi}$.

Lemma 1.2 *Let $T_1 \sim$ Weibull(γ, λ) and let the proportional hazards assumption (1.37) be satisfied. Then $T_2 \sim$ Weibull $(\gamma, \psi^{1/\gamma}\lambda)$, for some constant ψ.*

Proof. T_1 has hazard function $h_1(t) = \lambda\gamma(\lambda t)^{\gamma-1}$. Then T_2 has hazard function,

$$\begin{aligned} h_2(t) &= \psi h_1(t) \\ &= \psi\lambda\gamma(\lambda t)^{\gamma-1} \\ &= \psi\lambda^{\gamma}\gamma t^{\gamma-1} \\ &= \psi^{1/\gamma}\psi^{(\gamma-1)/\gamma}\lambda\lambda^{\gamma-1}\gamma t^{\gamma-1} \\ &= \psi^{1/\gamma}\lambda\gamma[\psi^{1/\gamma}\lambda t]^{\gamma-1} \\ &= \lambda^*\gamma[\lambda^* t]^{\gamma-1} \end{aligned}$$

Lemma 1.2 shows that proportional hazards under the Weibull distribution is satisfied if and only if the shape parameter γ stays the same.

Lemma 1.3 *Let* $T_1 \sim Weibull(\gamma, \lambda)$, $T_2 = cT_1$. *Since* T_1 *and* T_2 *satisfy the proportional hazards assumption with* $h_2 = \psi h_1(t)$ *and* $S_2(t) = [S_1(t)]^{\psi}$ *then* $T_2 \sim Weibull(\gamma, \lambda/c)$.

Proof.

$$\begin{aligned} S_2(t) &= P[T_2 \geq t] \\ &= P[cT_1 \geq t] \\ &= P\left[T_1 \geq \frac{t}{c}\right] \\ &= S_1\left(\frac{t}{c}\right) \end{aligned}$$

Lemma 1.3 shows that scalar transformation under the Weibull distribution maintains the shape parameter γ and only change the scale parameter λ thus satisfiying the proportional hazards assumption.

Chapter 2

Log-rank Test Versus Wilcoxon Test

2.1 Some Known Results

Tarone and Ware (1977) demonstrated that the test statistics for the Log-rank and Wilcoxon differ only in the choice of weights, which are functions of the total number of individuals at risk at each event time. In discussing the weights utilized by the Log-rank and generalized Wilcoxon statistics Tarone and Ware (1977) have noted that the generalized Wilcoxon statistic gives more weight to early events and the Log-rank statistic is more sensitive to late occurring events. The Log-rank test, proposed by Mantel (1966) and discussed extensively by others [Cox (1972), Breslow (1975), Peto (1972), Peto and Peto (1972) and Peto and Pike (1973)], is known to be fully efficient rank test under Lehmann alternatives, $S_2(t) = [S_1(t)]^{\psi}$ (alternatives in which the relative hazard is constant) when censoring distributions are equal. In general, the Log-rank test tends to be sensitive to distributional differences which are most evident late in time. In comparison, a generalized Wilcoxon procedure proposed by Gehan (1965) has been found to be more powerful in detecting differences very evident early in time (Lee, Desu and Gehan (1975);

Prentice and Marek (1979); also some results by Lee and Thomas presented at meetings in Boston in 1976).

Tarone and Ware (1977) and Lee, Desu and Gehan (1975) indicate that while tests based on the Logrank statistic perform extremely well when the underlying distributions have constant hazard ratios, they may lose considerable power under deviations from the proportional hazard model. The assumption of proportional hazards is that the hazard of death at any given time for an individual in one group is proportional to the hazard at that time for a similar individual in the other group.

Collett (1994) suggested that in order to help decide which test is the more suitable in any given situation, one make use of the result that if the hazard functions are proportional, the survival functions for the two groups do not cross. Since the survivor function takes values between 0 and 1, it can be shown under Lehmann-alternative (1.21) that $S_1(t)$ is greater than or less than $S_2(t)$, depending on whether ψ is less than or greater than 1 at any time t. This means that if two hazard functions are proportional, the true survivor functions do not cross. This is a necessary, but not a sufficient condition for proportional hazards.

There have been several graphical methods suggested for assessing the proportional hazards assumption (Hess, 1995). The commonly used graphical method that is available on most statistical software is the plotting of the log of the cumulative hazard function against log time and checking for parallelism. A plot of the values of log-log S(t) against log t is known as the *log-cumulative hazard plot.*

Although there have been many numerical goodness-of-fit statistics proposed to detect violations of the PH assumption, none of these gained wide use, and they are limited both by statistical power and the family of alternatives con-

sidered. (Hess, 1995)

Fleming et al (1980) found out on the course of their research that both the Log-rank test and the Gehan-Wilcoxon test may be insensitive to certain commonly occurring departures from the null hypothesis. They have seen that substantial differences between two survival distributions may be apparent at one point in time, but fail to exist elsewhere. Stablein and Koutrouvelis (1985) mentioned in their paper the concern expressed by practitioners that alternative distributions may not be from a set of location shift alternatives. This departure from the null hypothesis fall within the class commonly referred to as "crossing-hazards alternatives". Careful inspection of the Log-rank and Gehan-Wilcoxon test statistics suggests that test procedures based upon these statistics may be insensitive to such departures. Neither of these tests described is designed to detect differences where survival curves cross or differ in other more general ways, because the tests are based on the weighted integral of estimated difference between the survival curves.

2.2 A Simulation Study

To compare the performance of the Log-rank and Wilcoxon tests, data were generated by computer simulation from Weibull, Lognormal and Log-logistic distribution. The 50 survival times for each group are observed to fail (uncensored case). Let T_1 denote the Group 1 (or control group) random variable. In all simulation cases, T_1 will have mean 100 and distribution either Weibull, Lognormal or Log-logistic.

Let T_2 denote the Group 2 (or treatment group) random varible. We will compare the power of Log-rank and Wilcoxon under the following alternative hypothesis treatment structures:

1. $T_2 = cT_1$, where $c > 1$ (scale transformation)

2. $T_2 = T_1^c$, where $c > 1$ (this is equivalent to $\log T_2 = c \log T_1$, scalar transformation in the log scale)

Two additional treatment effects under the Weibull distribution were also investigated (Cases 3 and 4).

Simulation were done 10,000 times for each case of treatment effects to show the size and power of the Log-rank and Wilcoxon tests. The alternative hypothesis used is that the treatment group have a higher survival times than the control group.

- **WEIBULL Distribution**

  Case 1:

  ⋆ $T_1 \sim \text{Weibull}(\gamma, \lambda)$

  ⋆ $T_2 = cT_1,\ c > 1$

  $T_2 \sim \text{Weibull}(\gamma, \lambda/c)$ and proportional hazards assumption is satisfied

  Case 2:

  ⋆ $T_1 \sim \text{Weibull}(\gamma, \lambda)$

  ⋆ $T_2 = T_1^c,\ c > 1$

  $T_2 \sim \text{Weibull}(\gamma/c, \lambda^c)$ and proportional hazards assumption is not satisfied since shape parameter changed

  Case 3:

  ⋆ $T_1 \sim \text{Weibull}(\gamma, \lambda)$

  ⋆ $T_2 \sim \text{Weibull}(c\gamma, \lambda/c),\ c > 1$

  Proportional hazards assumption is not satisfied since shape parameter changed

  Case 4:

  ⋆ $T_1 \sim \text{Weibull}(\gamma_1, \lambda_1)$

  ⋆ $T_2 \sim \text{Weibull}(\gamma_2, \lambda_2)$, where $\gamma_1 < \gamma_2$ and $\lambda_1 > \lambda_2$

  Proportional hazards assumption is not satisfied since shape parameter changed

- **LOGNORMAL Distribution**

  Case 5:

  ⋆ $T_1 \sim \text{Lognormal}(\mu, \sigma)$

  ⋆ $T_2 = cT_1,\ c > 1$

  $T_2 \sim \text{Lognormal}(\ln + \mu, \sigma)$ and proportional hazards assumption is not satisfied

  Case 6:

  ⋆ $T_1 \sim \text{Lognormal}(\mu, \sigma)$

  ⋆ $T_2 = T_1^c,\ c > 1$

  $T_2 \sim \text{Lognormal}(c\mu, c\sigma)$ and proportional hazards assumption is not satisfied

- **LOG-LOGISTIC Distribution**

  Case 7:

  ⋆ $T_1 \sim \text{Log-logistic}(\alpha, \lambda)$

  ⋆ $T_2 = cT_1,\ c > 1$

  $T_2 \sim \text{Log-logistic}(\alpha, \lambda/c^\alpha)$ and proportional hazards assumption is not satisfied

  Case 8:

  ⋆ $T_1 \sim \text{Log-logistic}(\alpha, \lambda)$

  ⋆ $T_2 = T_1^c,\ c > 1$

  $T_2 \sim \text{Log-logistic}(\alpha/c, \lambda)$ and proportional hazards assumption is not satisfied

2.3 Simulation Results Under Weibull Distribution (Cases 1, 2, 3 and 4)

Case 1

If survival time T_1 of individuals in Group 1 have a Weibull distribution with shape parameter γ, and the hazard for an individual in the second group is proportional to that of an individual in the first, the survival times for those in the second group will also have a Weibull distribution with shape parameter γ (Collett, 1994). Multiplying a constant to the survival time only changes the scale parameter λ and maintains the same value of the shape parameter γ. Therefore, when control T_1 has a Weibull distribution and treatment $T_2 = cT_1$, then T_1 and T_2 satisfy the proportional hazards assumption.

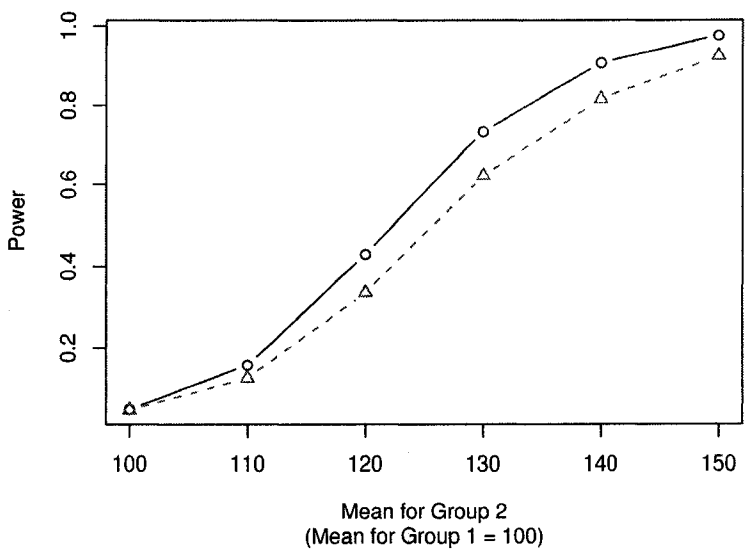

Figure 2.1: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with equal shape parameters $\gamma = 2$ and treatment effect cT: ○, Log-rank; △, Wilcoxon

Since the Log-rank test is efficient under proportional hazards, the Log-rank is expected to have higher power than Wilcoxon. This is confirmed by our simulation results for Case 1, presented in Figure 2.1.

Case 2

If $T_2 = T_1^c$, then T_2 has a Weibull distribution with scale parameter $\lambda^* = \lambda^c$ and shape parameter $\gamma^* = \gamma/c$. Therefore, proportional hazards assumption is not satisfied since $\gamma^* \neq \gamma$. However, Log-rank test has higher power than the Wilcoxon test (Figure 2.2) even if the proportional hazards assumption is violated.

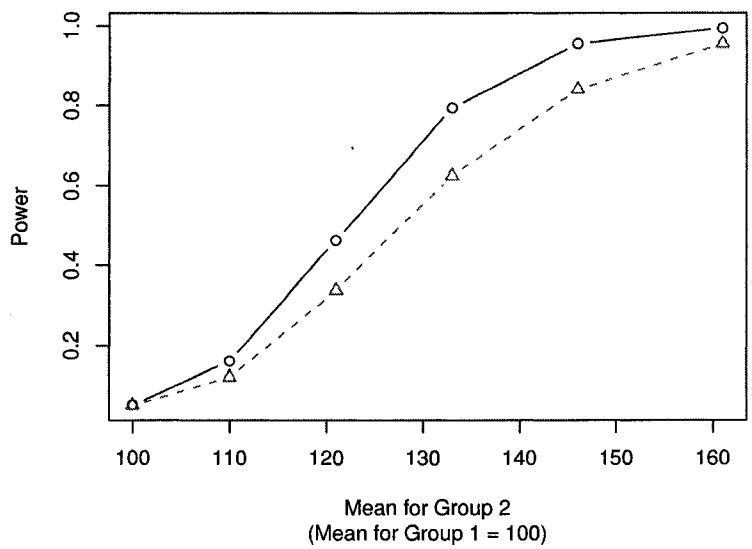

Figure 2.2: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with control group parameters γ =2, λ=0.008862 and treatment effect T^c: ○, Log-rank; △, Wilcoxon

Case 3

Another treatment effect is when survival times were generated under the Weibull distribution where the shape parameter is increased by multiplying it by c ($c\gamma$) and the scale parameter was decreased by dividing it by c (λ/c). This has the effect of increasing the mean survival time for the treatment group. In this case proportional hazards assumption is not satisfied and it is confirmed in our simulation that the Wilcoxon test has higher power than the Log-rank test, Figure 2.3.

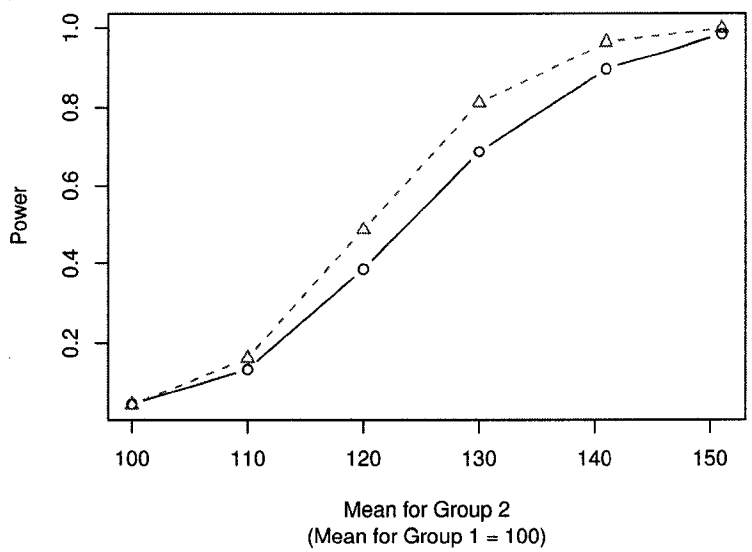

Figure 2.3: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with shape parameters $\gamma_1 = 2$ and shape is increased by c and scale is decreased by c: ◯, Log-rank; △, Wilcoxon

Case 4

In this case the shape parameters of the two groups are different and the mean of the treatment group is higher. For this simulation the shape parameter for Group 1 is 2 and the shape parameter for Group 2 is 3. I adapted the shape parameter values from Ng'Andu(1997) and the mean of the control group from Lee, Desu and Gehan (1975). Proportional hazards assumption is not satisified since the shape parameters of the two groups are not the same. In Figure 2.4 we can see that the power of the Wilcoxon test is higher than the Log-rank test.

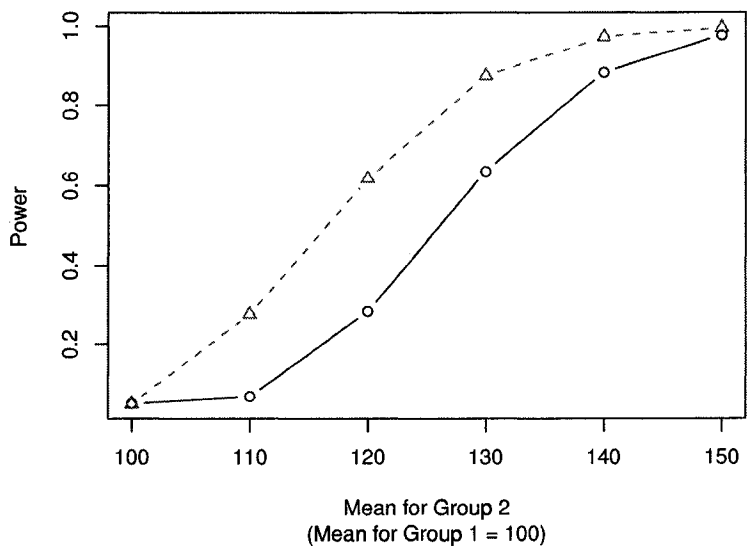

Figure 2.4: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with shape parameters γ_1=2 and γ_2=3 for Group 1 and Group 2, respectively: ○, Log-rank; △, Wilcoxon

2.4 Simulation Results Under Lognormal Distribution (Cases 5 and 6)

Case 5

Let T_1 follow a Lognormal distribution with parameters μ and σ. If $T_2 = cT_1$, then T_2 will follow a Lognormal distribution with parameters $\ln c + \mu$ and σ. Here the Wilcoxon test has a higher power than the Log-rank test as can be seen in Figure 2.5.

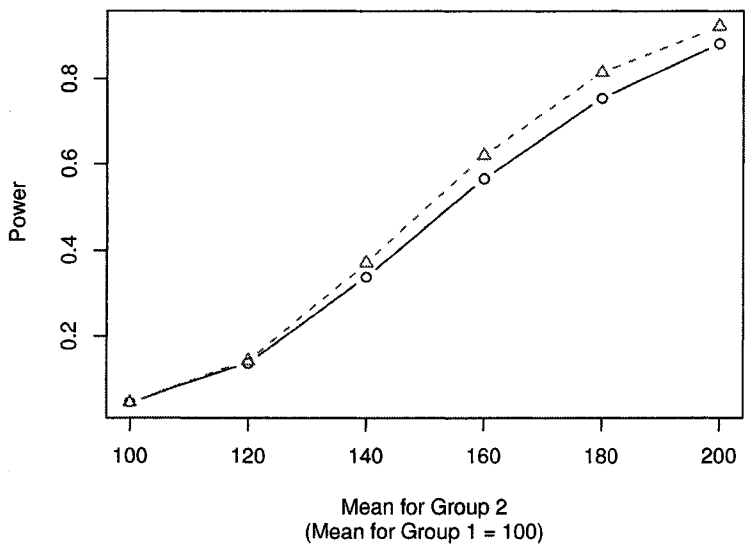

Figure 2.5: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Lognormal distribution with parameters μ=4.1052 and σ=1 for Group 1 and treatment effect cT for Group 2: ◯, Log-rank; △, Wilcoxon

Case 6

On the other hand, if $T_2 = T_1^c$, then T_2 will follow a Lognormal distribution with parameters $c\mu$ and $c\sigma$. In this case the Log-rank test has higher power than the Wilcoxon test as can be seen in Figure 2.6.

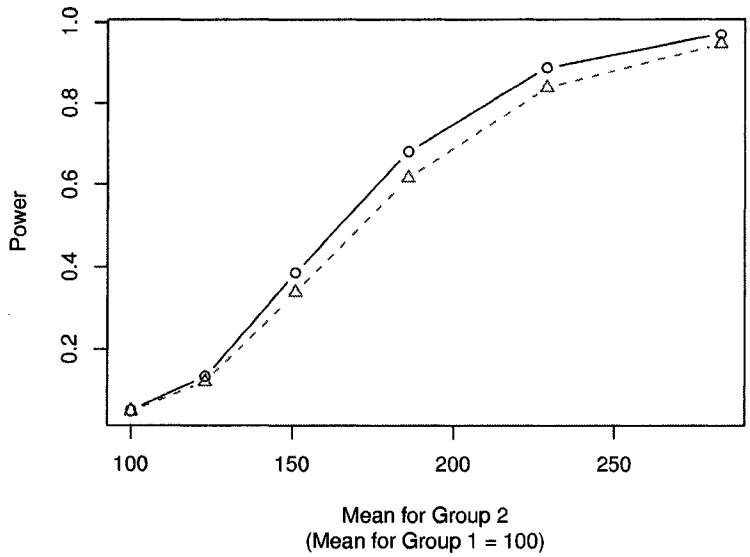

Figure 2.6: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Lognormal distribution with parameters μ=4.1052 and σ=1 for Group 1 and treatment effect T^c for Group 2: ○, Log-rank; △, Wilcoxon

2.5 Simulation Results Under Log-logistic Distribution (Cases 7 and 8)

Case 7

Let the survival time T_1 follow a Log-logistic distribution with parameters α and λ. If $T_2 = cT_1$, then T_2 will follow a Log-logistic distribution with parameters α and λ/c^{α}. Our simulation shows that the Wilcoxon test has a higher power than the Log-rank test (Figure 2.7).

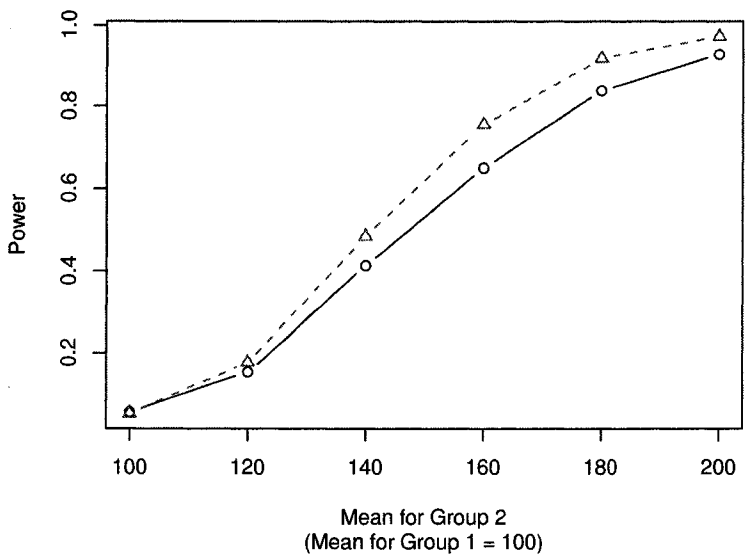

Figure 2.7: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Log-logistic distribution with parameters μ=4.1536 and σ=0.5 for Group 1 and treatment effect cT for Group 2: ○, Log-rank; △, Wilcoxon

Case 8

The survival time is raised to a constant ($T_2 = T_1^c$) and $T_2 \sim$ Log-logistic$(\alpha/c, \lambda)$. Our simulation shows that not much difference is seen between the powers of the Log-rank test and Wilcoxon test (Figure 2.8).

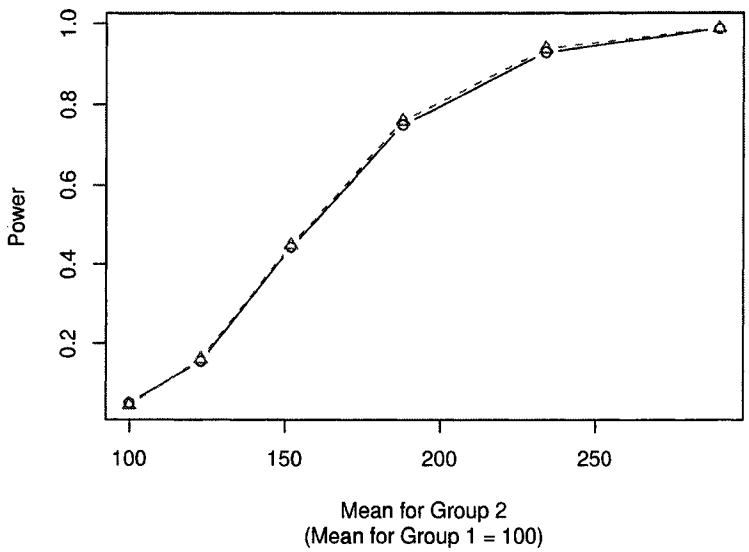

Figure 2.8: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Log-logistic distribution with parameters μ=4.1536 and σ=0.5 for Group 1 and treatment effect T^c for Group 2: ◯, Log-rank; △, Wilcoxon

2.6 Summary

Case 1 satisfy the proportional hazards assumption. For this case, the Log-rank test out performs the Wilcoxon test. In Cases 3, 4 and 5 the proportional hazards assumption is violated. In these cases, the Log-rank test performs worst than the Wilcoxon test. In Cases 2, 6, and 7 the proportional hazards assumption is violated, but the Log-rank test still out performs the Wilcoxon test. In Case 8 the proportional hazards assumption is not satisfied and the power of Log-rank test and Wilcoxon test shows not much difference.

Chapter 3

Diagnostics for Choosing Between Log-rank Test and Wilcoxon Test

3.1 Log-cumulative hazard plot

It is a common practice on dealing with two-sample problem in survival analysis that Log-rank test is used under the assumption of proportional hazards and the Wilcoxon test otherwise. Several graphical methods were suggested for assessing the proportional hazards assumption and the commonly used graphical method that is available on most statistical software is the plotting of the log of the cumulative hazard functions against log time and checking for parallelism (Hess, 1995). A plot of the values of log-log S(t) against log t is known as the *log-cumulative hazard plot.*

It can be shown using (1.21) under the two-sample problem,

$$\begin{aligned} S_2(t) &= [S_1(t)]^{\exp(\beta)} \\ \log S_2(t) &= \exp(\beta) \log S_1(t) \\ -\log S_2(t) &= -\exp(\beta) \log S_1(t) \end{aligned}$$

$$\log\{-\log S_2(t)\} = \beta + \log\{-\log S_1(t)\}$$
$$\log[H_2(t)] = \beta + \log[H_1(t)]$$

Thus, plot of the two log cumulative hazard functions will be parallel under the proportional hazards assumption against log t.

When $T_1 \sim$ Weibull(γ, λ), the log-cumulative hazard plot will give a straight line with intercept $\gamma \log \lambda$ and slope γ since the survival function for Weibull is $\exp[-(\lambda t)^\gamma]$. In the Weibull model, the assumption of proportional hazards between two groups corresponds to the assumption that the shape parameter γ is the same in each group. It then follows that if $T_2 \sim$ Weibull(γ, λ^*), as they would in the proportional hazards model in (1.18), the log-cumulative hazard plot will give a straight line, also of slope γ, but with intercept $\gamma \log \lambda^*$. Parallel straight lines would mean that the assumption of proportional hazards model and Weibull survival time are satisfied (Figure 3.1). If the two lines in a log-cumulative hazard plot are straight but not parallel, this means that the shape parameter γ is different in the two groups, and the hazards are no longer proportional. If lines are not particularly straight, the Weibull model may not be appropriate. However, if the curves can be taken to be parallel, this would mean that the proportional hazards model is valid, and the Cox regression model might be satisfactory (Figure 3.2).

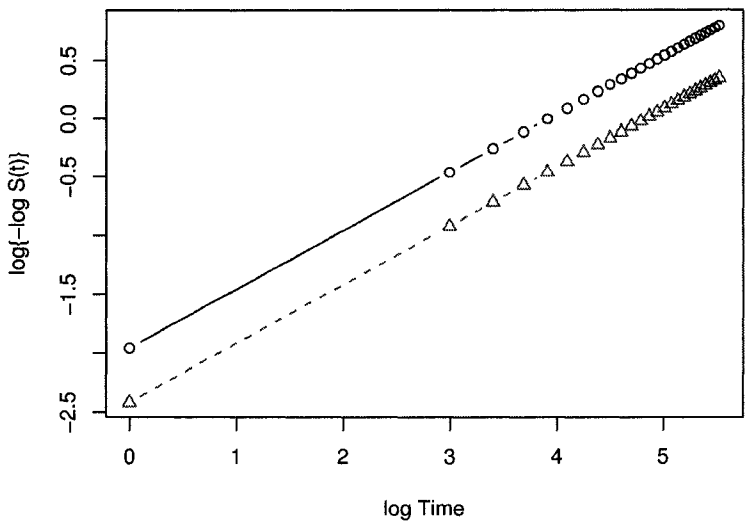

Figure 3.1: Log-cumulative hazards plot for two groups with Weibull survival times and satisfies the proportional hazards model assumption

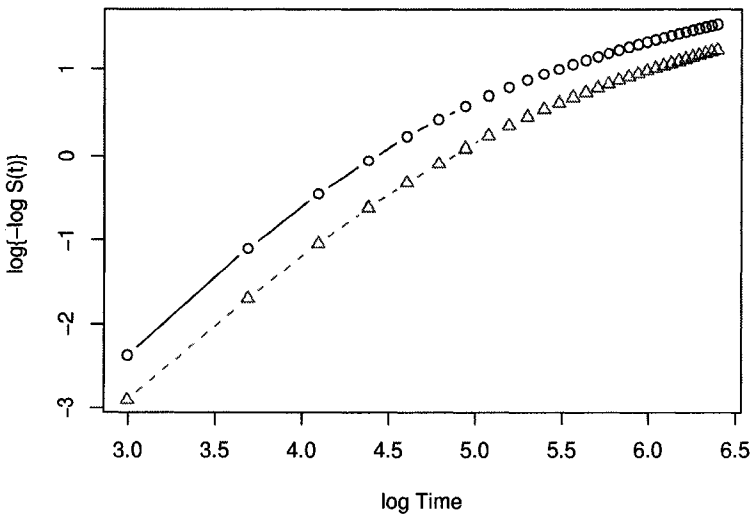

Figure 3.2: Log-cumulative hazards plot for two groups that satisfies the proportional hazards model assumption

The log-cumulative hazard plot will correctly identify the more powerful test in simulation Cases 1 and 4, but fail in simulation Case 2.

For simulation Case 2, the log-cumulative hazard plots are diverging (Figure 3.3). This tells the user to use the Wilcoxon test. However, our simulation shows (Figure 2.2) that Log-rank is the better test for this case.

The lesson here is to detect not just whether proportional hazards assumption is violated, but how it is violated. In simulation Case 2, proportional hazards assumption is violated but treatment effect is late rather than early in the survival range (Figure 3.4). This suggest diagnostics not just for proportional hazards assumption but for whether separation between the two curves is early or late.

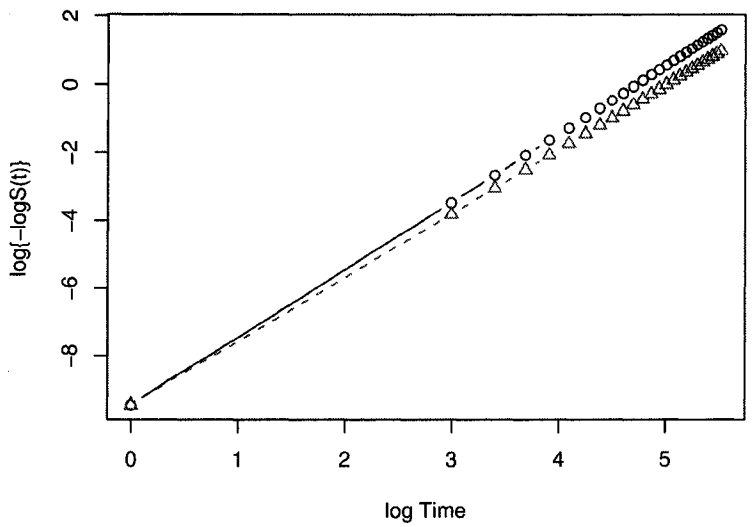

Figure 3.3: Log-cumulative hazard plot for Weibull distribution with control group parameters ($\gamma = 2$, $\lambda = 0.008862$) and treatment effect T^c: ○, Control Group; △, Treatment Group

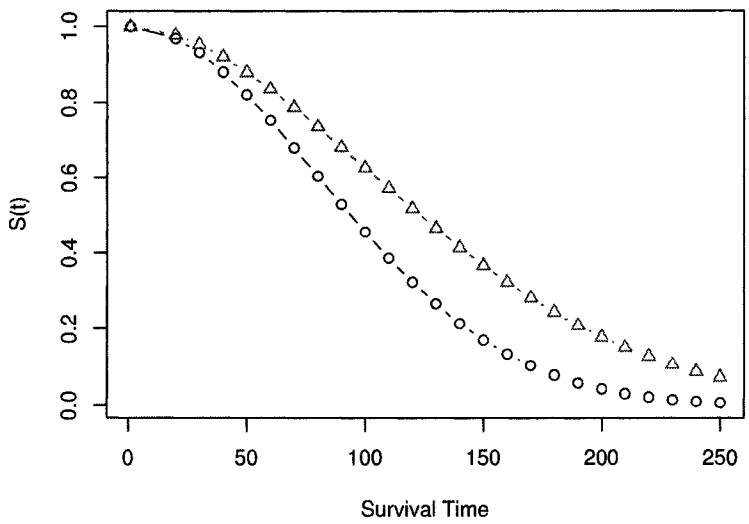

Figure 3.4: Survival functions for Weibull distribution with control group parameters ($\gamma = 2$, $\lambda = 0.008862$) and treatment effect T^c: ◯, Control Group; △, Treatment Group

In addition, it can be shown that the log cumulative hazard function of the two groups is,

$$\text{Group 1: } \gamma \log(\lambda) + \gamma \log(t)$$
$$\text{Group 2: } \gamma \log(\lambda) + \frac{\gamma \log(t)}{c}$$

which also confirms that they are not parallel.

3.2 Lehmann Alternative and Early-Late Treatment Differences

The Log-rank test has more power than the Wilcoxon test on situations where there is later separation.

Lemma 3.1 (Late Treatment Differences) *Under the Lehmann alternative (1.21), the maximum difference between the survival functions $S_1(t)$ and $S_2(t)$ will occur at T such that $S_1(t) < 0.4$.*

Proof. Let,

$$\begin{aligned} S_2(t) - S_1(t) &= [S_1(t)]^{\psi} - S_1(t) \\ &= p^{\psi} - p \\ f(p) &= p^{\psi} - p \end{aligned}$$

where p $= S_1(t)$

The first and second derivative of the function are,

$$\frac{d}{dp}\left[p^{\psi} - p\right] = \psi p^{\psi-1} - 1 \tag{3.1}$$

$$\frac{d^2}{dp^2}\left[p^{\psi} - p\right] = \psi(\psi - 1)p^{\psi-2} \tag{3.2}$$

The second derivative will always be negative since $\psi p^{\psi-2}$ is positive and $(\psi - 1)$ is negative for $0 < \psi < 1$. This implies that the function is concave down and from the first derivative the maximum value of p was computed to be

$$S_1(t) = p = \left[\frac{1}{\psi}\right]^{1/(\psi-1)} \tag{3.3}$$

Using Equation (3.3), for $0 < \psi < 1$, the value of $S_1(t)$ that will give the maximum difference for $[S_1(t)]^\psi - S_1(t)$ are values less than 0.4. The values are tabulated in Table 3.1 and plotted in Figure 3.5.

Table 3.1: Different values of ψ and the corresponding $S_1(t)$

ψ	**Maximum difference** $[S_1(t)]^\psi - S_1(t)$ achieved at $S_1(t)$ equal to
0.05	0.0427
0.10	0.0774
0.15	0.1073
0.20	0.1337
0.25	0.1575
0.30	0.1791
0.35	0.1989
0.40	0.2172
0.45	0.2341
0.50	0.2500
0.55	0.2649
0.60	0.2789
0.65	0.2921
0.70	0.3046
0.75	0.3164
0.80	0.3277
0.85	0.3384
0.90	0.3487
0.95	0.3585
0.99	0.3660

Observe that for all ψ between 0 and 1, the maximum difference $[S_1(t)]^\psi - S_1(t)$ is achieved at later event times, i.e. late enough so that $S_1(t) < 0.40$. For example, if $\psi = 0.5$, then $S_2(t) - S_1(t) = [S_1(t)]^\psi - S_1(t)$ is largest at $S_1(t) = \left(\frac{1}{0.5}\right)^{1/(0.5-1)} = 0.25$. The maximum of course depends on the value of ψ.

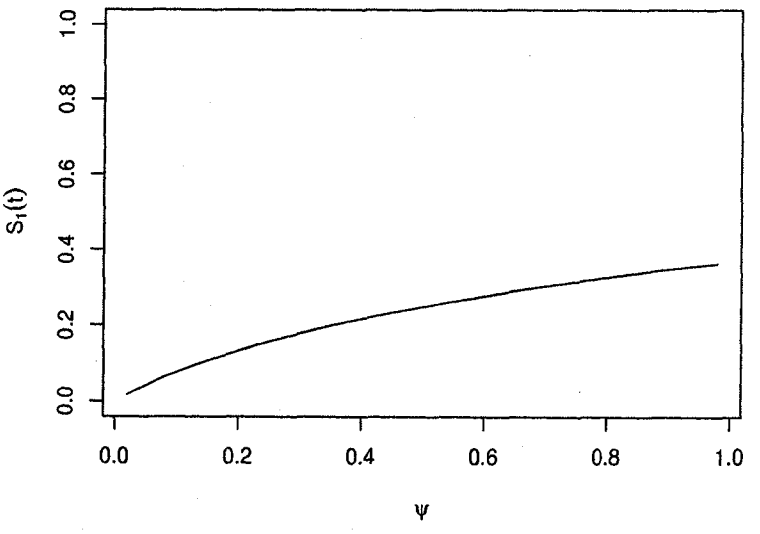

Figure 3.5: Plot of ψ versus $S_1(t) = \left[\frac{1}{\psi}\right]^{1/(\psi-1)}$

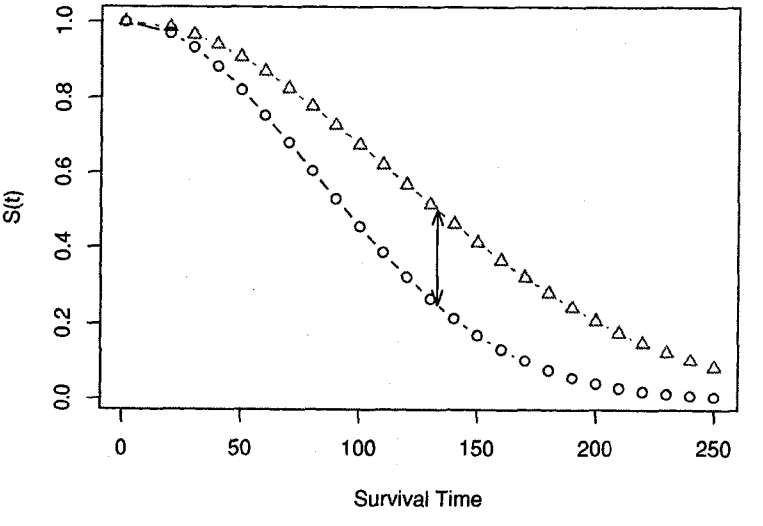

Figure 3.6: Plot of $S_1(t)$ and $S_2(t) = [S_1(t)]^{0.5}$. Maximum separation occurs at $S_1(t) = 0.25$: ○, $S_1(t)$; △, $S_2(t)$

3.3 Treatment Effects on the Survival Function

In this section we will show that simulation Cases 1, 2, 6 and 8 in Chapter 2 where Log-rank test beat Wilcoxon test correspond to late separation between survival curves.

In the following three subsections corresponding to the three survival time distribution being studied we will see how the survival function behaves on the three different treatment scenarios described on each graph.

3.3.1 Weibull Distribution

Figure 3.7 give plots of $S_1(t)$ where $T_1 \sim$Weibull$(\gamma = 2, \lambda = 0.008862)$ against the following $S_2(t)$:

(a) $T_2 = cT_1$ (Case 1 and Lehmann alternative)

(b) $T_2 \sim$ Weibull$(\gamma = 3, \lambda = 0.0069)$ (Case 4)

(c) $T_2 = T_1^c$ (Case 2)

In Figure 3.7 we can clearly see that survival function for Case 1 have wider difference from the control group on the lower half of the graph. The survival function for Case 4 clearly shows the wide difference from the survival function of the control group on the upper half of the two curves.

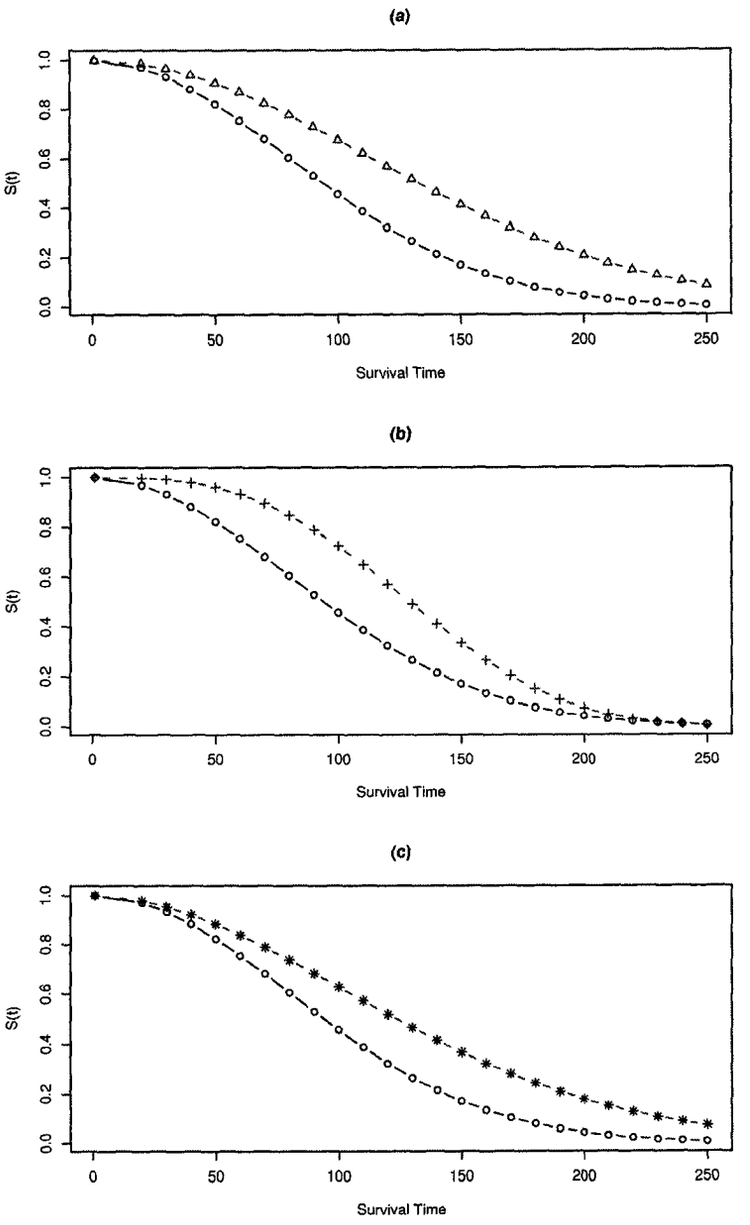

Figure 3.7: ○, Control group with survival times following a Weibull distribution with parameters $\gamma = 2$, $\lambda = 0.008862$; (a) Case 1: △, Lehmann alternatives with $\psi = 1/2$, (b) Case 4: +, Treatment group with parameters $\gamma = 3$, $\lambda = 0.006869$, (c) Case 2: $*$, Treatment effect T^c

3.3.2 Lognormal Distribution

Figure 3.8 give plots of $S_1(t)$ where T_1 ~Lognormal($\mu = 4.1052, \sigma = 1$) against the following $S_2(t)$:

(a) $S_2(t) = [S_1(t)]^{0.5}$ (Lehmann alternative)

(b) $T_2 = cT_1$ (Case 5)

(c) $T_2 = T_1^c$ (Case 6)

In Figure 3.8 we can clearly see that survival function for the Lehmann alternatives with ψ=1/2 (a) have wider difference from the control group on the lower half of the graph. On the other hand, it is a little hard to tell from the graphs of the survival functions of the control group and treatment group for Case 5 and Case 6 that (b) shows a wider difference on the upper half and (c) shows a wider difference on the lower half of the graph.

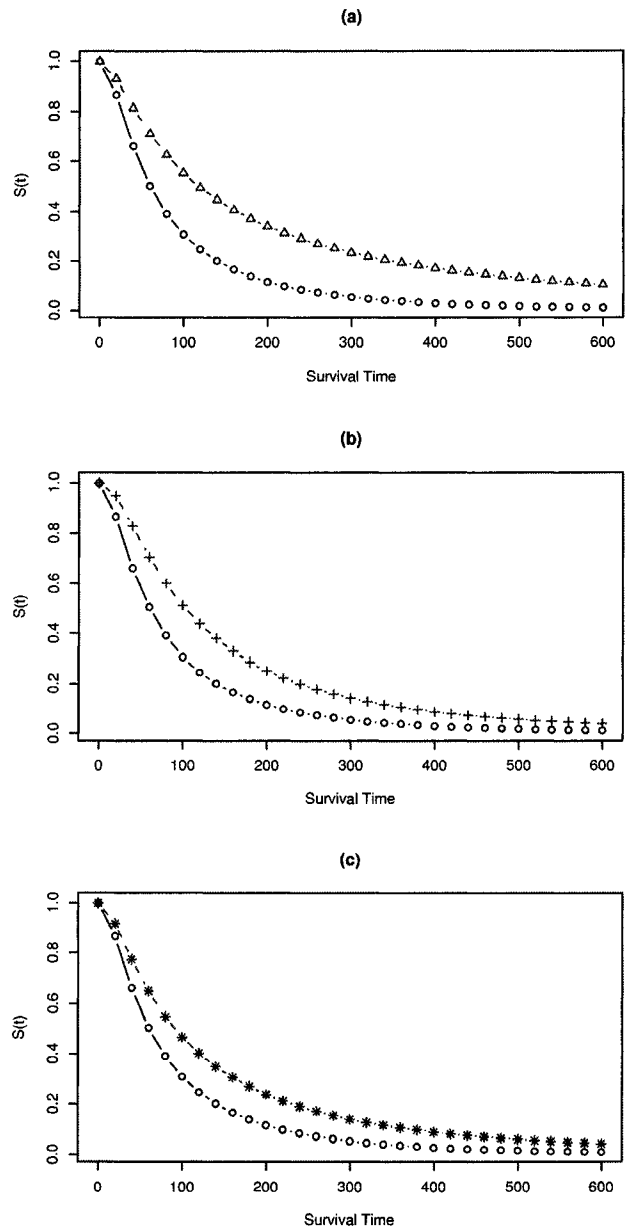

Figure 3.8: ◯, Control group with survival times following a Lognormal distribution with parameters $\mu = 4.1052$, $\sigma = 1$; (a) △, Lehmann alternatives with $\psi = 1/2$, (b) Case 5: +, Treatment effect cT, (c) Case 6: $*$, Treatment effect T^c

3.3.3 Log-logistic Distribution

Figure 3.9 give plots of $S_1(t)$ where T_1 ~Log-logistic($\mu = 4.1536, \sigma = 0.5$) against the following $S_2(t)$:

(a) $S_2(t) = [S_1(t)]^{0.5}$ (Lehmann alternative)

(b) $T_2 = cT_1$ (Case 7)

(c) $T_2 = T_1^c$ (Case 8)

In Figure 3.9 we can clearly see that survival function for the Lehmann alternatives with ψ=1/2 (a) have wider difference from the control group on the lower half of the graphs. It seems like the graphs (b) and (c) corresponding to Cases 5 and 6 looks the same.

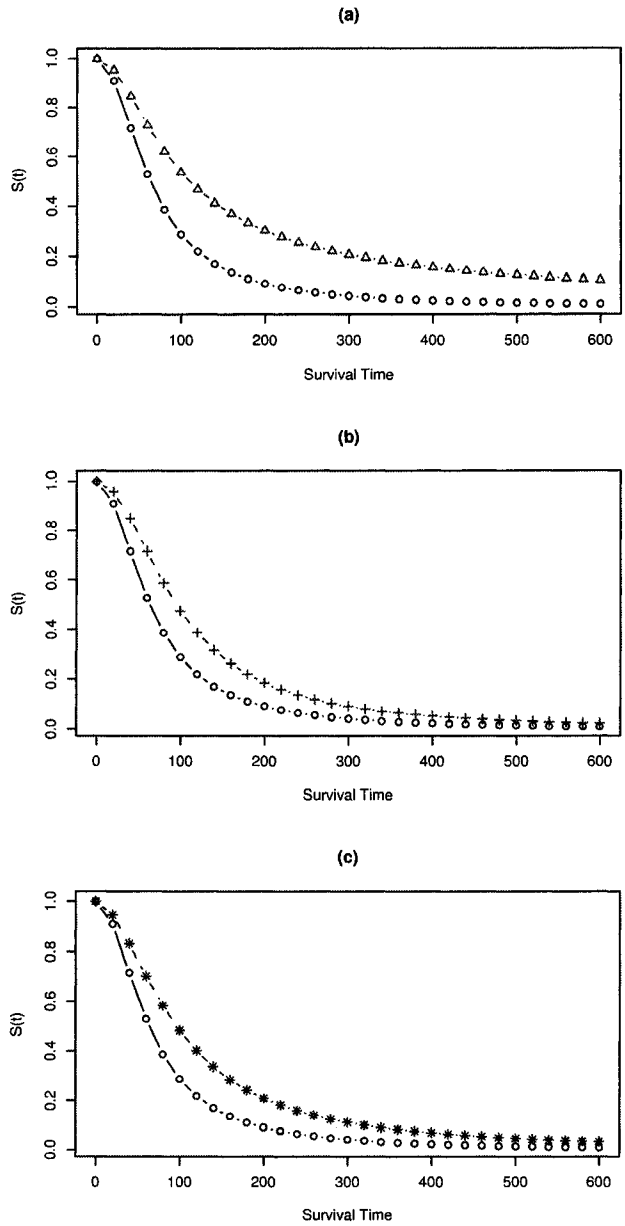

Figure 3.9: ◯, Control group with survival times following a Log-logistic distribution with parameters $\mu = 4.1536$, $\sigma = 0.5$; (a) $\triangle$, Lehmann alternatives with $\psi = 1/2$, (b) Case 7: $+$, Treatment effect cT, (c) Case 8: $*$, Treatment effect T^c

Chapter 4

Proposed Diagnostic for Early Versus Late Treatment Differences

Our simulation results show that power comparison between Log-rank and Wilcoxon tests does not depend entirely on proportional hazards assumption but more on early and later treatment effects. Motivated by this idea we propose a way to quantify the separation between the two survival functions.

4.1 The M Test

We propose as diagnostic the comparison of the lengths of the two vertical lines in Figure 4.1, as follows. If separation between the two curves occur early, then the vertical line on the left is expected to be longer than that one on the right. If separation is late, then the one on the left should be shorter. This is quantified in the following statistic M.

$$\mathrm{M}=\overbrace{[\widetilde{S}_2(t_{0.5,\,1}) - \underbrace{\widetilde{S}_1(t_{0.5,\,1})}_{0.5}]}^{early\ deaths} - \overbrace{[\underbrace{\widetilde{S}_2(t_{0.5,\,2})}_{0.5} - \widetilde{S}_1(t_{0.5,\,2})]}^{late\ deaths} \quad (4.0)$$

where

$t_{0.5,\ 1}$ is the time in Group 1 with $\widetilde{S}_1(t) = 0.5$

$t_{0.5,\ 2}$ is the time in Group 2 with $\widetilde{S}_2(t) = 0.5$

$\widetilde{S}_2(t_{0.5,\ 1})$ is the survival estimate of $t_{0.5,\ 1}$ in Group 2

$\widetilde{S}_1(t_{0.5,\ 2})$ is the survival estimate of $t_{0.5,\ 2}$ in Group 1

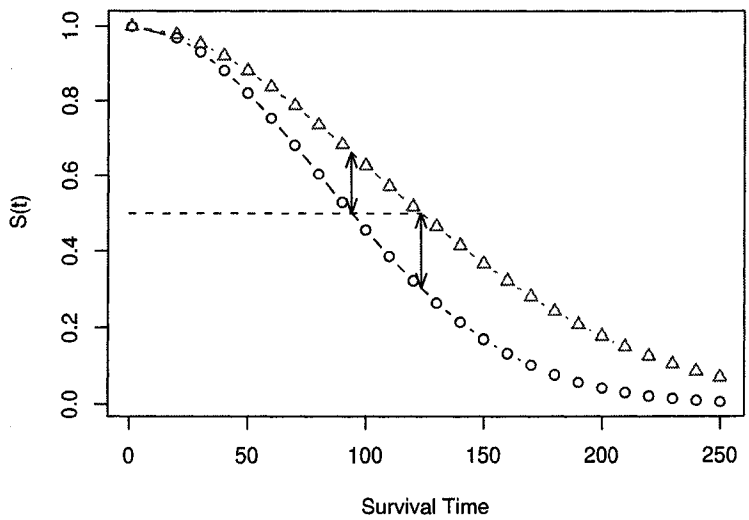

Figure 4.1: M test on survival functions ○, Control Group; △, Treatment Group

If separation is early, then we expect $M \geq 0$. If separation is late we expect $M < 0$. In Chapter 5, we propose the following pretest procedure:

1. If $M < 0$, then use Log-rank test

2. If $M \geq 0$, then use Wilcoxon test

Now, we will show that the Lehmann alternative, and hence proportional hazards assumption, implies that $M < 0$. This result is in Theorem 4.2. The proof will need the following Lemma.

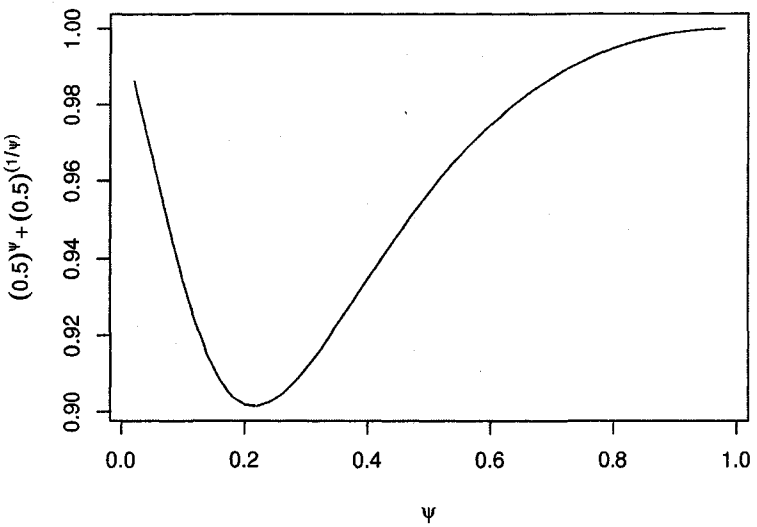

Figure 4.2: Plot of ψ versus $(0.5)^{\psi} + (0.5)^{1/\psi}$

Lemma 4.1 *Let* $f(\psi) = (0.5)^{\psi} + (0.5)^{1/\psi}$. *Then*

1. $f(1) = 1$ *and* $f'(1) = 0$
2. $f(\psi) = f(1/\psi)$
3. $\lim_{\psi \to 0^+} f(\psi) = \lim_{\psi \to \infty} f(\psi) = 1$
4. $f''(1)$ *is negative*
5. f' *has exactly three zeroes on* $(0,\infty)$;
6. f *achieves its maximum value at* $\psi = 1$

Proof.

(a) $f(1) = 0.5 + 0.5 = 1$

$$f'(\psi) = \ln(0.5)\left((0.5)^{\psi} - \frac{1}{\psi^2}(0.5)^{1/\psi}\right)$$

$$f'(1) = \ln(0.5)(0.5 - 0.5) = 0$$

(b) $f(\psi) = f\left(\frac{1}{\psi}\right)$

Let $y = \frac{1}{\psi} \Longrightarrow \psi = \frac{1}{y}$, $y > 0$

$$f(y) = (0.5)^{y} + (0.5)^{1/y}$$

(c) $\lim_{\psi \to 0^+} (0.5)^{\psi} + (0.5)^{1/\psi} = 1 + 0 = 1$

$\lim_{\psi \to \infty} (0.5)^{\psi} + (0.5)^{1/\psi} = 0 + 1 = 1$

(d) Taking the second derivative

$$f''(\psi) = \ln(0.5)\left((0.5)^{\psi}\ln(0.5) + \frac{1}{\psi^4}(0.5)^{1/\psi}\ln(0.5) + \frac{2}{\psi^3}(0.5)^{1/\psi}\right)$$

and plugging in $\psi = 1$ we get

$$f''(1) = \ln(0.5)[0.5\ln(0.5) + 0.5\ln(0.5) + 2(0.5)] = \underbrace{-0.2127}_{negative}$$

(e) The following substitution were made to turn the unusual symmetry of (b) as shown in Figure 4.2 into a genuine symmetry around the *y-axis*:

$$a = \ln(0.5) \quad \text{and} \quad t = \ln(\psi)$$

Then we have $f(\psi) = g(t)$, where

$$g(t) = e^{ae^{t}} + e^{ae^{-t}}$$

$$g'(t) = e^{ae^{t}}\, ae^{t} - e^{ae^{-t}}\, ae^{-t}$$

The domain $(0, \infty)$ for f has been stretched to $(-\infty, \infty)$ for g. Now $g'(t) = 0$ precisely when $f'(\psi) = 0$, so we work on the equation $g'(t) = 0$. This

happens when

$$\begin{aligned} e^{ae^t}\, ae^t &= e^{ae^{-t}}\, ae^{-t} \\ ae^{ae^t+t} &= ae^{ae^{-t}-t} \\ \ln(e^{ae^t+t}) &= \ln(e^{ae^{-t}-t}) \\ ae^t + t &= ae^{-t} - t \\ a(e^t - e^{-t}) &= -2t \\ (-a)\sinh(t) &= t \end{aligned}$$

Using the definition below

$$\begin{aligned} \sinh(t) &= \frac{1}{2}(e^t - e^{-t}) \\ \cosh(t) &= \frac{1}{2}(e^t + e^{-t}) \end{aligned}$$

we take the first derivative of the left hand-side of the last equation above to get

$$\frac{d}{dt}(-a\ \sinh(t)) = -a\ \cosh(t) = \frac{1}{2}(e^t + e^{-t}).$$

When t = 0,

$$\begin{aligned} -a\ \cosh(0)) &= \frac{1}{2}(e^0 + e^0) \\ &= 1 \end{aligned}$$

Now $-a$ = - ln(0.5) is positive, from the previous information we can conclude that, $(-a)\ \sinh(t) = t$ has three solutions since $-\ln(0.5) < 1$.

It follows from (d) and (e) that f has three local maxima: at 0, ∞ and 1; the values at 0, ∞ and 1 are all equal to 1. Thus, f achieves a maximum value of 1 at $\psi = 1$.

Theorem 4.2 *If* $S_2(t) = [S_1(t)]^{\psi}$, *then* $M < 0$.

Proof. Let $S_2(t) = [S_1(t)]^{\psi}$, then

$$\begin{aligned} M &= [S_2(S_1^{-1}(0.5)) - 0.5] - [0.5 - S_1(S_2^{-1}(0.5))] \\ &= [[S_1(S_1^{-1}(0.5))]^{\psi} - 0.5] - [0.5 - [S_2(S_2^{-1}(0.5))]^{1/\psi}] \\ &= [(0.5)^{\psi} - 0.5] - [0.5 - (0.5)^{1/\psi}] \end{aligned}$$

Therefore, $M < 0$ if and only if

$$[(0.5)^{\psi} - 0.5] - [0.5 - (0.5)^{1/\psi}] < 0 \tag{4.1}$$

or equivalently ,

$$[(0.5)^{\psi} + (0.5)^{1/\psi}] < 1 \tag{4.2}$$

This is true as shown in Lemma 4.1.

Theorem 4.2 says that the Lehmann alternative implies $M < 0$. Since proportional hazards assumption implies the Lehmann alternative, the proportional hazards assumption implies $M < 0$. In Chapter 5, we propose using $M < 0$ as a pretest for using Log-rank test.

4.2 The Q Test

Since the values of $S_1(t)$ that will give the maximum difference between two survival functions under the Lehmann-alternative for $0 < \psi < 1$ are less than 0.4, we tried the cut-offs that capture 0.4 in the middle, for example 0.2 and 0.6. Measuring the vertical distance from the two points to their corresponding survival function to the treatment group is another way to quantify the separation between the two survival functions (Figure 4.3). Let,

$$Q = [\widetilde{S}_1(t_{0.6}) - \widetilde{S}_0(t_{0.6})] - [\widetilde{S}_1(t_{0.2}) - \widetilde{S}_0(t_{0.2})] \tag{4.3}$$

where

$t_{0.6,\ 1}$ is the time in Group 1 with $\widetilde{S}_1(t) = 0.6$

$t_{0.2,\ 1}$ is the time in Group 1 with $\widetilde{S}_1(t) = 0.2$

$\widetilde{S}_2(t_{0.6,\ 1})$ is the survival estimate of $t_{0.6,\ 1}$ in Group 2

$\widetilde{S}_2(t_{0.2,\ 1})$ is the survival estimate of $t_{0.2,\ 1}$ in Group 2

The Q test directs the user to use Log-rank when $Q < 0$, thus implying a late separation in the survival curves which also says differences between groups are occurring at later points in time. When $Q \geq 0$, thus implying early separation in the survival curves, the Q test directs the user to use the Wilcoxon test instead. Since the neither Log rank nor Wilcoxon are particularly good at detecting differences when survival curves cross we can only limit the use of the proposed test where the Log rank and Wilcoxon test are particularly good.

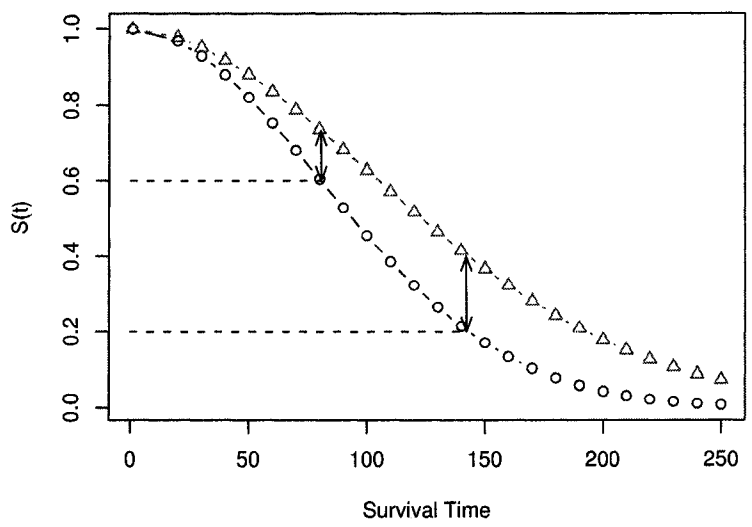

Figure 4.3: Q test on survival functions: ◯, Control Group; △, Treatment Group

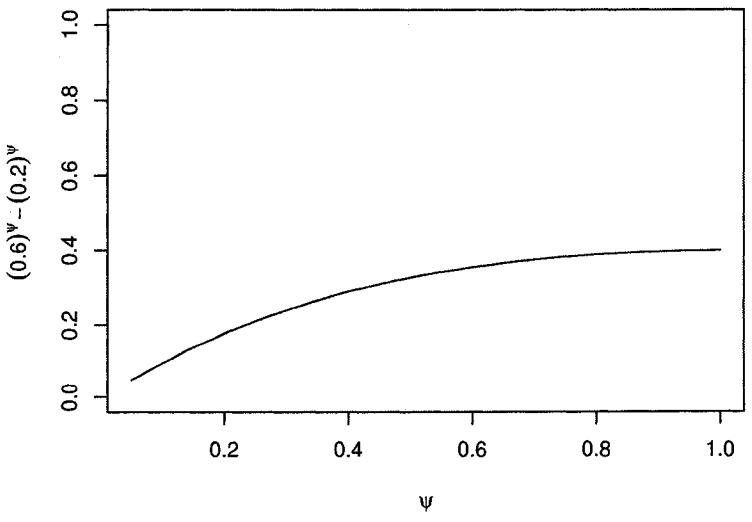

Figure 4.4: Plot of ψ versus $(0.6)^\psi - (0.2)^\psi$

Theorem 4.4 will show that Lehmann alternative also implies that $Q < 0$. The proof will need the following Lemma.

Lemma 4.3 *Let* $f(\psi) = (0.6)^{\psi} - (0.2)^{\psi}$, *for* $0 \leq \psi \leq 1$, $f(\psi) < 0.4$

Proof.

$$\begin{aligned} f(\psi) &= (0.6)^{\psi} - (0.2)^{\psi}, \quad 0 < \psi < 1 \\ f'(\psi) &= (0.6)^{\psi} \ln(0.6) - (0.2)^{\psi} \ln(0.2) \\ &= (0.2)^{\psi}[3^{\psi} \ln(0.6) - \ln(0.2)] \end{aligned}$$

Note: $(0.2)^{\psi} > 0$ for all ψ

$\Longrightarrow f'(\psi) > 0$ if $3^{\psi} \ln(0.6) - \ln(0.2) > 0$

$\Longleftrightarrow 3^{\psi} \ln(0.6) - \ln(0.2) > 0$

$\Longleftrightarrow 3^{\psi} < \dfrac{\ln(0.2)}{\ln(0.6)}$ since $\ln(0.6) < 0$

$\Longleftrightarrow 3^{\psi} < \log_3 (3.15)$ *Note* : $\log_3(3.15) > 1$

Therefore, $f'(\psi) > 0$ on $0 < \psi < 1$

f is increasing in $0 < \psi < 1$

It then follows that, $f(\psi) < f(1) = 0.4$

Theorem 4.4 *If* $S_2(t) = [S_1(t)]^{\psi}$, *then* $Q < 0$.

Proof. If $S_2(t) = [S_1(t)]^{\psi}$, then

$$\begin{aligned} Q &= [S_2(S_1^{-1}(0.6)) - 0.6] - [S_2(S_1^{-1}(0.2)) - 0.2] \\ &= [[S_1(S_1^{-1}(0.6))]^{\psi} - 0.6] - [[S_1(S_1^{-1}(0.2))]^{\psi} - 0.2] \\ &= [(0.6)^{\psi} - 0.6] - [(0.2)^{\psi} - 0.2] \end{aligned}$$

Therefore, $Q < 0$ if and only if

$$[(0.6)^{\psi} - 0.6] - [(0.2)^{\psi} - 0.2] < 0 \tag{4.4}$$

or equivalently ,

$$[(0.6)^{\psi} - (0.2)^{\psi}] < 0.4 \tag{4.5}$$

This was proved in Lemma 4.3.

Theorem 4.4 says that the Lehmann alternative implies $Q < 0$. Since proportional hazards assumption implies the Lehmann alternative, the proportional hazards assumption implies $Q < 0$. In Chapter 6, we propose using $Q < 0$ as a pretest for using Log-rank test.

Chapter 5

Pretest Based on M: Simulation

Here we investigate the power of an adaptive procedure that uses either Log-rank or Wilcoxon based on a pretest using M as follows:

1. If $M < 0$, then use Log-rank test
2. If $M \geq 0$, then use Wilcoxon test

The following sections show the power curves of the Log-rank test, Wilcoxon test and adaptive M test for the Weibull, Lognormal and Log-logistic distributions. In all the 8 cases, the adaptive M test approximates the power of the more efficient test.

5.1 Simulation Results Under Weibull Distribution (Cases 1, 2, 3 and 4)

Case 1

In this case Log-rank test is the optimal test. Figure 5.1 shows the M test is less powerful than Log-rank test, but more powerful than Wilcoxon test.

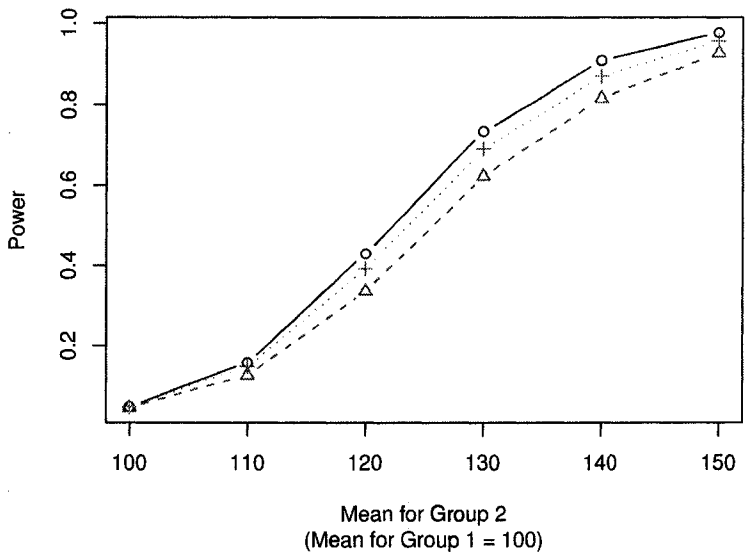

Figure 5.1: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with equal shape parameters ($\gamma = 2$) and treatment effect cT: ○, Log-rank; △, Wilcoxon; +, M Test

Case 2

Log-rank is the optimal test in this case even if the proportional hazards assumption is not satisfied. Again, as Figure 5.2 show, M test is less powerful than Log-rank test, but more powerful than Wilcoxon test.

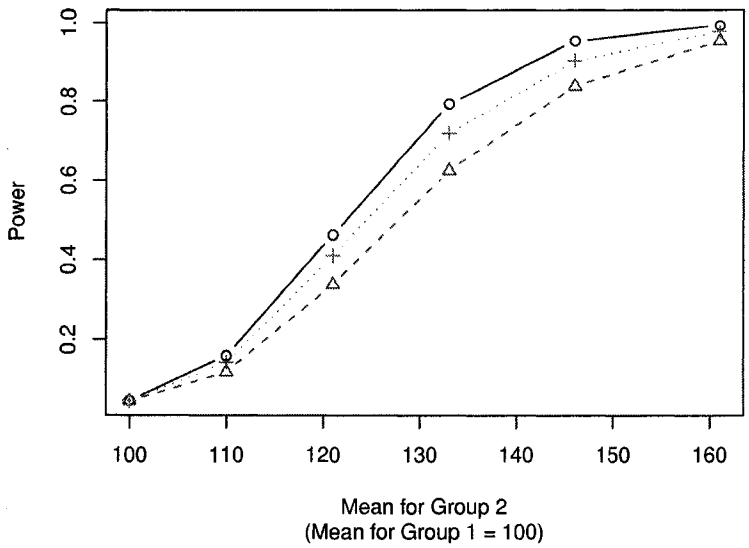

Figure 5.2: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with control group parameters (γ =2, λ=0.008862) and treatment effect T^c: ◯, Log-rank; △, Wilcoxon; +, M Test

Case 3

In this case the proportional hazards assumption is violated and Wilcoxon test is the more efficient test. The power after using the adaptive test is very close to the power of the Wilcoxon test.

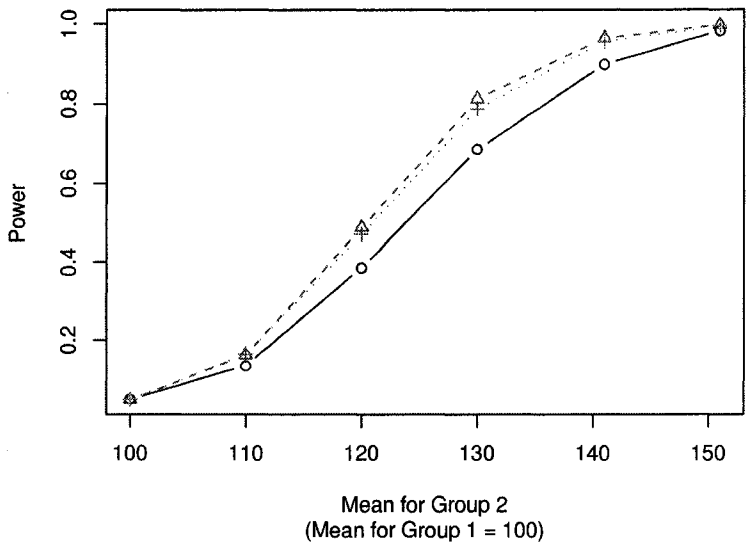

Figure 5.3: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with shape parameters $\gamma_1 = 2$ and shape is increased by c and scale is decreased by c: ◯, Log-rank; △, Wilcoxon; +, M Test

Case 4

This is another case that the proportional hazards assumption is violated and Wilcoxon test is the more efficient test. The power after using the adaptive test is not far behind the power of the Wilcoxon test.

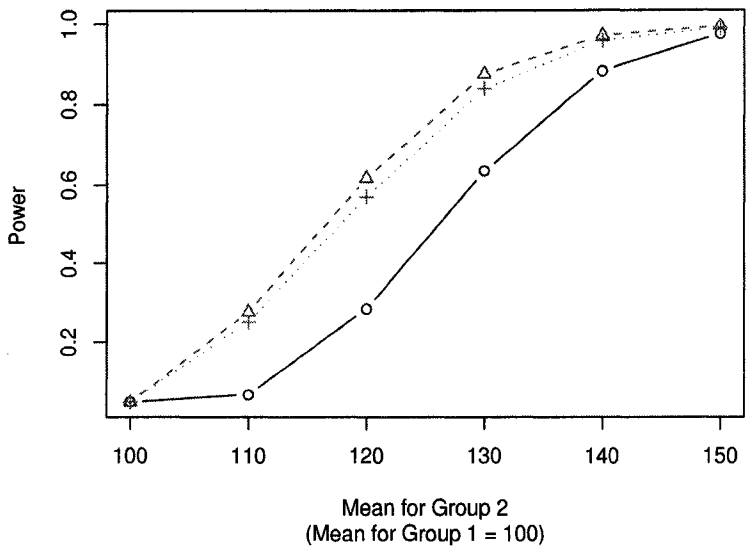

Figure 5.4: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with shape parameters γ_1=2 and γ_2=3 for Control group and Treatment group, respectively: ◯, Log-rank; △, Wilcoxon; +, M Test

5.2 Simulation Results Under Lognormal Distribution (Cases 5 and 6)

Case 5

The M test also performed well for the survival times generated from the Lognormal distribution. Again it closely follows the test with the higher power. Figure 5.5 shows that the power of using the M test as a pretest is very close to the power of Wilcoxon test.

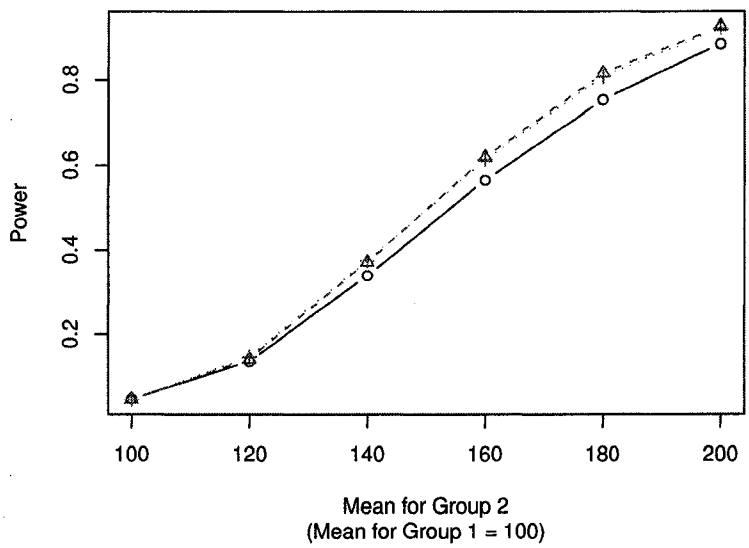

Figure 5.5: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Lognormal distribution with parameters μ=4.1052 and σ=1 for Group 1 and treatment effect cT for Group 2: ○, Log-rank; △, Wilcoxon; +, M Test

Case 6

On the otherhand for Case 6, Figure 5.6 shows that the M test follows the Logrank test.

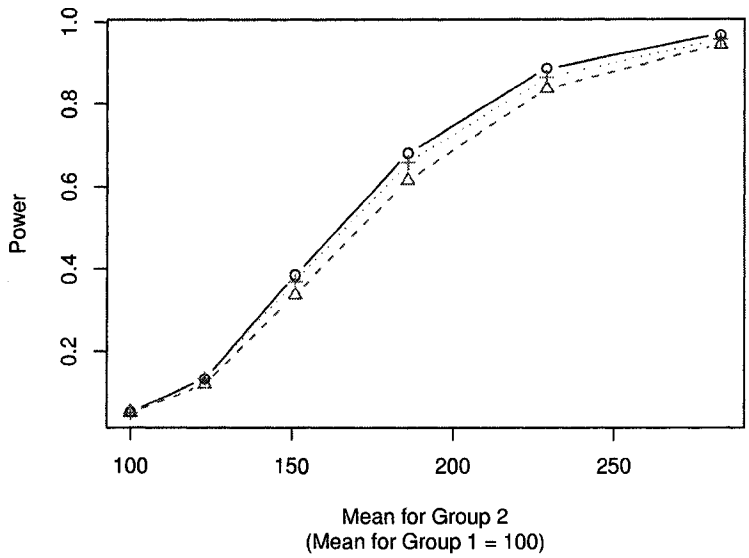

Figure 5.6: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Lognormal distribution with parameters μ=4.1052 and σ=1 for Group 1 and treatment effect T^c for Group 2: ◯, Log-rank; △, Wilcoxon; +, M Test

5.3 Simulation Results Under Log-logistic Distribution (Cases 7 and 8)

Case 7

In Case 7, Wilcoxon test has a higher power than the Log-rank test. The power after using the adaptive test is very close to the power of the Wilcoxon test (Figure 5.7).

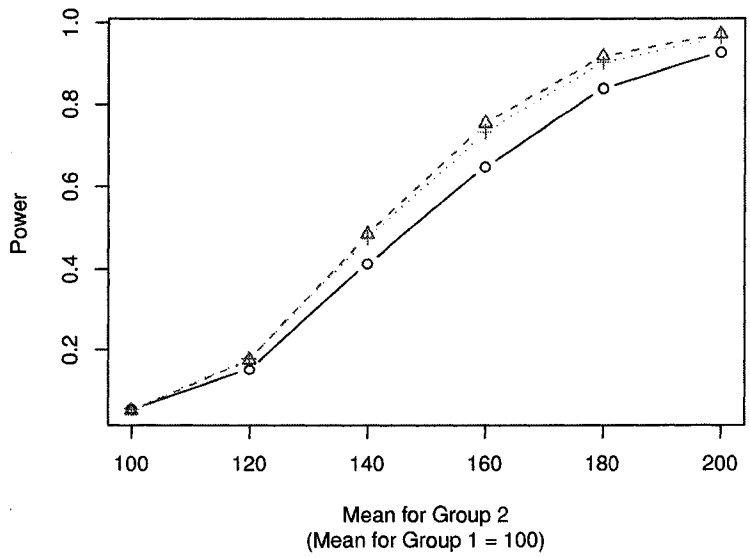

Figure 5.7: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Log-logistic distribution with parameters μ=4.1536 and σ=0.5 for Group 1 and treatment effect cT for Group 2: ○, Log-rank; △, Wilcoxon; +, M Test

Case 8

In Case 8, not much difference is seen between the power curves of Log-rank test and Wilcoxon test. So is the power after using the adaptive test (Figure 5.8).

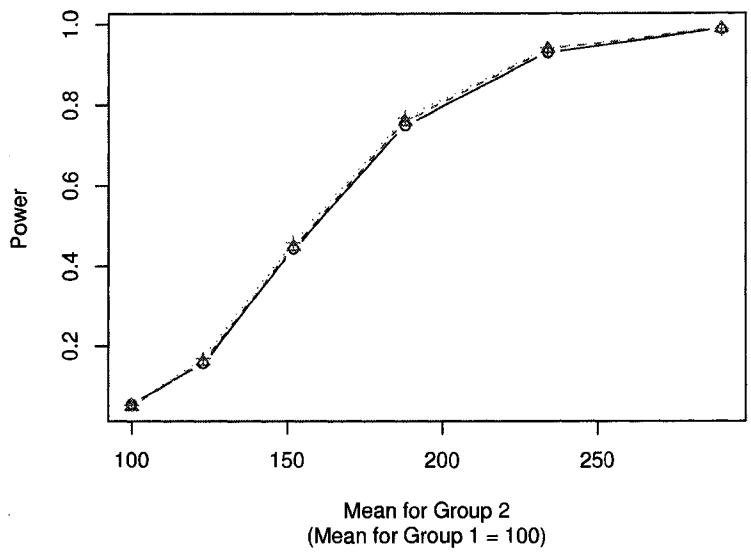

Figure 5.8: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Log-logistic distribution with parameters μ=4.1536 and σ=0.5 for Group 1 and treatment effect T^c for Group 2: ◯, Log-rank; △, Wilcoxon; +, M Test

Chapter 6

Pretest Based on Q: Simulation

Here we investigate the power of an adaptive procedure that uses either Log-rank or Wilcoxon, based on a pretest using Q as follows:

1. If $Q < 0$, then use Log-rank test

2. If $Q \geq 0$, then use Wilcoxon test

The following sections show the power curves of the Log-rank test, Wilcoxon test and adaptive Q test for the Weibull, Lognormal and Log-logistic distributions. In all the 8 cases, the adaptive Q test approximates the power of the more efficient test.

6.1 Simulation Results Under Weibull Distribution (Cases 1, 2, 3 and 4)

Case 1

In this case proportional hazards assumption is valid and Log-rank is the optimal test. Figure 6.1 shows the power after using Q as the pretest is very close to the power of Log-rank test.

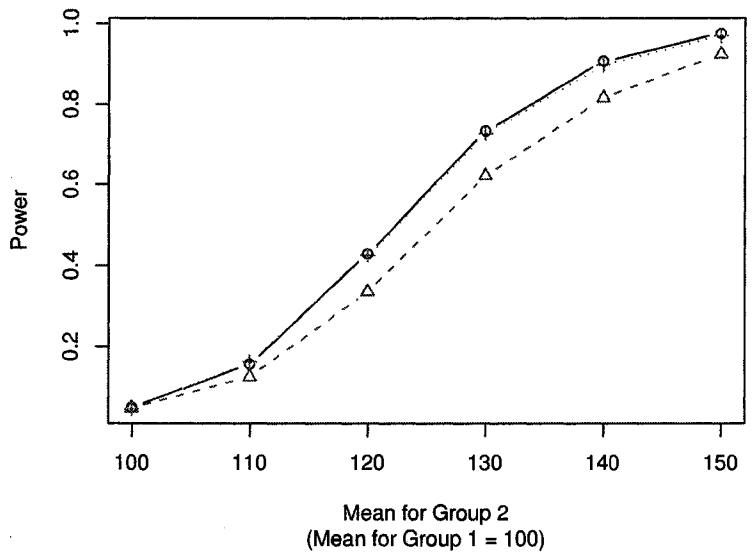

Figure 6.1: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with equal shape parameters ($\gamma = 1$) and treatment effect cT: ◯, Log-rank; △, Wilcoxon; +, Q Test

Case 2

In this case proportional hazards assumption is violated but Log-rank is the optimal test. Figure 6.2 shows the power after using Q as the pretest is very close to the power of Log-rank test.

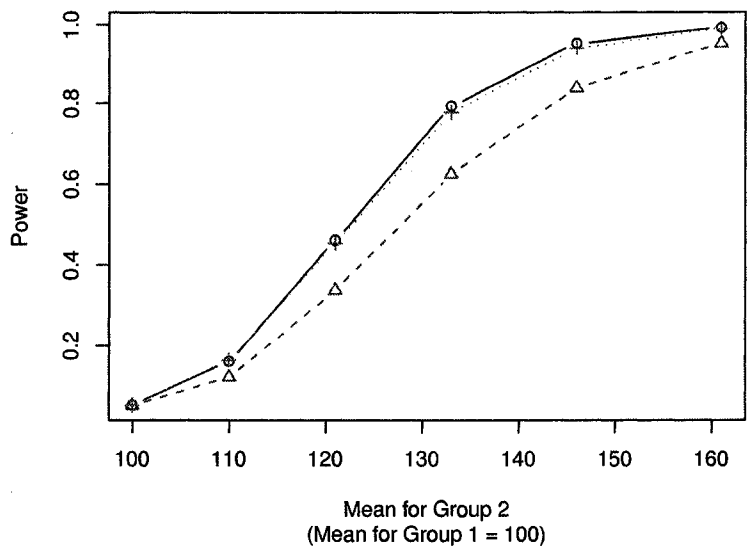

Figure 6.2: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with control group parameters (γ =2, λ=0.008862) and treatment effect T^c: ○, Log-rank; △, Wilcoxon; +, Q Test

Case 3

In this case proportional hazards assumption is violated and Wilcoxon is the optimal test. Figure 6.3 shows the power after using Q as the pre-test is very close to the power of Wilcoxon test.

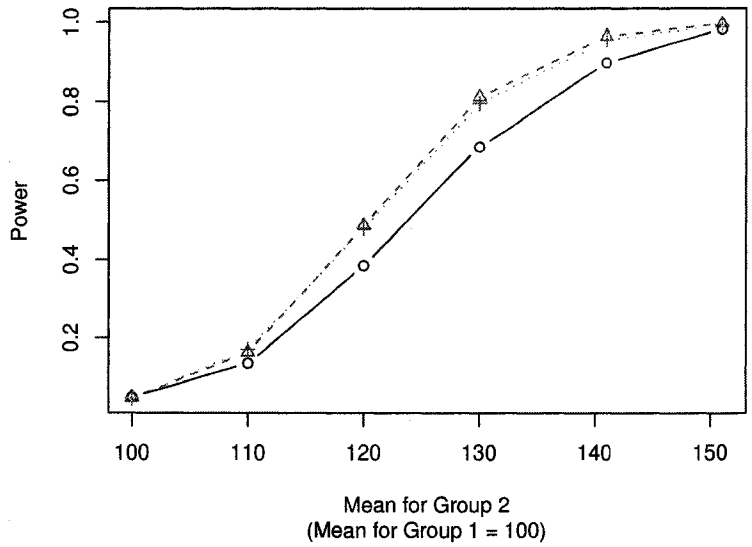

Figure 6.3: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with shape parameters $\gamma_1 = 2$ and shape is increased by c and scale is decreased by c: ◯, Log-rank; △, Wilcoxon; +, Q Test

Case 4

This is another case that the proportional hazards assumption is violated and Wilcoxon test is the more efficient test. The power after using the adaptive test is not far behind the power of the Wilcoxon test.

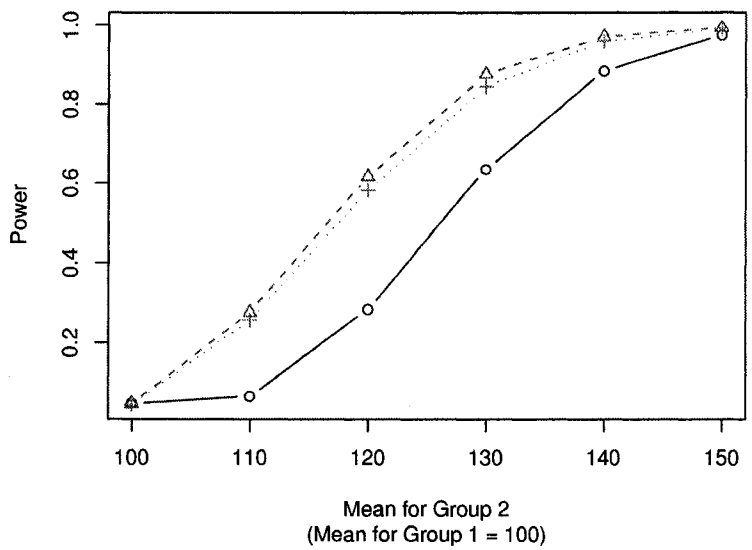

Figure 6.4: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with shape parameters γ_1=2 and γ_2=3 for Control group and Treatment group, respectively: ◯, Log-rank; △, Wilcoxon; +, Q Test

6.2 Simulation Results Under Lognormal Distribution (Cases 5 and 6)

Case 5

The Q test performed well too for the survival times generated from the Lognormal distribution. Again it closely follows the test with the higher power. In Figure 6.5 Q test have the higher power than the Log-rank test and is very close to the Wilcoxon test.

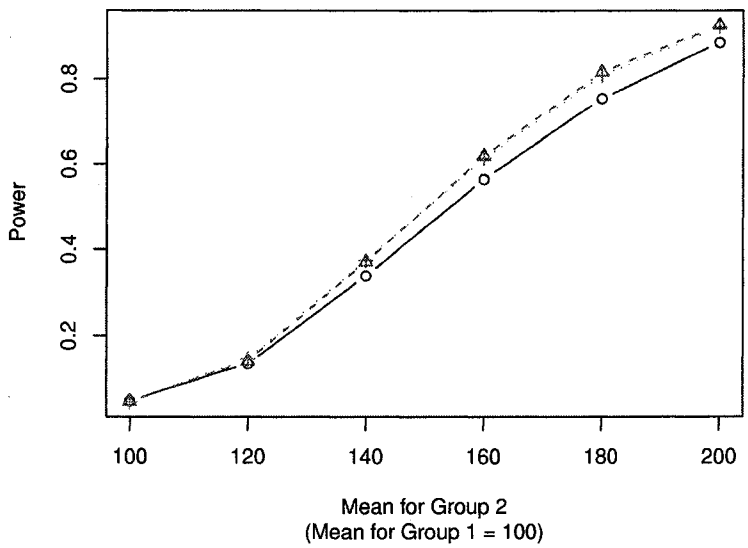

Figure 6.5: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Lognormal distribution with parameters μ=4.1052 and σ=1 for Group 1 and treatment effect cT for Group 2: ○, Log-rank; △, Wilcoxon; +, Q Test

Case 6

Figure 6.6 shows that the power using Q as the pre-test is very close to the power of Logrank test.

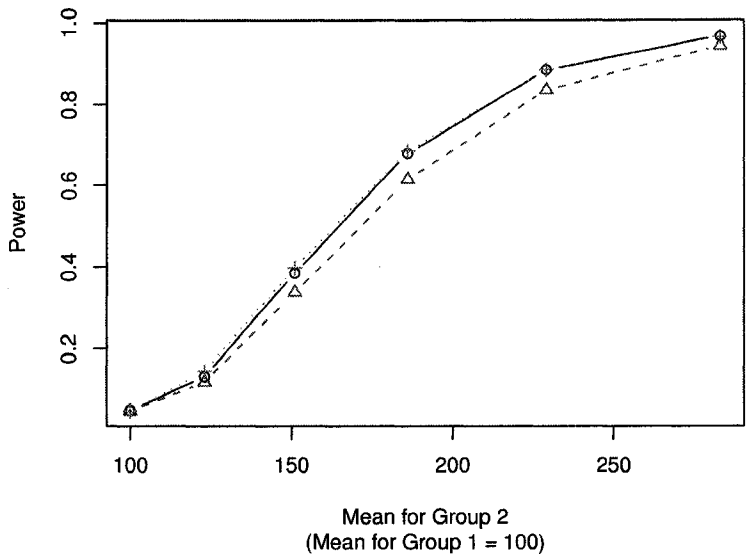

Figure 6.6: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Lognormal distribution with parameters μ=4.1052 and σ=1 for Group 1 and treatment effect T^c for Group 2: ◯, Log-rank; △, Wilcoxon; +, Q Test

6.3 Simulation Results Under Log-logistic Distribution (Cases 7 and 8)

Case 7

For the Log-logistic survival times, Wilcoxon test has a higher power than the Log-rank test for Case 7 which is closely reflected by using Q as the pre-test (Figure 6.7).

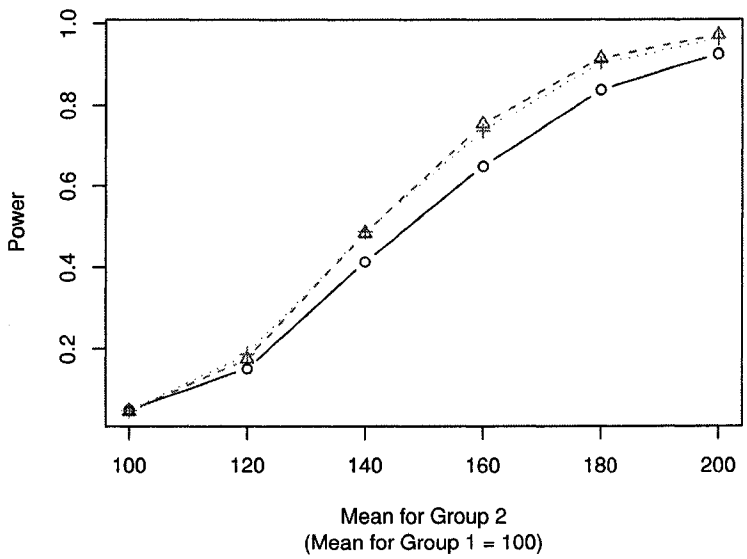

Figure 6.7: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Log-logistic distribution with parameters μ=4.1536 and σ=0.5 for Group 1 and treatment effect cT for Group 2: ○, Log-rank; △, Wilcoxon; +, Q Test

Case 8

Not much difference is seen on the power between Log-rank test and Wilcoxon test, but using Q as the pre-test gave the highest power (Figures 6.8) among the three.

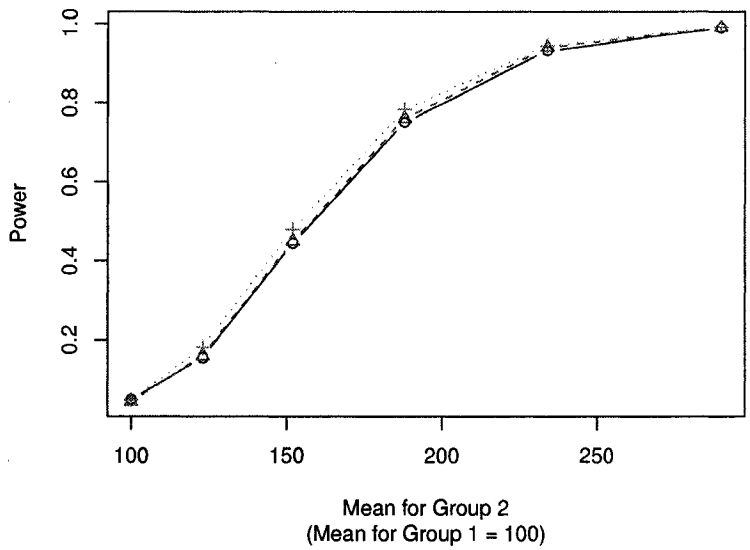

Figure 6.8: Power curves for the tests in samples of size $n_1 = n_2 = 50$ from a Log-logistic distribution with parameters μ=4.1536 and σ=0.5 for Group 1 and treatment effect T^c for Group 2: ◯, Log-rank; △, Wilcoxon; +, Q Test

Chapter 7

Conclusion and Recommendations

We have shown that standard diagnostics based on log-cumulative hazard plot to choose the more appropriate test between the Log-rank and Wilcoxon tests fail in some instances. An example is Case 2 where the Log-rank test has higher power than the Wilcoxon test even if the proportional hazards assumption is not satisfied. We have shown that relative performance of the two test depend not just on the proportional hazards assumption but also on the pattern of differences between the two survival curves. The crucial factor is whether the differences tend to occur early or late in time. This is evident in the structure of the test statistics themselves, with Wilcoxon giving more weight to earlier events and Log-rank to later events.

In this dissertation we propose diagnostics to measure early-or-late differences between two survival curves. The two adaptive tests were able to approximate the power of the more efficient test. Thus, it will help the user choose the more efficient test between Log-rank and Wilcoxon under various patterns of treatment differences.

Future studies are going to be devoted on, first extending the adaptive tests

to survival data with various censoring patterns. Secondly, we will allow for other types of covariates.

Appendix A

Power Simulation Tables for Weibull Distribution

Table A.1: Power for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with equal shape parameters ($\gamma = 2$) and treatment effect cT.

c	Mean of Group 2	Always Log rank	Always Wilcoxon	Power of M Test	Power of Q Test
1.00	100	0.0496	0.0481	0.0493	0.0458
1.10	110	0.1582	0.1266	0.1484	0.1638
1.20	120	0.4301	0.3364	0.3924	0.4258
1.30	130	0.7341	0.6224	0.6900	0.7275
1.40	140	0.9084	0.8168	0.8704	0.8987
1.50	150	0.9763	0.9260	0.9553	0.9708

Table A.2: Power for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with control group parameters (γ =2, λ=0.008862) and treatment effect T^c.

c	Mean of Group 2	Always Log rank	Always Wilcoxon	Power of M Test	Power of Q Test
1.00	100	0.0506	0.0494	0.0495	0.0498
1.02	110	0.1610	0.1207	0.1445	0.1634
1.04	121	0.4625	0.3379	0.4105	0.4550
1.06	133	0.7937	0.6250	0.7201	0.7778
1.08	146	0.9541	0.8397	0.9032	0.9406
1.10	161	0.9938	0.9542	0.9779	0.9910

Table A.3: Power for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with shape parameters $\gamma_1 = 2$ and shape is increased by c $(c\gamma)$ and scale is decreased by c, (λ/c).

c	Mean of Group 2	Always Log rank	Always Wilcoxon	Power of M Test	Power of Q Test
1.00	100	0.0496	0.0481	0.0493	0.0458
1.10	110	0.1354	0.1628	0.1633	0.1708
1.20	120	0.3854	0.4871	0.4698	0.4799
1.30	130	0.6870	0.8121	0.7871	0.7960
1.40	141	0.8979	0.9649	0.9564	0.9548
1.50	151	0.9834	0.9977	0.9953	0.9949

Table A.4: Power for the tests in samples of size $n_1 = n_2 = 50$ from a Weibull distribution with shape parameters γ_1=2 and γ_2=3 for Control group and Treatment group, respectively.

Mean of Group 2	Always Log rank	Always Wilcoxon	Power of M Test	Power of Q Test
100	0.0496	0.0481	0.0493	0.0458
110	0.0673	0.2771	0.2514	0.2576
120	0.2844	0.6179	0.5706	0.5819
130	0.6360	0.8761	0.8394	0.8450
140	0.8840	0.9727	0.9591	0.9601
150	0.9762	0.9952	0.9915	0.9926

Appendix B

Power Simulation Tables for Lognormal Distribution

Table B.1: Power for the tests in samples of size $n_1 = n_2 = 50$ from a Lognormal distribution with parameters μ=4.1052 and σ=1 for Group 1 and treatment effect cT for Group 2.

c	Mean of Group 2	Always Log rank	Always Wilcoxon	Power of M Test	Power of Q Test
1.00	100	0.0496	0.0482	0.0475	0.0478
1.20	120	0.1374	0.1431	0.1477	0.1597
1.40	140	0.3392	0.3718	0.3751	0.3909
1.60	160	0.5655	0.6199	0.6142	0.6312
1.80	180	0.7546	0.8165	0.8093	0.8168
2.00	200	0.8843	0.9250	0.9203	0.9237

Table B.2: Power for the tests in samples of size $n_1 = n_2 = 50$ from a Lognormal distribution with parameters μ=4.1052 and σ=1 for Group 1 and treatment effect T^c for Group 2.

c	Mean of Group 2	Always Log rank	Always Wilcoxon	Power of M Test	Power of Q Test
1.00	100	0.0513	0.0501	0.0494	0.0490
1.04	123	0.1328	0.1195	0.1334	0.1461
1.08	151	0.3855	0.3382	0.3709	0.3983
1.12	186	0.6804	0.6155	0.6579	0.6867
1.16	229	0.8870	0.8373	0.8648	0.8873
1.20	283	0.9689	0.9451	0.9575	0.9671

Appendix C

Power Simulation Tables for Log-logistic Distribution

Table C.1: Power for the tests in samples of size $n_1 = n_2 = 50$ from a Log-logistic distribution with parameters μ=4.1536 and σ=0.5 for Group 1 and treatment effect cT for Group 2.

c	Mean of Group 2	Always Log rank	Always Wilcoxon	Power of M Test	Power of Q Test
1.00	100	0.0535	0.0498	0.0524	0.0517
1.20	120	0.1528	0.1769	0.1790	0.1876
1.40	140	0.4140	0.4853	0.4751	0.4872
1.60	160	0.6495	0.7549	0.7334	0.7376
1.80	180	0.8384	0.9180	0.9024	0.9051
2.00	200	0.9275	0.9725	0.9633	0.9632

Table C.2: Power for the tests in samples of size $n_1 = n_2 = 50$ from a Log-logistic distribution with parameters μ=4.1536 and σ=0.5 for Group 1 and treatment effect T^c for Group 2.

c	Mean of Group 2	Always Log rank	Always Wilcoxon	Power of M Test	Power of Q Test
1.00	100	0.0533	0.0476	0.0505	0.0511
1.04	123	0.1571	0.1622	0.1699	0.1822
1.08	152	0.4438	0.4491	0.4603	0.4794
1.12	188	0.7504	0.7603	0.7692	0.7828
1.16	234	0.9293	0.9380	0.9397	0.9419
1.20	290	0.9875	0.9888	0.9889	0.9905

Bibliography

Allison, P.D., *Survival Analysis Using SAS: A Practical Guide,* (Cary, NC: SAS Institute Inc., 1995).

Bender, R., Augustin, T., and Blettner, M., "Generating survival times to simulate Cox proportional hazards model," *Statistics in Medicine,* **24**, (Feb., 2005): 1713-1723.

Breslow, N., "A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship," *Biometrika,* **57** (3), (1970): 579-594.

Collett, D., *Modelling Survival Data in Medical Research,* (Great Britain: Chapman & Hall, 1994).

Cox, D.R., "Regression Models and Life-tables (with discussion)," *Journal of the Royal Statistical Society,* **34** (2), (1972): 187-220.

Efron, B., "The Two Sample Problem With Censored Data, in" *Proceedings of the 5th Berkeley Symposium* **4**, (Berkeley: University of California Press, 1967): 831-853.

Fleming, T.R., O'Fallon, J.R., O'Briend, P.C., and Harrington, D.P., "Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right-censored data," *Biometrics,* **36**, (1980): 607-625.

Gehan, E.A., "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples," *Biometrika,* **52**(1/2), (Jun., 1965): 203-223.

Gehan, E.A., "A Generalized Two-Sample Wilcoxon Test for Doubly Censored Data," *Biometrika,* **52**(3/4), (Dec., 1965): 650-653.

Gehan, E.A. and Thomas, D.G., "The Peformance of Some Two-Sample Tests in Small Samples With and Without Censoring," *Biometrika,* **56** (1), (Mar., 1969): 127-132.

Gill, R. and Schumacher, M., "A Simple Test of the Proportional Hazards Assumptions," *Biometrika* **74** (2), (Jun., 1987):289-300.

Hess, K., "Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression," *Statistics in Medicine,* **14**, (1995): 1707-1723.

Kaplan, E.L. and Meier, P., "Nonparametric Estimation From Incomplete Observations," *Journal of American Statistical Association,* **53** (1958): 457-481.

Klein, J.P. and Moeschberger, M.L, *Survival Analysis: Techniques for Censored and Truncated Data,* (New York: Springer-Verlag, 2003).

Latta, R., "A Monte Carlo Study of Some Two-Sample Rank Tests with Censored Data," *Journal of American Statistical Association,* **76** (375), (Sep., 1981): 713-719.

Lee, E.T., Desu, M.M., and Gehan, E.A., "A Monte Carlo Study of the Power of Some Two-Sample Tests," *Biometrika,* **62** (2), (Aug., 1975): 425-432.

Lee, E.T. and Go, O.T., "Survival Analysis in Public Health Research," *Annual Review of Public Health,* **18**, (1997):105-134.

Leemis, L.M., *Reliability-Probabilistic Models and Statistical Methods,* (Englewood Cliffs, NJ: Prentice Hall, Inc, 1995)

Lehmann, E.L., "The power of rank tests," *Ann. Math. Statist.,* **24**, (1953):23-43

Mantel, N., "Evaluation of survival data and two new rank order statistics arising in its consideration," *Cancer Chemotherapy Rep.,* **50**, (1966):163-170.

Mantel, N., "Ranking procedures for arbitrarily restricted observation," *Biometrics,* **23**, (1967):65-78.

Mantel, N. and Haenszel, W., "Statistical aspects of the analysis of data from retrospective studies of disease," *J. Nat. Cancer Inst.,* **22**, (1959):719-748.

Ng'Andu N.H., "An Empirical Comparison of Statistical Tests for Assessing the Proportional Hazards Assumption of Cox's Model," *Statistics in Medicine,* **16**, (1997): 611-626.

Peto, R., "Rank tests of maximum power against Lehmann type alternatives," *Biometrika* **59**, (1972):472-474.

Peto, R. and Pike, M., "352. Note: Conservatism of the Approximation $\Sigma(O-E)^2/E$ in the Logrank Test for Survival Data or Tumor Incidence Data," *Biometrics,* **29** (3), (1973): 579-584.

Peto, R. and Peto, J., "Asymptotically Efficient Rank Invariant Test Procedures," *Journal of the Royal Statistical Society. Series A (General),* **135** (2), (1972):185-207.

Prentice, R.L. and Marek, P., "A qualitative discrepancy between censored data rank tests," *Biometrics* **35**, (1979): 861-867.

Stablein, D.M. and Koutrouvelis, I.A, "A two-sample test sensitive to crossing hazards in uncensored and singly censored data," *Biometrics* **41**, (1985):643-652.

Tarone, R.E. and Ware, J., "On Distribution-Free Tests for Equality of Survival Distributions," *Biometrika* **64** (1), (Apr., 1977): 156-160.