Choosing the Cut Point for a Restricted Mean in Survival Analysis, a Data Driven Method

Emily H. Sheldon
Virginia Commonwealth University

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

Part of the Biostatistics Commons

© The Author

Downloaded from
https://scholarscompass.vcu.edu/etd/3121

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.
Choosing the Cut Point for a Restricted Mean in Survival Analysis, a Data Driven Method

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

By

Emily Hocknell Sheldon
B.S., St. Lawrence University, 2006

Director
Leroy R. Thacker
Assistant Professor
Department of Biostatistics

Virginia Commonwealth University
Richmond, VA
April, 2013
Acknowledgements

“No man is an island, entire of itself,” John Donne, 1624. If it was not for the vast support network that is made up by my friends and family, this paper would not be available, it would not even exist.

My friends come from a variety of backgrounds and I have known some of them my entire life, and others for a few short months. No matter the duration of our friendship I am grateful to all of you for listening to my worries and reassuring me when I needed it most. So I thank you, in no particular order, Taya, Denise, Kate, Alicia, John, Gabby, Miles, Caroline, Adam, Stephanie, Rebecca, Katie, Amy, Marlowe, Cathy, Kristian, Mackenzie, Jennie, Dr. Thacker, and Dr. Gennings.

Mom and Dad, thank you for your unending love, words of wisdom, encouragement, and patience. You provided me with the skills I needed to survive on my own, and allowed me to keep my sanity in a barn just a few miles up the highway. This paper is as much a result of your work as it is mine.

I would also like to take this chance to recognize some of the wonderful teachers I have had over the years. A lot of people tell me that they don’t like math, and I can only surmise that this is due to a lack of proper instruction in the subject. I was fortunate to have wonderful teachers, not only in math but in all subjects, from a very early age. Maegan, thank you for recruiting me to be a math major at St. Lawrence University. Dr. Schuckers, thank you for
working with me my senior year on a project that led to my choice of graduate work. Dr. Bebo, thank you for assigning us reading every night and checking our notes the next day. Young Mr. Fronhofer, thank you for telling stories to keep us awake during long math lessons. Younger Mr. Fronhofer, thank you for assigning problem sets that required thinking beyond the classroom. Mrs. Hudy-Luzadis, thank you for being an example of a strong woman with excellent math skills. Mrs. Jantson, thank you for instilling in me a love of biology. Ms. Artz and Mr. Lundgren, thank you for keeping my creative side satiated.

“No man is an island, entire of itself,” and neither is a woman working on her doctoral degree.
# Table of Contents

List of Tables ........................................................................................................................................ vi

List of Figures ......................................................................................................................................... viii

Abstract ................................................................................................................................................ xi

Chapter 1  Introduction.......................................................................................................................... 1
  1.1 Introduction................................................................................................................................... 1
  1.2 Prospectus .................................................................................................................................... 4

Chapter 2  Overview of Survival Analysis............................................................................................ 6
  2.1 Time to Event Data, and the General Mathematical Model for Survival Analysis ............ 6
  2.2 Censoring and Truncation.............................................................................................................. 9
  2.3 Non Parametric Procedures.......................................................................................................... 11
  2.4 Hypothesis Testing Using the Log-Rank Test Statistic ............................................................... 17
  2.5 Cox-Proportional Hazards Model – A Semi-parametric Procedure ..................................... 21
  2.6 Parametric Proportional Hazards Model.................................................................................... 24
  2.7 Accelerated Failure Time Models (AFT) .................................................................................... 26

Chapter 3  The Restricted Mean ............................................................................................................ 29
  3.1 Basic Calculation of the Mean...................................................................................................... 29
  3.2 Advantages of Using the Mean.................................................................................................... 31
  3.3 Current uses of Mean survival time ............................................................................................ 32
  3.4 Restricted Mean ........................................................................................................................... 34
3.5 Methods Used to Calculate the Restricted Mean

3.5.1 Discounting Future Years

3.5.2 Mixture Cure Models

3.5.3 Pseudo-Values

Chapter 4 Several New Methods for Determining a Cut-Point for the Restricted Mean

4.1 A Mathematical Definition of the Problem at Hand

4.2 Methods Developed to Address the Problem

4.2.1 Maximum Distance Method

4.2.2 Renyi Type Test

4.2.3 Estimated Hazards Method

Chapter 5 Simulations and Results

5.1 Simulation Set-Up

5.2 Simulation Results

5.2.1 Simulation Results for a Cut-Point of 10

5.2.2 Simulations Results for a Cut-Point of 5

5.2.3 Simulation Results when the Cut-Point is 40

Chapter 6 Conclusion and Future Work

6.1 Conclusion

6.2 Future Work

Bibliography

Appendices SAS Programs for Computing Estimates

A.1 Simulating the data from a piecewise exponential
A.2 SAS® Program Calculating $\tau$ Using the Maximum Distance Method ..................................97

A.3 SAS® Program Calculating $\tau$ Using the Renyi Method......................................................100

A.4 SAS® Program Calculating $\tau$ Using the Hazard Method.....................................................103
List of Tables

Table 1.1 The mean, or area under the Kaplan Meier curve, for two treatment groups, calculated from zero up to the time given.................................................................4

Table 2.1 The number of individuals at risk, and the number of individuals who experience the event, by group, at an event time. ........................................................................19

Table 5.1 Mean, time, hazard, and restricted mean scenarios for the treatment group. ..............59

Table 5.2 Simulation Results for the Maximum Distance Method with no censoring τ=10 ........65

Table 5.3 Results from the Maximum Distance Method with 25% censoring τ=10 ..................66

Table 5.4 Results from the Renyi Method with no censoring τ=10 ........................................68

Table 5.5 Results from the Renyi Method with 25% censoring τ=10 ....................................69

Table 5.6 Results from the Hazard Method with no censoring τ=10 ......................................71

Table 5.7 Results from the Hazard Method with 25% censoring τ=10 ...................................71

Table 5.8 Results from the Maximum Distance Method with no censoring τ=5 ....................74

Table 5.9 Results from the Maximum Distance Method with 25% censoring τ=5 ..................74

Table 5.10 Results from the Renyi Method with no censoring τ=5 ...........................................77

Table 5.11 Results from the Renyi Method with 25% censoring τ=5 .......................................78

Table 5.12 Results from the Hazard Method with no censoring τ=5 ...........................................80

Table 5.13 Results from the Hazard Method with 25% censoring τ=5 .....................................81

Table 5.14 Results from the Maximum Distance Method with no censoring τ=40 .................83

Table 5.15 Results from the Maximum Distance Method with 25% censoring τ=40 .............83

Table 5.16 Results from the Renyi Method with no censoring τ=40 .......................................84

Table 5.17 Results from the Renyi Method with 25% censoring τ=40 .....................................84

Table 5.18 Results from the Hazard Method with no censoring τ=40 .....................................85
Table 5.19 Results from the Hazard Method with no censoring $\tau = 40$ ........................................85
List of Figures

Figure 1.1 The Kaplan Meier Survival Curves for Two Groups ............................................. 3

Figure 2.1 A plot of the Kaplan Meier Survival curves for two groups of women, the outcome of interest is pregnancy and they are grouped by level of their mothers’ exposure to PCB chemicals........................................................................................................................................ 12

Figure 4.1 Example of survival functions of the two groups with equal hazards after time 10, and all observations censored after time 20.................................................................................................................. 45

Figure 4.2 Example of cumulative hazard function of the two groups with equal hazards after time 10, and all observations censored after time 20............................................................................................................ 46

Figure 4.3 Example of hazard functions of the two groups with equal hazards after time 10, and all observations censored after time 20.................................................................................................................. 46

Figure 4.4 The survival curves for a simulated piecewise exponential model where \( h_1 t = 0.05, t < 10 \) and \( h_1 t = 0.20, t \geq 10 \), and \( h_2 t = 0.20 \) for all \( t \)................................................................. 49

Figure 4.5 A plot of the absolute value for the distance between the two estimated survival curves for the two groups. Note that the maximum occurs right around \( t=10 \). ................. 50

Figure 4.6 A plot of the absolute value of the \( Z(t) \) statistic, or the numerator of the Log-Rank statistic, versus time. The horizontal line indicated the maximum value of this statistic...... 52

Figure 4.7 A hockey stick regression model is fit and plotted over the \( Z(t) \) statistics. The vertical reference line is at the point where \( \delta \) was estimated using an NLIN Procedure....................... 54

Figure 4.8 Possible time values for \( \delta \) were fit from 5 to 25 by 0.1 and then NLIN models were fit for each value. The plot shows the SSE for the NLIN model that corresponds to the value for \( \delta \). Note that the smallest values occur around time 10................................................................. 56
Figure 4.9 The NLIN Procedure model fit of the estimated hazards from time 0 to $\delta$, and $\delta$ to 30. .......................................................... 57

Figure 5.1 The cumulative hazard function for two groups, generated from a piecewise
exponential distribution where $\lambda=0.10$ and $\lambda^*=0.05$, and $\tau_1=10$ .................. 60

Figure 5.2 The cumulative hazard function for two groups, generated from a piecewise
exponential distribution where $\lambda=0.10$ and $\lambda^*=0.05$, and $\tau_1=10$. Censoring has been built
into the data.................................................................................................................. 61

Figure 5.3 The cumulative hazard function for two groups, generated from a piecewise
exponential distribution where $\lambda=0.10$ and $\lambda^*=0.05$, and $\tau_1=10$. Censoring has been built
into the data, and all observations after time 20 have been censored, $\tau_2 = 20$. .............. 61

Figure 5.4 Looking at the cumulative hazard function when the events prior to time 10, $\tau_1$, have
been removed .................................................................................................................. 62

Figure 5.5 MSE versus Squared Bias for the Maximum Distance Method when $\tau=10$. ............. 67

Figure 5.6 MSE versus Squared Bias for the Maximum Distance Method when $\tau=10$ with 25%
censoring....................................................................................................................... 68

Figure 5.7 MSE versus Squared Bias for the Renyi Method when $\tau=10$ ............................... 70

Figure 5.8 MSE versus Squared Bias for the Renyi Method with 25% censoring when $\tau=10$ ..... 70

Figure 5.9 MSE versus Squared Bias for the Hazard Method when $\tau=10$ ............................ 73

Figure 5.10 MSE versus Squared Bias for the Maximum Distance Method when $\tau=5$, no
censoring....................................................................................................................... 76

Figure 5.11 MSE versus Squared Bias for the Maximum Distance Method when $\tau=5$, with 25%
censoring....................................................................................................................... 77

Figure 5.12 MSE versus Squared Bias for the Renyi Method when $\tau=5$, with no censoring...... 79
Figure 5.13 MSE versus Squared Bias for the Renyi Method when $\tau = 5$, with 25% censoring .. 80
Figure 5.14 MSE versus Squared Bias for the Hazard Method when $\tau = 5$ .......................... 82
Figure 5.15 MSE versus Squared Bias for the Hazard Method when $\tau = 40$, 25% censoring ...... 86
Choosing the Cut Point for a Restricted Mean in Survival Analysis, a Data Driven Method

By Emily H. Sheldon

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2013

Director:
Dr. Leroy Thacker
Assistant Professor
Department of Biostatistics

Survival Analysis generally uses the median survival time as a common summary statistic. While the median possesses the desirable characteristic of being unbiased, there are times when it is not the best statistic to describe the data at hand. Royston and Parmar (2011) provide an argument that the restricted mean survival time should be the summary statistic used when the proportional hazards assumption is in doubt. Work in Restricted Means dates back to 1949 when J.O. Irwin developed a calculation for the standard error of the restricted mean using Greenwood’s formula. Since then the development of the restricted mean has been thorough in the literature, but its use in practical analyses is still limited. One area that is not well developed in the literature is the choice of the time point to which the mean is restricted. The aim of this dissertation is to develop a data driven method that allows the user to find a cut-point to use to restrict the mean.
Three methods are developed. The first is a simple method that locates the time at which the maximum distance between two curves exists. The second is a method adapted from a Renyi-type test, typically used when proportional hazards assumptions are not met, where the Renyi statistics are plotted and piecewise regression model is fit. The join point of the two pieces is where the meant will be restricted. Third is a method that applies a nonlinear model fit to the hazard estimates at each event time, the model allows for the hazards between the two groups to be different up until a certain time, after which the groups hazards are the same. The time point where the two groups’ hazards become the same is the time to which the mean is restricted. The methods are evaluated using MSE and bias calculations, and bootstrap techniques to estimate the variance.
Chapter 1  Introduction

1.1 Introduction

In survival analysis the most common summary statistic is median survival time. The median is often preferred over mean survival time since it has the attractive quality of being unbiased. Often when censoring is present, the mean is underestimated, or if the distribution of survival time is highly positively skewed the mean will not be a good representation of the overall data. However, there are many situations in which the advantages of using the mean survival time outweigh the nuisance of a biased estimator. For example, when it comes to interpreting results, comparison of treatments is more clearly interpreted as length of time lost or gained using mean survival, than in the case of the hazard ratio or 5-year survival (Maetani, 2004). Mean survival can help individualize therapeutic decisions, which is currently a prominent topic in statistics literature; for example, younger patients may want a riskier surgery if the chance of long term survival is high.

A median is not the best estimator in cases when fifty percent of the sample has not yet experienced the event of interest. Projecting half-lives can result in over-estimation, in that real half-lives could be significantly shorter, and relies on the assumption that the slope of the survival curve over time follows a predictable function. This point was made in a paper by Meier-Kriesche, et al, which compared projected median kidney transplant survival from previous studies performed during 1988 to 1995 with what the actual median was found to be when they examined follow up data. The authors argue that using the difference in the area
under the K-M curve between the two groups for the amount of time studied would provide a
better, more reliable estimate for the treatment difference than a prediction of the median
differences, especially during early study times.

On the other side of the argument, when a study has had well over fifty percent of
patients experience the event of interest, using the median doesn’t utilize all of the information
available; it only uses half of the sample. For example, a study by Moore (2007), illustrates a
pancreatic cancer treatment where the medians between the two groups were identical, but prior
to the median one group performed significantly better. This treatment difference is picked up
by the difference in the areas under the K-M curves. So using the mean survival, (or area under
the curve) to describe a group, incorporates all of the data at hand, unlike the median.

To solve the problem of censoring, the restricted mean and its standard error were
developed in 1949 by J.O. Irwin. A restricted mean can be used where either the last observation
is treated as an event (we will call this LOT) or the investigator can assign an interval $[0, \tau]$ where $\tau$ is assumed to be the longest possible survival time for that study. Another version of the
restricted mean is to assume the last event time is the last observed time regardless of later
censored observations; we will call this LET. There is very little advice on how to choose $\tau$
given in the literature. With the choice of $\tau$ left to be arbitrary, using the restricted mean could
yield misleading results.

To illustrate how important the choice of $\tau$ could be, consider a study where the restricted
mean survival time of two groups is being compared to determine if the new treatment improves
survival over the standard treatment. Take a simple case of choosing $\tau$ to either be LET or LOT.
If both groups experience long right term censoring the area between the two curves is
continuing to grow so long as the study continues. Looking at the information from Figure 1.1 and Table 1.1, note that using LOT and LET does not lead to the same time point for both groups. If we wish to compare mean survival time between groups then a common time point should be used to restrict the mean. The differences calculated in Table 1 were simply the differences of the means given, a common cut-point wasn’t used. Also note that the difference in the area between the curves is much larger if the LOT option is used, this difference is inflated since there is no more treatment effect being seen when the survival curves of both groups have flattened.

![Product-Limit Survival Estimates](image)

**Figure 1.1** The Kaplan Meier Survival Curves for Two Groups
Table 1.1 The mean, or area under the Kaplan Meier curve, for two treatment groups, calculated from zero up to the time given.

<table>
<thead>
<tr>
<th>Trt Group</th>
<th>Choice of ( \tau )</th>
<th>Time</th>
<th>Restricted Mean</th>
<th>SE</th>
<th>Mean Difference</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LET</td>
<td>24.07</td>
<td>19.52</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>LET</td>
<td>17.16</td>
<td>11.58</td>
<td>0.88</td>
<td>6.906</td>
<td>1.22</td>
</tr>
<tr>
<td>1</td>
<td>LOT</td>
<td>60.63</td>
<td>38.95</td>
<td>3.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>LOT</td>
<td>56.09</td>
<td>26.92</td>
<td>3.79</td>
<td>12.03</td>
<td>5.17</td>
</tr>
</tbody>
</table>

We hypothesize that it is possible that the treatment effect ends earlier than the last event time. If the treatment difference only exists for a certain amount of time, then we do not wish to amplify the treatment difference by including area under the curves beyond that time. The two groups will appear to be different only because of what occurred at the beginning of the study. The goal of this research is to find a data driven method that allows us to find a common cut-point in order to calculate the restricted means and difference between the restricted means for the two groups. This dissertation outlines the background for this work and 3 methods that could be possible ways to find an appropriate cut-point based on the data.

1.2 Prospectus

Chapter 2 gives an overview of Survival Analysis including the relationships between the survivor function and hazard function, types of censoring, non-parametric procedures, the Log-Rank test, the Cox Proportional Hazards model, and the accelerated failure time model. Chapter 3 reviews existing work on the restricted mean survival statistic. Comparisons between the mean and median are made, and current work that applies to restricted means is summarized. Chapter 4 gives a mathematical definition for our proposed scenario and describes the three methods that
have been developed for this research. Chapter 5 contains the results of the simulations for the three proposed methods, including plots that compare MSE and bias of the estimators. Chapter 6 outlines possible extensions of this research and ideas for future work in restricted mean survival times.
Chapter 2  Overview of Survival Analysis

The purpose of this chapter is to provide the reader with an overview of the subject of survival analysis. It is not assumed that the reader has prior knowledge or experience with the topic. This framework will be used in Chapter 3 when means and medians are examined. And the mathematical underpinnings of survival analysis will be used to develop the methodology in Chapter 4.

2.1 Time to Event Data, and the General Mathematical Model for Survival Analysis

Survival Analysis, or Time to Event Analysis, deals with studies where the main outcome variable is the time until some event of interest occurs. Examples of events of interest include death, recurrence of cancer, time to pregnancy, and time until failure of a device. Since time only moves forward, survival data is often not symmetric and may be positively skewed. Another reason time to event data is unique and thus “standard” statistical practices cannot be used is the presence of censoring. An observation is censored when a patient is lost to follow up or they haven’t experienced the event of interest by the end of the study. The survival times of \( n \) individuals in the population of interest are assumed to be independent. The main goals of survival analysis are to assess the dependence of survival times on explanatory variables and to be able to estimate and model the underlying survival distribution.

Let \( T \) be a nonnegative random variable representing the event time of an individual from the population of interest. The random variable \( T \) can be discrete or continuous; we will be
focusing on the continuous case for our purposes. Three specifications of the probability
distribution of $T$ are particularly useful, the survivor function, the probability density function
(p.d.f.), and the hazard function. It is assumed that the random variable $T$ has a p.d.f. $f(t)$ and a
cumulative distribution function (CDF)

$$
F(t) = P(T < t) = \int_0^t f(u)du,
$$

(2.1.1)

which can be interpreted as the probability that the time to the event of interest is less than some
value $t$. The following properties apply to all CDFs: they are monotonic, non-decreasing, right
continuous, and the limit of the CDF as $t$ approaches infinity is equal to one. Note that by
properties of integrals the derivative of the CDF is equal to the p.d.f.

$$
f(t) = \frac{d}{dt} F(t).
$$

A survivor function is defined to be the probability that the survival time is greater than or equal
to $t$, which is the complement if the cumulative distribution function,

$$
S(t) = P(T \geq t) = 1 - F(t), \quad 0 < t < \infty.
$$

(2.1.2)

So the survivor function is equal to one minus the CDF for the random variable $T$. Now the
following relationship between the p.d.f. and the survivor function can be drawn,

$$
f(t) = \frac{d}{dt} F(t) = \frac{d}{dt} (1 - S(t)) = - \frac{d}{dt} S(t),
$$

(2.1.3)

the p.d.f. is equal to the negative derivative of the survivor function with respect to $t$.

So far we have defined functions when $T < t$, and $T \geq t$, now the hazard function gives
us the instantaneous rate of failure at $T = t$ given the individual has survived up to time $t$. The
hazard function (sometimes called the hazard rate, instantaneous death rate, intensity rate, or force of mortality) is defined as follows

\[
h(t) = \lim_{\delta t \to 0} \frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t}.
\]  
(2.1.4)

Then Equation (2.1.4) can be simplified to the ratio of the p.d.f. over the survivor function using the following definition of the derivative of the CDF which is equal to the p.d.f.’

\[
f(t) = \lim_{\delta t \to \infty} \frac{P(t \leq T < t + \delta t)}{\delta t} = -\frac{dS(t)}{dt}.
\]  
(2.1.5)

Resulting in

\[
h(t) = -\frac{d \log[S(t)]}{dt} = \frac{f(t)}{S(t)}.
\]  
(2.1.6)

From here we can derive \(H(t)\) the cumulative hazard function. First off, note that

\[
H(t) = \int_0^t h(u) du.
\]  
(2.1.7)

Then we can use Equation (2.1.6) to say

\[
H(t) = \int_0^t \left\{ -\frac{d \log[S(t)]}{dt} \right\} du = -\log S(t).
\]  
(2.1.8)

Also note that

\[
S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t h(u) du \right\}.
\]  
(2.1.9)

In this section, the relationships between the p.d.f., the survivor function, and the hazard function have been illustrated. These are important connections that will be called on throughout this work.
2.2 Censoring and Truncation

Censoring occurs when the investigator does not observe the event of interest for a particular subject, either because the subject left the study or did not experience the event during the observation period. For example, a doctor observes a patient who is participating in a cancer drug study to see when the cancer reoccurs, the patient dies of a heart attack in the middle of the study and doesn’t experience a recurrence of cancer, so this patient is a censored observation at the time of his death. In another scenario a patient could be cured from the cancer being studied and thus she will not experience the event of interest (recurrence of cancer), so she is a censored observation recorded at the end of the study. In both scenarios we do not want to delete these patients from the study, or record them as missing value, since they still provide information about survival times. Both patients survived the event of interest up to the time they were censored and this information can be incorporated into the survival model. These scenarios are both examples of right censoring (as opposed to left censoring or interval censoring) and we will be assuming from here on out that we only have right censored observations in our data sets.

Left censoring often occurs when the event of interest has already happened before a study begins. For example research regarding a link between smoking and heart attacks may follow patients who smoke to see if they experience myocardial infarction. However, if any patients have already had a heart attack before the beginning of the study, they would be left censored observations. Again, the observations should not be thrown out simply because the heart attack occurred before the study, this information can still be incorporated into a likelihood function that uses an indicator as to whether the observation is an event time or a censored time.
As the name implies, interval censoring occurs when a patient is known to have had an event within an interval of time, but the specific time of the event is not known.

While less common than censoring, truncation of data is another action that can be factored into a likelihood of survival time. Truncation of data occurs when the investigator or analyst sets an observational window \((Y_L, Y_R)\) in which time only those patients who experience an event will be observed. When \(Y_R = \infty\) the data is considered to be left truncated, and only individuals who experienced the event after truncation time \(Y_L\) are included in the study, or analysis. This differs from left censoring where some information about the patients is known and that information is included in the likelihood. For right truncation, \(Y_L = 0\) and we examine all event times less than or equal to \(Y_R\).

Now we discuss how to construct a likelihood function for censored and truncated data. First assume that event times and censoring times are independent. We know that a right censored observation gives information up to the censoring time (the event did not occur up to and somewhere beyond that time). A left censored observation gives us information that can be added to the cumulative distribution at the beginning of the study time. And truncated data leads to a conditional distribution. (page 74, yellow book) All of this can be put into symbols used to construct the likelihood function.
Exact lifetimes - \( f(x) \)

Right-censored observations - \( S(C_R) \)

Left-censored observations - \( 1 - S(C_L) \)

Interval-censored observations - \( [S(L) - S(R)] \)

Left-Truncated observations - \( f(x)/S(Y_L) \)

Right-Truncated Observations - \( f(x)/[1 - S(Y_R)] \).

Thus the likelihood can be expressed as

\[
L \propto \prod_{i \in D} f(x_i) \prod_{i \in R} S(C_i) \prod_{i \in L} (1 - S(C_i)) \prod_{i \in I} [S(L_i) - S(R_i)],
\]

(2.2.1)

where \( D \) is the set of death times, \( R \) is the set of right censored observations, \( L \) is the set of left censored observations, and \( I \) is the set of interval censored observations. For a given situation the likelihood can be altered to fit the type of data at hand.

No matter what type, censoring needs to be accounted for in an analysis. When there is a large amount of right censoring at the end of the study, or even if only the last observation is censored, this affects the calculation of mean survival since by definition the mean is undefined if the last observation is censored.

### 2.3 Non Parametric Procedures

As detailed in Section 2.1 the survivor function gives us the probability that an individual survives up to or beyond time \( t \). A basic estimate of this probability is the empirical survivor function

\[
\hat{S}(t) = \frac{\text{Number of individuals with survival times} \geq t}{\text{Number of individuals in the data set}}.
\]

(2.3.1)
The empirical survivor function $\hat{S}(t)$ is equal to one when all individuals in the data set are still alive and is equal to zero when the last individual in the data set has experienced the event. A plot of $\hat{S}(t)$ vs. $t$ is a step function, (see Figure 2.1).

![Product-Limit Survival Estimates](image)

**Figure 2.1** A plot of the Kaplan Meier Survival curves for two groups of women, the outcome of interest is pregnancy and they are grouped by level of their mothers’ exposure to PCB chemicals.

The empirical survivor function is not adequate when the data contains censored observations because it does not allow for the information from censored observations to be included. The Kaplan-Meier (K-M) estimate of the survivor function allows the use of censored observations and is often used as a preliminary step in survival analyses (Kaplan & Meier, 1958). This estimator is also known as the product limit estimator as it takes the product of the
probability of survival at each unique event time. Suppose there are \( n \) individuals with observed survival times \( t_1, t_2, \ldots, t_n \). Observations may be right censored and there may be individuals with the same observed event time. Thus suppose there are \( r \) distinct event times where \( r \leq n \).

Now let \( j = 1, 2, \ldots, r \) where \( t_{(1)} < t_{(2)} < \cdots < t_{(r)} \) are the \( r \) ordered event times. The number of individuals at risk, those who have not yet experienced the event of interest, just before time \( t_{(j)} \) will be denoted \( n_j \), and the number of individuals who fail at time \( t_{(j)} \) will be represented by \( d_j \). The term \( (n_j - d_j)/n_j \) can be thought of as the an estimate of the probability of surviving the interval from \( t_{(j)} \) to \( t_{(j+1)} \). We will assume all events in the sample occur independently from one another.

The K-M estimate is expressed as follows,

\[
\hat{S}(t) = \prod_{j=1}^{k} \left( \frac{n_j - d_j}{n_j} \right),
\]

for \( t_{(k)} \leq t < t_{(k+1)} \), where \( k = 1, 2, \ldots, r \) ordered survival times, and \( \hat{S}(t) = 1 \) for \( t < t_{(1)} \), and \( t_{(r+1)} \) is defined as \( \infty \). If the last observed survival time \( t_{(r)} \) is an uncensored observation then \( \hat{S}(t) = 0 \) for \( t \geq t_{(r)} \). If the largest observed survival time, \( t^* \), is censored then \( \hat{S}(t) \) is undefined for \( t > t^* \). A plot of the K-M estimate of the survivor function is a step function much like the empirical survival function, see Figure 1. Censored observations are incorporated into the K-M estimate since they contribute to the risk set \( n_j \) until the observation is censored. If an observation is censored at the same time as an event time, the censored observation is included in \( n_j \) and then removed from the risk set at the next event time; that is, the censored survival time is assumed to have occurred immediately after the observed event time.
Now that the K-M estimate has been defined, we can derive the standard error (se) of the non-parametric estimate. In order to derive the standard error of the K-M estimate first define
\[ \hat{p}_j = \frac{n_j - d_j}{n_j} \] so that
\[ \hat{S}(t) = \prod_{j=1}^{k} \hat{p}_j. \]

Now take the log of both sides so that we may invoke the useful property that the log of the products equals the sum of the logs
\[ \log\left(\hat{S}(t)\right) = \sum_{j=1}^{k} \log(\hat{p}_j), \]
so the variance of \(\log(\hat{S}(t))\) is given by
\[ \text{Var}\{\log(\hat{S}(t))\} = \sum_{j=1}^{k} \text{Var}\{\log(\hat{p}_j)\}. \]

If we assume individuals surviving through the interval beginning at \(t_j\) have a binomial\( (n_j, p_j) \) distribution where \(p_j\) is the true probability of survival through that interval, and \(n_j - d_j\) is the observed number of individuals who survived. Based on properties of the binomial distribution we know
\[ \text{Var}\left(n_j - d_j\right) = n_j p_j (1 - p_j). \] (2.3.3)

We will use this to find the variance of \(\hat{p}_j\)
\[
\text{Var}\{p_j\} = \text{Var}\left\{ \frac{n_j - d_j}{n_j} \right\}
= \left( \frac{1}{n_j} \right)^2 \text{Var}\{n_j - d_j\}
= \left( \frac{1}{n_j} \right)^2 n_j p_j (1 - p_j)
= \frac{p_j (1 - p_j)}{n_j}.
\]

Thus the estimate of the variance of \( \hat{p}_j \) is expressed as

\[
\text{Var}(\hat{p}_j) = \frac{\hat{p}_j (1 - \hat{p}_j)}{n_j}.
\] (2.3.4)

Now we find the \( \text{var}\{\log(\hat{p}_j)\} \) using the Taylor Series Approximation, also known as the Delta Method, which states

\[
\text{Var}\{g(X)\} \approx \left( \frac{dg(X)}{dX} \right)^2 \text{Var}(X),
\] (2.3.6)

So using equations (2.3.6) and (2.3.5) we find

\[
\text{var}\{\log(\hat{p}_j)\} \approx \frac{\text{var}(\hat{p}_j)}{\hat{p}_j^2}
\approx \frac{(1 - \hat{p}_j)}{\hat{p}_j n_j}.
\] (2.3.7)

Substituting back in the original terms for \( \hat{p}_j \) we get

\[
\text{Var}\left\{ \log\left( \hat{S}(t) \right) \right\} \approx \sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}.
\]

Applying the definition of the Taylor Series Approximation again gives us
\[ \text{Var} \{ \log (\hat{S}(t)) \} \approx \left( \frac{1}{\hat{S}(t)} \right)^2 \text{Var} \{ \hat{S}(t) \}, \]

So we can find the variance for K-M estimate

\[ \text{Var} \left( \hat{S}(t) \right) \approx \left( \hat{S}(t) \right)^2 \sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}. \]

The standard error is defined to be the square root of the estimated variance of the K-M estimate which leads us to

\[ \text{se} \{ \hat{S}(t) \} \approx \hat{S}(t) \left( \sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)} \right)^{1/2}, \] (2.3.8)

This estimate is known as Greenwood’s Formula.

One common summary statistic in survival analysis is the median survival time. This is defined to be the time at which 50% of the population is expected to experience the event of interest. This number is expressed as \( t(50) \) where \( S(t(50)) = 0.5 \). Deriving the median from an estimated survival function can pose a problem since these tend to be step functions. The estimated mean survival time, \( \hat{t}(50) \), is defined as the smallest observed survival time for which \( \hat{S}(t) < 0.5 \). In equation form,

\[ \hat{t}(50) = \min \left\{ t_{(j)} \mid \hat{S}(t_{(j)}) < 0.5 \right\}, \] (2.3.9)

Where \( t_{(j)} \) is the jth ordered event time, \( j = 1, 2, \ldots, r \). The standard error of the estimate for the median is calculated by applying the delta method (which was also used for Greenwood’s formula) to be
\[
\text{se}\{\hat{S}\{50\}} = \frac{1}{\text{f}\{\hat{S}\{50\}}\text{se}\{\hat{S}\{50\}}\}.
\]

(2.3.10)

Using the standard error of the estimate of the median, confidence intervals can be built if desired.

### 2.4 Hypothesis Testing Using the Log-Rank Test Statistic

Generally statistical analyses are used to make decisions about 2 or more independent samples in regards to an outcome variable of interest. In the survival setting, often we compare 2 or more groups receiving different treatments and wish to determine which group has a longer survival time. Hypothesis testing is used to draw conclusions about such comparisons. A hypothesis is a statement about a population which we wish to prove or disprove looking at our sample. In order to make such a decision, two complementary statements are made. For the purpose of this argument, assume we are trying to decide if a difference exists between two groups. The null hypothesis, denoted \(H_0\), is generally the argument that there is no difference between the groups. While the alternative hypothesis, denoted \(H_1\), states that there is a difference between at least two of the groups. In order to quantify the decision making process we calculate a test statistic from the observed data. From the test statistic we calculate a p-value which allows us to say how likely it is that we observe our data, given that the null hypothesis is true. So if we have a large test statistic, which corresponds to a small p-value, we know it is very unlikely that we would observe a difference at least as extreme as our data, given the null hypothesis is true, thus we favor the alternative hypothesis. This process is known as hypothesis testing.

The underlying distribution of a test statistic allows us to determine the p-value. Say we have a test statistic, \(W\), whose realized value is \(w\), then the p-value will be
P(W ≥ w) = 1 − F(w), where F(w) is the distribution function for W under the null hypothesis. Often times the test statistic takes the form of a standard normal distribution, or the square of the standard normal which is a chi-squared distribution with \( n - 1 \) degrees of freedom, \( \chi^2_{n-1} \), where \( n \) is the number of groups we are comparing. To decide what p-value allows us to reject the null hypothesis we set a significance level, \( \alpha \), which is the probability of making a type I error, or the probability of rejecting the null hypothesis when it is in fact true. Typically, \( \alpha = 0.05 \), thus we have a five percent significance level test. So if our p-value is less than 0.05 we reject the null hypothesis. It is important to note that a statistically significant result does not imply a clinically significant result. A test may be powered to find a very small difference which is statistically significant, however in terms of scientific results that difference may be meaningless. For this reason, a measure of the size of the treatment difference along with a confidence interval can be more useful than a hypothesis test in describing discrepancies between groups. We will be examining the Log-Rank (L-R) Test, which is a non-parametric test of the difference between two survival groups.

When constructing the L-R test for two groups, we need the ordered event times for the pooled data, and the overall numbers at risk and failing for each event time. Let there be \( r \) distinct death times, \( t_{(1)} < t_{(2)} < \cdots < t_{(r)} \), across the two groups. At time \( t_{(j)} \), the number of individuals who experience the event in Group 1 is \( d_{1j} \), out of a total number at risk \( n_{1j} \), and the number of individuals who experience the event in Group 2 is \( d_{2j} \), out of a total number at risk \( n_{2j} \). Therefore at each distinct event time, a 2x2 table such as Table 2.1 can be constructed.
Table 2.1 The number of individuals at risk, and the number of individuals who experience the event, by group, at an event time.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of events at $t_{(j)}$</th>
<th>Number surviving past $t_{(j)}$</th>
<th>Number at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$d_{1j}$</td>
<td>$n_{1j} - d_{1j}$</td>
<td>$n_{1j}$</td>
</tr>
<tr>
<td>2</td>
<td>$d_{2j}$</td>
<td>$n_{2j} - d_{2j}$</td>
<td>$n_{2j}$</td>
</tr>
<tr>
<td>Total</td>
<td>$d_j$</td>
<td>$n_j - d_j$</td>
<td>$n_j$</td>
</tr>
</tbody>
</table>

Now to construct a hypothesis test of independence, where the null hypothesis states that there is no difference in the survival experience of individuals between the two groups (or that group and event time are independent), and conversely for the alternative hypothesis that there is a difference in the survival experience of individuals between the two groups. In order to evaluate this hypothesis we consider the number of individuals who experience the event at each event time, versus the number of individuals we would expect to experience the event, given the null hypothesis. Then combine the information for each death time to form a single value.

If the event times and groups are assumed to be independent, and the marginal totals in Table 2.1 are held fixed, the numbers in the table can be determined just by $d_{1j}$, the number of events at time $t_{(j)}$ in group 1. Thus $d_{1j}$ is a random variable with values ranging from 0 to $\min(d_j, n_{1j})$, and has a hypergeometric distribution of the form

$$
\binom{d_j}{d_{1j}} \binom{n_j - d_j}{n_{1j} - d_{1j}} \binom{n_j}{n_{1j}}.
$$

(2.4.1)
The expected value of \(d_{1j}\) is

\[
e_{ij} = \frac{n_{ij}d_j}{n_j},
\]

or the expected number of individuals who experience the event at time \(t\).

We now sum the difference between the observed number of events and the expected number of events over all distinct event times, yielding the following statistic,

\[
U_L = \sum_{j=1}^{r} (d_{ij} - e_{ij}).
\]

Note that the mean for this statistic is zero, since \(E(d_{1j}) = e_{1j}\). The variance for \(U_L\) will be the sum of the variances for the \(d_{1j}\)’s, since the event times are independent of one another. Thus the variance can be expressed as

\[
V_L = \text{var}(U_L) = \sum_{j=1}^{r} \sum_{j=1}^{r} n_{ij}n_{2j}d_j(n_j - d_j) / n_j^2(n_j - 1).
\]

If we assume that \(U_L\) has an asymptotically normal distribution with an adequate number of death times, then

\[
\frac{U_L}{\sqrt{V_L}} \sim \mathcal{N}(0,1),
\]

or, in words, \(U_L\) divided by the square of its variance has a standard normal distribution. We know that the square of a standard normal is a chi-square distribution with one degree of freedom, thus

\[
\frac{U_L^2}{V_L} \sim \chi_1^2.
\]

This number is the L-R test statistic, \(W_L = \frac{U_L^2}{V_L}\), which evaluates how far the expected number of events deviates from the observed number under the null hypothesis of no difference between
the two groups. The L-R test will be revisited in the section of Chapter 4 when the Renyi-type tests are discussed.

2.5 Cox-Proportional Hazards Model – A Semi-parametric Procedure

While nonparametric methods are attractive because of the lack of assumptions and the simplicity of the calculations, nonparametric modeling doesn’t allow for covariates to be included in the model. Often when comparing survival times of two treatment groups, factors such as age, weight, and gender need to be included in the analysis. While the K-M estimate is frequently used for an initial look at the dataset, it is too simplistic to take into account demographic variables and other possibly important explanatory variables. The relationship between survival time and explanatory variables cannot be ignored and so statistical modeling techniques are employed. Often in survival analysis the risk or hazard of the event occurring after the onset of the study is of particular interest. So we aim to determine the effect of treatments on the form of the hazard function as well as any role explanatory variables may play. Since the hazard function is directly related to the survival function (see equations (2.1.6) and (2.1.9)) once the hazard function is determined, an estimate of the survival function can be found along with median survival time which will be a function of the explanatory variables.

It is important to note that the proportional hazards model operates under the proportional hazards assumption, that the hazard for an individual in one treatment group at a given time is proportional to the hazard of a similar individual in the other group, and this proportion remains constant over time. Suppose there are two treatments being compared, the standard treatment and a new treatment with \( n \) individuals in each group. Let \( h_i(t) \), \( i = 1, 2, ..., n \) denote the hazard function for an individual in the study. Let \( h_0(t) \) denote the hazard function for an individual on the standard treatment with covariate values all equal to zero, this is also known as
the baseline hazard function. Given the proportional hazards assumption the hazard function for an individual in the new treatment group will be $\psi h_0(t)$, where the relative hazard $\psi = \exp(\eta_i)$ is a non-negative function of a set of parameters. Let $x_1, x_2, \ldots, x_p$ be the realized values of $p$ explanatory variables, $X_1, X_2, \ldots, X_p$. Let $x_1$ be an indicator variable equal to 1 if the patient is on the new treatment and zero if the patient is on the standard treatment. A set of values of the explanatory variables will be represented by the vector $x = (x_1, x_2, \ldots, x_p)'$. So we say that $\eta_i$ is a linear combination of the $p$ explanatory variables in $x_i$

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}, \quad (2.5.1)$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ is a vector of coefficients for the explanatory variables in the model.

Putting all of these elements together we can write the general form of the hazard function for the $i$th individual

$$h_i(t) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) h_0(t) = \exp(\eta_i) h_0(t) \quad (2.5.2)$$

In order to determine the relationship of the explanatory variables, $X_1, X_2, \ldots, X_p$, to the hazard or risk of an event we need to estimate the unknown values for the parameters, $\beta_1, \beta_2, \ldots, \beta_p$, and we also may need to estimate $h_0(t)$. Generally the $\beta$’s are estimated first and then those estimates are used to calculate an estimate of the baseline hazard function. The method of maximum likelihoods is used to find the estimates of the $\beta$’s. We use the Cox partial likelihood function for the proportional hazards model,

$$L(\beta) = \prod_{j=1}^{n} \left( \frac{\exp \left( \beta' x_{(j)} \right)}{\sum_{i \in R(t_{(j)})} \exp \left( \beta' x_{(i)} \right)} \right)^{\delta_j}, \quad (2.5.3)$$
where $R(t_{(j)})$ is the risk set, or group of individuals who still have not experienced the event of interest and are eligible to do so, just before time $t_{(j)}$. And $\delta_i$ is an indicator variable which equals zero if an observation is censored and equals one otherwise. In order to find the estimates for the $\beta$’s we first take the log of the partial likelihood function (equation (2.5.3))

$$
\log L(\beta) = \sum_{i=1}^{n} \delta_i \left\{ \beta' x_i - \log \sum_{l \in R(t_{(j)})} \exp(\beta' x_{(j)}) \right\}.
$$

(2.5.4)

Then we take the derivative of the partial log likelihood (equation (2.5.4)) with respect to each parameter, set that derivative equal to zero and solve for each parameter. When there are multiple parameters involved in an analysis an iterative procedure is needed to accomplish this estimation, the Newton-Raphson procedure is commonly used. While these parameter estimates are derived from a partial likelihood function rather than a likelihood function, they still have the nice properties of maximum likelihood estimators.

The standard errors for the $\hat{\beta}$’s can be found by taking the square root of the diagonal elements of the inverse information matrix $I^{-1}(\hat{\beta})$. The information matrix is a $p \times p$ matrix of the negative second derivatives of the log likelihood.

The standard error for the hazard ratio $\psi$ can be calculated using the Taylor Series Approximation, also known as the Delta Method (see equation (2.3.6)). We know that $\psi$ is a function of $\beta$, where $\psi = \exp(\beta)$, so we can use the estimate $\hat{\psi} = \exp(\hat{\beta})$ to calculate an estimate of the variance

$$
\text{var}(\hat{\psi}) = \text{var}\left(\exp(\hat{\beta})\right) = \left\{\exp(\hat{\beta})\right\}^{2} \text{var}(\hat{\beta}).
$$

(2.5.5)
So to find the standard error for the estimate of the hazard ratio we take the square root of the variance giving us

\[ se(\hat{\psi}) = \hat{\psi} \cdot se(\hat{\beta}). \]

### 2.6 Parametric Proportional Hazards Model

If a data set has a probability distribution associated with it we can apply that probability distribution to obtain more precise estimates of the parameters for the coefficients than one would get with a Cox PH model. The most commonly used distribution in survival models is the Weibull distribution

\[ f(t) = \lambda \gamma t^{\gamma - 1} \exp(-\lambda t^\gamma), \]

for \( 0 \leq t < \infty \), with scale parameter \( \lambda \) and shape parameter \( \gamma \). When \( \gamma = 1 \), the p.d.f. becomes the exponential distribution. The hazard function for the Weibull distribution is

\[ h(t) = \lambda \gamma t^{\gamma - 1}, \]

and \( \gamma \) directly influences the shape of the hazard function. So if \( 0 < \gamma < 1 \) the hazard function is decreasing, if \( \gamma = 1 \) the hazard function is constant, and if \( \gamma > 1 \) then the hazard function is increasing over time. Also note that the survival function for the Weibull distribution is

\[ S(t) = \exp(-\lambda t^\gamma). \]

In order to fit this model to a single sample we use the method of maximum likelihood. Unlike other forms of data, survival data often include right censored observations and these can be included in the likelihood. A censored observation will have survived up to time \( t^* \) so the
probability of survival is beyond this time or \( P(T \geq t^*) \) which is equal to \( S(t^*) \) and we include this information in the likelihood. Given that there are \( r \) death times and \( n - r \) right censored observations the likelihood becomes

\[
\prod_{j=1}^{r} f(t_j) \prod_{l=1}^{n-r} S(t_l^*). 
\]

Now using the indicator variable, \( \delta_i \), for censored observations the likelihood becomes

\[
\prod_{i=1}^{n} \left\{ f(t_i) \right\}^{\delta_i} \left\{ S(t_i) \right\}^{1-\delta_i},
\]

(2.6.4)

Which can be simplified to

\[
\prod_{i=1}^{n} \left\{ \frac{f(t_i)}{S(t_i)} \right\}^{\delta_i} S(t_i) = \prod_{i=1}^{n} \left\{ \frac{h(t_i)}{S(t_i)} \right\}^{\delta_i} S(t_i).
\]

(2.6.5)

Now apply this to the Weibull distribution using equations (2.6.2) and (2.6.3) to obtain the likelihood function

\[
L(\lambda, \gamma) = \prod_{i=1}^{n} \left\{ \lambda \gamma t_i^{\gamma-1} \right\}^{\delta_i} \exp\left(-\lambda t_i^\gamma \right).
\]

(2.6.6)

Since maximizing the likelihood is easier with logarithmic we write

\[
\log L(\lambda, \gamma) = \sum_{i=1}^{n} \delta_i \log(\lambda \gamma) + (\gamma - 1) \sum_{i=1}^{n} \delta_i \log(t_i) - \lambda \sum_{i=1}^{n} t_i^\gamma,
\]

And noting that \( \sum_{i=1}^{n} \delta_i = r \), the log-likelihood simplifies to

\[
\log L(\lambda, \gamma) = r \log(\lambda \gamma) + (\gamma - 1) \sum_{i=1}^{n} \delta_i \log(t_i) - \lambda \sum_{i=1}^{n} t_i^\gamma.
\]
2.7 Accelerated Failure Time Models (AFT)

In some cases the hazard function is not a monotonic increasing or decreasing function, in other words, the proportional hazards assumption is not valid. Thus an accelerated failure time (AFT) model may be an appropriate way to represent the survival data. The Weibull distribution can be expanded to fit an AFT model which we will illustrate. Other distributions that can be used to fit AFT models include, but are not limited to, the log-logistic distribution, the lognormal distribution, the gamma distribution, as well as the inverse Gaussian distribution. With the accelerated failure time model the hazard for the $i$th individual at time $t$ can be written in the general form

$$h_i(t) = e^{\eta_i} h_0(t \mid e^{\eta_i}),$$

(2.7.1)

where $\eta_i = \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_p x_{pi}$ is the linear component of the model for the $i$th subject with $p$ explanatory variables for the covariate vector $x = (x_1, x_2, ..., x_p)'$. The baseline hazard function, $h_0(t)$, is the hazard function for an individual whose values for the explanatory variables are all equal to zero.

The survival function for the $i$th individual can be written as

$$S_i(t) = S_0\{t / \exp(\eta_i)\},$$

where $S_0(t)$ is the baseline survival function.

Now apply this to the Weibull distribution to develop the Weibull AFT model. Assume the survival times have a Weibull distribution with shape parameter $\gamma$ and scale parameter $\lambda$; so the baseline hazard function is

$$h_0(t) = \lambda \gamma t^{\gamma - 1}.$$

(2.7.2)
Applying this form of the baseline hazard function to equation (2.5.2) the hazard function for the
ith individual becomes

\[ h_i(t) = e^{-\eta_i} \lambda y \left(e^{-\eta_i} t\right)^{y-1} = \left(e^{-\eta_i} \right)^y \lambda y t^{y-1}. \] (2.7.3)

Thus the scale parameter becomes \( \lambda e^{-\eta_i} \) and the shape parameter remains \( y \) for the Weibull
distribution. So the Weibull distribution has both the accelerated failure time property and
proportional hazards property, and it is the only distribution to possess both.

Now to expand the survival function to accommodate the AFT property we first define a
log-linear model for the random variable \( T_i \) associated with the event time for the ith individual

\[ \log T_i = \mu_i + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_p x_{pi} + \sigma \epsilon_i. \] (2.7.4)

Here \( \alpha_1, \alpha_2, \ldots, \alpha_p \) are the unknown parameters associated with the \( p \) explanatory variables. Two
other parameters are now involved, an intercept parameter \( \mu \), and a scale parameter \( \sigma \). The
random variable \( \epsilon_i \) represents the variation of the values of \( \log T_i \) from the linear portion of the
model, and \( \epsilon_i \) has its own underlying distribution. Given this model the survivor function then
becomes

\[ S_i(t) = S_{\epsilon_i} \left( \frac{\log t - \mu - \alpha_1 x_{i1} - \alpha_2 x_{i2} - \ldots - \alpha_p x_{pi}}{\sigma} \right) = S_{\epsilon_i}(z_i). \] (2.7.5)

To get the p.d.f. \( f_i(t_i) \) we differentiate the above equation with respect to \( t \) which gives us

\[ f_i(t_i) = \frac{1}{\sigma t_i} f_{\epsilon_i}(z_i), \] (2.7.6)

where \( z_i = \frac{\log t - \mu - \alpha_1 x_{i1} - \alpha_2 x_{i2} - \ldots - \alpha_p x_{pi}}{\sigma} \). Now plug this information into the general form of the
likelihood (equation (2.6.4)) so that it applies to AFT model giving us
\[
L(\alpha, \mu, \sigma) = \prod_{i=1}^{n} \left( \sigma t_i \right)^{-\delta_i} \left( f_{\varepsilon_i}(z_i) \right)^{\delta_i} \left( S_{\varepsilon_i}(z_i) \right)^{1-\delta_i}.
\]

For ease of maximum likelihood estimation procedures we then take the log of the likelihood

\[
\log L(\alpha, \mu, \sigma) = \sum_{i=1}^{n} \left\{ -\delta_i \log(\sigma t_i) + \delta_i \log f_{\varepsilon_i}(z_i) + (1 - \delta_i) \log S_{\varepsilon_i}(z_i) \right\},
\]

where the MLEs of \(\mu, \sigma, and \alpha_1, \alpha_2, ..., \alpha_p\) are found using an iterative method such as the Newton-Raphson procedure.

In this chapter we have provided a broad overview of survival analysis topics. Non-parametric estimates, the Log-Rank Test, Cox proportional hazards models, and accelerated failure time models all were discussed. The relationships between the hazard functions, survival functions and p.d.f.s were established. Chapter 3 will use some of the concepts outlined here to take an in depth look at mean survival time, restricted mean survival time, and how the median compares to these lesser known summary statistics in survival analysis.
Chapter 3  The Restricted Mean

In this chapter we will discuss mean survival time, censoring, restricted mean survival time, and some of the existing work and methods for calculating the restricted mean. Restricted mean survival time is calculated as the area under a survival curve up to a point \( \tau \) and currently the choice of this point appears to be arbitrary. In all of the literature researched for this work, very little guidance is given or attention is paid to the choice of \( \tau \).

3.1 Basic Calculation of the Mean

The median is most often preferred as the summary statistic for survival time since it is not influenced by censored observations. However, there are still times when a mean summary statistic could be appropriate and informative. The mean survival time for a given population is calculated by integrating the survival function over the interval \([0, \infty)\), or in equation form

\[
\mu = E(x) = \int_0^\infty tf(t) \, dt = \int_0^\infty S(t) \, dt. \tag{3.1.1}
\]

This equality is proven using integration by parts.

As we know from Chapter 2,

\[
S(t) = 1 - F(t), \tag{3.1.2}
\]

Also,

\[
\frac{\partial S(t)}{\partial t} = -f(t). \tag{3.1.3}
\]
Now

\[ \int_0^\infty f(t)\,dt = -\int_0^\infty \frac{\partial S(t)}{\partial t}\,dt. \quad (3.1.4) \]

Integrating by parts, let \( u = t, \, du = dt, \, dv = \frac{\partial S(t)}{\partial t}\,dt, \) and \( v = S(t), \) so we get

\[ -\int_0^\infty \frac{\partial S(t)}{\partial t}\,dt = -\left[ tS(t) - \int_0^\infty S(t)\,dt \right]_0^\infty \]
\[ = -\left[ 0 - \int_0^\infty S(t)\,dt \right] \]
\[ = \int_0^\infty S(t)\,dt. \quad (3.1.5) \]

Thus the expected value of survival time is equal to the area under the survival curve. This relationship is useful in parametric modeling since the forms of \( E(x) \) are known for all distributions whose moments exist. For example, we know \( E(x) = \lambda \) for an Exp \((1/\lambda)\) distribution, and this parameter is easily estimated by computer packages that perform survival modeling, minimizing the need to calculate integrals by hand.

An initial estimate of the mean can be calculated from the K-M Estimate of the survival function as described in Chapter 2 equation (2.3.2). So we substitute this estimate in our equation for the mean giving us

\[ \hat{\mu} = \int_0^\infty \hat{S}(t)\,dt. \quad (3.1.6) \]

The K-M estimate is a step function and thus there is no mathematical function to integrate. However, the integral of the K-M survival estimate is the same as dividing the area under the curve into rectangles, computing the area of the rectangles, and then summing all of the areas together. In symbols,
\[
\hat{\mu} = \sum_{i=1}^{D} \hat{S}(t_{i-1})(t_i - t_{i-1}),
\]

where \( t_0 \) is defined to be zero, and \( D \) is the last observed event time.

This estimate is only valid if the largest event time is observed not censored. If the largest observed event time is censored then the estimate of the mean is undefined, because the function does not converge to zero. To correct this problem some simple solutions have been developed as well as some more complicated procedures, all of which will be discussed later in this chapter.

### 3.2 Advantages of Using the Mean

In survival analysis the mean survival time is often cast aside in favor of the median survival time because the mean is a biased statistic in most cases. Often when censoring is present, equation (3.1.6) underestimates the mean, or if the distribution of survival time is highly positively skewed the mean will not be a good representation of the overall data. However, there are many situations in which the advantages of using the mean survival time outweigh the nuisance of a biased estimator. For example, when it comes to interpreting results, comparison of treatments is more clearly interpreted as length of time lost or gained using mean survival, than in the case of the hazard ratio or 5-year survival. When examining “Extending life” (living longer with the disease) versus “Saving life” (being cured of the disease), a hazard ratio can be misleading by indicating that a treatment may save a percentage of lives where it really only is extending them, mean survival helps correct this misconception. (Maetani, 2004) Mean survival can help individualize therapeutic decisions, which is currently a prominent topic in statistics literature; for example, younger patients may want a riskier surgery if the chance of long term survival is high.
A median is not the best estimator in cases when fifty percent of the sample has not yet experienced the event of interest. Projecting half-lives can result in over-estimation, in that real half-lives could be significantly shorter, and relies on the assumption that the slope of the survival curve over time follows a predictable function. This point was made in a paper by Meier-Kriesche, et al, which compared projected median kidney transplant survival from previous studies performed during 1988 to 1995 with what the actual median was found to be when they examined follow up data. The authors argue that using the difference in the area under the K-M curve between the two groups for the amount of time studied would provide a better, more reliable estimate for the treatment difference than a prediction of the median differences, especially during early study times. Also, projecting the median doesn’t take other factors into account, such as success due to re-transplant in a renal allograft study. In a situation such as this, mean survival estimation would allow for use of the data at hand, rather than making predictions on future data, and for confounding factors to be included in the analysis.

On the other side of the argument, when a study has had well over fifty percent of patients experience the event of interest, using the median doesn’t utilize all of the information available; it only uses half of the sample. For example, a study by Moore (2007), illustrates a pancreatic cancer treatment where the medians between the two groups were identical, but prior to the median one group performed significantly better. This treatment difference is picked up by the difference in the areas under the K-M curves. So using the mean survival, (or area under the curve) to describe a group, incorporates all of the data at hand, unlike the median.

3.3 Current uses of Mean survival time

While median survival time is one of the most commonly cited statistics used to summarize survival data, there are instances when mean survival time is more descriptive and
useful to investigators. According to Maetani et al (2004), mean survival time is currently used for the estimation of quality-adjusted life years, technology assessment, and economic analysis (such as cost effectiveness analysis). Meier-Kriesche (2004) cites examples of using the mean to describe life years gained, and years of graft life gained for a renal transplant. Seruga (2009) describes comparisons of medical therapies in terms of absolute benefit between therapies as the area between time-to-event curves in phase III clinical trials for breast and colorectal cancer. Tanju (2010) discusses examining extended resections of pulmonary metastases using the K-M estimate method, where mean survival is calculated from date of first metastasectomy to date of last follow-up or death.

The following example is provided to give a detailed illustration of a current use of mean survival time. **Example:** One prominent area where the mean is in use is pharmacoeconomic methodology, which examines the cost of a treatment to extend a life by one year, or cost per life year saved. This is a good measurement when comparing a new, more expensive treatment to a standard treatment. Does the new treatment save more years than the old treatment and if so does that make the cost worthwhile? This method, presented by Messori (1997) fits a Gompertz function to the survival data and integrates the function over \((0, \infty)\) to calculate the area under the curve (AUC). A separate curve is fit for each treatment, and then the difference in AUC is calculated. From this a cost-effectiveness analysis can be performed,

\[
\frac{\text{incremental cost of new treatment}}{\text{incremental benefit of improved survival}} = \text{cost per life year gained.}
\]

The incremental cost is the difference in the cost of treating 100 patients with treatment A versus treating 100 patients with treatment B. Incremental benefit is the difference in the AUC between treatment A and treatment B corrected for any difference in sample size of the two groups. AUC measurements are much more reliable when the experimental portion of the curve
has a larger area than the extrapolated portion of the survival curve. In other words, when the function is fit to the data, it is not desirable to have much of the curve extend beyond the last observed time. The Gompertz function used in calculations is

\[ SP = f(t) = 100s^t g^e^t \]

where SP stands for survival percentage, t is time, and s, g, and c are the model parameters. The measurement of cost per life year gained provides a numerical way to compare costs across treatments and determine pharmacoeconomic benefit. (Messori, 1997) End Example.

While mean survival time does have its place in survival analysis, caution should be taken when it is used. The mean is a biased estimator and it often underestimates the true population mean survival time, this should be stated when a mean is cited in a study. To not acknowledge the bias or account for it in some way in a study raises a flag to the educated reader. As stated previously, the mean is biased due to the positive nature of survival time data, as well as the common presence of censored observations. While the nature of time cannot be changed, there are ways to account for censoring, and this is the focus of this thesis.

3.4 Restricted Mean

To solve the problem of censoring, the restricted mean and its standard error were developed in 1949 by J.O. Irwin. He used an actuarial life table for survival estimates, and developed the basis for the restricted mean standard error estimate that is still in practice today. (Irwin, 1948) Irwin’s work is cited in Kaplan and Meier’s 1958 paper which develops their namesake nonparametric estimate for survival time, also known as the Product-Limit Estimate. Kaplan and Meier make note of a restricted mean saying, “The estimate of the mean life may have to be truncated at the greatest of the observations limits,” which would be Last Observed Time (LOT) in the terminology used in this work. The K-M estimate has become the favored
nonparametric measure in survival literature, and it is a maximum likelihood estimator. It is also the estimator used when defining the simplest version of a restricted mean. (Kaplan and Meier, 1958)

The simplest way of calculating mean survival time in the presence of censoring is to use a restricted mean. A restricted mean can be used where either the last observation is treated as an event (we will call this LOT) or the investigator can assign an interval \([0, \tau]\) where \(\tau\) is assumed to be the longest possible survival time for that study. Another version of the restricted mean is to assume the last event time is the last observed time regardless of later censored observations; we will call this LET. The abbreviations LOT and LET are used by SAS® in the Lifetest Procedure as options for calculating mean survival time. Define the restricted mean survival time, \(\mu(\tau)\), for random variable \(T\) to be the expected value of \(\min(T, \tau)\), which is evaluated as the area under the survival curve \(S(t)\) up to \(\tau\)

\[
\mu(\tau) = E\left( \min(T, \tau) \right) = \int_0^\tau S(t) dt
\]  
(3.4.1)

Thus the restricted mean estimator is given by

\[
\hat{\mu}_\tau = \int_0^{\hat{\tau}} \hat{S}(t) dt.
\]  
(3.4.2)

When the event of interest is death, \(\mu(\tau)\) can be thought of as the “\(\tau\)-year life expectancy.” We use the restricted mean in place of the regular (unrestricted) mean as right censoring often leaves the regular mean to be undefined. As \(\tau \to \infty\) the unrestricted mean will exceed the restricted
mean for any \( \tau \), and \( \mu(\tau) \) is a monotonically increasing function of \( \tau \). (Royston and Parmar, 2011)

The variance for the restricted mean is

\[
\hat{\nu}[\hat{\mu}_\tau] = \sum_{i=1}^{d} \left[ \int_{0}^{\tau} \hat{S}(t) dt \right]^2 \frac{d_i}{Y_i(Y_i - d_i)}
\]

Where \( Y_i \) is the number of individuals who are at risk at time \( t_i \), and \( d_i \) is the number of events at time \( t_i \), and in this case \( \hat{S}(t) \) is the KM estimate (Moeschberger and Klein section 4.5).

It has been shown by Gill (1983) that using the K-M estimator in estimating \( \mu \) has attractive large sample properties. The K-M estimate of the mean has the property of weak convergence, where the K-M estimator converges to an underlying distribution on the interval \([0, \tau]\), so it has applications to the mean lifetime. As \( n \) approaches infinity the mean converges to a standard normal. (Gill, 1983) So

\[
\hat{\mu}(t^*) = \int_{0}^{t^*} \hat{S}(t) dt,
\]

and \( \hat{\mu} = \hat{\mu}(\infty) = \hat{\mu}(x_{(n)}) \) where \( x_{(n)} \) is the largest observation time.

### 3.5 Methods Used to Calculate the Restricted Mean

In this section we review methods which calculate a restricted mean, or which are an alternate method to calculating a restricted mean. The method of discounting future years, mixture cure models, and using pseudo-observations to calculate a restricted mean in the presence of censoring will be summarized. Note that in none of these methods do we have guidance for how to choose \( \tau \), the cut-point at which the mean should be restricted.
3.5.1 Discounting Future Years

Discounting future years is a method that emphasizes the importance of survival at the beginning of the study. We anticipate that our method will also place the focus on early study times, so we go into some detail here. This method is useful in regards to making organ transplant decisions such as whether or not to accept a transplant and what patient should be offered a transplant. The concept of discounting future years comes from the idea that patients do not value years in the distant future the same as they do the current or upcoming year. Often a patient is willing to sacrifice a longer survival time if it means his or her quality of life will improve in the near future. This is often the case with kidney failure, where a person can live for quite a long time on dialysis; however, a new kidney is preferable since it eliminates the need for dialysis, but also increases risks associated with organ transplants. In order to help in the decision making process, survival curves can be generated from past patients’ experiences and estimates of expected lifetimes can be made for both cases; either the patient has the transplant or does not. A restricted mean can be used, as illustrated above, but imposing a cut-point, \( \tau \), on the data does not allow for comparing the entire lifetime which may pose complications in this case. Discounting future years is a way to manage the problem of poorly estimated tail distributions. Tail distributions are used when the last observation is censored and the investigator wished to estimate the remaining survival distribution after the study has ended. Discounted lifetimes can be calculated using the following equation

\[
\int_0^\infty S(u)\exp(-D(u))du
\]

(3.4.3)

Where \( D(u) \) is a positive increasing discount function of time in the future, \( u \). As \( D(u) \) increases, the value \( \exp(-D(u)) \) becomes smaller, thus the extrapolated values at the longer time points have less influence on the integral. The discounted expected lifetimes are then
where $W_i$ indicates data without transplants, and $T_i$ indicates data from those with transplants. A “quality of life adjustment” could also be made and combined with this calculation but that is not currently pertinent to our task at hand. (SRTR Working Paper, 2007) Again, discounting future years illustrates one way emphasis can be placed on early treatment time, which is valued by patients.

### 3.5.2 Mixture Cure Models

The next method is mentioned for thoroughness. Mixture Cure Models provide an alternative to a restricted mean, since long term right censored patients are assumed to be “cured”, but this assumption can be misused and lead to inaccurate results.

Farewell’s paper (1982) arose out of observations by Pierce, Stewart, and Kopecky (1979) where a proportion of animals in a study did not experience the event of interest during the allotted time. The investigators didn’t feel it was appropriate to model the animals that didn’t experience the event with the same Weibull or lognormal model and consider the animals censored as those animals that did experience the event. They believed there were actually two underlying populations, thus a mixture cure model should be used to illustrate this situation.

These models are of particular interest in the case of cancer studies where some patients die from the cancer, some patients are in remission and the cancer comes back, while other patients are considered “cured.” Unsusceptible is the term used for cured patients, and susceptible refers to patients who can die from cancer. A brief background to the model is given

egin{align*}
E(W_i|D(u)) &= \int_0^\infty S_i^W(u) \exp(-D(u)) \, du,
\quad \\
E(T_i|D(u)) &= \int_0^\infty S_i^T(u) \exp(-D(u)) \, du
\end{align*}
as illustration; however, we do not go into detail as this method will not be explored further in this work. First off, define $Y$ to be a binary variable indicating long term survival

$$Y = \begin{cases} 0 & \text{if unsusceptible} \\ 1 & \text{if susceptible} \end{cases}$$

For a person with covariate vector $\mathbf{x} = (x_0, x_1, \ldots, x_p)$, where $p$ is the total number of covariates, $Y$ can be modeled using a logistic distribution,

$$P(Y = 1) = \frac{\exp(\beta' \mathbf{x})}{1 + \exp(\beta' \mathbf{x})}. \tag{3.4.4}$$

The time to event for patients where $Y = 1$ can be modeled with a variety of distributions, we will follow the example of Farwell and use the Weibull distribution,

$$f(t | Y = 1, \mathbf{x}) = \delta \lambda (\lambda t)^{\delta - 1} \exp\left\{- (\lambda t)^{\delta}\right\}, \tag{3.4.5}$$

Where $\delta$ is the shape parameter and $\lambda = \exp(-\mathbf{y}' \mathbf{x})$ is the scale parameter for the distribution. The combination of equations (3.4.4) and (3.4.5) will yield a mixture cure model.

Farewell warns in a follow up paper (1986) that mixture cure models are appealing but they carry the danger of necessary assumptions. One must be sure the assumption of two independent underlying populations is valid; otherwise a mixture cure model will give misleading results since the subpopulation proportions will be imprecise.

### 3.5.3 Pseudo-Values

Recently a method which creates pseudo-values in the presence of censoring has been developed. Pseudo-values replace observed event times and censored observations with a “leave-one-out” estimator. A restricted mean can then be calculated from the pseudo-values.
For the work done in this area, the choice of the cut-point, $\tau$, for the restricted mean is still arbitrary.

Royston and Parmar (2011) provide a convincing argument for the use of a restricted mean when the proportional hazards (PH) assumption cannot be upheld or is in question. The hazard ratio is commonly considered the appropriate measure of difference between two treatment groups in randomized clinical trials. However, if the proportional hazards assumption does not hold, then the hazard ratio is a misleading if not useless description. The Log-Rank test can be used in the presence of non-PH, but doesn’t work effectively for extreme cases, such as when the survival curves cross. Note that when the survival curves cross, the PH assumption does not hold. The authors also suggest the restricted mean would be a good secondary measure when the PH assumption does hold and using a hazard ratio as a primary measure is justified. The restricted mean is a good secondary measure since its interpretation is intuitive to most audiences and adds to the description of treatment differences. One method Royston and Parmar explore to calculate a restricted mean is pseudo-values, or pseudo-observations.

As was described in Chapter 1, in survival analysis regression models are often specified using the hazard function and relationships are expressed using hazard ratios or hazard differences. However, in cases when the PH assumption is in question, it would be useful to be able to express the effect of covariates on a mean survival time, in a manner similar to classical regression analysis which is focused on the mean of an outcome variable or some transformation of the mean. Pseudo-values allow for this by replacing censored observations and event times with “leave-one-out” estimates of the probability of survival at a given time. (Anderson, 2004)

Using restricted mean survival time, instead of the median or hazard ratio, has the following advantages: it has a straightforward interpretation; it uses the entire range of data up to
τ instead of a single point in time; it can be used to model covariates; used with pseudo-values it has a the ability to model data using linear regression; and the model assumptions are minimal. The disadvantage lies in its dependence on the choice of τ, the wrong choice can produce misleading results. (Royston and Parmar, 2011) (Anderson, 2004) In Chapter 4 we present three methods developed to calculate an estimate of τ under a specific case of non-proportional hazards using the K-M estimates of survival time.
Chapter 4

Several New Methods for Determining a Cut-Point for the Restricted Mean

4.1 A Mathematical Definition of the Problem at Hand

So far we have discussed the restricted mean and its properties as well as some relevant current work in pseudo-observations. In all of the literature that was examined there was relatively little advice for choosing a cut-point, \( \tau \), for the restricted mean. The most commonly used options for the cut-point are the last observed time (LOT) or the last event time (LET). We argue that it is possible that neither of these choices is optimal for the goal of this work which is to calculate the restricted mean during the time where a true treatment difference exists.

According to SAS® Lifetest Procedure documentation, the options for choosing a cut-point when requesting mean survival time in the output are TIME\( \text{LIM} = \) LOT, LET (default), or a user defined time so long as it is after LET. That is to say, there is not an option for the SAS® package that allows the user to define a cut-point for calculating the restricted mean prior to the last event time. One problem with these options is that if the user employs the Strata option, to test for homogeneity of two groups, and requests mean survival time at LOT or LET, the cut-points could be different for the two groups. For purposes of comparing area under the survival curves of two groups it is logical that the areas must be over the same time interval. Therefore it is advised that the user be aware of the restrictions when using computer software to calculate the difference in mean survival time between two groups.
Other than LOT or LET, the only advice found for choosing a cut-point to restrict the mean is in Karrison (1997). He recommends finding the largest time point for which the standard error of the survival estimate is within reasonable limits. He cites the following equation for calculating standard error

\[ SE(\hat{S}(t)) \approx \hat{S}(t) \sqrt{\frac{(1-\hat{S}(t))}{n_t}}, \]  

(4.1.1)

where \( n_t \) is the number still at risk at time \( t \). Karrison states that a “reasonable” limit for SE is between five and ten percent. While this method at least provides a guideline for choosing \( \tau \) we argue that it does not take into account that the treatment difference could have ended long before the end of the study. This leads us to the mathematical definition of the scenario we are proposing for estimating \( \tau \).

Assume we have two treatment groups with survival functions

\[ S_1(t) = P(X > t) \]
\[ S_2(t) = P(Y > t) \]  

(4.1.2)

which encompass the survival times for both groups. Now we define functions with only the individuals surviving past a specific time, \( \tau \), so the functions will then be

\[ S_1(t; \tau) = P(X > t | t \geq \tau) \]
\[ S_2(t; \tau) = P(Y > t | t \geq \tau). \]  

(4.1.3)

We aim to find \( \tau \) such that the remaining cumulative difference in the area between the survival curves is less than or equal to a tolerance level, \( \delta \). Recall that the formula for the mean survival time is the area under the survival curve from zero to infinity, and the restricted mean is the area evaluated from zero to \( \tau \)

\[ \mu = \int_0^\infty S(t)dt, \]  

(4.1.4)
\[ \mu(\tau) = \int_0^\tau S(t)dt. \] (4.1.5)

So our goal is to find \( \tau \) such that,

\[ \int_\tau^\infty \left| S_i(t; \tau) - S_z(t; \tau) \right| dt \leq \epsilon, \] (4.1.6)

implying that area between the survival curves becomes negligible or zero after time \( \tau \), thus there is no longer a difference in treatment effect. Once \( \tau \) is determined, the restricted mean difference between the two groups can be calculated as

\[ \mu(\tau) = \int_0^\tau \left| S_i(t) - S_z(t) \right| dt. \] (4.1.7)

An alternate way of expressing this is to define

\[ \mu(\tau) = \int_0^\tau \left| S_i(t; \tau) - S_z(t; \tau) \right| dt. \] (4.1.8)

Thus

\[ \mu(\tau) = \mu(\tau) - \int_\tau^\infty \left| S_i(t) - S_z(t) \right| dt. \] (4.1.9)

We can also write this definition in terms of the hazard function since there is a direct relationship between the survival function and the hazard function. Here we do not write an equal statement to equation (4.1.9) but one that parallels the same logic, that after time \( \tau \) the hazards of the two groups are the same and there is no longer a treatment difference. It is arguably easier to visualize the problem at hand by looking at a cumulative hazard plot rather than a survival function. Figure 0.1 is a manufactured sample which mimics a piecewise
exponential example where if we were to move the origin to time 10, the survival curves for the 2 groups would be the same. This is easier to see in Figure 0.2, which is the cumulative hazard plot of the same data, where after time 10 the slopes for the two groups are the same. Since the hazard function is essentially the slope of the cumulative hazard function we can then plot that and see that after time 10 the hazards are indeed the same, shown in Figure 0.1.

![Figure 0.1 Example of survival functions of the two groups with equal hazards after time 10, and all observations censored after time 20.](image)
Figure 0.2 Example of cumulative hazard function of the two groups with equal hazards after time 10, and all observations censored after time 20.

Figure 0.3 Example of hazard functions of the two groups with equal hazards after time 10, and all observations censored after time 20.
Notice that in Figure 0.3, after time 10 the hazards do not go to zero immediately, but the area under the two curves becomes the same. This data was made specifically to illustrate this point, however we hypothesize that similar scenarios could, and do, happen in actual time to event studies.

Finally we put our mathematical definition in terms of the hazard function. Now we are looking for a \( \tau \) such that

\[
\int_{\tau}^{\infty} \left| h_1(t) - h_2(t) \right| dt \leq c^*,
\]  

(4.1.10)

where \( c^* \) is some tolerance level such that the area between the two hazard curves is negligible or zero.

While the area under the hazard curves is not equal to the mean survival time, we can still give this area a value. Define

\[
\omega_\Delta(t) = \int_{0}^{\tau} \left| h_1(t) - h_2(t) \right| dt,
\]

(4.1.11)

and

\[
\omega_\Delta' = \int_{0}^{\infty} \left| h_1(t) - h_2(t) \right| dt,
\]

(4.1.12)

thus

\[
\omega_\Delta(\tau) = \omega_\Delta' - \int_{\tau}^{\infty} \left| h_1(t) - h_2(t) \right| dt.
\]

(4.1.13)
4.2 Methods Developed to Address the Problem

In order to simulate scenarios as described in the previous section we will use a piecewise exponential framework. This allows user defined hazards to change at specified time periods during the simulated study time. We will test our methods to see how well they find the times where the hazards change, this will be our estimated cut-point,  \( \hat{T} \). The following methods will be developed and examined: the maximum distance method, the Renyi-type test method, and the estimated hazards method.

4.2.1 Maximum Distance Method

The maximum distance method is an intuitive approach to the problem at hand. The method is based on the idea that the separation between Kaplan-Meier curves after a certain time is due to the point where the largest difference between the survival probabilities occurs. Another way to think of this idea is that the largest distance between the curves illustrates the point where the treatment difference ends and it is no longer beneficial to be on one treatment versus the other. The separation of the curves after this point is simply due to the early treatment difference, thus we don’t want to include the area between the curves after the treatment difference has ended and inflate the effect of one treatment versus another.
Figure 0.4 The survival curves for a simulated piecewise exponential model where $h_1(t) = 0.05, t < 10$ and $h_1(t) = 0.20, t \geq 10$, and $h_2(t) = 0.20$ for all $t$.

In order to calculate the maximum distance between the two survival curves, assume we have $t_{(i)}$ ordered event times where $i = 1, \ldots, D$. We define our estimate of the cut-point to be

$$\hat{\tau} = \max\{\left|\hat{S}_1(t_{(i)}) - \hat{S}_2(t_{(i)})\right|\}$$

(4.2.1)

where $\hat{S}(t)$ is the usual K-M estimate of survival. The plot of the difference between the two survival curves can be found in Figure 0.5.
Figure 0.5 A plot of the absolute value for the distance between the two estimated survival curves for the two groups. Note that the maximum occurs right around t=10.

4.2.2 Renyi Type Test

The Renyi type test was developed from the idea that traditional tests, such as the Log Rank test, have little power to detect differences between groups when the hazard rates cross. This is due to the fact that early differences, in which one group has a higher hazard than the other, are negated when that relationship is inverted as time continues. This instance is a clear violation of the PH assumption. Thus the Renyi Type test was developed, and comes from an extension of the Kolmogrov-Smirnov test, but accounts for censoring. The basic premise is to take sequential evaluations of the absolute value of the following equation

\[ Z_j(t_{\text{max}}) = \sum_{i=1}^{p} W(t_i) \left[ d_{ij} - Y_{ij} \left( \frac{d_i}{Y_i} \right) \right], \quad j = 1, 2. \]  

And then find the maximum of those statistics. Equation (4.2.2) is analogous to the numerator of the L-R statistic with \( W(t_i) = 1 \), or the Wilcoxon test with \( W(t_i) = n_i \), the number at risk at
time \( t_i \). In hypothesis test framework, large values of the maximum value favor the alternative hypothesis that there is a difference in the hazards between the two groups. Notation is similar to that for the Log Rank test as outlined in Chapter 2. Take two independent samples of size \( n_1 \), and \( n_2 \) for a total sample size if \( n = n_1 + n_2 \). Given all event times, let \( t_1 < t_2 < \cdots < t_p \) be the distinct ordered death times for the total sample. As before \( d_{ij} \) represents the number of events at time \( t_i, i = 1, ..., D \), and group \( j = 1, 2 \), and \( Y_{ij} \) is the total number at risk for the given event time.

So, at each distinct event time point the following is calculated

\[
Z(t_i) = \sum_{t_k \leq t_i} W(t_k) \left[ d_{k1} - Y_{k1} \left( \frac{d_k}{Y_k} \right) \right]
\]

Where \( W(t_k) \) is a weight function and, as is the case with the Log Rank test statistic, here is equal to one. The standard error for \( Z \) is calculated using the following equation

\[
\sigma^2(t_{\text{max}}) = \sum_{t_k \leq t_{\text{max}}} W(t_k)^2 \left( \frac{Y_{k1}}{Y_k} \right) \left( \frac{Y_{k2}}{Y_k} \right) \left( \frac{1}{Y_k - 1} \right) (d_k)
\]

Where \( t_{\text{max}} \) is the last event time at which both groups still have at least one person at risk of experiencing the event. In order to test for a difference in hazards between groups the test statistic for this hypothesis is

\[
Q = \sup \{|Z(t)|, t < t_{\text{max}} \} / \sigma(t_{\text{max}})
\]

Or the supremum of the absolute value of \( Z(t) \) divided by the variance. Under the null hypothesis \( Q \) is approximately distributed as the supremum of the absolute value of a standard Brownian motion process, or \( \sup(|B(x)|, 0 \leq x \leq 1) \).
It is from this test that the second method for estimating the cut-point, $\tau$, is developed. The time when supremum of the absolute value of $Z(t)$ occurs is a logical place to set $\hat{\tau}$. This makes intuitive sense because the supremum is the point at which the test is performed, thus it takes into account the information in the data up to that time point. Thus we can calculate a restricted mean from the time origin to the time where the supremum occurs. This estimate of $\tau$ will be the same for both groups and occurs at a time when one or both groups has an event time. However, this statistic may be subject to small changes toward the end of the data, especially if no difference is occurring after a certain time point and the values for $|Z(t)|$ plateau, see Figure 0.6.

![Method looking at the change in the log-rank test at each failure point](image)

**Figure 0.6** A plot of the absolute value of the $Z(t)$ statistic, or the numerator of the Log-Rank statistic, versus time. The horizontal line indicated the maximum value of this statistic.

Since the supremum is sensitive to small changes, we fit a hockey stick model and search for the join point which most adequately fits the data. This will likely produce a conservative
value for \( \hat{\tau} \); however, in our scenario we prefer to be conservative in our estimate of the difference of the restricted means. This preference is due to the idea that we don’t want to overestimate the difference of the restricted means between two treatment groups and make a new treatment appear more effective than it actually is.

One we have the \( Z(t) \) statistic for the entire dataset, a hockey stick regression model is used to fit two regression lines to one sample with a point joining the two lines. In our case we are fitting the two lines to the plot of \( |Z(t)| \) and we are interested in using the point where the two curves join as our \( \hat{\tau} \). The SAS® NLIN Procedure is used to fit the model

\[
|Z(t)| = \begin{cases} 
\beta_0 + \beta_1 t, & t < \delta \\
\beta_0 + \beta_1 t + \beta_2 (t - \delta), & t \geq \delta 
\end{cases}
\]  

(4.2.6)

where \( \delta \) is an estimated point where the two lines join. We then set \( \hat{\tau} = \delta \). Note that because the \( |Z(t)| \) statistics are not independent we do not wish to evaluate the SE of the estimate for \( \delta \) based on the output from the parameter estimates in NLIN. Instead we will use a nonparametric bootstrap in order to obtain an estimate of the SE for \( \hat{\tau} \).
Figure 0.7 A hockey stick regression model is fit and plotted over the Z(t) statistics. The vertical reference line is at the point where $\delta$ was estimated using an NLIN Procedure.

4.2.3 Estimated Hazards Method

As we illustrated in Chapter 1, the hazard, cumulative hazard and survival functions are all related. Therefore if we have an estimate of a survival function, we can estimate the cumulative hazard and the hazard functions. In the case of the K-M survival estimates, we can take the negative log of these estimates to get the cumulative hazards estimates and then transform those points into hazard estimates. So taking $\hat{S}(t)$, the K-M estimator, we can write

$$\hat{H}(t_i) = -\log(\hat{S}(t_i)).$$

(4.2.7)

Then since we don’t have a function to derivate to find the hazard estimate, we calculate the slope of the cumulative hazard between ordered event times giving us

$$\hat{h}(t_i) = \frac{\hat{H}(t_{i+1}) - \hat{H}(t_i)}{t_{i+1} - t_i}$$

(4.2.8)
where there are \( i = 1, \ldots, D \) ordered event times. Based on the premise that the treatment differences only exist up to a certain point, we then take the hazard estimates and fit a nonlinear model that allows for two different hazards up to a certain time, \( \delta \), after which the hazards for the two groups are the same. It is this point, \( \delta \), which we will use as our estimate of \( \tau \) for the hazard estimates model. In order to find the estimate we create a list of possible time points, \( \delta \), at appropriate small intervals, and fit a nonlinear model for each possible point. Then the \( \delta \) which corresponds to the model with the smallest sums of squared error term, or SSE, is the cut-point we chose (See Figure 0.8). We will then use bootstrap methods to obtain the SE of \( \delta \).

The constant hazard piecewise model fit using the NLIN procedure is as follows

\[
x = \begin{cases} 
0 & \text{if Trt}=0 \\
1 & \text{if Trt}=1 
\end{cases}
\]

(4.2.9)

\[
h(t) = \begin{cases} 
\beta_1 + \beta_2 x, & t < \delta \\
\beta_3 & t \geq \delta 
\end{cases}
\]

(4.2.10)

with the indicator variable \( x \) allowing for two hazards to be modeled in the time prior to \( \delta \). Figure 0.9 illustrates an example of this model.
Figure 0.8 Possible time values for $\delta$ were fit from 5 to 25 by 0.1 and then NLIN models were fit for each value. The plot shows the SSE for the NLIN model that corresponds to the value for $\delta$. Note that the smallest values occur around time 10.
In this chapter we have illustrated the three methods we are proposing to find a data driven cut-point for the calculation of the restricted mean in the situation where we want to determine the true mean difference in treatments. All of these proposed methods are easily implemented using standard statistical software. In the next chapter we will evaluate these methods through a series of simulation studies to see how the methods perform and we will report the results of the simulation studies.

Figure 0.9 The NLIN Procedure model fit of the estimated hazards from time 0 to $\delta$, and $\delta$ to 30.
Chapter 5  Simulations and Results

5.1 Simulation Set-Up

Based on our mathematical definition we set up our simulations so that before time $\tau_1$ we have two different hazard functions at play in the two groups. After $\tau_1$ we have the same hazard function in the two groups, and after $\tau_2$ the hazard functions continue to be equal, and could likely be zero if all observations beyond that point are censored. While we will be working in a non-parametric framework for our methods in estimating the cut-point, it is advantageous to simulate data from known distributions so that we may compare our findings with what we know to be true. In this framework we are working with a piecewise exponential distribution. In equation form, for group 1

$$h_1(t) = \begin{cases} \lambda, & 0 \leq t < \tau_2 \\ \lambda_f, & t \geq \tau_2 \end{cases} \quad (5.1.1)$$

$$S_1(t) = \begin{cases} \exp(-\lambda t), & 0 \leq t < \tau_2 \\ \exp(-\lambda \tau_2 - \lambda_f (t - \tau_2)), & t \geq \tau_2 \end{cases} \quad (5.1.2)$$

And for group 2 we have

$$h_2(t) = \begin{cases} \lambda^*, & 0 \leq t < \tau_1 \\ \lambda, & \tau_1 \leq t < \tau_2 \\ \lambda_f, & t \geq \tau_2 \end{cases} \quad (5.1.3)$$
\[ S_2(t) = \begin{cases} \exp(-\lambda^* t) & 0 \leq t < \tau_1 \\ \exp(-\lambda^* \tau_1 - \lambda (t - \tau_1)) & \tau_1 \leq t < \tau_2 \\ \exp(-\lambda^* \tau_1 - \lambda (\tau_2 - \tau_1) - \lambda^*(t - \tau_2)) & t \geq \tau_2 \end{cases} \]  

(5.1.4)

The equations for the survivor functions result from the relationship

\[ S(t) = \exp(-\int_0^t h(u)du). \]  

(5.1.5)

Now we can calculate the restricted mean based on (5.1.4)

\[
\mu_1(\tau) = \int_0^{\tau} \exp(-\lambda t)dt = \frac{-\exp(-\lambda \tau)}{\lambda} \bigg|_0^{\tau} = \frac{1 - \exp(-\lambda \tau_1)}{\lambda} \\
\mu_2(\tau) = \int_0^{\tau} \exp(-\lambda^* t)dt = \frac{-\exp(-\lambda^* t)}{\lambda^*} \bigg|_0^{\tau} = \frac{1 - \exp(-\lambda^* \tau_1)}{\lambda^*} 
\]  

(5.1.6)

So by setting values for \( \lambda, \lambda^* \), and \( \tau_1 \) we can calculate the true restricted mean and compare it to our model.

We wish to simulate data such that the control group, \( \text{trt}=0 \), has a constant hazard, \( \lambda = 0.20 \), which corresponds to an overall mean survival of 5.0 years. We will set \( \tau_1 = 10 \), thus for the restricted mean we have \( \mu_1(\tau) = 4.323 \), based on equation (5.1.6). For the treatment group, or \( \text{trt}=1 \), we will look at three different \( \lambda^* \)'s, where \( \lambda^* = 0.15, 0.10, \) and \( 0.05 \), which correspond to overall mean survivals of 5.856, 6.998, and 8.546 respectively, in the piecewise framework, see Table 5.1.

**Table 5.1 Mean, time, hazard, and restricted mean scenarios for the treatment group.**

<table>
<thead>
<tr>
<th>Overall ( \mu )</th>
<th>( \tau_1 )</th>
<th>( \lambda^* )</th>
<th>( \mu_2(\tau) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.856</td>
<td>10</td>
<td>0.15</td>
<td>5.179</td>
</tr>
<tr>
<td>6.998</td>
<td>10</td>
<td>0.10</td>
<td>6.321</td>
</tr>
<tr>
<td>8.546</td>
<td>10</td>
<td>0.05</td>
<td>7.869</td>
</tr>
</tbody>
</table>
We then wish to examine different censoring patterns. Figure 5.1 shows a case when there is no censoring present in the model for the cumulative hazard function. Figure 5.2 illustrates when censoring occurs 25% of the time in both groups. Figure 5.3 illustrates the same censoring percentages as Figure 5.2 with long term right censoring included after time 20. Figure 5.4 illustrates that the cumulative hazard plots are not significantly different after time 10 ($\tau_1$).

**Figure 5.1** The cumulative hazard function for two groups, generated from a piecewise exponential distribution where $\lambda=0.10$ and $\lambda^*=0.05$, and $\tau_1=10$
Figure 5.2 The cumulative hazard function for two groups, generated from a piecewise exponential distribution where $\lambda=0.10$ and $\lambda^*=0.05$, and $\tau_1=10$. Censoring has been built into the data.

Figure 5.3 The cumulative hazard function for two groups, generated from a piecewise exponential distribution where $\lambda=0.10$ and $\lambda^*=0.05$, and $\tau_1=10$. Censoring has been built into the data, and all observations after time 20 have been censored, $\tau_2 = 20$. 
Figure 5.4 Looking at the cumulative hazard function when the events prior to time 10, $\tau_1$, have been removed.

Note that there is not a significant difference, p-value=0.2445, according to the log-rank test, for the remaining event times in Figure 5.4.

Simulations were run for $n_j = 100$, 250, and 500 per group, and included the various hazards outlined in Table 5.1, and various censoring schemes illustrated in Figure 5.1 and Figure 5.2. For the scenarios when censoring is present, censored observations were generated using a univariate random number generator and if Ranuni<0.25 then the observation was censored. The number of simulations run for each of the scenarios is 1,000.

We will not be concerned with the case when there is long right tail censoring, as we are concerned with what is happening in the data before that point. We recommend if data has a long right tail and the restricted means are being compared to determine treatment benefit, and the proportional hazards assumption is upheld, that the mean be restricted to the minimum LET between the two groups.
5.2 Simulation Results

The results of the 27 different scenarios with no censoring, and then with 25% censoring incorporated, are presented in the next pages. Tables 5.2 through 5.19 give the estimates found from the three different methods, as well as the bootstrap variance estimates, 90th percentile intervals, and MSE and bias values. Plots of MSE versus squared bias are examined to determine what scenarios provide estimates with the smallest MSE and bias.

An estimate of the variance of the estimate for \( \tau \) was calculated using bootstrap sampling techniques. In order to get an estimate of the variance, \( B=100 \) bootstrap replicates were performed on each simulated dataset. The bootstrap variance estimate is calculated as

\[
    V_{boot} = V(\hat{\tau}^{(b)}) = \frac{1}{100-1} \sum_{b=1}^{100} (\hat{\tau}^{(b)} - \hat{\tau}_{boot})^2,
\]

(5.2.1)

And

\[
    \hat{\tau}_{boot} = \frac{\sum_{b=1}^{100} \hat{\tau}^{(b)}}{100}.
\]

(5.2.2)

Efron (1993, p.52) states that between 50 to 200 replicates are sufficient for calculating a variance estimate for the estimator, thus our chosen replicate number of 100.

The following SAS® code was used to create bootstrap sample estimates. Sampling was done with replacement using the same sample size as the original data.
We also use these bootstrap estimates to calculate a 90th percentile interval to provide an idea of the spread of our estimates. The percentiles are calculated simply by ordering the bootstrap estimates and identifying the \( \hat{r}^{(b)} \) where 5% of the estimates are below, and another where 95% of the estimates are below. It is recognized that 100 replicates is not sufficient to get a reliable estimate of the percentile intervals as Efron recommends at least 1,000 replicates to obtain a representation of the distribution of the estimates. The percentile estimates are given with that limitation recognized, since they still provide descriptive value.

In addition to bootstrapping to get the estimated variance of the estimate \( \hat{\theta} \), we examine the mean squared error, MSE, and squared bias of \( \hat{\theta} \) graphically to determine under which scenario our methods are best behaved. A consistent estimator is one which converges to the true value as \( n \) approaches infinity. (Casella and Berger, 2002) This definition can be expressed as

\[
\lim_{n \to \infty} P_{\theta}(|\hat{\theta} - \theta| < c) = 1
\]  
(5.2.3)
where \( c \) is an arbitrarily small constant. The definition of MSE is the squared difference between the estimate and the true value of the parameter, or

\[
\text{MSE} = E((\hat{\tau} - \tau)^2) = \text{Var}(\hat{\tau}) + \text{Bias}(\hat{\tau})^2
\]

where the bias is defined as the difference between the estimate and the true value of the parameter. It is desirable to see a trend develop in the simulations where the sample size increases, the bias decreases, as this indicates a consistent estimator.

**5.2.1 Simulation Results for a Cut-Point of 10**

The first set of results examine an estimator for \( \tau_1 = 10 \). In these simulations this time point occurs roughly one third of the way to midway through the total on study time.

**Table 5.2 Simulation Results for the Maximum Distance Method with no censoring \( \tau=10 \)**

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>%Significant</th>
<th>( \hat{\tau} )</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>8.071</td>
<td>1.388</td>
<td>5.660</td>
<td>9.748</td>
<td>-1.929</td>
<td>5.966</td>
</tr>
<tr>
<td>200</td>
<td>0.10</td>
<td>96.3</td>
<td>6.383</td>
<td>1.913</td>
<td>3.495</td>
<td>9.288</td>
<td>-3.617</td>
<td>17.227</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>39.4</td>
<td>5.179</td>
<td>2.493</td>
<td>1.714</td>
<td>9.232</td>
<td>-4.821</td>
<td>30.213</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>8.566</td>
<td>1.087</td>
<td>6.500</td>
<td>9.710</td>
<td>-1.434</td>
<td>3.339</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>100.0</td>
<td>6.793</td>
<td>1.562</td>
<td>4.279</td>
<td>9.023</td>
<td>-3.207</td>
<td>13.132</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>75.2</td>
<td>5.518</td>
<td>2.129</td>
<td>2.375</td>
<td>8.775</td>
<td>-4.482</td>
<td>24.759</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>8.741</td>
<td>0.898</td>
<td>7.066</td>
<td>9.737</td>
<td>-1.259</td>
<td>2.543</td>
</tr>
<tr>
<td>1000</td>
<td>0.10</td>
<td>100.0</td>
<td>6.817</td>
<td>1.347</td>
<td>4.760</td>
<td>8.887</td>
<td>-3.183</td>
<td>11.994</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>95.9</td>
<td>5.704</td>
<td>1.829</td>
<td>2.947</td>
<td>8.543</td>
<td>-4.296</td>
<td>22.287</td>
</tr>
</tbody>
</table>
Table 5.3 Results from the Maximum Distance Method with 25% censoring $\tau=10$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>% Significant</th>
<th>$\hat{\tau}$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>8.364</td>
<td>1.118</td>
<td>6.502</td>
<td>9.775</td>
<td>-1.636</td>
<td>4.391</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>98.5</td>
<td>6.671</td>
<td>1.719</td>
<td>4.055</td>
<td>9.194</td>
<td>-3.329</td>
<td>14.776</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>46.1</td>
<td>5.632</td>
<td>2.267</td>
<td>2.912</td>
<td>9.669</td>
<td>-4.368</td>
<td>25.058</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>8.637</td>
<td>0.974</td>
<td>6.836</td>
<td>9.714</td>
<td>-1.363</td>
<td>3.062</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>6.824</td>
<td>1.492</td>
<td>4.492</td>
<td>9.028</td>
<td>-3.176</td>
<td>12.679</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>83.5</td>
<td>5.768</td>
<td>1.900</td>
<td>3.077</td>
<td>8.774</td>
<td>-4.232</td>
<td>22.292</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>8.785</td>
<td>0.858</td>
<td>7.163</td>
<td>9.737</td>
<td>-1.215</td>
<td>2.383</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>6.780</td>
<td>1.314</td>
<td>4.839</td>
<td>8.875</td>
<td>-3.220</td>
<td>12.215</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>98.1</td>
<td>5.668</td>
<td>1.693</td>
<td>3.307</td>
<td>8.460</td>
<td>-4.332</td>
<td>22.134</td>
</tr>
</tbody>
</table>

The numbers given in Table 5.2 reveal that the maximum distance method underestimates the true value of $\tau$, here set to be 10. Looking at the combinations of treatment hazards during the first time period and total sample size, N, as the hazard remains constant and sample size increases, the bias does decrease. The column labeled “% Significant” indicates the total percentage of datasets per scenario which had a significant difference between the two groups according to the Log Rank test, $\alpha=0.05$. Figure 5.5 shows the relationship between squared bias and MSE, taking the hazards and sample sizes into account. The increasing sizes of the dots correspond to the increasing sample sizes. Note that the hazards form small clusters, and that the smallest hazard corresponds to the smallest bias and MSE. This indicates that the estimator performs best when the early differences in the hazards between the two groups is the largest (control hazard=0.20, and treatment hazard=0.05).

Table 5.3 shows the maximum distance method estimates when 25% censoring is incorporated into the data. When censoring is present the method does not seem to perform differently. The estimates are still below the true value of $\tau$. The relationship between MSE and squared bias is not quite as straightforward as in the non-censoring case, see Figure 5.6. However,
the smallest hazard still corresponds to the smallest squared bias and MSE. The clustering of the points with the larger treatment hazards (0.10, and 0.15) seems to indicate that at a certain point, increasing the sample size does not affect bias.

Figure 5.5 MSE versus Squared Bias for the Maximum Distance Method when \( \tau=10 \).
Figure 5.6 MSE versus Squared Bias for the Maximum Distance Method when $\tau=10$ with 25% censoring

Table 5.4 Results from the Renyi Method with no censoring $\tau=10$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>% Significant</th>
<th>$\tau^*$</th>
<th>SD*</th>
<th>L $90%$*</th>
<th>U $90%$*</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>7.359</td>
<td>1.267</td>
<td>5.186</td>
<td>9.220</td>
<td>-2.641</td>
<td>8.937</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>96.3</td>
<td>6.570</td>
<td>1.979</td>
<td>3.721</td>
<td>9.945</td>
<td>-3.430</td>
<td>17.027</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>39.4</td>
<td>5.263</td>
<td>2.700</td>
<td>2.278</td>
<td>10.641</td>
<td>-4.737</td>
<td>31.024</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>7.545</td>
<td>0.847</td>
<td>6.115</td>
<td>8.853</td>
<td>-2.455</td>
<td>6.790</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>6.975</td>
<td>1.349</td>
<td>4.799</td>
<td>9.109</td>
<td>-3.025</td>
<td>11.317</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>75.2</td>
<td>5.934</td>
<td>2.282</td>
<td>2.929</td>
<td>10.092</td>
<td>-4.066</td>
<td>23.517</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>7.607</td>
<td>0.583</td>
<td>6.672</td>
<td>8.576</td>
<td>-2.393</td>
<td>6.092</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>7.125</td>
<td>1.035</td>
<td>5.407</td>
<td>8.729</td>
<td>-2.875</td>
<td>9.388</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>95.9</td>
<td>6.264</td>
<td>1.975</td>
<td>3.514</td>
<td>9.800</td>
<td>-3.736</td>
<td>18.508</td>
</tr>
</tbody>
</table>

*Estimates here are from 200 simulated datasets with $B=100$ replicates.
Table 5.5 Results from the Renyi Method with 25% censoring $\tau=10$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>%Significant</th>
<th>$\hat{\tau}$</th>
<th>SD*</th>
<th>L 90%*</th>
<th>U 90%*</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>7.461</td>
<td>1.042</td>
<td>5.709</td>
<td>9.014</td>
<td>-2.539</td>
<td>7.790</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>98.5</td>
<td>6.074</td>
<td>1.662</td>
<td>4.000</td>
<td>9.180</td>
<td>-3.926</td>
<td>19.859</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>46.1</td>
<td>5.146</td>
<td>2.320</td>
<td>3.177</td>
<td>10.340</td>
<td>-4.854</td>
<td>30.254</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>7.668</td>
<td>0.721</td>
<td>6.421</td>
<td>8.742</td>
<td>-2.332</td>
<td>5.976</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>6.750</td>
<td>1.321</td>
<td>4.698</td>
<td>8.854</td>
<td>-3.250</td>
<td>13.016</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>83.5</td>
<td>5.719</td>
<td>2.078</td>
<td>3.562</td>
<td>10.048</td>
<td>-4.281</td>
<td>24.477</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>7.734</td>
<td>0.491</td>
<td>6.904</td>
<td>8.509</td>
<td>-2.266</td>
<td>5.404</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>7.014</td>
<td>1.105</td>
<td>5.138</td>
<td>8.661</td>
<td>-2.986</td>
<td>10.231</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>98.1</td>
<td>6.342</td>
<td>1.900</td>
<td>3.899</td>
<td>9.861</td>
<td>-3.658</td>
<td>18.269</td>
</tr>
</tbody>
</table>

*Estimates here are from 200 simulated datasets with $B=100$ replicates.

Table 5.4 provides summary of the simulations for the Renyi Method when the cut-point is set to 10. Due to the large amount of computing required to obtain a Renyi Method estimate for $\tau$, the bootstrap estimates of the standard deviation and 90th percentile interval are from a subset of the 1,000 simulated datasets. Like the Maximum distance method, the Renyi method underestimates the true value of $\tau$. Again, examining a plot of the MSE versus squared bias, Figure 5.7, reveals that the estimator has the smallest bias and MSE when the sample size is N=1000 and the treatment hazard is 0.05 during the first time period.

The Renyi method doesn’t seem to perform any differently when 25% of the observations are censored, Table 5.5. This intuitively makes sense since the Renyi statistic only looks at event times, and by censoring twenty five percent of the data we are eliminating some event times, but the rest of the data remains the same. The method still underestimates the true value of $\tau$. The relationship between MSE and squared bias also appears to be similar as is seen in Figure 5.8.
Figure 5.7 MSE versus Squared Bias for the Renyi Method when $\tau=10$

Figure 5.8 MSE versus Squared Bias for the Renyi Method with 25% censoring when $\tau=10$
Table 5.6 Results from the Hazard Method with no censoring $\tau=10$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt Hazard</th>
<th>% Significant</th>
<th>$\hat{\tau}$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>9.808</td>
<td>2.090</td>
<td>6.259</td>
<td>12.605</td>
<td>-0.192</td>
<td>1.977</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>96.3</td>
<td>10.760</td>
<td>3.862</td>
<td>4.484</td>
<td>16.575</td>
<td>0.760</td>
<td>16.057</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>39.4</td>
<td>11.420</td>
<td>4.115</td>
<td>3.652</td>
<td>16.029</td>
<td>1.420</td>
<td>22.861</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>9.903</td>
<td>0.651</td>
<td>8.686</td>
<td>10.463</td>
<td>-0.097</td>
<td>0.193</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>10.228</td>
<td>3.150</td>
<td>5.930</td>
<td>15.902</td>
<td>0.228</td>
<td>8.849</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>75.2</td>
<td>12.426</td>
<td>4.975</td>
<td>4.098</td>
<td>19.114</td>
<td>2.426</td>
<td>36.802</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>9.952</td>
<td>0.265</td>
<td>9.473</td>
<td>10.201</td>
<td>-0.048</td>
<td>0.039</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>9.966</td>
<td>1.665</td>
<td>7.513</td>
<td>12.279</td>
<td>-0.034</td>
<td>1.698</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>95.9</td>
<td>12.392</td>
<td>5.004</td>
<td>4.600</td>
<td>19.918</td>
<td>2.392</td>
<td>39.338</td>
</tr>
</tbody>
</table>

Table 5.7 Results from the Hazard Method with 25% censoring $\tau=10$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt Hazard</th>
<th>%Significant</th>
<th>$\hat{\tau}$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>9.808</td>
<td>2.362</td>
<td>6.802</td>
<td>14.064</td>
<td>-0.192</td>
<td>1.977</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>98.5</td>
<td>10.760</td>
<td>3.905</td>
<td>4.241</td>
<td>16.395</td>
<td>0.760</td>
<td>16.057</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>46.1</td>
<td>11.420</td>
<td>4.148</td>
<td>3.514</td>
<td>15.910</td>
<td>1.420</td>
<td>22.861</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>9.903</td>
<td>0.866</td>
<td>8.683</td>
<td>10.818</td>
<td>-0.097</td>
<td>0.193</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>10.228</td>
<td>3.365</td>
<td>5.618</td>
<td>16.357</td>
<td>0.228</td>
<td>8.849</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>83.5</td>
<td>12.426</td>
<td>5.031</td>
<td>3.950</td>
<td>19.021</td>
<td>2.426</td>
<td>36.802</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>9.952</td>
<td>0.278</td>
<td>9.483</td>
<td>10.229</td>
<td>-0.048</td>
<td>0.039</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>9.966</td>
<td>1.851</td>
<td>7.456</td>
<td>12.773</td>
<td>-0.034</td>
<td>1.698</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>98.1</td>
<td>12.392</td>
<td>5.067</td>
<td>4.444</td>
<td>19.921</td>
<td>2.392</td>
<td>39.338</td>
</tr>
</tbody>
</table>

In Table 5.6 the results from the Hazard Method are presented when the cut point is defined at time 10. Here looking at the same treatment hazards as the sample size increases we see the bias decreases except for the case when the treatment hazard is 0.15. This is unusual behavior, and no explanation can be made at the moment, thus this should be looked into further in the future. Other than the strange trend in the bias, the hazard estimate for the cut point behaves rather nicely. These set of estimates are the closest to the true $\tau$, and the 90th percentile intervals all cover the true value, although some are rather wide. Again we note that the
percentile intervals are provided for descriptive purposes only, as the number of bootstrap replicates is not large enough to give reliable estimates of the distribution of $\hat{\tau}$. Figure 5.9 shows the relationship between MSE and squared bias, and again the smaller treatment hazard yields and estimate with smallest MSE and bias squared. In fact, all of the points for the different sample sizes for $h=0.05$ are so close together they appear as one dot near the origin in the plot.

Table 5.7 displays the results from the hazard method with 25% of the observations censored. Oddly enough, the averages of the estimates, which is expressed in the tables, are the same for the scenarios with no censoring and with 25% censoring. Examining the estimates from a single simulated dataset, one without censoring and the other with censoring but otherwise the same data, does not always provide the same estimate. So the fact that the averages come out to be identical is likely because the estimate is calculated to the nearest tenth decimal place. The bootstrap estimates of the standard deviation and 90th percentile interval are not identical between the two censoring schemes. Of the estimates available it looks like the standard deviation estimates are slightly higher for the data with censoring. Since the estimates are the same the bias and MSE calculations are also the same for the tables, so a second plot is not necessary.
Figure 5.9 MSE versus Squared Bias for the Hazard Method when $\tau=10$
5.2.2 Simulations Results for a Cut-Point of 5

Now an earlier time point is set, $\tau_1 = 5$, to see how the methods perform when the cut point occurs sooner, which means that fewer events have occurred.

Table 5.8 Results from the Maximum Distance Method with no censoring $\tau = 5$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>% Significant</th>
<th>$\hat{\tau}$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>4.839</td>
<td>0.636</td>
<td>3.880</td>
<td>5.767</td>
<td>-0.161</td>
<td>0.268</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>99.8</td>
<td>4.524</td>
<td>1.299</td>
<td>2.759</td>
<td>6.649</td>
<td>-0.476</td>
<td>1.385</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>46.6</td>
<td>4.149</td>
<td>2.224</td>
<td>1.538</td>
<td>8.371</td>
<td>-0.851</td>
<td>4.650</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>4.959</td>
<td>0.326</td>
<td>4.356</td>
<td>5.313</td>
<td>-0.041</td>
<td>0.062</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>4.729</td>
<td>0.753</td>
<td>3.381</td>
<td>5.650</td>
<td>-0.271</td>
<td>0.482</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>87.8</td>
<td>4.303</td>
<td>1.733</td>
<td>2.010</td>
<td>7.275</td>
<td>-0.697</td>
<td>2.636</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>4.979</td>
<td>0.185</td>
<td>4.611</td>
<td>5.160</td>
<td>-0.021</td>
<td>0.015</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>4.802</td>
<td>0.515</td>
<td>3.798</td>
<td>5.329</td>
<td>-0.198</td>
<td>0.217</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.4</td>
<td>4.397</td>
<td>1.258</td>
<td>2.450</td>
<td>6.223</td>
<td>-0.603</td>
<td>1.569</td>
</tr>
</tbody>
</table>

Table 5.9 Results from the Maximum Distance Method with 25% censoring $\tau = 5$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>% Significant</th>
<th>$\hat{\tau}$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>5.138</td>
<td>0.436</td>
<td>4.836</td>
<td>6.031</td>
<td>0.138</td>
<td>0.129</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>100.0</td>
<td>4.679</td>
<td>1.005</td>
<td>3.464</td>
<td>6.416</td>
<td>-0.321</td>
<td>0.879</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>58.9</td>
<td>4.667</td>
<td>2.038</td>
<td>2.732</td>
<td>8.833</td>
<td>-0.333</td>
<td>4.017</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>5.059</td>
<td>0.196</td>
<td>4.928</td>
<td>5.468</td>
<td>0.059</td>
<td>0.024</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>4.712</td>
<td>0.644</td>
<td>3.657</td>
<td>5.603</td>
<td>-0.288</td>
<td>0.400</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>93.9</td>
<td>4.371</td>
<td>1.455</td>
<td>2.732</td>
<td>7.099</td>
<td>-0.630</td>
<td>2.004</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>5.032</td>
<td>0.100</td>
<td>4.964</td>
<td>5.240</td>
<td>0.032</td>
<td>0.006</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>4.770</td>
<td>0.470</td>
<td>3.900</td>
<td>5.317</td>
<td>-0.230</td>
<td>0.218</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.9</td>
<td>4.372</td>
<td>1.047</td>
<td>2.869</td>
<td>6.022</td>
<td>-0.628</td>
<td>1.322</td>
</tr>
</tbody>
</table>

The results from Table 5.8 show the maximum distance method results when trying to estimate a cut-point equal to 5. This method works quite well with the early time point. The bias is small, and while all the estimates are slightly below the true value, the bootstrap 90th
percentile interval easily covers the cut-point of 5. Again the estimate appears to be consistent as the bias decreases with an increase in sample size, looking at the same hazards. Figure 5.10 MSE versus Squared Bias for the Maximum Distance Method when τ= 5, no censoring shows the relationship between the MSE and squared bias. As was the case when the cut-point was set to 10, the smallest treatment hazard and largest sample size correspond to the smallest squared bias and MSE.

Table 5.9 examines the Maximum Distance method when 25% of the observations are censored, and the cut-point is set to 5. There is not much change in the performance of the estimator as compared to when there is no censoring. Figure 5.11 also indicates the similar trend between the two censoring schemes. However the estimator does not perform as well with censoring present when the treatment hazard is larger; or the difference between treatment and control hazards is small.
Figure 5.10 MSE versus Squared Bias for the Maximum Distance Method when $\tau = 5$, no censoring
Figure 5.11 MSE versus Squared Bias for the Maximum Distance Method when $\tau= 5$, with 25% censoring

Table 5.10 Results from the Renyi Method with no censoring $\tau= 5$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>Significant</th>
<th>$\tau^*$</th>
<th>SD*</th>
<th>90%* L</th>
<th>90%* U</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>4.582</td>
<td>0.828</td>
<td>3.394</td>
<td>6.103</td>
<td>-0.418</td>
<td>0.825</td>
</tr>
<tr>
<td>200</td>
<td>0.10</td>
<td>99.8</td>
<td>4.506</td>
<td>1.429</td>
<td>2.789</td>
<td>7.244</td>
<td>-0.494</td>
<td>2.189</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>46.6</td>
<td>4.395</td>
<td>2.402</td>
<td>2.096</td>
<td>9.719</td>
<td>-0.605</td>
<td>5.681</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>4.584</td>
<td>0.471</td>
<td>3.869</td>
<td>5.400</td>
<td>-0.416</td>
<td>0.395</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>100.0</td>
<td>4.528</td>
<td>0.794</td>
<td>3.376</td>
<td>5.920</td>
<td>-0.472</td>
<td>0.841</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>87.8</td>
<td>4.439</td>
<td>1.790</td>
<td>2.453</td>
<td>8.140</td>
<td>-0.561</td>
<td>3.534</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>4.573</td>
<td>0.331</td>
<td>4.092</td>
<td>5.177</td>
<td>-0.427</td>
<td>0.280</td>
</tr>
<tr>
<td>1000</td>
<td>0.10</td>
<td>100.0</td>
<td>4.495</td>
<td>0.551</td>
<td>3.668</td>
<td>5.465</td>
<td>-0.505</td>
<td>0.521</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.4</td>
<td>4.389</td>
<td>1.373</td>
<td>2.772</td>
<td>7.059</td>
<td>-0.611</td>
<td>1.962</td>
</tr>
</tbody>
</table>

*Estimates here are from 200 simulated datasets with $B=100$ replicates.
Table 5.11 Results from the Renyi Method with 25% censoring $\tau = 5$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>Significant</th>
<th>$\tau^\wedge$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>4.849</td>
<td>0.632</td>
<td>4.029</td>
<td>6.045</td>
<td>-0.151</td>
<td>0.442</td>
</tr>
<tr>
<td>200</td>
<td>0.10</td>
<td>100.0</td>
<td>4.201</td>
<td>1.183</td>
<td>3.140</td>
<td>6.787</td>
<td>-0.799</td>
<td>1.736</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>58.9</td>
<td>4.303</td>
<td>2.086</td>
<td>2.957</td>
<td>9.492</td>
<td>-0.697</td>
<td>4.269</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>4.846</td>
<td>0.376</td>
<td>4.324</td>
<td>5.529</td>
<td>-0.154</td>
<td>0.160</td>
</tr>
<tr>
<td>500</td>
<td>0.10</td>
<td>100.0</td>
<td>4.254</td>
<td>0.654</td>
<td>3.380</td>
<td>5.451</td>
<td>-0.746</td>
<td>0.997</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>93.9</td>
<td>4.298</td>
<td>1.675</td>
<td>2.927</td>
<td>8.185</td>
<td>-0.702</td>
<td>2.932</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>4.825</td>
<td>0.268</td>
<td>4.461</td>
<td>5.332</td>
<td>-0.175</td>
<td>0.089</td>
</tr>
<tr>
<td>1000</td>
<td>0.10</td>
<td>100.0</td>
<td>4.223</td>
<td>0.483</td>
<td>3.549</td>
<td>5.090</td>
<td>-0.777</td>
<td>0.821</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.9</td>
<td>4.272</td>
<td>1.444</td>
<td>3.003</td>
<td>7.427</td>
<td>-0.728</td>
<td>1.957</td>
</tr>
</tbody>
</table>

*Estimates here are from 200 simulated datasets with $B=100$ replicates.

In Table 5.10 the Renyi Method estimates the earlier cut-point of 5 reasonably well. While the estimates are all below the true value of $\tau$, the bias is small. The bias is roughly constant as the sample size increases. Figure 5.12 shows the relationship between MSE and squared bias for these scenarios. In this case it seems that a sample size of 500 corresponds to the smallest bias in each hazard group, and the smallest hazard results in the best performance of the estimator.

The results in Table 5.11 do not indicate that 25% censoring has much of an effect on the performance of the Renyi Method when the cut-point is set to 5. Figure 5.13 continues to indicate the estimator does well when the treatment hazard is 0.05.
Figure 5.12 MSE versus Squared Bias for the Renyi Method when $\tau = 5$, with no censoring.
**Figure 5.13** MSE versus Squared Bias for the Renyi Method when $\tau=5$, with 25% censoring

**Table 5.12** Results from the Hazard Method with no censoring $\tau=5$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt Hazard</th>
<th>% Significant</th>
<th>$\hat{r}$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>5.062</td>
<td>2.043</td>
<td>3.658</td>
<td>9.404</td>
<td>0.062</td>
<td>2.881</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>99.8</td>
<td>6.467</td>
<td>3.529</td>
<td>3.528</td>
<td>14.035</td>
<td>1.467</td>
<td>21.364</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>46.6</td>
<td>8.653</td>
<td>3.857</td>
<td>3.501</td>
<td>15.060</td>
<td>3.653</td>
<td>42.931</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>4.950</td>
<td>0.572</td>
<td>4.211</td>
<td>5.561</td>
<td>-0.050</td>
<td>0.088</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>5.323</td>
<td>2.946</td>
<td>3.802</td>
<td>12.144</td>
<td>0.323</td>
<td>6.344</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>87.8</td>
<td>8.870</td>
<td>4.923</td>
<td>3.614</td>
<td>18.021</td>
<td>3.870</td>
<td>54.262</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>4.960</td>
<td>0.203</td>
<td>4.601</td>
<td>5.170</td>
<td>-0.040</td>
<td>0.027</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>5.030</td>
<td>1.283</td>
<td>4.164</td>
<td>7.136</td>
<td>0.030</td>
<td>1.249</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.4</td>
<td>7.629</td>
<td>5.128</td>
<td>3.698</td>
<td>18.554</td>
<td>2.629</td>
<td>42.418</td>
</tr>
</tbody>
</table>
Table 5.13 Results from the Hazard Method with 25% censoring \( \tau = 5 \)

<table>
<thead>
<tr>
<th>N</th>
<th>Trt</th>
<th>Hazard</th>
<th>% Significant</th>
<th>( \hat{\tau} )</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>5.062</td>
<td>4.080</td>
<td>4.227</td>
<td>16.183</td>
<td>0.062</td>
<td>2.881</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>100.0</td>
<td>6.467</td>
<td>3.709</td>
<td>3.417</td>
<td>14.422</td>
<td>1.467</td>
<td>21.364</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>58.9</td>
<td>8.653</td>
<td>3.877</td>
<td>3.366</td>
<td>14.925</td>
<td>3.653</td>
<td>42.931</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>4.950</td>
<td>5.221</td>
<td>4.558</td>
<td>19.456</td>
<td>-0.050</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>5.323</td>
<td>3.961</td>
<td>3.652</td>
<td>15.119</td>
<td>0.323</td>
<td>6.344</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>93.9</td>
<td>8.870</td>
<td>5.015</td>
<td>3.500</td>
<td>18.140</td>
<td>3.870</td>
<td>54.262</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>4.960</td>
<td>5.764</td>
<td>4.715</td>
<td>21.054</td>
<td>-0.040</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>5.030</td>
<td>2.846</td>
<td>4.069</td>
<td>11.606</td>
<td>0.030</td>
<td>1.249</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.9</td>
<td>7.629</td>
<td>5.387</td>
<td>3.655</td>
<td>19.105</td>
<td>2.629</td>
<td>42.418</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.12 and Table 5.13 give the results of the Hazard Method with no censoring and 25% censoring, respectively. Again the averages of the estimates for \( \tau \) are the same for both censoring schemes. The Hazard Method estimate performs well when the treatment hazard is 0.05.
This section provides results to see how often the methods estimate a cut-point when one does not exist. In other words, when the hazard ratios between the two treatments remain constant over the course of the whole study period, do our methods still find a time that is prior to the end of the study?
Table 5.14 Results from the Maximum Distance Method with no censoring $\tau = 40$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>%Significant</th>
<th>$\hat{\epsilon}$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>8.854</td>
<td>2.242</td>
<td>5.937</td>
<td>12.724</td>
<td>-21.146</td>
<td>452.180</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>100.0</td>
<td>6.657</td>
<td>2.336</td>
<td>3.576</td>
<td>10.682</td>
<td>-23.343</td>
<td>550.883</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>49.2</td>
<td>5.354</td>
<td>2.654</td>
<td>1.722</td>
<td>9.795</td>
<td>-24.646</td>
<td>615.613</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>9.098</td>
<td>1.701</td>
<td>6.720</td>
<td>11.905</td>
<td>-20.902</td>
<td>439.495</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>6.919</td>
<td>1.844</td>
<td>4.322</td>
<td>9.939</td>
<td>-23.081</td>
<td>536.264</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>89.8</td>
<td>5.354</td>
<td>2.654</td>
<td>1.722</td>
<td>9.795</td>
<td>-24.646</td>
<td>615.613</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>9.168</td>
<td>1.386</td>
<td>7.199</td>
<td>11.426</td>
<td>-20.832</td>
<td>435.884</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>6.875</td>
<td>1.505</td>
<td>4.790</td>
<td>9.387</td>
<td>-23.125</td>
<td>536.906</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.4</td>
<td>5.759</td>
<td>2.012</td>
<td>2.979</td>
<td>9.095</td>
<td>-24.241</td>
<td>591.839</td>
</tr>
</tbody>
</table>

Table 5.15 Results from the Maximum Distance Method with 25% censoring $\tau = 40$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>%Significant</th>
<th>$\hat{\epsilon}$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>9.281</td>
<td>2.003</td>
<td>6.703</td>
<td>12.691</td>
<td>-20.719</td>
<td>434.002</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>100.0</td>
<td>6.936</td>
<td>2.116</td>
<td>4.118</td>
<td>10.479</td>
<td>-23.064</td>
<td>537.223</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>60.4</td>
<td>5.826</td>
<td>2.482</td>
<td>2.942</td>
<td>10.330</td>
<td>-24.174</td>
<td>592.006</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>9.211</td>
<td>1.595</td>
<td>7.019</td>
<td>11.859</td>
<td>-20.789</td>
<td>434.784</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>6.989</td>
<td>1.748</td>
<td>4.508</td>
<td>9.825</td>
<td>-23.012</td>
<td>532.863</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>94.4</td>
<td>5.863</td>
<td>2.118</td>
<td>3.142</td>
<td>9.498</td>
<td>-24.137</td>
<td>587.815</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>9.227</td>
<td>1.341</td>
<td>7.315</td>
<td>11.394</td>
<td>-20.773</td>
<td>433.343</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>6.851</td>
<td>1.458</td>
<td>4.840</td>
<td>9.295</td>
<td>-23.149</td>
<td>537.993</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.9</td>
<td>5.745</td>
<td>1.848</td>
<td>3.309</td>
<td>8.904</td>
<td>-24.255</td>
<td>592.141</td>
</tr>
</tbody>
</table>

Table 5.14 and Table 5.15 show the results for the Maximum Distance method when the cut-point is set beyond the range of the data, or no cut-point exists. This is the case when this estimator performs extremely poorly. Since the Maximum Distance Method picks up the largest distance between the two survival curves, it seems logical that it would find a time point in the data where this occurs and it wouldn’t always be at the end of the data. Here bias was calculated by subtracting the estimates from 30, this was done because the study end time was set to 30 in the simulations. In all scenarios the bias is quite large, and there doesn’t seem to be much
difference between censoring schemes. The bootstrap standard deviation estimates are all below 3, but in this case the possibility of a small variance does not make up for the large bias.

Table 5.16 Results from the Renyi Method with no censoring \( \tau = 40 \)

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>% Significant</th>
<th>( \tau^\tau )</th>
<th>SD*</th>
<th>( L_{90%}^* )</th>
<th>( U_{90%}^* )</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>7.317</td>
<td>1.609</td>
<td>4.794</td>
<td>9.914</td>
<td>-22.683</td>
<td>517.955</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>100.0</td>
<td>6.472</td>
<td>2.325</td>
<td>3.413</td>
<td>10.761</td>
<td>-23.528</td>
<td>560.482</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>49.2</td>
<td>5.195</td>
<td>2.839</td>
<td>2.228</td>
<td>11.061</td>
<td>-24.805</td>
<td>624.699</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>7.559</td>
<td>1.157</td>
<td>5.778</td>
<td>9.477</td>
<td>-22.441</td>
<td>505.050</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>6.934</td>
<td>1.693</td>
<td>4.488</td>
<td>9.872</td>
<td>-23.067</td>
<td>535.230</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>89.8</td>
<td>5.759</td>
<td>2.532</td>
<td>2.809</td>
<td>10.757</td>
<td>-24.241</td>
<td>596.009</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>7.637</td>
<td>0.824</td>
<td>6.418</td>
<td>9.093</td>
<td>-22.363</td>
<td>500.821</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>7.005</td>
<td>1.300</td>
<td>5.040</td>
<td>9.192</td>
<td>-22.995</td>
<td>530.561</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.4</td>
<td>6.061</td>
<td>2.249</td>
<td>3.262</td>
<td>10.393</td>
<td>-23.939</td>
<td>578.718</td>
</tr>
</tbody>
</table>

*Estimates here are from 200 simulated datasets with \( B=100 \) replicates.

Table 5.17 Results from the Renyi Method with 25% censoring \( \tau = 40 \)

<table>
<thead>
<tr>
<th>N</th>
<th>Trt hazard</th>
<th>% Significant</th>
<th>( \tau^\tau )</th>
<th>SD*</th>
<th>( L_{90%}^* )</th>
<th>( U_{90%}^* )</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>7.471</td>
<td>1.266</td>
<td>5.494</td>
<td>9.479</td>
<td>-22.529</td>
<td>509.807</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>100.0</td>
<td>5.942</td>
<td>1.920</td>
<td>3.793</td>
<td>9.739</td>
<td>-24.058</td>
<td>584.259</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>60.4</td>
<td>5.078</td>
<td>2.471</td>
<td>3.147</td>
<td>10.693</td>
<td>-24.922</td>
<td>628.062</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>7.711</td>
<td>0.959</td>
<td>6.169</td>
<td>9.232</td>
<td>-22.289</td>
<td>497.778</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>6.675</td>
<td>1.575</td>
<td>4.363</td>
<td>9.299</td>
<td>-23.325</td>
<td>547.229</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>94.4</td>
<td>5.567</td>
<td>2.346</td>
<td>3.439</td>
<td>10.720</td>
<td>-24.433</td>
<td>604.216</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>7.818</td>
<td>0.697</td>
<td>6.705</td>
<td>8.957</td>
<td>-22.183</td>
<td>492.570</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>6.777</td>
<td>1.351</td>
<td>4.728</td>
<td>9.012</td>
<td>-23.223</td>
<td>541.454</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.9</td>
<td>6.054</td>
<td>2.175</td>
<td>3.671</td>
<td>10.497</td>
<td>-23.947</td>
<td>579.563</td>
</tr>
</tbody>
</table>

*Estimates here are from 200 simulated datasets with \( B=100 \) replicates.

Table 5.16 and Table 5.17 give the results of the simulations for the Renyi Method when the cut-point is set beyond the range of the data. The Renyi Method performs very poorly in this
situation, estimating a cut-point very early on in the data. This means that the method is finding a point when in fact no point exists where the hazard changes for the treatment group.

Table 5.18 Results from the Hazard Method with no censoring $\tau= 40$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt Hazard</th>
<th>% Significant</th>
<th>$\tau^\wedge$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>21.520</td>
<td>3.901</td>
<td>10.661</td>
<td>22.294</td>
<td>-8.481</td>
<td>93.434</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>100.0</td>
<td>17.157</td>
<td>4.980</td>
<td>5.789</td>
<td>20.996</td>
<td>-12.843</td>
<td>199.178</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>49.2</td>
<td>10.852</td>
<td>4.469</td>
<td>3.801</td>
<td>17.358</td>
<td>-19.148</td>
<td>403.527</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>24.801</td>
<td>2.644</td>
<td>14.896</td>
<td>22.722</td>
<td>-5.199</td>
<td>39.996</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>21.003</td>
<td>4.484</td>
<td>8.859</td>
<td>22.369</td>
<td>-8.997</td>
<td>112.399</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>89.8</td>
<td>15.568</td>
<td>5.442</td>
<td>4.585</td>
<td>20.906</td>
<td>-14.432</td>
<td>245.797</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>26.651</td>
<td>1.906</td>
<td>17.343</td>
<td>22.875</td>
<td>-3.349</td>
<td>19.030</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>23.756</td>
<td>3.664</td>
<td>11.789</td>
<td>22.648</td>
<td>-6.244</td>
<td>62.797</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.4</td>
<td>17.964</td>
<td>5.538</td>
<td>5.489</td>
<td>22.000</td>
<td>-12.036</td>
<td>188.634</td>
</tr>
</tbody>
</table>

Table 5.19 Results from the Hazard Method with no censoring $\tau= 40$

<table>
<thead>
<tr>
<th>N</th>
<th>Trt Hazard</th>
<th>% Significant</th>
<th>$\tau^\wedge$</th>
<th>SD</th>
<th>L 90%</th>
<th>U 90%</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.05</td>
<td>100.0</td>
<td>21.520</td>
<td>3.892</td>
<td>10.550</td>
<td>22.292</td>
<td>-8.481</td>
<td>93.434</td>
</tr>
<tr>
<td>200</td>
<td>0.1</td>
<td>100.0</td>
<td>17.157</td>
<td>5.151</td>
<td>5.242</td>
<td>20.903</td>
<td>-12.843</td>
<td>199.178</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>60.4</td>
<td>10.852</td>
<td>4.518</td>
<td>3.623</td>
<td>17.184</td>
<td>-19.148</td>
<td>403.527</td>
</tr>
<tr>
<td>500</td>
<td>0.05</td>
<td>100.0</td>
<td>24.801</td>
<td>2.657</td>
<td>14.833</td>
<td>22.723</td>
<td>-5.199</td>
<td>39.996</td>
</tr>
<tr>
<td>500</td>
<td>0.1</td>
<td>100.0</td>
<td>21.003</td>
<td>4.602</td>
<td>8.521</td>
<td>22.360</td>
<td>-8.997</td>
<td>112.399</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>94.4</td>
<td>15.568</td>
<td>5.524</td>
<td>4.322</td>
<td>20.855</td>
<td>-14.432</td>
<td>245.797</td>
</tr>
<tr>
<td>1000</td>
<td>0.05</td>
<td>100.0</td>
<td>26.651</td>
<td>1.904</td>
<td>17.351</td>
<td>22.871</td>
<td>-3.349</td>
<td>19.030</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
<td>100.0</td>
<td>23.756</td>
<td>3.671</td>
<td>11.815</td>
<td>22.65</td>
<td>-6.244</td>
<td>62.797</td>
</tr>
<tr>
<td>1000</td>
<td>0.15</td>
<td>99.9</td>
<td>17.964</td>
<td>5.602</td>
<td>5.263</td>
<td>21.983</td>
<td>-12.036</td>
<td>188.634</td>
</tr>
</tbody>
</table>

Table 5.18 and Table 5.19 give the results of the Hazard Method. While the bias of these estimates is still large, they are much better than the other two methods for a cut-point beyond the range of the data. The method performs best when the hazard for the treatment is 0.05. Also the bias decreases as the sample size increases. Figure 5.15 shows the MSE versus squared bias relationship.
This chapter presented the simulations and results for the methods proposed in Chapter 4. All methods perform the best, in terms of bias, when the difference between the treatment hazard and control hazard is large during the first time period. The Maximum Distance Method constantly underestimates the true value of $\tau$. When $\tau$ occurs early the Maximum Distance Method does reasonably well. The Renyi Method also constantly underestimates $\tau$. This is likely due to the hockey stick, or piecewise regression model, being fit to the $|Z|$ statistics not being flexible enough. The Hazard Nlin method seemed to perform the best across all scenarios. While it does have consistency issues when the treatment hazard is large, the Hazard Nlin method overall comes the closest in for all three values of $\tau_1$. 

Figure 5.15 MSE versus Squared Bias for the Hazard Method when $\tau=40$, 25% censoring
Chapter 6 Conclusion and Future Work

6.1 Conclusion

The goal of this work was to establish a case for using the restricted mean survival time when the proportional hazards assumption was in doubt. Specifically, the aim was to find a time point in the data beyond which no treatment difference occurs. Irwin (1949) established the standard error of a restricted mean estimate in the non-parametric case, but he did not provide guidance as to how to choose the point of restriction. It is common practice to restrict the mean to either the last event time or the last observed time. Royston and Parmar (2011) argue the case for using the restricted mean as the primary summary statistic when the proportional hazards assumption is violated. This dissertation extends that argument, supposing that at some point in time the hazards for the two groups could become equal. We developed data driven methods to find that time. This work is important since the difference between the areas under the survival curves will continue to grow for as long as the follow up period extends, however the treatment difference could have ended long before. If the restricted mean is going to be a useful estimator it is important that treatment benefit is not exaggerated.

The methods developed in this work tended to underestimate the true value of $\tau$. The Maximum Distance Method, and Renyi Method estimates were always below the cut-point. The Hazard Method estimates appeared to be the closest to the true value of $\tau$ across all scenarios.
Chapter 2 was an overview of Survival Analysis. Censoring, non-parametric methods, the log-rank test, the cox proportional hazards model, parametric models and the accelerated failure time model were all reviewed. Chapter 3 reviewed literature on current uses of the restricted mean, as well as comparing the mean and median survival statistics. One method discussed for calculating the mean, discounting future years (SSTR Working Paper, 2007), gives greater weight to earlier event times when calculating a restricted mean. Another method utilizing pseudo-observations (Anderson, 2004) creates values for all observed times, event and censored, using leave one out estimators and a restricted mean can then be calculated from these numbers. In all of the methods reviewed, only one provided a guideline for choosing a time point for restricting the mean. Karrison (1997) suggested examining standard errors of survival probability estimates, however this choice seemed arbitrary.

In Chapter 4 the mathematical definition of the problem was presented, as well as the proposed methods for choosing a cut-point for restricting mean survival time. Given a situation when there is evidence that the difference in the treatments may only occur at the beginning of a study, we aimed to find the time when the treatment difference ends. The first method proposed was simply finding the time at which the maximum distance between the two survival curves occurs. The second proposed method was based on a Renyi-type test which plots the absolute value of the numerator of a log-rank statistic, Z, and looks for the supremum. We fit a piecewise regression, or hockey stick model, to the values of |Z|, to control for small changes at the end of the study. The third proposed method estimated the hazard at each event time, and then searched for a point which gave the smallest sums of squared error for a model which allowed for different hazards between the two groups and the same hazard after the chosen time point.

Chapter 5 presented the results of our methods applied to simulated datasets.
6.2 Future Work

This dissertation can be the basis for work expanding to parametric methods, as well as incorporating covariates in modeling the restricted mean. Karrison (1997) and Zucker (1998) have already developed models for including covariates, and their work could be combined with the methods presented here to see how a choice of \( \hat{\beta} \) effects modeling covariates.

Another aspect of this dissertation that could be expanded on are the models used to fit the Renyi-type statistics, since the simple piecewise regression model didn’t necessarily provide the best fit of the statistics. Also, the hazard non-linear estimate method could be expanded to be more flexible, allowing for non-constant hazards over time.

The major shortcoming of the work presented here is the size of the bootstrap replicates. With more computing time, a larger number of bootstrap replicates should be used to estimate the underlying distribution of the cut-point estimates.
Bibliography


Farewell, V.T. The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors. *Biometrics.* 1982; 38:1041-1046


Hudson D. Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association.* 1966; 61(316):1097-&.


Royston P, Parmar MKB. The Use of Restricted Mean Survival Time to Estimate the Treatment Effect in Randomized Clinical Trials when the Proportional Hazards Assumption is in Doubt. *Statistics in Medicine.* 2011; 30(19):2409-2421.


Appendices

SAS Programs for Computing Estimates

A.1 Simulating the data from a piecewise exponential

/*The following code was modified from a SAS(R) user group article
written by Iza Peszek, 2004*/
/*Prepared by: Emily Sheldon*/

/*A dataset to incorporate all of the scenarios of sample size,
hazard ratios, and preset deltas*/
data simulate.combination2;
   input n   n1   n2   hctrl1   hctrl2   htrt1 htrt2 del;
datalines;
200 100 100 0.2 0.2 0.05 0.2 10
200 100 100 0.2 0.2 0.1  0.2 10
200 100 100 0.2 0.2 0.15 0.2 10
200 100 100 0.2 0.2 0.05 0.2 5
200 100 100 0.2 0.2 0.1  0.2  5
200 100 100 0.2 0.2 0.15 0.2 5
200 100 100 0.2 0.2 0.05 0.2 40
200 100 100 0.2 0.2 0.1  0.2 40
500 250 250 0.2 0.2 0.05 0.2 10
500 250 250 0.2 0.2 0.1  0.2 10
500 250 250 0.2 0.2 0.15 0.2 10
500 250 250 0.2 0.2 0.05 0.2 5
500 250 250 0.2 0.2 0.1  0.2  5
500 250 250 0.2 0.2 0.15 0.2 5
500 250 250 0.2 0.2 0.05 0.2 40
500 250 250 0.2 0.2 0.1  0.2 40
500 250 250 0.2 0.2 0.15 0.2 40
500 250 250 0.2 0.2 0.05 0.2 40
500 250 250 0.2 0.2 0.1  0.2 40
500 250 250 0.2 0.2 0.15 0.2 40
1000 500 500 0.2 0.2 0.05 0.2 10
1000 500 500 0.2 0.2 0.1  0.2 10
1000 500 500 0.2 0.2 0.15 0.2 10
1000 500 500 0.2 0.2 0.05 0.2 5
1000 500 500 0.2 0.2 0.1  0.2  5
1000 500 500 0.2 0.2 0.15 0.2 5
1000 500 500 0.2 0.2 0.05 0.2 40
1000 500 500 0.2 0.2 0.1  0.2 40
1000 500 500 0.2 0.2 0.15 0.2 40
1000 500 500 0.2 0.2 0.05 0.2 40
; run;

data combination;
   set simulate.combination2;
   l1=hctrl1*1000;
   l2=htrt1*1000;
   obs=_n_; run;

************************************************
* Get the total number of simulation scenarios *
proc means data=combination n noprint;
  var n;
  output out=max n=num;
run;

data max;
set max;
  call symput('max', trim(num));
run;

* How many simulations at each combination *
**************************************************************************************
%Let sim=1000;

* Simulation Macro *
**********************
%Macro Sim(n=, n1=, n2=, hctrl1=, hctrl2=, htrt1=, htrt2=, del=, l1=, l2=);
proc iml;
  n  = &n; * total sample size;
  n1 = &n1; * control sample size;
  n2 = &n2; * treatment sample size;
  hctrl1 = &hctrl1; * hazard, period 1, control grp;
  hctrl2 = &hctrl2; * hazard, periods 2-4, control grp;
  htrt1 = &htrt1; * hazard, period 1, trt grp;
  htrt2 = &htrt2; * hazard, period 2, trt grp;

* initialize a dummy used for random number generation;
  myran1=j(n1, 1, &i);
  myran2=j(n2, 1, (&i+10));

* create treatment groups and study duration matrices;
  trt=j(n2, 1, 1);
  control=j(n1, 1, 0);
  studyduration=j(n, 1, 30);
  plduration = j(n, 1, &del); * duration of period 1;

* generate time of event according to 2-stage distribution;
  * time of event according to initial distribution (in period1);
    period1 = (-1*log(1-uniform(myran1))/hctrl1)
    // (-1*log(1-uniform(myran2))/htrt1);
  * time of event according to 2nd stage distribution,
    applies only if no event in period1 (conditional on no event in period1);
    period2 = plduration +((-1*log(1-uniform(myran1))/hctrl2)
    //(-1*log(1-uniform(myran2))/htrt2));
  * finally, “absolute” time of event;
    teventBX = period1#(period1<=plduration)+period2#(period1>plduration);
* time of event is censored if it occurs study ended;
censor = (studyduration<teventBX);

* final time of event is the smaller of study ending time and event time;
event = studyduration#censor+teventBX#(1-censor);

* final matrix has time of event/censoring in 1st column,
censoring indicator in the second column and treatment group indicator in the third column;
times0 = event||censor||(control//trt);

create times0 from times0 [colname={'Time' 'Censor' 'Group'}];
append from times0;
run;
quit;

/*Add a column of a random uniform number so that censoring may be
incorporated later*/
data times0;
  set times0;
  rnum=ranuni(&i);
run;

data simulate.sim&i._n&n._trt&l2._del&del;
  set times0;
run;
%Mend Sim;
%Macro RunIt(simnum=);
  %do j=1 %to &max;
    data comb&j;
      set combination;
      if (_n_ eq &j);
      call symput('ng', trim(n));
      call symput('n1g', trim(n1));
      call symput('n2g', trim(n2));
      call symput('hctrl1g', trim(hctrl1));
      call symput('hctrl2g', trim(hctrl2));
      call symput('htrt1g', trim(htrt1));
      call symput('htrt2g', trim(htrt2));
      call symput('delg', trim(del));
      call symput('l1g', trim(l1));
      call symput('l2g', trim(l2));

    run;
  %end;
%end RunIt;
%Mend RunIt;
%RunIt(simnum=&sim);

A.2 SAS® Program Calculating $\hat{\tau}$ Using the Maximum Distance Method

/*A SAS Macro to calculate an estimate of $\tau$ using the maximum distance method*/
/*Prepared by: Emily Sheldon*/

%Macro Maxdist;

%do i=1 %to 1000;

/*Bring in simulated dataset*/
/*Here data was simulated from a piecewise exponential model with a standard treatment hazard of 0.20 for the whole study and the new treatment has a hazard of 0.05 up to time 10, and then a hazard of 0.20 after, sample size=200 total*/

data distance;
    set simulate.sim&i._n200_trt50_del10;
    trt=group;
run;

/*Have SAS create dataset with survival KM estimates*/
ods listing close;
proc lifetest data=distance outsurv=survdist;
    strata trt;
    time time*censor(1);
run;
ods listing;

/*Separate survival estimates by group*/
data trt0 trt1;
    set survdist;
    if (trt eq 0) then output trt0;
else if (trt eq 1) then output trt1;
    keep time _censor_ survival;
run;

/*Sort and merge data so that distances between the curves at each event time can be calculated*/
proc sort data=trt0; by time;
run;
proc sort data=trt1; by time;
run;
proc sort data=survdist out=eventsing nodupkey; by time;
run;
data all;
    merge eventsing (keep=time)
trt0 (rename=(_censor_=cens0 survival=surv0))
trt1 (rename=(_censor_=cens1 survival=surv1)); by time;
run;

data all;
  set all;
  retain lasts0 lasts1;
  if (_n_ eq 1) then do;
    lasts0=surv0;
    lasts1=surv1;
  end;
  else do;
    if (surv0 eq .) then surv0=lasts0;
    if (surv1 eq .) then surv1=lasts1;
    lasts0=surv0;
    lasts1=surv1;
  end;
  drop cens0 cens1 lasts0 lasts1;
run;

/*Calculate the distance between the two curves*/
data all2;
  set all;
  distance=surv0-surv1;
  if distance<0 then absdist=-1*distance;
  if distance>=0 then absdist=distance;
run;

/*Find the maximum distance between the curves*/
ods listing close;
proc means data=all2 max ;
  var absdist;
  output out=distmax max=distmax;
run;
ods listing;

data all2max;
  set all2;
  if (_n_ eq 1) then set distmax;
  distmax=distmax;
run;

/*Identify the time where the maximum distance occurs*/
data ctptdist;
  set all2max;
  if distmax=absdist then cutpoint=time;
  else cutpoint = .;
  if cutpoint=. then delete;
  keep time distmax cutpoint;
run;

/***If there is a tie in max distance, take the larger time point***/
proc sort data=ctptdist nodupkey; by time;
run;

proc means data=ctptdist n noprint;
  var time;
  output out=max max=time;
run;

data max;
  set max;
  call symput('max', trim(time));
run;

data result;
  set ctptdist;
  time=trim(time);
  if time=&max;
run;

data result&i;
  set result;
  id=&i;
  keep distmax cutpoint id;
run;

proc append base=results.MD1result data=result&i force;
run;

proc datasets;
  delete distance survdist;
run;

%end;
%Mend Maxdist;

%Maxdist;
A.3 SAS® Program Calculating \( \hat{\tau} \) Using the Renyi Method

/*SAS Macros to get the sup|Z| and the hockey stick NLIN model to find an estimate of \( \tau \)*/
/*Prepared by: Emily Sheldon*/

/*Macro to the |Z| values*/
%Macro Getem;

/*Step through each event time*/
%do i=1 %to &max;

data tm&i;
   set t2;
   if (_n_ eq &i);
      call symput('tm', trim(time));
run;

data chemo&i;
   set chemo;
   if (time gt &tm) then censor=1;
run;

ods listing close;

/*Get the log rank chi-sq value at each event time*/
/*It was found that using the whole log-rank test statistic produced the same sup|Z| time as just using the numerator*/

proc lifetest data=chemo&i method=km;
   time time*censor(1);
   strata group / test=logrank;
   ods output homstats=lr&i;
run;

ods listing;

data lr&i;
   set lr&i;
   if (_n_ eq 1);
      cuttime=&tm;
      z=abs(logrank);
run;

proc append base=lrsummary data=lr&i force;
run;
%end;

%Mend Getem;
%Macro GetLRRenyi;
%do j=1 %to 1000;
data t1;
   set simulate.sim&j._n200_trt50_del10;
run;

data chemo;
   set t1;
run;

*****************************************************************************
* Get distinct failure times *
*****************************************************************************;
data t2;
   set chemo;
if (censor eq 1) then delete;
   keep time;
run;

proc sort data=t2 nodupkey; by time;
run;

/*Use proc means to identify the max event time*/
proc means data=t2 n noprint;
   var time;
   output out=max n=num;
run;

data max;
   set max;
   call symput('max', trim(num));
run;

%GetEm;

/*Graph the |Z| statistics*/
symbol1 value=none interpol=j;
proc gplot data=lrsummary;
   plot z*cuttime;
title1 'Method looking at the change in the log-rank test at each failure point';
run;
quit;

proc means data=lrsummary max noprint;
   var z;
   output out=cutpoint max=zmax;
run;

data lrsummary;
   set lrsummary;
if (_n_ eq 1) then set cutpoint;
time=cuttime;
run;

data compare;
    set lrsummary;
if (z eq zmax);
run;

proc print data=compare noby;
    var cuttime z;
run;

proc print data=lrsummary;
    var cuttime z;
run;

/*Now fit a hockey stick regression model*/
/*Also known as a piecewise regression model*/

proc nlin data=lrsummary;
    parms b0=0
        b1=0
        b2=0
        cut=5;
    if time<cut then do;
        mu=b0+b1*(time);
    end;
    if time>=cut then do;
        mu=b0+b1*time+b2*(time-cut);
    end;
    model z=mu;
ods output parameterestimates=parms;
run;

/*Identify the join point of the hockey stick model*/
data parms&j;
    set parms;
    if (parameter eq 'cut');
        keep estimate;
        id=&j;
run;

proc datasets;
    delete lrsummary parms t1;
run;

proc append base=test.ren4 data=parms&j force;
run;

%end;
%Mend GetLRRenyi;

%GetLRRenyi;
A.4 SAS® Program Calculating \( \hat{\tau} \) Using the Hazard Method

/*SAS Macro to compute the hazard estimates and then find a nonlinear model that best fits the estimates, with the parameter of interest being del, our estimate of \( \tau \)*/
/*Prepared by: Emily Sheldon*/

%Macro Hazard;
/*Read in the simulated dataset for given scenario in this example, total n=200, treatment hazard for the first period=0.05 and del was set to be 10, or the first time period ends at 10*/
data t1;
  set simulate.sim&i._n200_trt50_del10;
  trt=group;
  if rnum<0.25 then censor=1;
run;
proc sort data=t1; by time;
run;
/*Get K-M survival estimates so that we may estimate the hazard at each time point*/
ods listing close;
proc lifetest data=t1 method=KM outsurv=survdist;
  strata trt;
  time time*censor(1);
run;
ods listing;
/*Calculate hazard estimates at each event time by treatment*/
data hazards;
  set survdist;
  if trt=0 then do;
    cumhaz=-log(survival);
    lagtime=lag(time);
    laghaz=lag(cumhaz);
    oldtime=time;
    newtime=lagtime;
    oldhaz=cumhaz;
    newhaz=laghaz;
    hazard=(oldhaz-newhaz)/(oldtime-newtime);
  end;
  if trt=1 then do;
    cumhaz=-log(survival);
    lagtime=lag(time);
    laghaz=lag(cumhaz);
    oldtime=time;
    newtime=lagtime;
  end;
run;
oldhaz = cumhaz;
newhaz = laghaz;
hazard = (oldhaz - newhaz) / (oldtime - newtime);
end;
run;

data hazardsest;
  set hazards;
  time = newtime;
  cumhaz = newhaz;
  if hazard = 0 then delete;
  lhazest = log(hazard);
  keep time cumhaz hazard lhazest trt;
run;

data one;
  set hazardsest;
run;

/* Make a variable del, think of this as the potential values for the estimator */
/* This expands the dataset so that we may fit a nlin model for each value of del */
/* The del with the smallest SSE for the model will be our estimator */
data both;
  set one;
  do del = 5 to 40 by .1;
    output;
  end;
proc sort data = both; by del;
run;
ods listing close;
proc nlin data = both method = newton outest = betas;
  by del;
  parms b0 = 2 b1 = 1 b2 = 2 b3 = 0 ; *del = 9 to 30 by .2;
  h1 = b0 + b1 * trt;
  h2 = b2 + b3 * (time - del);
  mu = h1 * (time < del) + h2 * (time >= del);
  id del;
  model lhazest = mu;
  output out = outp p = pred;
  title 'Constant piecewise Hazards';
run;
ods listing;

/* Make a dataset of SSEs from the model */
data betas;
  set betas;
  where _type_ = 'FINAL';
run;

/* Graph SSE vs. Del */
axis1 label = (a = 90);
proc gplot data = betas;
  plot _sse_ * del / vaxis = axis1;
  label del = 'Delta';
run;
quit;

/*Now find the del that corresponds to the smallest SSE*/
proc sort data=betas;
  by _sse_;  
run;

proc means data=betas min ;
  var _sse_;  
  output out=ssemin min=ssemin;
run;

data ssemin;
  set ssemin;
  keep ssemin;
run;

data delfind;
  set betas;
  if (_n_ eq 1) then set ssemin;
  ssemin=ssemin;
  if _sse_=ssemin then output;
  keep del ssemin;
run;

/*Since ties are likely, this will find the minimum del of those with the smallest SSE*/
proc means data=delfind min;
  var del;
  output out=delmin min=delmin;
run;
data delmin1;
  set delmin;
  keep delmin;
run;
data delfind;
  set delfind;
  if (_n_ eq 1) then set delmin1;
  if del=delmin then output;
    id=&i;
  keep del ssemin id;
run;

data onedel;
  set one;
  if (_n_ eq 1) then set delfind;
run;

/*If desired, fit the NLIN with the chosen del, then plot*/
proc nlin data=onedel method=newton;
  parms b0=2 b1=1 b2=2;
  h1 = b0+ b1*trt;
  h2 = b2  ;
  mu = h1*(time<del) + h2*(time>=del);
  id del;
  model lhazest=mu;
output out=outp p=pred;
title 'Constant piecewise Hazards';
run;

proc datasets;
   delete t1;
run;

proc append base=results.haz1rescen data=delfind force;
run;
%Mend Hazard;

%MMacro Getit;
/*Pulls in the 1000 simulated datasets for the given scenario*/
%do i=1 %to 1000;
   %hazard;
%end;
%Mend Getit;

%Getit;