INFERENCE ON QUANTILE RESIDUAL LIFE FOR LENGTH-BIASED SURVIVAL DATA

by

Samia H Lopa

M.A in Mathematical Statistics, Ball State University, 2000

M.Sc. in Statistics, University of Dhaka, Bangladesh, 1995

B.Sc. (Honors) in Statistics, University of Dhaka, Bangladesh, 1994

Submitted to the Graduate Faculty of

the Department of Biostatsitcs

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Samia H Lopa

It was defended on

December 2, 2014

and approved by

Jong-Hyeon Jeong, PhD
Professor of Biostatistics
Graduate School Public Health
University of Pittsburgh

Maria Mori Brooks, PhD
Associate Professor of Epidemiology
Graduate School Public Health
University of Pittsburgh

Ying Ding, PhD
Assistant Professor of Biostatistics
Graduate School Public Health
University of Pittsburgh

Sally C. Morton, PhD
Professor of Biostatistics
Graduate School Public Health
University of Pittsburgh

Dissertation Director: Jong-Hyeon Jeong, PhD
Professor of Biostatistics
Graduate School Public Health
University of Pittsburgh

INFERENCE ON QUANTILE RESIDUAL LIFE FOR LENGTH-BIASED SURVIVAL DATA

Samia H Lopa, PhD

University of Pittsburgh, 2014

ABSTRACT

Length biased data occurs when a prevalent sampling is used to recruit subjects into a study that investigates the time from an initial event to a terminal event. Such data are usually left-truncated and right-censored. While there have been accurate and efficient methods to estimate the survival function, not much work has been done regarding the estimation of the residual life time distribution or the summary parameters such as the median and quantiles of the residual life. This dissertation proposes to make two new contributions. In the first part of the dissertation, we propose two ways to estimate the quantiles of the residual life time at fixed time points accounting for the length biased and censored nature of the data. We provide the asymptotic properties of these estimators and investigate them through simulation studies. Considering that the variances of these estimators require density estimation, we suggest an alternate approach taken by Jeong and others to obtain the confidence interval for the available residual function. We apply these methods to a breast cancer dataset from the National Surgical Adjuvant Breast and Bowel Project (NSABP).

In the second part of the dissertation, we propose a method for testing the equality of quantile residual life times from two different populations under prevalence sampling. This test can also be inverted to construct confidence intervals for the ratio or difference of two quantile residual life between two populations. We compare the performance of two methods, namely, the TPL and Huang and Qin methods via simulation. The results show that the

proposed tests maintain Type I error. The test based on Huang and Qin survival estimator is more powerful than that of based on the TPL estimator. We apply our methods to test the equality of median residual life of breast cancer patients having recurrence and undergoing two different treatments.

Public health significance of this research is enormous. For a population experiencing certain disease such as cancer, it is important to estimate the quantiles of the residual life time at specific time points to assess the impact of a disease and an intervention strategy on the population. This dissertation will provide accurate and efficient methods for estimating these quantiles.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 LENGTH-BIASED DATA AND QUANTILE RESIDUAL LIFE

1.1 LENGTH-BIASED DATA

In epidemiological studies, it is often of interest to study the time between the occurrence of an initial event and a terminating event. For example, in breast cancer studies it is of interest to study time from recurrence to death. In mental health studies, it is important to estimate the time between the onset of depression and the remission to understand the disease history and its evolution over time. In the ideal scenario, every subject in the population who experienced the initial event would be followed until the occurrence of the terminating event and the difference between the times of two events would be computed, which would result in the distribution of the time to the terminating event from the initiating event. In reality it is not possible to follow each member of the cohort for the occurrence of these paired events as it requires unlimited follow-up time. Two practical strategies are usually taken: an incident cohort study or a prevalent cohort study.

An incident cohort in epidemiology is formed by individuals who experience the initial events within a specified calendar time interval. These cases are then followed for a further fixed time period until failure, loss to follow-up or end of the study. For example, a group of women aged 18 or older may be followed for breast cancer occurrence over a period of five years and then those who have breast cancer would be followed until death, recurrence, or termination of the study, A prevalent cohort is a group of individuals who have experienced the initial event but have not experienced the terminating event at the time of recruitment into the study. Women diagnosed with breast cancer and alive during the recruitment period would enter the cohort for follow-up until death, recurrence, or termination of the study.

In a prevalent cohort study, cross-sectional sampling is used to identify the cases first and then the cases are followed until failure or censoring. Note that individuals who have already experienced both the initial and failure event will not be eligible to enter the study. Thus the prevalent cohort will be biased against individuals with shorter failure event times. Consider individuals 1 and 5 in Figure 1, both of whom experienced the initial and failure events before the study recruitment and hence will not be enrolled in the cohort. However, individuals 2 and 3 have not experienced the failure event at the study recruitment and will be enrolled in the cohort. Similarly, patient 4 would be included as the patient had experienced the initial event before recruitment. However this patient becomes right censored at the end of the study. The exclusion of individuals 1 and 5 occurs because of the length of time of the study, and the bias towards the longer times between the two events is referred to as length bias.

As mentioned earlier, in prevalence sampling, cases have already experienced their initial event. These initiating times are assumed to form a stationary poisson process and is referred to as stationarity. Under stationarity the incidence rate of the disease is approximately constant over time. The stationarity assumption is not valid when there is an epidemic of disease before the study starts. But it holds in situations where the disease is stable, that is, the rate of occurrence of disease (initial event) remains constant over time.

When prevalence sampling is used for recruiting a cohort, left truncation arises. In the presence of left truncation, we can only observe those individuals whose event time is longer than a given time. In addition to left truncation, survival data may also be subject to right censoring. Right censoring occurs when the event occurs after the study ends or a subject leaves the study (loss-to follow-up). Thus, if the data are right censored, the exact event time is not known, but it is known to be greater than certain time (e.g. time of last contact).

Failure to account for left-truncation properly results in biased estimation (overestimation) of the survival function. Wolfson et al. (2001) reported that when left-truncation was not accounted for, the survival rates for patients with dementia from the onset of the disease was overestimated such that the estimated median survival time was 6.6 years, twice as high compared to the left-truncation adjusted median of 3.3 years. Thus a proper method of analysis is essential for making valid inference from length-biased data.

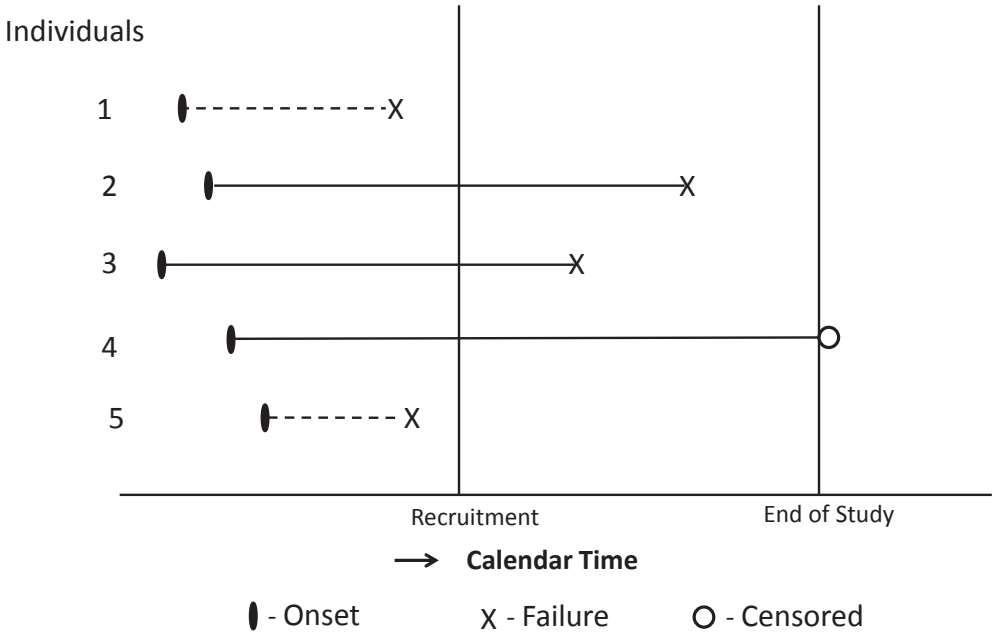

Figure 1: Prevalent cohort study with follow up.

In the last several decades, many authors have addressed the issues of left-truncation and right censoring in survival analysis (Wang et al., 1986[37]; Tsai et al., 1987[29]; and references therein). Although the concept of a truncation product limit estimator was first introduced by Lynden-Bell (1971)[22] to account for left truncation, it was only given attention in the 80's. This estimator is the analogue of the product limit estimator of Kaplan and Meier (1958)[20] for randomly censored data, and hence is referred to as the truncation product limit estimator. Wang, Jewel, and Tsai (1986)[37] discussed the asymptotic properties of this estimator. When the distribution of left truncation time is unspecified, a conditional analysis is preferred, conditioning on the truncation times (see Turnbull, 1976; Wang et al., 1986, 1993; Tsai et al., 1987; Lagakos et al., 1988; Wang, 1991, Wang et al. (1993), Anderson et al. (1993)[1, 21, 29, 30, 35, 36, 36, 37]) as this approach provides a simple and easy-to-implement expression for the estimator. However, if the onset of a disease follows a stationary Poisson process such that the incidence rate remains constant over time (Wang, 1991)[35], i.e., if the stationarity assumption holds, the truncation time will follow a Uniform distribution and an estimator that incorporates the truncation time distribution is generally more efficient (Vardi, 1982[31], 1985[32], 1989[32], Gill et al., 1988[15], Vardi and Zhang (1992)[33], Asgharian et al., 2002[2], and Asgharian and Wolfson, 2005[3]) than truncation product limit estimator. Vardi (1982[31], 1985[32]) and Gill, Vardi, and Wellner (1988)[15] extensively discussed the issues of length-bias and proposed a nonparametric maximum likelihood estimator (NPMLE) of the underlying non-length-biased survival distribution. The unconditional estimator, though more efficient, is not easy to implement as its large-sample properties involve mathematically intractable integral equations (Huang and Qin, 2011)[17]. Wang (1987)[8] showed that truncation product limit estimator is the nonparametric maximum likelihood estimator of the full likelihood. Wang (1991)[35] maintained that the truncation product limit estimator can be derived by maximizing the conditional likelihood and it has no information loss when the distribution of the truncation time is not specified. Huang and Qin (2011)[17] provided a new non-parametric estimator incorporating the distribution of truncation times under stationarity assumption. The new estimator has a closed-form expression and almost as efficient as the NPMLE.

1.2 RESIDUAL LIFE TIME

Breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer deaths in women. With early detection or screening for breast cancer and advances in medical care and treatments, the mortality rate from breast cancer has significantly reduced in recent years, yet an astonishing number of women still die from breast cancer every year. According to American Cancer Society data, 232,340 women will be diagnosed with invasive breast cancer and 39,620 women will die from breast cancer in 2013. Discovery of new treatments for both prevention and cure of breast cancer relies on reducing the incidence of breast cancer and mortality rate respectively in addition to prolonging survival. Thus the role of survival analysis remains as important as ever in breast cancer research. With the introduction of new treatment, often the question arises as to which patient is going to respond (e.g. live longer) to with which treatment. This is because survival time varies substantially across individuals, even under the same treatment strategies.

From a patient's perspective, one of the most important questions a woman diagnosed with breast cancer would like to be answered is about her predicted survival time (time from diagnosis to death), or even more pragmatically, her remaining life time. A breast cancer patient deciding to go through a rigorous treatment sequence naturally would like to know how long her life expectancy would be extended by this treatment. One way to answer such questions is to provide her an estimated mean survival for patients undergoing such treatment or an estimate of the probabilities of survival beyond certain times (e.g., 3 years, 5 years, etc.), or to provide her with an estimate of the reduction in risk of death through the application of the proposed treatment given her physical, demographic, and other clinical characteristics. These estimates are generally found by analyzing data from clinical trials, primarily employing statistical tools such as Kaplan-Meier curves (Kaplan and Meier, 1958[20]), Cox proportional hazards models (Cox, 1972[10]), or accelerated failure time models (Wei, 1992[38]).

Often a new drug is given in the middle of a follow up period to the patients because of its potential benefit, and the patient may wish to know her residual life expectancy (how much the new drug can extend her life expectancy, given that she has survived up to that time).

Kaplan-Meier plots cannot be used to infer the remaining life years of a patient when the patients are being followed after the first diagnosis. In recent years, breast cancer patients who remain recurrence free for several years after initial treatment have been offered new drugs. It is important for those patients to know the efficacy of the new drug and how it can prolong their remaining life time. Similarly, it is equally important for the oncologist to have the ability to explain or predict the remaining lifetime of the patients who are in the study.

The residual lifetime is the remaining survival time of a person, given that s/he has already survived up to a certain time. Suppose that a patient has survived up to time t_0. The residual survival time at t_0 for this patient is then defined as $R^0 = T - t_0$, where T is his or her actual survival time. The distribution of the random variable R^0 is of interest. In particular, there are two quantitative features of this distribution that are used to summarize the distribution of the residual life time R^0. They are mean residual life (MRL) function (Chiang, 1960[9]) and median residual life (MERL) function (Schmittlein and Morrison, 1981[27]). These two functions summarize the residual lifetime distribution in a way which is easily interpretable to the patients and oncologists alike.

1.2.1 Mean Residual Life Function

The MRL function at a fixed time t_0 is defined to be the expected remaining life at that time t_0. Using the notation above, MRL, denoted by $\mu(t_0)$, is defined as

$$\mu(t_0) = E[R^0 | T > t_0].$$

For example, if $\mu(6) = 12$ months for treatment and $\mu(6) = 6$ months for placebo, we can say for patients who are alive at six months, their average remaining life time will be six months longer on treatment than on placebo.

The MRL function has been extensively studied by many authors over the last several decades. Yang (1978[40]) considered a natural nonparametric estimator of MRL on a fixed finite interval and showed that the estimator was uniformly consistent and weakly convergent to a Gaussian process. Hall and Wellner (1981[16]) extended Yang's result by constructing

simultaneous confidence bands for the mean residual life. Oakes and Dasu (1990[25]) proposed a proportional mean residual life model, which is an alternative to the Cox (Cox, 1972[10]) regression model, to study the association between life expectancy and covariates. Maguluri and Zhang (1994[23]) extended this model with explanatory variables. Chen and Cheng (2005[7]) used counting process theories to develop semiparametric regression model with censored data. Chen and Zhao (2004[42]) proposed an empirical likelihood ratio method for semiparametric inference procedures.

1.2.2 Quantile Residual Life Function

When the data contains censored observations, a very common phenomenon in survival analysis is that estimated mean residual life function is not reliable. Moreover, even in the case of complete data, the estimated MRL function can be unfavorable if the underlying distribution is highly skewed or heavy tailed because of the outliers. As an alternative, one can consider the median or any other quantile of the residual lifetime. The $100q$th $(0 \leq q \leq 1)$ quantile residual life function, denoted by $Q(q, t_0)$, is defined as the $100q$th percentile of the distribution of residual life $T - t_0$ among patients in the population who survived longer than t_0. For example, for $q = 0.5$, $Q(0.5, t_0)$ represents the median residual time at t_0 and indicates that among patients who are survivors at time t_0, 50% will live longer than $Q(0.5, t_0)$.

While inference about the MRL has been studied by many authors, inference about the MERL, or more generally the quantile residual lifetime (QRL) are limited in the literature. A few works studied the estimation and testing of median residual life (MERL) from censored survival data. For example, Berger et al. (1988[5]) proposed a test Fligner and Rust's (1982[13]) approach to compare two median residual lifetimes under censoring. Wang and Hettmansperger's (1990[34]) work focused on testing the equality of two quantiles from failure time distributions under censoring by constructing a confidence interval for the differences between the two quantiles. Su and Wei (1993[28]) introduced a nonparametric test statistic to improve upon Wang and Hettmansperger's test, which requires estimation of the probability density function of failure times under censoring to evaluate the variance of the median failure

time. Both methods, however, were intended only for comparing the remaining lifetimes of patients at the origin of the follow-up period.

Gelfand and Kottas (2003[14]) proposed a Baysian semiparametric approach to the median residual life regression model. Jeong et al. (2008[18]) proposed a method to estimate median residual lifetime through Kaplan-Meier survival estimator in a single group. Jung et al. (2009[19]) proposed a log-linear regression model based on the median residual lifetime which is an extension of Ying, Jung & Wei (1995[41])'s work.

1.3 SPECIFIC AIMS

The goal of this dissertation is two-fold. First, we propose two consistent and efficient nonparametric estimators for the quantile residual life function based on length-biased survival data. We derive the asymptotic properties of the proposed estimators and compare the proposed estimators in terms of efficiency and coverage probabilities. Furthermore, we derive test statistics to test the equality of quantiles of residual life from two populations with length-biased samples. We also provide the asymptotic properties of the proposed test statistics, and evaluate the proposed test based on Type I error and Power.

2.0 NONPARAMETRIC ESTIMATION OF QUANTILE RESIDUAL LIFE FOR LENGTH-BIASED SURVIVAL DATA

In this chapter we propose non-parametric estimators for the quantile residual life function based on length-biased survival data. Specifically, we propose two estimators of quantile residual life function based on survival functions estimated using Truncation Product limit method and Huang and Qin (2011) method, respectively, and show that the latter is more efficient than the former. We will build upon the techniques presented in Jeong et al. (2008[18]) to construct confidence intervals for the quantiles of length-biased right-censored survival data.

2.1 DATA SET-UP AND NOTATION

Let $T_1^0, T_2^0, T_3^0, \ldots$ be i.i.d. positive random variables representing the time from disease incidence to failure event. Denote by $S(t)$ and $f(t)$ the survival and probability density function of T^0 (note that we drop the subscript i to represent a generic copy of the i.i.d variables) respectively. Let W^0 denote the calender time of disease incidence and ξ be the potential recruitment time. A prevalent population would include individuals with the disease who have not experienced the failure event at the sampling time, that is, the prevalent population would consist of failure times for which $T^0 \geq \xi - \boldsymbol{W}^0 > 0$. We will use W and T to indicate the 'observed' W^0 and T^0 respectively in the prevalent population.

For stable disease, we assume two conditions : (i) the probability distribution of survival time T^0 is independent of W^0, and, (ii) the occurrence rate of disease remains constant over calender time, that is, W^0 has a constant density function.

Let $A = \xi - W$ be the truncation time from disease incidence to sampling time and let $V = T - A$ be the residual survival time from sampling time. Let $f_T(t)$, $f_A(t)$ and $f_V(t)$ be the marginal density functions and $S_T(t)$, $S_A(t)$ and $S_V(t)$ be the survival functions of the corresponding the random variables T, A and V.

When conditions (i) and (ii) are satisfied, (A, T) has the joint density function:

$$f_{A,T}(a,t) = cf(t)I(t > a > 0), \tag{2.1}$$

where c is a constant. This means that,

$$\begin{aligned} & \int_0^\infty \int_0^t cf(t)dadt = 1 \\ \Leftrightarrow \quad & c\int_0^\infty f(t)tdt = 1 \\ \Leftrightarrow \quad & c = \tfrac{1}{\int_0^\infty tf(t)dt} = \tfrac{1}{\mu}. \end{aligned}$$

where $\mu = E(T^0)$. Therefore, the joint distribution of A and T is given by,

$$f_{A,T}(a,t) = \frac{1}{\mu} f(t)I(t > a > 0). \tag{2.2}$$

The marginal distribution of length-biased survival time T is then obtained as

$$f_T(t) = \int_0^t \frac{1}{\mu} f(t)da = \frac{1}{\mu} tf(t)I(t > 0) \tag{2.3}$$

The marginal distribution of truncation time A is then

$$f_A(a) = \int_a^\infty \frac{1}{\mu} f(t)dt = \frac{1}{\mu} S(a)I(a > 0). \tag{2.4}$$

The conditional distribution of T given A is

$$f_{(T|A=a)}(t) = \frac{f_{A,T}\,(t,\, a)}{f_A\,(a)} = \frac{\frac{1}{\mu} f(t)}{\frac{1}{\mu} S(a)} = \frac{f(t)}{S(a)}, \tag{2.5}$$

and the corresponding survival function is

$$S_{(T|A=a)}(t) = P(T > t|A = a) = \frac{1}{S(a)} \int_t^\infty f(t)\, dt = \frac{S(t)}{S(a)}. \tag{2.6}$$

The marginal distribution of $V = T - A$ can also be derived as

$$f_V(v) = \frac{1}{\mu} S(v) I(v > 0). \tag{2.7}$$

The second assumption of the occurrence rate of disease being constant over calender time is often referred to as stationarity. The consequence of this assumption is that the truncation time A given T is uniformly distributed. This follows from that fact that if a person lives up to time T, then because the initial event occurs at a constant rate, it could occur at any time before T with equal probability. That the marginal distribution of V and A are identical provides us a way to test the stationarity assumption. Asgharian et al. (2006[4]) suggested checking this in the data graphically by comparing the estimated distributions (Kaplan-Meier curves) of V and A.

Our interest lies in statistical inference for the distribution of residual lifetime at a specific timepoint s of the population represented by the survival time T^0. Specifically our objective is to estimate the 100qth quantile residual lifetime function $Q(q, s)$ at time s, defined by,

$$Pr(T^0 - s > Q(q, s) | T^0 > s) = 1 - q, \tag{2.8}$$

where $0 \leq q \leq 1$. Or, equivalently, $Q(q, s)$ is the solution to the equation

$$\frac{S(s + Q)}{S(s)} = 1 - q \tag{2.9}$$

The 100qth quantile residual lifetime for the prevalent population $Q^*(q, s)$ is defined as

$$\begin{aligned} P(T - s > Q^*(q, s) | T > s) &= 1 - q, \\ \Leftrightarrow \frac{S_T(s + Q^*)}{S_T(s)} &= 1 - q, \end{aligned}$$

where

$$\begin{aligned} S_T(s) &= \int_s^\infty \frac{t f(t)}{\mu} I(t > 0) dt \\ &= \frac{1}{\mu} \int_s^\infty t f(t) dt. \end{aligned}$$

Since the survival functions S(.) and $S_T(.)$ are not same, Q^* is not necessarily the same as Q, and hence the estimator for Q^* based on the estimated survival function of T might not be a good estimator of Q.

Suppose that individual i has censoring time $C_i^* = A_i + C_i$, where C_i is the residual censoring time, the time from recruitment until the individual is censored and $T_i = A_i + V_i$, is the overall survival time, such that V_i is the time from recruitment to failure event. We can observe only $min(C_i^*, T_i)$, We also assume that C_i is independent of (ξ_i, W_i, T_i). However, the total survival time T_i and the total censoring time C_i^* have A_i in common and are dependent. Thus the total survival time T_i is subject to informative censoring (Vardi, 1989). Informative censoring occurs when the censoring time, C_i^* gives more information on the survival time T_i, in addition to the knowledge that $T_i > C_i^*$.

Under right censoring of the survival time, we observe $(W_i, A_i, \widetilde{V}_i, \Delta_i)$, where $\widetilde{V}_i = min(V_i, C_i)$ and $\Delta_i = I(V_i \leq C_i)$, for $i = 1, ..., n$ where n is the number of individuals in the sample. Define $Y_i = A_i + \widetilde{V}_i = min(T_i, C_i^*)$.

2.2 ESTIMATION OF THE QUANTILE RESIDUAL FUNCTION

As defined in the previous section (Equation (2.8)), the 100qth quantile residual lifetime function $Q(q, s)$ at time s is given by

$$P(T^0 - s \geq Q(q,s) | T^0 > s) = 1 - q.$$

The left hand side of the above equation can be expressed as

$$\begin{aligned} & P(T^0 - s \geq Q(q,s) | T^0 > s) \\ &= \frac{P(T^0 - s \geq Q(q,s), T^0 > s)}{P(T^0 > s)} \\ &= \frac{P(T^0 \geq s + Q(q,s)), P(T^0 > s)}{P(T^0 > s)} \\ &= \frac{P(T^0 \geq s + Q(q,s))}{P(T^0 > s)} \\ &= \frac{S(s + Q(q,s))}{S(s)}. \end{aligned} \tag{2.10}$$

Thus the 100qth quantile of the residual lifetime distribution at a fixed time t_0 can be calculated as a solution to the equation

$$u(Q) = S(t_0 + Q) - (1 - q)S(t_0) = 0. \tag{2.11}$$

Note that for simplicity we have used Q for $Q(q, t_0)$. To estimate Q, one would replace the survival function $S(.)$ by its sample estimate, and solve the same equation for Q. In other words, Q would be estimated by solving the equation $\hat{u}(Q) = 0$ where

$$\hat{u}(Q) = \hat{S}(t_0 + Q) - (1 - q)\hat{S}(t_0), \tag{2.12}$$

where $\hat{S}(t)$ is a uniformly consistent estimator of the survival function in the prevalent population [35].

Different choices of $\hat{S}(.)$ might lead to different estimates of Q. Here we will consider two estimates of $S(.)$. One is the traditionally used truncation product limit (TPL) estimator by Tsai et, al. (1987[29]), and the other is a recently proposed estimator by Huang and Qin (2011[17]).

The TPL estimator of $S(.)$ based on the length-biased data is

$$\hat{S}(t) = \prod\nolimits_{u \in [0,t]} \{1 - d\hat{\Lambda}(u)\} \tag{2.13}$$

where $\hat{\Lambda}$ is the estimated cumulative hazard function

$$\hat{\Lambda}(t) = \int_0^t \frac{d\bar{N}(u)}{\bar{R}(u)} \tag{2.14}$$

with $\bar{N}(u) = n^{-1} \sum_{j=1}^{n} \Delta_j I(Y_j \leq u)$ and $\bar{R}(u) = n^{-1} \sum_{j=1}^{n} I(Y_j \geq u \geq A_j)$

The Truncation product-limit estimator ignores the information in the marginal distribution of A_j, which can be very inefficient in the situation where the data comes from a length-biased sample. Huang and Qin (2011) argued that the truncation time A has the same marginal distribution as the residual survival time V under length-biased sampling. They estimate S_A and Λ by combining the information from both A and V which leads to a more efficient estimator than TPL estimator.

The survival estimator proposed by Huang and Qin (2011) is given by

$$\hat{S}^*(t) = \prod_{u\in[0,t]}\{1 - d\tilde{\Lambda}(u)\} \tag{2.15}$$

with the corresponding estimated cumulative hazard function defined as

$$\tilde{\Lambda}(t) = \int_0^t \frac{d\bar{N}(u)}{\tilde{R}(u)}, \tag{2.16}$$

where

$$\tilde{R}(u) = n^{-1}\sum_{i=1}^{n} I(Y_i \geq u) - \tilde{S}_A(u),$$

$$\tilde{S}_A(u) = \prod_{t\in[0,u]}\left\{1 - \frac{d\tilde{B}(t)}{\tilde{K}(t)}\right\},$$

$$\tilde{B}(t) = n^{-1}\sum_{i=1}^{n}\{I(A_i \leq t) + \Delta_i I(\tilde{V}_i \leq t)\},$$

and

$$\tilde{K}(t) = n^{-1}\sum_{i=1}^{n}\{I(A_i \geq t) + I(\tilde{V}_i \geq t)\}.$$

Our first proposed estimator uses $\hat{S}(t)$ for $S(t)$ in Equation (2.12) and solves the equation for Q to obtain an estimator $\hat{Q}$ of Q. This means that

$$\hat{S}(t_0 + \hat{Q}) = (1-q)\hat{S}(t_0)$$

$$\Rightarrow \hat{Q} = \hat{S}^{-1}((1-q)\hat{S}(t_0)) - t_0,$$

where $\hat{S}^{-1}(.)$ is the inverse of the estimated survival function defined as

$$\hat{S}^{-1}(p) = inf\{t : \hat{S}(t) < p\} \tag{2.17}$$

for $0 < p < 1$. Note that, this implies that $\hat{Q}$ exists only if $\hat{S}^{-1}(p)$ exists at $p = (1-q)\hat{S}(t_0)$; or equivalently, there exists a t for which $\hat{S}(t) < (1-q)\hat{S}(t_0)$.

In a similar fashion, we define the estimator $\hat{Q}^*$ of Q. Specifically, $\hat{Q}^*$ is the solution of the equation $\hat{u}^*(Q) = 0$ for Q where

$$\hat{u}^*(Q) = \hat{S}^*(t_0 + Q) - (1-q)\hat{S}^*(t_0).$$

In other words,

$$\hat{Q}^* = [\hat{S}^*((1-q)\hat{S}^*(t_0))]^{-1} - t_0,$$

and $\hat{Q}^*$ exists only when there is at least one t for which $\hat{S}^*(t) < (1-q)\hat{S}^*(t_0)$.

2.3 ASYMPTOTIC PROPERTIES AND INFERENCE

It has been shown that $\hat{S}(t)$ is uniformly consistent and asymptotically normal (Wang, 1991) for $0 \leq t \leq \tau$, where τ is such that a positive proportion in the population is alive beyond τ. Moreover, this estimator is asymptotically linear such that

$$\sqrt{n}\left\{\hat{S}(t) - S(t)\right\} = n^{-1/2}\sum_{i=1}^{n}\phi_i(t) + o_p(1), \tag{2.18}$$

where

$$\phi_i(t) = \left[\int_0^t R(u)^{-2} I(Y_i \geq u \geq A_i) dF^u(u) - \frac{\Delta_i I(Y_i \leq t)}{R(Y_i)}\right] S(t).$$

$R(u) = \text{pr}(Y \leq u \leq A)$ and $F^u(t) = \text{pr}(\Delta = 1, Y \leq t)$ is the subdistribution function of complete observations. The variables ϕ_i, $i = 1, 2...., n$ are i.i.d. with $E(\phi_i) = 0$ and variance covariance function $\Sigma(t_1, t_2) = E\{\phi_i(t_1)\phi_i(t_2)\}$, $0 \leq t_1, t_2 \leq \tau$, and are referred to as the influence function of $\hat{S}(t)$.

Huang and Qin (2011) established that $\hat{S}^*(t)$ is uniformly consistent and asymptotically normal with

$$\sqrt{n}\left\{\hat{S}^*(t) - S(t)\right\} = n^{-1/2}\sum_{i=1}^{n}\phi_i^*(t) + o_p(1), \tag{2.19}$$

where

$$\phi_i^*(t) = \phi_i(t) + \int_0^t R(u)^{-2}\{I(A_i > u) - S_A(u) - S_A(u)\phi_i(u)\}dF^u(u),$$

$i = 1, 2....., n$, are i.i.d random variables with mean zero and the covariance function

$$\Sigma^*(t_1, t_2) = E\left[\phi_i^*(t_1)\phi_i^*(t_2)\right].$$

Now, the estimator $\hat{Q}$ is a solution to the estimating equation

$$\hat{u}(Q) = 0, \tag{2.20}$$

where $\hat{u}(.)$ is defined in (2.12), with $\hat{S}(t)$ calculated using Equation (2.13).

That is,

$$\hat{u}(Q) = \hat{S}(t_0 + Q) - (1 - q)\hat{S}(t_0). \tag{2.21}$$

Since $\hat{S}(t)$ is uniformly consistent over $t \in [0, \tau]$, $\hat{u}(Q)$ uniformly converges to

$$u(Q) = S(t_0 + Q) - (1 - q)S(t_0) \tag{2.22}$$

for $t \in [0, \tau]$. Since $u(Q_0) = 0$ at the true value $Q_0 = Q(t_0, q)$, the estimated 100qth quartile residual life of T^0 at time t_0, $\hat{Q}$ is a consistent estimator of Q_0.

As with all quantile estimators, finding the asymptotic distribution of $\hat{Q}$ through the expansion of $\hat{u}(Q)$ involves the estimation of the density function of T, which appears in the denominator of the variance expression. Since density estimators are highly variable and can often take values that are close to zero, the estimated variance of $\hat{Q}$ calculated in this manner is highly unstable. To overcome this, Jeong et al. (2008) suggested directly obtaining the confidence interval for Q by inverting a standardized statistic based on $\hat{u}(Q)$. This is a procedure similar to that suggested in Brookmeyer and Crowley (1982[6]) for obtaining confidence interval for quantiles. Thus, one would need to find the asymptotic distribution of $\hat{u}(Q_0)$. We note that

$$\begin{aligned}
\sqrt{n}\hat{u}(Q_0) &= \sqrt{n}\{\hat{u}(Q_0) - u(Q_0)\} \\
&= \sqrt{n}\{\hat{S}(t_0 + Q_0) - (1 - q)\hat{S}(t_0) - S(t_0 + Q_0) + (1 - q)S(t_0)\} \\
&= \sqrt{n}\{\hat{S}(t_0 + Q_0) - S(t_0 + Q_0)\} - \sqrt{n}(1 - q)\{\hat{S}(t_0) - S(t_0)\},
\end{aligned} \tag{2.23}$$

Using (2.18),

$$\begin{aligned}
\sqrt{n}\hat{u}(Q_0) &= n^{-\frac{1}{2}} \sum \phi_i(t_0 + Q_0) - n^{-\frac{1}{2}}(1 - q) \sum \phi_i(t_0) + o_p(1) \\
&= n^{-\frac{1}{2}} \sum \phi_i^u + o_p(1),
\end{aligned} \tag{2.24}$$

where $\phi_i^u = \phi_i(t_0 + Q_0) - (1 - q)\phi_i(t_0)$. Therefore, $\sqrt{n}\hat{u}(Q_0)$ is asymptotically distributed as normal with mean zero and variance

$$\begin{aligned} V_0(Q_0) &= \text{var}(\phi_i^u) \\ &= \text{var}[\phi_i(t_0+Q_0) - (1-q)\phi_i(t_0)] \\ &= \Sigma(t_0+Q_0, t_0+Q_0) + (1-q)^2\Sigma(t_0,t_0) - 2(1-q)\Sigma(t_0+Q_0,t_0), \quad (2.25) \end{aligned}$$

where $\Sigma(t_1,t_2) = E\{\phi_i(t_1)\phi_i(t_2)\}$.

For known Q_0 this variance can be estimated by

$$\hat{V}_0(Q_0) = \hat{\Sigma}(t_0+Q_0, t_0+Q_0) + (1-q)^2\hat{\Sigma}(t_0,t_0) - 2(1-q)\hat{\Sigma}(t_0+Q_0,t_0),$$

where

$$\hat{\Sigma}(t_1,t_2) = \frac{1}{n}\Sigma_{i=1}^n \hat{\phi}_i(t_1)\hat{\phi}_i(t_2)$$

with

$$\hat{\phi}_i(t) = \hat{S}(t)\left[\int_0^t \frac{I(Y_i \geq u \geq A_i)d\bar{N}(u)}{\hat{R}^2(u)} - \frac{\Delta_i I(Y_i \leq t)}{\hat{R}(Y_i)}\right].$$

A $100(1-\alpha)\%$ confidence interval for Q can then be constructed by inverting the Wald statistic based on the asymptotic distribution of $\sqrt{n}\hat{u}(Q_0)$. More explicitly, a $100(1-\alpha)\%$ confidence interval for Q is given by

$$\{Q : [\hat{V}_0(Q)]^{-1} n[\hat{u}(Q)]^2 < \chi^2_{1,1-\alpha}\}, \quad (2.26)$$

where $\chi^2_{1,1-\alpha}$ is the $100(1-\alpha)$th percentile of a χ^2 distribution with 1 degree of freedom.

A similar argument can be used to construct a confidence intervals for Q using the other estimator $\hat{Q}^*$. Explicitly, a $100(1-\alpha)\%$ confidence interval for Q using this method is given by,

$$\{Q : [\hat{V_0}^*(Q)]^{-1} n[\hat{u}^*(Q)]^2 < \chi^2_{1,1-\alpha}\},$$

where

$$\hat{u}^*(Q) = \hat{S}^*(t_0+Q) - (1-q)\hat{S}^*(t_0),$$

$$\hat{V_0}^*(Q) = \hat{\Sigma}^*(t_0+Q, t_0+Q) + (1-q)^2\hat{\Sigma}^*(t_0,t_0) - 2(1-q)\hat{\Sigma}^*(t_0+Q,t_0),$$

$$\hat{\Sigma}^*(t_1,t_2)=\frac{1}{n}\sum_{i=1}^{n}\hat{\phi_i}^*(t_1)\hat{\phi_i}^*(t_2),$$

with

$$\hat{\phi_i}^*(t)=\hat{S}^*(t)\left[\int_0^t\frac{I(Y_i\geq u\geq A_i)d\bar{N}(u)}{\tilde{R}^2(u)}-\frac{\Delta_i I(Y_i\leq t)}{\tilde{R}(Y_i)}\right.$$
$$\left.+\int_0^t\left\{\frac{I(A_i>u}{\tilde{R}^2(u)}-\tilde{S_A}(u)-\tilde{S_A}(u)\tilde{\psi_i}(u)\right\}d\bar{N}(u)\right],$$

where

$$\tilde{\psi_i}(u)=\int_0^t\frac{\{I(A_i\geq u)+I(\tilde{V_i}\geq u)\}d\tilde{B}(u)}{\tilde{K}^2(u)}-\frac{I(A_i\leq t)}{\tilde{K}(A_i)}-\frac{\Delta_i I(\tilde{V_i}\leq t)}{\tilde{K}(\tilde{V_i})}.$$

2.4 SIMULATION STUDY

We conducted a simulation study to evaluate the finite-sample performance of the two estimators of quantile residual life functions discussed in the previous sections, namely, $\hat{Q}^*$ calculated based on Huang Qin method, and $\hat{Q}$ based on the truncation product-limit estimator. We generated 1000 Monte-Carlo samples of sizes n (200 or 400) from the target population. The population parameters were chosen to be similar to simulation conducted by Huang and Qin (2011). More explicitly, we first set the sampling time (recruiting time) ξ to be 100. The time of disease onset W^0 was generated from a uniform distribution over the interval $[0,100]$. The survival time T^0 was independently generated from a Weibull distribution with the survival function $S(t)=\exp(-t^2/4)$. To form a prevalent cohort of sample size n, the pair (W^0,T^0) were generated repeatedly until there were n pairs of observations satisfying the sampling constraint $W^0+T^0\geq\xi$. The residual censoring time C from enrollment to loss to follow-up was generated from a uniform distribution with a support $(1,2)$ or $(0,2)$ resulting respectively in 30% and 51% censoring rates.

The true 100qth quantile residual life function at time t_0 for the Weibull distribution above is given by

$$\exp\{-(t_0+Q(q,t_0))^2/4\}-(1-q)\times\exp(-t_0^2/4)=0,$$

which is equivalent to the positive root of the quadratic equation

$$Q^2(q, t_0) + 2Q(q, t_0)t_0 + 4\ln(1-q) = 0.$$

Thus the true 100qth quantile residual life function at time t_0 is

$$Q(q, t_0) = \sqrt{t_0^2 - 4\ln(1-q)} - t_0. \tag{2.27}$$

Figure 2 shows the true QRL function for various values of q, specifically, the 25^{th}, 50^{th}, and 75^{th} percentile of residual life function.

We show the numerical results for estimating these residual life functions at fixed time points 0.0, 0.5, and 1.0. We present the bias, average length and coverage probability of 95% confidence intervals for the two estimates, $\hat{Q}$ and $\hat{Q}^*$.

Table 1 the summarizes mean of 75^{th} quantile residual life function, bias, the average length of 95% confidence interval, and the coverage probability of 95% confidence interval for both estimators at different time points $(0.0, 0.5, 1.0)$ for two sample sizes $n = (200, 400)$ based on 1000 Monte-Carlo samples. The true 75^{th} quantile residual life functions at the selected time points $(0.0, 0.5, 1.0)$ are respectively $(2.35, 1.91, 1.56)$. When censoring was moderate (30%), both estimators were approximately unbiased with bias being relatively smaller for the $\hat{Q}^*$ for the larger sample size of 400; the bias ranged from -0.0032 to 0.0013 for $\hat{Q}$ and -0.0006 to 0.0064 for the $\hat{Q}^*$. In general, the coverage probabilities of the 95% confidence intervals met the nominal level, specifically for the larger sample size of 400; $(91.6\% - 94.5\%)$ for $\hat{Q}$ and $(93.0\% - 95.1\%)$ for $\hat{Q}^*$. Average length of these confidence intervals was generally wider for $\hat{Q}$. For example, at $n = 200$ the average length of the 95% confidence intervals was 0.47 for $\hat{Q}$ versus 0.41 for $\hat{Q}^*$, showing a 12.8% reduction. This reduction in the length of the confidence interval is equivalent to an increase in efficiency for $\hat{Q}^*$. When sample size was increased from 200 to 400, the confidence intervals became narrower as shown by the decrease in the average length of the intervals for both methods. However $\hat{Q}^*$ still resulted in narrower confidence intervals with a 9.4% to 15.4% reduction in average length as compared to $\hat{Q}$. This is equivalent to an efficiency gain of 21.8% to 39.7% for $\hat{Q}^*$ over $\hat{Q}$.

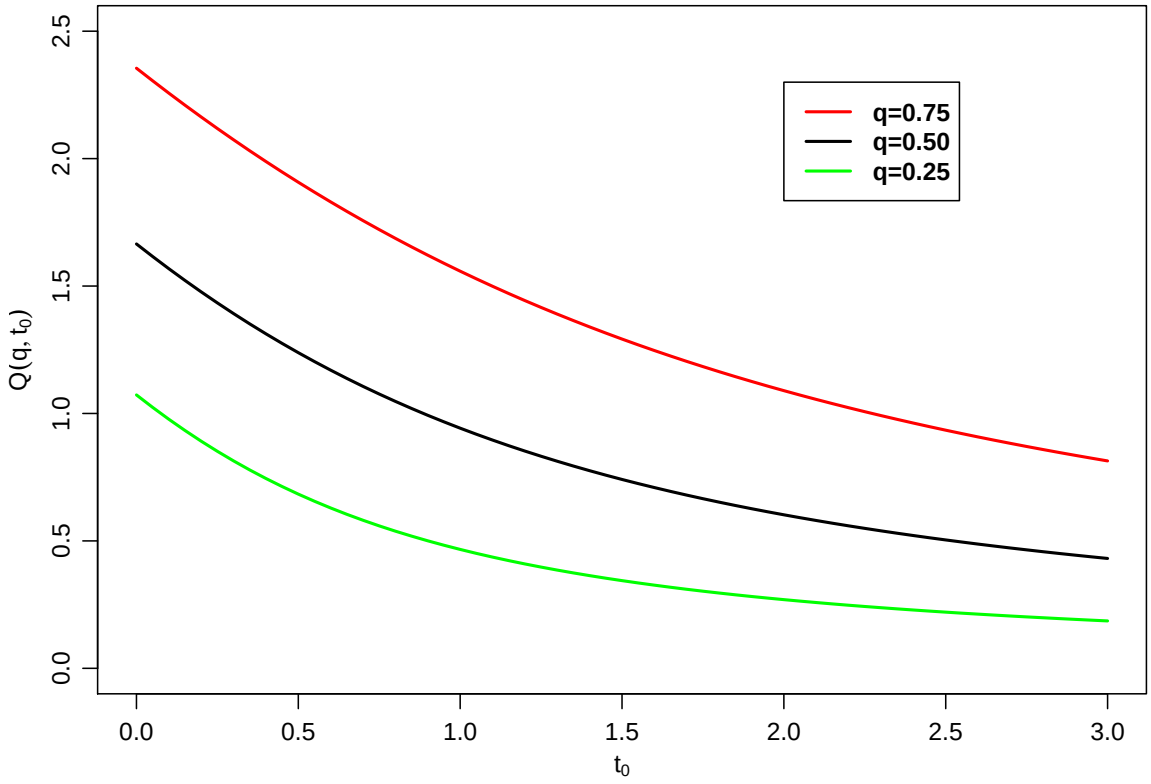

Figure 2: Graphical representation of the true quantile residual life function for various values of q.

Table 1: Simulation results for estimating the 75^{th} quantile residual life function depicted in Figure 2. EST = Monte Carlo mean of the estimator, BIAS = absolute bias of the estimator, AL = Monte Carlo average of the length of 95% confidence interval, CP = Coverage probability of 95% confidence interval.

% cen	t_0	n	$Q(0.75, t_0)$	$\hat{Q}$	BIAS	AL	CP	$\hat{Q}^*$	BIAS	AL	CP
	0.0	200	2.35	2.35	-0.0003	0.47	91.6	2.36	0.0059	0.41	93.7
30%	0.5		1.91	1.91	0.0013	0.45	91.7	1.91	0.0064	0.40	93.0
	1.0		1.56	1.56	0.0013	0.45	92.0	1.56	0.0040	0.40	93.6
	0.0	400	2.35	2.35	-0.0032	0.33	92.5	2.36	0.0004	0.29	93.6
	0.5		1.91	1.91	-0.0016	0.32	94.5	1.91	0.0006	0.28	93.7
	1.0		1.56	1.56	-0.0027	0.32	94.1	1.56	-0.0006	0.29	95.1
	0.0	200	2.35	2.36	0.0005	0.56	93.1	2.36	0.0039	0.48	93.1
51%	0.5		1.91	1.91	0.0027	0.55	92.5	1.91	0.0066	0.47	93.1
	1.0		1.56	1.56	0.0062	0.55	92.9	1.57	0.0086	0.49	95.1
	0.0	400	2.35	2.35	-0.0038	0.40	92.4	2.35	-0.0001	0.34	93.1
	0.5		1.91	1.91	-0.0018	0.39	92.2	1.91	0.0009	0.33	94.0
	1.0		1.56	1.56	-0.0022	0.39	93.3	1.56	-0.0008	0.35	93.8

Table 2: Simulation results for estimating the 50^{th} quantile residual life function depicted in Figure 2. EST = Monte Carlo mean of the estimator, BIAS = absolute bias of the estimator, AL = Monte Carlo average of the length of 95% confidence interval, CP = Coverage probability of 95% confidence interval.

% Cen	t_0	n	$Q(0.50, t_0)$	$\hat{Q}$	BIAS	AL	CP	$\hat{Q}^*$	BIAS	AL	CP
	0.0	200	1.67	1.66	-0.0004	0.45	92.3	1.67	0.0034	0.40	93.7
30%	0.5		1.24	1.24	0.0024	0.39	93.5	1.24	0.0059	0.36	94.3
	1.0		0.94	0.94	0.0016	0.35	92.2	0.94	0.0026	0.33	93.9
	0.0	400	1.67	1.66	-0.0040	0.31	92.4	1.67	0.0008	0.28	93.2
	0.5		1.24	1.24	0.0004	0.28	93.0	1.24	0.0024	0.25	92.8
	1.0		0.94	0.94	-0.0029	0.25	91.9	0.94	-0.0009	0.23	92.4
	0.0	200	1.67	1.67	0.0023	0.51	92.6	1.67	0.0049	0.46	93.1
51%	0.5		1.24	1.24	0.0042	0.47	93.9	1.25	0.0065	0.42	93.8
	1.0		0.94	0.95	0.0029	0.43	92.4	0.95	0.0053	0.39	93.9
	0.0	400	1.67	1.66	-0.0037	0.36	91.9	1.67	0.0010	0.32	93.8
	0.5		1.24	1.24	0.00003	0.33	92.0	1.24	0.0009	0.29	92.8
	1.0		0.94	0.94	-0.0011	0.30	91.9	0.94	0.0007	0.27	92.3

Table 3: Simulation results for estimating the 25^{th} quantile residual life function depicted in Figure 2. EST = Monte Carlo mean of the estimator, BIAS = absolute bias of the estimator, AL = Monte Carlo average of the length of 95% confidence interval, CP = Coverage probability of 95% confidence interval.

%Cen	t_0	n	$Q(0.25, t_0)$	$\hat{Q}$	BIAS	AL	CP	$\hat{Q}^*$	BIAS	AL	CP
	0.0	200	1.07	1.07	-0.0013	0.49	90.8	1.08	0.0042	0.47	93.1
30%	0.5		0.68	0.69	0.0049	0.37	91.4	0.69	0.0074	0.36	91.6
	1.0		0.47	0.47	0.0024	0.29	93.8	0.47	0.0041	0.27	93.9
	0.0	400	1.07	1.07	-0.0006	0.35	92.1	1.08	0.0026	0.34	92.8
	0.5		0.68	0.69	0.0034	0.27	91.4	0.69	0.0048	0.25	91.2
	1.0		0.47	0.47	-0.0011	0.20	93.4	0.47	-0.0001	0.19	93.6
	0.0	200	1.07	1.07	0.0016	0.53	92.2	1.08	0.0053	0.50	93.2
51%	0.5		0.68	0.69	0.0063	0.42	91.4	0.69	0.0068	0.40	91.9
	1.0		0.47	0.47	0.0039	0.34	93.6	0.47	0.0051	0.33	93.8
	0.0	400	1.07	1.07	-0.0005	0.38	93.5	1.08	0.0028	0.36	93.2
	0.5		0.68	0.69	0.0035	0.30	92.2	0.69	0.0050	0.28	92.4
	1.0		0.47	0.47	0.0007	0.24	93.6	0.47	0.0022	0.23	93.4

When the censoring rate was increased to 51%, the biases were increased for both estimators but they remained small. With the increased censoring, $\hat{Q}^*$ almost always provided better coverage for the 95% confidence interval for both smaller (200) and larger (400) sample sizes, compared to $\hat{Q}$.

Table 2 shows the results for estimating the median residual lifetimes at times $(0.0, 0.5, 1.0)$. The results are similar to those described for 75^{th} quantile estimators. Specifically both $\hat{Q}$ and $\hat{Q}^*$ are approximately unbiased and maintain the nominal coverage probability at moderate to extreme levels of censoring and sample sizes. However, $\hat{Q}^*$ is more efficient (has narrower confidence intervals). Also, the coverage probabilities are almost uniformly better for $\hat{Q}^*$ as compared to $\hat{Q}$. The same trend followed for the 25^{th} percentile residual estimator [Table 3].

In summary, the two proposed estimators of quantile residual lifetime provide reasonable estimates with $\hat{Q}^*$ being more efficient than $\hat{Q}$, in general.

2.5 ANALYSIS OF NSABP B-20 DATA

In this section we illustrate our method using the data from the National Surgical Adjuvant Breast and Bowel Project (NSABP) protocol B-20 clinical trial. Protocol B-20 was designed to determine whether addition of chemotherapy to tamoxifen would improve the beneficial effect of tamoxifen in the treatment of estrogen receptor (ER) positive patients with axillary lymph node-negative breast cancer. Women at participating NSABP clinical centers in the USA and Canada who had primary breast cancer, histologically node-negative and ER positive breast cancer were eligible for this trial. Patients fulfilling eligibility criteria underwent surgery (total mastectomy and lymph node dissection or lumpectomy and lymph node dissection followed by breast irradiation). Following surgery, patients providing written consent to participate in the trial were randomized to one of the three treatment groups: tamoxifen(T) alone (TAM), T plus sequential methotrexate (M) and fluorouracil (F) (MFT), T plus cyclophosphamide (C), M, and F (CMFT) stratified by age, tumor size, and tumor ER level. Between Oct 17, 1988, and March 5, 1993, a total of 2363 women were enrolled in this study

(788 randomly assigned to TAM, 786 to MFT, and 789 to CMFT). Earlier findings from this clinical trial (Fisher et al., 1997[11]) analyzing 5-year follow-up data showed that addition of chemotherapy to tamoxifen significantly improved the disease-free survival (DFS) rate by at least 4%, MFT vs. TAM (90% vs. 85%, p = .01) and CMFT vs. TAM (89% vs. 85%, p = .001). Chemotherapy plus Tamoxifen groups (MFT and CMFT) perform significantly better in DFS than tamoxifen alone(Fisher et al., 1997). Similar results were observed for distant disease-free survival and overall survival. In a follow-up publication, Fisher et al., (2004[12]) showed that CMFT-treated women had significantly better recurrence-free survival (89% vs. 79%, p<0.0001) and better overall survival (87% vs. 83%, p = 0.063) than women treated with tamoxifen alone over 12 years of follow-up.

This difference might prompt one to characterize the pattern of survival among patients with recurrence. Recurrence of cancer is not uncommon among breast cancer patients treated with tamoxifen. In the B-20 trial, over 18 years of follow-up, it has been observed that 17% of the women experienced recurrence before death. Thus it might be of interest to estimate the residual survival following a recurrence in histologically node-negative and ER-positive breast cancer patients who are being treated with tamoxifen with or without chemotherapy. Thus, our initial event will be recurrence, and the primary endpoint will be death.

A total of 788 women were randomly assigned to receive Tamoxifen. Out of these 788 women, 170 women experienced recurrence. To construct a length-biased sample of patients, we identified the date of recurrence for the first patient who experienced recurrence and set the recruitment time to be at 5 years following this first recurrence date. By the recruitment date, 120 patients had died or been lost-to-follow-up and hence would not be eligible to be in the sample. (Alternatively, a window can be considered for recruitment, but for simplicity, we will assume that all patients are recruited at once.) Thus, our length-biased sample consists of 50 patients who had recurrence prior to September 17, 1994 and were still being followed at the same date. We will estimate the quantile residual lifetimes of these patients.

Table 4 presents the estimated 25^{th} percentile and median for the residual life times for patients who experienced recurrence following treatment with tamoxifen based on the NSABP B-20 data at times 0, 0.5 and 1 years after recurrence. We also provide 95% confidence intervals for both estimates. For this specific dataset the TPL estimates of quantile residuals

Table 4: Estimated quantile residual lifetimes and 95% confidence interval in the Tamoxifen arm of the NSABP B-20 data

q	t_0	$\hat{Q}(q,t_0)$	95% CL	$\hat{Q}^*(q,t_0)$	95% CL
	0.0	1.48	(1.27, 3.29)	1.37	(1.26, 2.59)
.25	0.5	0.98	(0.77, 2.79)	0.87	(0.76, 2.09)
	1.0	0.48	(0.27, 2.29)	0.37	(0.26, 1.59)
	0.0	3.11	(1.49, 10.32)	2.36	(1.42, 4.34)
.50	0.5	2.61	(0.99, 9.82)	1.86	(0.92, 3.84)
	1.0	2.11	(0.49, 9.32)	1.36	(0.42, 3.34)

$(\hat{Q})$ were generally larger than $\hat{Q}^*$, which employs the distribution of truncation time. For example, at the time of recurrence, the estimated median residual lifetime estimated by $\hat{Q}^*$ is approximately 2.4 years compared to 3.1 years for $\hat{Q}$. At 1 year, the median residual life time (95% CI) for patients with recurrence after tamoxifen treatment is approximately $1.36(0.42, 3.34)$ years by $\hat{Q}^*$ compared to $2.11(0.49, 9.32)$ years by $\hat{Q}$. Similar results follow for the 25^{th} percentile.

In Figure 3, we present the estimated median residual lifetimes and their pointwise confidence intervals for the NSABP B-20 patients experiencing recurrence. Consistent with the results shown in Table 4, the $\hat{Q}^*$ estimates are uniformly smaller than the $\hat{Q}$ estimates. Median residual lifetime first decreases until between 1 to 1.2 years after recurrence and then increases. This could potentially be the effect of post-recurrence treatment which the patients may have received outside the protocol. The confidence intervals are wider for $\hat{Q}$ estimators compared to $\hat{Q}^*$ estimators, as expected from the simulation results.

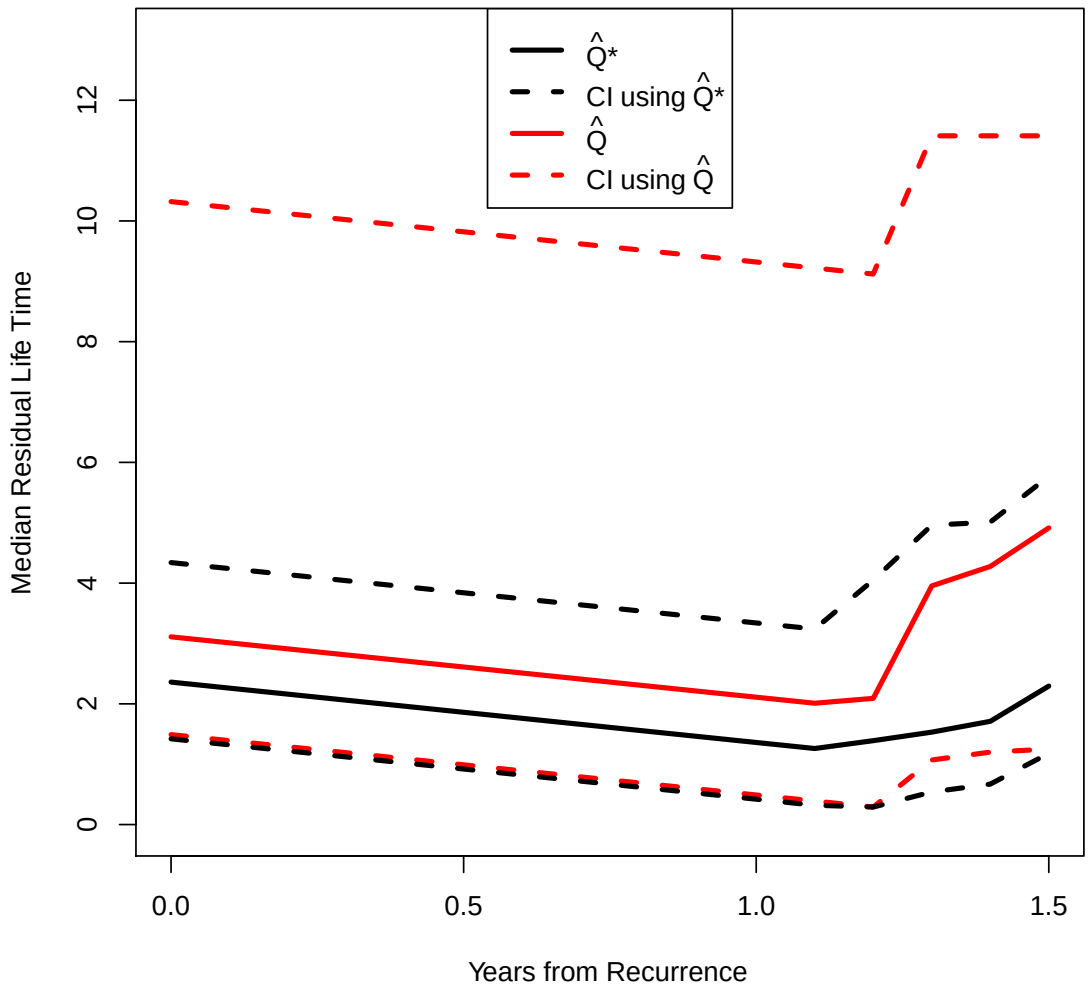

Figure 3: Estimated median residual lifetimes and 95% confidence intervals for time to death from recurrence for the NSABP B-20 data.

2.6 DISCUSSION

Length-biased data are common in epidemiological prevalent cohort studies where the time between an initial event and a terminal event is of interest. In this chapter, we have proposed two different approaches to estimate the length-bias corrected quantile residual life (QRL) function for a prevalence sample. The first estimator is based on the truncation product limit estimator (Wang et al., 1986) and the second is based on the Huang and Qin (2011) survival estimator. We have developed the asymptotic properties procedures for the two estimators and demonstrated how the confidence interval for QRL can be constructed by inverting a proposed test statistic.

Simulation results showed that both methods empirically lead to consistent QRL estimates, but the QRL estimator based on Huang and Qin method is more efficient than TPL estimator as shown by the narrower confidence intervals for the former compared to the latter. Both estimators maintain nominal coverage levels, though the coverage is often smaller than the nominal 95%.

We have illustrated our methods with a breast cancer dataset from one of the NSABP studies, estimating the quantiles of residual lifetime for breast cancer patients experiencing a recurrence.

3.0 TEST OF EQUALITY OF QUANTILE RESIDUAL LIFES FOR LENGTH-BIASED SURVIVAL DATA

3.1 INTRODUCTION

In epidemiological studies, it is often of interest to compare two populations with respect to the time between the occurrence of an initial event and a terminating event. In breast cancer studies, it may be of interest to compare time from recurrence to death among patients who are treated with tamoxifen to that for patients who were additionally treated with other chemotherapy agents. As discussed in Chapter 1, such times are length-biased due to prevalence sampling and hence regular Kaplan-Meier estimators are not suitable (Klein and Moeschberger, 2003[24]). Consequently, the comparison of the quantile residual life based on such estimators will also be biased. In this part of the dissertation, we propose methods for unbiased testing of the equality of quantile residual lifes for two different populations under prevalence sampling. As a byproduct we will be able to construct confidence intervals for the ratio or difference of two quantile residual lifes between two populations by inverting the corresponding tests.

Testing the equality of medians from censored survival data has been of interest to researchers for a long time. The famous Brookmeyer and Crowley paper (Brookmeyer and Crowley, 1982; Wang and Hettmansperger, 1990) discusses a test statistic that requires estimation of the density function, which is often not straightforward in the presence of censoring. This limitation was overcome by Su and Wei (1993) , who proposed a nonparametric test statistic for comparing two median failure times based on the minimum dispersion statistic (Basawa and Koul, 1988). Jeong et al. (2008) extended this to the case of equality testing of two median residual life functions for censored survival data. However, these methods are

not applicable when data are collected through prevalence sampling. We propose two test statistics based on the two estimation approaches discussed in Chapter 2. We compare the performance of the two statistics based on the Type I error and power.

3.2 THE TWO-SAMPLE TEST FOR QUANTILE RESIDUAL LIFE

3.2.1 Notation

Suppose $T_j^0, j = 1, 2$, denotes the variables representing the time between the two events (initial and terminating) in the j^{th} parent population. Suppose $Q_j \equiv Q_j(q, t_0)$ denote the $100q$th $(0 \leq q \leq 1)$ quantile residual life function for the j^{th} population at time t_0. In other words, Q_j is the $100q$th percentile of the distribution of residual life $T - t_0$ among patients in the j^{th} population who survived longer than t_0. Our goal is to test the null hypothesis

$$H_0 : Q_1 = Q_2$$

for fixed t_0 and to construct confidence intervals for functions of Q_1 and Q_2, $g(Q_1, Q_2)$, such as the difference of the two QRLs

$$g(Q_1, Q_2) = Q_1 - Q_2,$$

or the ratio of two QRLs

$$g(Q_1, Q_2) = \frac{Q_1}{Q_2}.$$

The data set-up is similar to that in Chapter 2, except that all the variables will be indexed by j to indicate the respective population the data are being sampled from. In brief, for the $j^{th}(j = 1, 2)$ population,we assume:

Symbol		Description
T_j^0	=	Time between initial event and terminating event in population j
$S_j(t)$	=	Survival function of T_j^0 at time t
Q_j	=	$Q_j(q, t_0)$ QRL of population j at time t_0
W_j	=	Calender time of initial event
A_j	=	Time between initial event and recruitment
V_j	=	Time between recruitment and terminating event
C_j	=	Time from recruitment until the individual is censored
n_j	=	Sample size for the jth group

Thus the observed data in this two-sample case consist of the following random vectors, $(W_{ji}, A_{ji}, \widetilde{V_{ji}}, \Delta_{ji})$ where $\widetilde{V_{ji}} = min(V_{ji}, C_{ji})$ and $\Delta_{ji} = I(V_{ji} \leq C_{ji})$, $j = 1, 2; i = 1, ..., n_j$. We make the following assumptions:

1. The samples are independent. That is, $(W_{1i}, A_{1i}, \widetilde{V_{1i}}, \Delta_{1i}, i = 1, ..., n_1)$ is independent of $(W_{2i}, A_{2i}, \widetilde{V_{2i}}, \Delta_{2i}, i = 1, ..., n_2)$.
2. The Residual censoring times C_{ji}'s are independent of (W_{ji}, T_{ji}) for $j = 1, 2$.

As in Chapter 2, define $Y_{ji} = A_{ji} + \widetilde{V_{ji}} = min(T_{ji}, C_{ji}^*)$ to be the observed time to the terminating event or censoring, whichever occurs first, from the initial event.

3.2.2 Estimators for Q_j

Following the derivations in Chapter 2, $Q_j, j = 1, 2$, satisfies the equation

$$u_j(Q_j) = S_j(t_0 + Q_j) - (1 - q)S_j(t_0) = 0, j = 1, 2. \tag{3.1}$$

Or, equivalently,

$$Q_j = S_j^{-1}((1 - q)S_j(t_0)) - t_0, j = 1, 2,$$

where $S_j^{-1}(.)$ is the inverse function of $S_j(.)$. Thus as long as $S_j(t)$ can be consistently estimated, we can estimate Q_j by plugging it into the right hand side of the above equation. In Chapter 2 we have shown two methods (truncation product limit estimator and the

length-bias corrected estimator) for consistently estimating the survival distribution from length-biased censored survival data. We restate the two estimators here, except that they are now indexed by $j(j = 1, 2)$ to indicate respective populations.

The TPL estimator of $S_j(.)$ from the length-biased data is calculated as

$$\hat{S}_j(t) = \prod_{u\in[0,t]}\{1 - d\hat{\Lambda}_j(u)\}, \tag{3.2}$$

where

$$\hat{\Lambda}_j(t) = \int_0^t \frac{d\bar{N}_j(u)}{\bar{R}_j(u)} \tag{3.3}$$

with $\bar{N}_j(u) = n_j^{-1}\sum_{k=1}^{n_j}\Delta_{jk}I(Y_{jk} \le u)$ and $\bar{R}_j(u) = n_j^{-1}\sum_{k=1}^{n_j} I(Y_{jk} \ge u \ge A_{jk})$.

On the other hand, the survival estimator proposed by Huang and Qin (2011) is given by

$$\hat{S}_j{}^*(t) = \prod_{u\in[0,t]}\{1 - d\tilde{\Lambda}_j(u)\} \tag{3.4}$$

with

$$\tilde{\Lambda}_j(t) = \int_0^t \frac{d\bar{N}_j(u)}{\tilde{R}_j(u)}, \tag{3.5}$$

where

$$\tilde{R}_j(u) = n_j^{-1}\sum_{i=1}^{n_j} I(Y_{ji} \ge u) - \tilde{S}_A^j(u),$$

$$\tilde{S}_A^j(u) = \prod_{t\in[0,u]}\left\{1 - \frac{d\tilde{B}_j(t)}{\tilde{K}_j(t)}\right\},$$

$$\tilde{B}_j(t) = n_j^{-1}\sum_{i=1}^{n_j}\{I(A_{ji} \le t) + \Delta_{ji}I(\tilde{V}_{ji} \le t)\},$$

and

$$\tilde{K}_j(t) = n_j^{-1}\sum_{i=1}^{n_j}\{I(A_{ji} \ge t) + I(\tilde{V}_{ji} \ge t)\}.$$

The above two estimators are then plugged into the Equation (3.1) and solved for Q_j to obtain an estimator of Q_j. Thus the TPL estimator of Q_j is the solution of the equation

$$\hat{u_j}(Q_j) = 0,$$

where

$$\hat{u_j}(Q_j) = \hat{S}_j(t_0 + \hat{Q}_j) - (1-q)\hat{S}_j(t_0).$$

Or, equivalently,

$$\hat{Q}_j = \hat{S}_j^{-1}((1-q)\hat{S}_j(t_0)) - t_0$$

, where $\hat{S}_j^{-1}(.)$ is the inverse of the estimated survival function defined in Chapter 2. Similarly, we define the estimator $\hat{Q}_j^*$ of Q_j as the solution of the equation,

$$\hat{u}^*(Q_j) = 0$$

for Q_j where

$$\hat{u_j}^*(Q_j) = \hat{S}_j^*(t_0 + Q_j) - (1-q)\hat{S}_j^*(t_0).$$

In other words,

$$\hat{Q}_j^* = \hat{S}_j^{*-1}((1-q)\hat{S}_j^*(t_0)) - t_0.$$

According to the results provided in Chapter 2, $\hat{Q}_j$ and $\hat{Q}_j^*$ are consistent and asymptotically normal for large n_j, for $j = 1, 2$. Therefore, for testing $H_0 : Q_1 = Q_2$ one may wish to construct a usual Wald test such as

$$\frac{\hat{Q}_1 - \hat{Q}_2}{SE(\hat{Q}_1 - \hat{Q}_2)}. \tag{3.6}$$

Unfortunately estimating standard error of $\hat{Q}_1 - \hat{Q}_2$ is not straightforward as it involves density estimates from censored survival data that are highly unstable (Padgett, 1984 [26]). Here we propose to test the hypothesis based on the distribution of $\hat{u}_j(Q_j)$, which does not require density estimates. Our derivation of the test statistic follows that of Su and Wei (1993) and Jeong et al. (2008).

3.2.3 Proposed Statistic

We recall from Chapter 2 that $\sqrt{n_j}\hat{u}_j(Q_{j0})$ at the true value Q_{j0} is asymptotically distributed as Normal with mean zero and variance

$$V_j(Q_{j0}) = \sigma_j^2(t_0 + Q_{j0}) + (1-q)^2\sigma_j^2(t_0) - 2(1-q)\Sigma_j(t_0 + Q_{j0}, t_0), j = 1, 2, \tag{3.7}$$

where $\Sigma_j(t_1, t_2) = E\{\phi_i^j(t_1)\phi_i^j(t_2)\}$, $0 \leq t_1, t_2 \leq \tau$ with $\sigma_j^2(t) = \Sigma_j(t, t)$. The function $\phi_i^j(t)$ is the influence function of the estimator $\hat{S}_j(t)$, and as shown in Chapter 2 is given by

$$\phi_i^j(t) = \left[\int_0^t R_j(u)^{-2} I(Y_{ji} \geq u \geq A_{ji}) dF_j^u(u) - \frac{\Delta_{ji} I(Y_{ji} \leq t)}{R_j(Y_{ji})}\right] S_j(t),$$

where $F_j^u(t) = \text{pr}(\Delta_j = 1, Y_j \leq t)$ is the sub distribution function of complete observations.

For known Q_{j0}, this variance can be estimated by

$$\hat{V}_j(Q_{j0}) = \hat{\Sigma}(t_0 + Q_{j0}, t_0 + Q_{j0}) + (1-q)^2\hat{\Sigma}(t_0, t_0) - 2(1-q)\hat{\Sigma}(t_0 + Q_{j0}, t_0),$$

where

$$\hat{\Sigma}(t_1, t_2) = \frac{1}{n_j}\Sigma_{i=1}^{n_j}\hat{\phi_i^j}(t_1)\hat{\phi_i^j}(t_2)$$

with

$$\hat{\phi_i^j}(t) = \hat{S}_j(t)\left[\int_0^t \frac{I(Y_{ji} \geq u \geq A_{ji}) d\bar{N}_j(u)}{\tilde{R_j}^2(u)} - \frac{\Delta_{ji} I(Y_{ji} \leq t)}{\tilde{R}_j(Y_i)}\right]$$

.

Similarly, $\sqrt{n_j}\hat{u}^*(Q)_{j0}$ is asymptotically distributed as Normal with mean zero and variance

$$V_j^*(Q_{j0}) = \sigma_j^{*2}(t_0 + Q_{j0}) + (1-q)^2\sigma_j^{*2}(t_0) - 2(1-q)\Sigma_j^*(t_0 + Q_{j0}, t_0), j = 1, 2, \tag{3.8}$$

where $\Sigma_j^*(t_1, t_2) = E\{\phi_i^{j^*}(t_1)\phi_i^{j^*}(t_2)\}$, $0 \leq t_1, t_2 \leq \tau$ with $\sigma_j^{*2}(t) = \Sigma_j^*(t, t)$. The function $\phi_i^{j^*}(t)$ is the influence function of the estimator $\hat{S}_j^*(t)$, and is given by

$$\phi_i^{j^*}(t) = \phi_i^j(t) + \int_0^t R_j(u)^{-2}\{I(A_{ji} > u) - S_A^j(u) - S_A^j(u)\phi_i^j(u)\} dF_j^u(u),$$

$i = 1, 2....., n$, are i.i.d random variable with mean zero and the covariance function

$$\Sigma_j^*(t_1, t_2) = E\left[\phi_i^{j^*}(t_1)\phi_i^{j^*}(t_2)\right].$$

$$\hat{V_j}^*(Q_{j0}) = \hat{\Sigma}^*(t_0 + Q_{j0}, t_0 + Q_{j0}) + (1-q)^2\hat{\Sigma}^*(t_0, t_0) - 2(1-q)\hat{\Sigma}^*(t_0 + Q_{j0}, t_0),$$

$$\hat{\Sigma}^*(t_1, t_2) = \frac{1}{n_j}\sum_{i=1}^{n_j} \hat{\phi_i^j}^*(t_1)\hat{\phi_i^*}^*(t_2),$$

with

$$\hat{\phi_i^j}^*(t) = \hat{S}_j^*(t)\left[\int_0^t \frac{I(Y_{ji} \geq u \geq A_{ji})d\bar{N}_j(u)}{\tilde{R_j}^2(u)} - \frac{\Delta_{ji}I(Y_j i \leq t)}{\tilde{R}_j(Y_{ji})} + \int_0^t \left\{\frac{I(A_{ji} > u}{\tilde{R_j}^2(u)} - \tilde{S_A^j}(u) - \tilde{S_A^j}(u)\tilde{\psi_i^j}(u)\right\} d\bar{N}_j(u)\right],$$

where

$$\tilde{\psi_i^j}(u) = \int_0^t \frac{\{I(A_{ji} \geq u) + I(\tilde{V_{ji}} \geq u)\}d\tilde{B}_j(u)}{\tilde{K_j}^2(u)} - \frac{I(A_{ji} \leq t)}{\tilde{K_j}(A_i)} - \frac{\Delta_{ji}I(\tilde{V_{ji}} \leq t)}{\tilde{K_j}(\tilde{V_{ji}})}.$$

We consider the TPL test statistic

$$W(Q_{10}, Q_{20}) = \frac{n_1\hat{u_1}^2(Q_{10})}{\hat{V}_1(Q_{10})} + \frac{n_2\hat{u_2}^2(Q_{20})}{\hat{V_2}(Q_{20})}. \tag{3.9}$$

Under $H_0 : \gamma = \gamma_0$ where, $\gamma = \frac{Q_{20}}{Q_{10}}$ and γ_0 is a known value,

$$W(Q_{20}) = \frac{n_1\hat{u_1}^2(\gamma_0 Q_{20})}{\hat{V_1}(\gamma_0 Q_{20})} + \frac{n_2\hat{u_2}^2(Q_{20})}{\hat{V_2}(Q_{20})}, \tag{3.10}$$

which depends on the true value Q_{20}. If Q_{20} were known, this statistic would asymptotically follow a χ^2 distribution with two degrees of freedom as it is the sum of squares of two independent standard Normal random variables. Since the true value Q_{20} is unknown, Su and Wei (1993) suggested minimizing $W(Q_{20})$ over the possible support of Q_{20}. Following their argument, under H_0,

$$W = \inf_{Q_{20}} W(Q_{20}) \tag{3.11}$$

follows a χ_1^2 distribution as one degree of freedom has been lost due to the minimization of Q_{20}.

Similarly, based on the Huang and Qin method, we define the statistic

$$W^* = \inf_{Q_{20}} W^*(Q_{20}), \tag{3.12}$$

where

$$W^*(Q_{20}) = \frac{n_1 \hat{u_1^*}^2(Q_{20})(\gamma_0 Q_{20})}{\hat{V_1}^*(Q_{20})(\gamma_0 Q_{20})} + \frac{n_2 \hat{u_2^*}^2(Q_{20})}{\hat{V_2}^*(Q_{20})} \tag{3.13}$$

and $W^*(Q_{20})$ follows a χ_1^2 distribution based on the same logic as before.

We investigate the properties of the two test statistics W and W^* in Equations (3.11) and (3.12) via simulation in the next section.

3.3 SIMULATION STUDY

3.3.1 Data Generation

Data were generated from a Weibull distribution with scale and shape parameters (α_1, β_1) and (α_2, β_2) for the two populations, where the values of $\alpha_1, \beta_1, \alpha_2$, and β_2 were varied to investigate various shapes of the distributions. The Weibull distribution with parameters $(\alpha_j, \beta_j), j = 1, 2$ is given by the survival function

$$S_j(t) = \exp\left\{-\left(\frac{t}{\alpha_j}\right)^{\beta_j}\right\}. \tag{3.14}$$

For all the simulation scenarios described below, we first generated 1000 pairs of Monte-Carlo samples of sizes n_1 and n_2. To generate n_1 observations from the first population we followed the following procedures. We set the sampling time (recruiting time) ξ to be 100. The time of disease onset W^0 was generated from a Uniform distribution over the interval $[0, 100]$. The survival time T^0 was independently generated from a Weibull distribution with the survival function

$$S_1(t) = \exp\left\{-\left(\frac{t}{\alpha_1}\right)^{\beta_1}\right\}. \tag{3.15}$$

To form a prevalent cohort of sample size n_1, the pair (W^0, T^0) were generated repeatedly until there were n_1 pairs of observations satisfying the sampling constraint $W^0 + T^0 \geq \xi$. The

residual censoring time C_1 from enrollment to loss to follow-up was generated from a Uniform distribution with a support (θ_1, θ_2), where θ_1 and θ_2 were varied to investigate varying censoring rates. We followed the same procedure to generate n_2 length-biased observations from the second population with Weibull parameters (α_2, β_2) .

We conducted the test of equality of median residual life at time points 0.0, 0.5, and 1.0. For each pair of Monte-Carlo samples, we calculated the test statistics W from Equation (3.11) and W^* from Equation (3.12). If the statistic is larger than the critical value $\chi^2_{1,\alpha}$ (at $100\alpha\%$ level of significance), we reject the null hypothesis. For the 1000 sample pairs, we count the number of sample pairs for which the null hypothesis is rejected. The proportion of times the null hypothesis is rejected provides an estimate of the Type I error of the test, if underlying populations satisfy the null hypothesis (median residual lifes for the two populations are identical). On the other hand, it provides an estimate of power of the test if the two population medians are not identical.

3.3.2 Simulation from the Null: Type I error

First we generated the data under the null hypothesis $H_0 : \gamma = 1$, that is the two population median residual life times are identical. To ensure this, in the first scenario we chose $\alpha_1 = \alpha_2 = 2$, and $\beta_1 = \beta_2 = 2$ so that in the population median residual lifes for both populations are the same. These medians are 1.67, 1.24, and .97 at 0.0, 0.5, and 1.0, respectively. Table 5 provides the estimated Type I errors for testing the equality of medians of the residual life at time points 0.0, 0.5, and 1.0 for both truncation product limit (TPL) and Huang-Qin (HQ) approaches. The censoring distribution parameters were taken to be $\theta_1 = 1$ and $\theta_2 = 2$ to give about 30% censored cases, and to be $\theta_1 = 0$ and $\theta_2 = 2$ to give about 51% censored cases for both samples. We conducted the tests at $\alpha = 0.05$ level of significance.

From the results in Table 5, we can see that empirical Type I errors are close to the nominal level of 0.05, ranging between 0.033 and 0.064 for the TPL statistic W, and between 0.026 and 0.068 for W^*. For earlier time points, both W and W^* are conservative in rejecting the null hypothesis with empirical Type I errors being lower than the nominal level. This may be due to the lack of variability in the survival estimates in the early time points.

The tests maintain Type I errors well irrespective of censoring percentages and the sample sizes, although sample sizes smaller than 100 tended to produce slightly unstable results (not shown here) under the same levels of censoring. As can be seen, empirical Type I errors are smaller for the test statistic W^* compared to W.

Table 5: Simulation results under the null hypothesis for testing $H_0 : \gamma = 1$ versus $H_1 : \gamma \neq 1$ under **moderate and heavy censoring**. Empirical Type I errors are based on 1000 Monte-Carlo sample pairs of sizes $n_1 = n_2 = n$ and $\alpha_1 = 2$, $\alpha_2 = 2$, and $\beta_1 = 2$, $\beta_2 = 2$ in samples 1 and 2 respectively.

Censored	t_0	n	W	W^*
	0.0		0.045	0.035
	0.5	200	0.047	0.052
	1.0		0.063	0.067
30%	0.0		0.040	0.041
	0.5	400	0.036	0.045
	1.0		0.056	0.066
	0.0		0.033	0.026
	0.5	200	0.040	0.036
	1.0		0.051	0.044
51%	0.0		0.037	0.035
	0.5	400	0.044	0.046
	1.0		0.064	0.068

We generated the data under other null distributions. For example, Table 6 provides the results for $\alpha_1 = \alpha_2 = 2.5$, and $\beta_1 = \beta_2 = 2$ so that the population median residual life for both populations are 2.08, 1.64, and 1.31 at 0.0, 0.5, and 1.0, respectively. The censoring distribution parameters were taken to be $\theta_1 = 1$ and $\theta_2 = 2$ to give about 40% censored cases, and to be $\theta_1 = 0$ and $\theta_2 = 2$ to give about 59% censored cases for both samples. We conducted the tests at $\alpha = 0.05$ level of significance. In this scenario, for the test statistic W, estimated Type I error ranged between 2.7% to 5.4% under 40% censoring and between

Table 6: Simulation results under the null hypothesis for testing $H_0 : \gamma = 1$ versus $H_1 : \gamma \neq 1$ under **moderate and heavy censoring**. Empirical Type I errors are based on 1000 Monte-Carlo sample pairs of sizes $n_1 = n_2 = n$ and $\alpha_1 = 2.5$, $\alpha_2 = 2.5$, and $\beta_1 = 2$, $\beta_2 = 2$ in samples 1 and 2 respectively.

Censored	t_0	n	W	W^*
	0.0		0.027	0.029
	0.5	200	0.029	0.029
	1.0		0.039	0.048
40%	0.0		0.038	0.035
	0.5	400	0.035	0.044
	1.0		0.054	0.069
	0.0		0.034	0.033
	0.5	200	0.035	0.039
	1.0		0.039	0.052
59%	0.0		0.039	0.041
	0.5	400	0.037	0.047
	1.0		0.047	0.053

3.4% to 4.7% under 59% censoring. For the test statistic W^*, estimated Type I error ranged between 2.9% to 6.9% for moderate censoring and between 3.3% to 5.3% for heavy censoring. An increase in sample size had no significant impact on the Type I error.

Table 7 provides the results for $\alpha_1 = \alpha_2 = 3$, and $\beta_1 = \beta_2 = 2$ so that the population median residual life for both populations are 2.5, 2.0, and 1.7 at 0.0, 0.5, and 1.0, respectively. The censoring distribution parameters were taken to be $\theta_1 = 1$ and $\theta_2 = 2$ to give about 48% censored cases, and to be $\theta_1 = 0$ and $\theta_2 = 2$ to give about 65% censored cases for both samples. As before, the level of significance was set at $\alpha = 0.05$. Note that the null survival distribution here is more skewed to the right than the scenarios presented in Table 5 or in Table 6, which is why censoring rates are higher than those two cases, even though we used

Table 7: Simulation results under the null hypothesis for testing $H_0 : \gamma = 1$ versus $H_1 : \gamma \neq 1$ under **moderate and heavy censoring**. Empirical Type I errors are based on 1000 Monte-Carlo sample pairs of sizes $n_1 = n_2 = n$ and $\alpha_1 = 3$, $\alpha_2 = 3$, and $\beta_1 = 2$, $\beta_2 = 2$ in samples 1 and 2 respectively.

Censored	t_0	n	W	W^*
	0.0		0.037	0.030
	0.5	200	0.029	0.022
	1.0		0.040	0.039
48%	0.0		0.041	0.036
	0.5	400	0.038	0.039
	1.0		0.056	0.055
	0.0		0.029	0.027
	0.5	200	0.028	0.026
	1.0		0.042	0.040
65%	0.0		0.040	0.038
	0.5	400	0.045	0.045
	1.0		0.048	0.048

the same censoring distributions across the tables.

The results are very similar to the ones shown in Tables 5 and 6. Basically both tests are conservative, with estimated Type I error being smaller than the nominal level of significance for smaller sample sizes. For larger sample sizes, the empirical rejection rates are very close to the nominal level, irrespective of the level of censoring. Generally tests of equality of median residual life at time point 1 showed better performance in terms of matching the nominal significance level.

3.3.3 Power of the Test: Simulation under the Alternative

To investigate the power of the proposed tests under various conditions, we have generated data from alternative hypotheses, varying the parameter values, censoring rates, and sample sizes. The data are generated from the Weibull distribution as outlined above, except that to investigate power we chose different parameter combinations for the two samples to ensure that γ_1 is not equal to 1.

Table 8: Simulation results under the alternative hypothesis for testing $H_0 : \gamma = 1$ versus $H_1 : \gamma = \gamma_1$ under **moderate censoring**. Empirical power (rejection rates) are based on 1000 Monte-Carlo sample pairs of sizes $n_1 = n_2 = n$ and $\alpha_1 = 2.5$, $\alpha_2 = 2$, and $\beta_1 = 2$, $\beta_2 = 2$. Approximately 40% and 30% observations are censored in samples 1 and 2 respectively.

n	t_0	$Q_1(0.5, t_0)$	$Q_2(0.5, t_0)$	γ_1	W	W^*
200	0.0	2.1	1.7	1.25	0.552	0.621
	0.5	1.6	1.2	1.32	0.618	0.689
	1.0	1.3	0.9	1.39	0.803	0.845
400	0.0	2.1	1.7	1.25	0.837	0.899
	0.5	1.6	1.2	1.32	0.895	0.940
	1.0	1.3	0.9	1.39	0.975	0.991

In Table 8 we present the empirical power (proportion of Monte-Carlo sample pairs for which the null hypothesis is rejected) for the first sample generated from the Weibull distribution with parameters $(\alpha_1 = 2.5, \beta_1 = 2)$, and the second sample generated from the Weibull $(\alpha_2 = 2, \beta_2 = 2)$. The true median residual life function at times $(0.0, 0.5, 1.0)$ were $(2.1, 1.6, 1.3)$ for the first population and $(1.7, 1.2, 0.9)$ for the second population, resulting in the true ratio $(1.25, 1.32, 1.39)$. For the results presented, the censoring distribution was assumed to be Uniform with parameters $(\theta_1 = 1, \theta_2 = 2)$ to give about 40% censored cases in the first sample and 30% in the second sample.

The results show that the power of the test increases as the ratio of the two median residual life (effect size) increases, which is expected by the theory of statistical power.

For example, with sample size 200 for each sample, the empirical power of the W test is $(0.55, 0.62, 0.80)$ for the QRL ratios $(1.25, 1.32, 1.39)$ at time points $(0.0, 0.5, 1.0)$ respectively. Corresponding empirical powers for W^* test are respectively $(0.62, 0.69, 0.85)$. The empirical power increases as the sample sizes are increased from 200 to 400. For this sample size under the same censoring rates, the power ranged between 0.84 and 0.98 for W test and between 0.90 and 0.99 for W^* test.

Table 9: Simulation results under the alternative hypothesis for testing $H_0 : \gamma = 1$ versus $H_1 : \gamma = \gamma_1$ under **heavy censoring**. Empirical powers (rejection rates) are based on 1000 Monte-Carlo sample pairs of sizes $n_1 = n_2 = n$ and $\alpha_1 = 2.5$, $\alpha_2 = 2$, and $\beta_1 = 2$, $\beta_2 = 2$. Approximately 59% and 51% observations are censored in samples 1 and 2 respectively.

n	t_0	$Q_1(0.5, t_0)$	$Q_2(0.5, t_0)$	γ_1	W	W^*
200	0.0	2.1	1.7	1.25	0.413	0.483
	0.5	1.6	1.2	1.32	0.448	0.522
	1.0	1.3	0.9	1.39	0.656	0.723
400	0.0	2.1	1.7	1.25	0.715	0.809
	0.5	1.6	1.2	1.32	0.775	0.849
	1.0	1.3	0.9	1.39	0.909	0.954

Table 9 presents the empirical power for the tests for samples generated under the same scenario as in Table 8 except that the censoring distribution was assumed to be uniform with parameters $(\theta_1 = 0, \theta_2 = 2)$ to have a heavier censoring rate of about 59% censored cases in the first sample and 51% in the second sample. Naturally, because of higher censoring rates, the power is smaller compared to that presented in Table 8. The results show a similar trend as in Table 8 that the power of the test increases as the ratio of the median residual lifes (effect size) increases and the empirical power increases as the sample sizes increase, except that both tests have reduced power due to the higher censoring rates. In both scenarios, the W^* test seems to consistently out-power the W test regardless of sample size, censoring rates, or effect size.

Table 10: Simulation results under the alternative hypothesis for testing $H_0 : \gamma = 1$ versus $H_1 : \gamma = \gamma_1$ under **moderate censoring**. Empirical power (rejection rates) is based on 1000 Monte-Carlo sample pairs of sizes $n_1 = n_2 = n$ and $\alpha_1 = 3$, $\alpha_2 = 2$, and $\beta_1 = 2$, $\beta_2 = 2$. Approximately 48% and 30% observations are censored in samples 1 and 2 respectively.

n	t_0	$Q_1(0.5, t_0)$	$Q_2(0.5, t_0)$	γ_1	W	W^*
200	0.0	2.5	1.7	1.50	0.949	0.966
	0.5	2.0	1.2	1.65	0.974	0.985
	1.0	1.7	0.9	1.80	0.999	1.000
400	0.0	2.5	1.7	1.50	>0.999	>0.999
	0.5	2.0	1.2	1.65	>0.999	>0.999
	1.0	1.7	0.9	1.80	>0.999	>0.999

Tables 10 and 11 shows the empirical power for first sample generated from the Weibull distribution with parameters $(\alpha_1 = 3.0, \beta_1 = 2)$, and the second sample generated from the Weibull $(\alpha_2 = 2, \beta_2 = 2)$. The true median residual life function at times $(0.0, 0.5, 1.0)$ was $(2.5, 2.0, 1.7)$ for the first population and $(1.7, 1.2, 0.9)$ for the second population, resulting in the ratio $(1.50, 1.65, 1.80)$. In Table 10, the censoring distribution was assumed to be Uniform with parameters $(\theta_1 = 1, \theta_2 = 2)$ to give about 48% censored cases in the first sample and 30% in the second sample. On the other hand, Table 11 represents a heavier censoring scenario where the censoring distribution was assumed to be Uniform with parameters $(\theta_1 = 0, \theta_2 = 2)$ to give about 65% censored cases in the first sample and 51% in the second sample.

These results show a similar trend as those seen in Tables 8 and 9 except that because of the larger effect sizes, the empirical power is larger than that presented in those tables. The power of the test increases as the ratio of the median residual lifes (effect size) or sample size increases. Similar to the results shown before, the W^* test seems to consistently out-power the W test regardless of the sample size, censoring rates, or effect size, but with larger sample size of 400, under moderate censoring, the effect sizes of $1.5 - 1.8$ resulted in almost identical empirical power.

Table 11: Simulation results under the alternative hypothesis for testing $H_0 : \gamma = 1$ versus $H_1 : \gamma = \gamma_1$ under **heavy censoring**. Empirical power (rejection rates) is based on 1000 Monte-Carlo sample pairs of sizes n_1 and n_2 and $\alpha_1 = 3$, $\alpha_2 = 2$, and $\beta_1 = 2$, $\beta_2 = 2$. C_1 and C_2 are percent censored in samples 1 and 2 respectively.

n	t_0	$Q_1(0.5, t_0)$	$Q_2(0.5, t_0)$	γ_1	W	W^*
200	0.0	2.5	1.7	1.50	0.866	0.914
	0.5	2.0	1.2	1.65	0.902	0.952
	1.0	1.7	0.9	1.80	0.975	0.991
400	0.0	2.5	1.7	1.50	0.993	0.997
	0.5	2.0	1.2	1.65	0.995	>.999
	1.0	1.7	0.9	1.80	>0.999	>0.999

3.3.4 Impact of Sample Size and Alternative Ratio on Power

To investigate the impact of sample size on the power of the two tests, we have fixed the two populations at Weibull $(2.5, 2)$ and Weibull $(2, 2)$ respectively with the alternative value γ_1 being equal to 1.25, 1.32, and 1.40 at times 0.0, 0.5, and 1.0 respectively. Then keeping the same censoring rates (40% in the first sample and 30% in the second) across all the sample sizes ranging from 50 to 400, we computed the empirical power based on 1000 Monte-Carlo pairs of samples. The resulting empirical power is plotted against the sample size in Figure 4. This figure demonstrates that the power of both tests increases as the sample size increases, and the power is relatively larger for the W^* test compared to the W test.

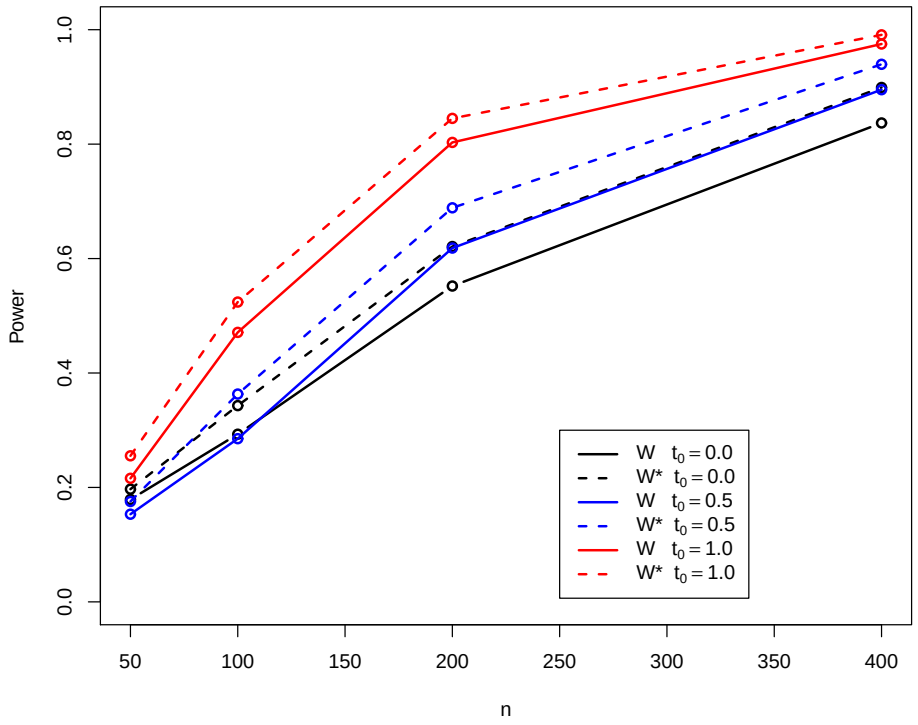

Figure 4: Empirical power of tests as a function of sample size. Empirical power is calculated by generating 1000 pairs of samples from Weibull $(2.5, 2)$ and Weibull $(2, 2)$ distributions respectively. The censoring rates are respectively 40% for the first sample and 30% for the second sample.

To investigate the impact of the effect size on the power of the two tests, we have fixed the sample size to 200 for each sample, however, the two populations were taken to be Weibull $(\alpha_1,\ \beta_1)$ and Weibull $(\alpha_2,\ \beta_2)$ respectively, where $(\alpha_j, \beta_j), j = 1, 2$ were chosen to obtain a range of values of the "effect size" γ_1 between 0 and 3. Then we computed the empirical powers based on 1000 Monte-Carlo pairs of samples. The resulting empirical power is plotted against the γ_1 in Figure 5. At all the three time points considered, the power shows a similar pattern across the valuses of γ_1, with values close to the null ($\gamma_1 = 1$) providing the lowest power (approximately equal to 0.05, the level of significance), and the power increasing as the γ_1 deviated from 1 in either direction. Again, at any given value of γ_1, the power is relatively larger for the W^* test compared to the W test.

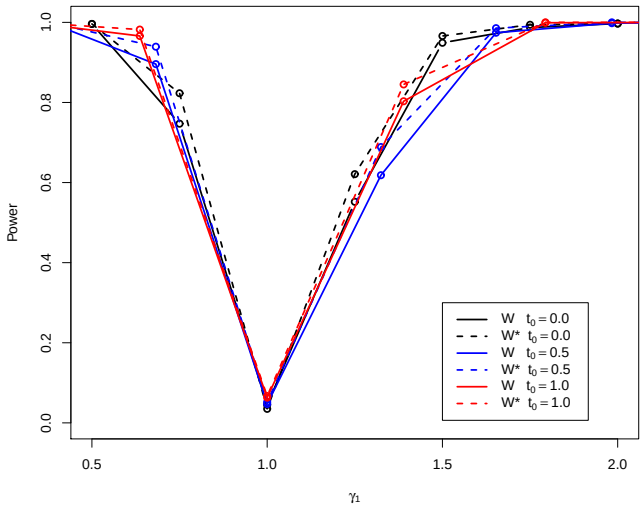

Figure 5: Empirical power of tests as a function of effect size.

3.4 DATA ANALYSIS

In this section we illustrate our method using the data from the National Surgical Adjuvant Breast and Bowel Project (NSABP) protocol B-20 clinical trial. The study is described in details in Chapter 2, where we used only one of the three arms (Tamoxifen only, TAM) to demonstrate the application of our methods to estimate the quantile residual life function after recurrence for breast cancer patients being treated with tamoxifen. Here we additionally included the Tamoxifen+ cyclophosphamide+sequential methotrexate+fluorouracil (following Fisher et al. 1997[11]) (CMFT arm) and compare the median residual life following recurrence across the two arms.

As described earlier, a total of 788 women were randomly assigned to receive Tamoxifen. Out of these 788 women, 170 women experienced recurrence. To construct a length-biased sample of patients, we identified the date of recurrence of the first patient who experienced recurrence and set the recruitment time to be at 5 years following the first recurrence date. By the recruitment date, 120 patients had died or lost-to-follow-up and hence would not be eligible to be in the sample. Thus, our length-biased sample consists of 50 patients who had recurrence prior to September 17, 1994 and were still being followed at the same date.

A total of 789 women were randomly assigned to receive CMFT. Out of these 789 women, 97 women experienced recurrence. Using the same recruitment date of September 17, 1994, we obtained a length-biased sample of 26 patients who had recurrence and were still being followed.

Table 12 provides the estimates of median residual lifes for the two arms at recurrence, half-year after recurrence, and one-year after recurrence using W as the test statistics. The median residual lifetimes are generally larger, by approximately a half-year, in the Tamoxifen arm compared to CMFT arm, however, the differences were not statistically significant.

Table 12: NSABP B-20 Data Analysis : Estimated median residual lifetimes along with 95% confidence intervals (using $\hat{Q}$) by treatment. The confidence interval for the ratio, and the p-value for test is computed using W.

	$\hat{Q}(0.5, t_0)$			
t_0	TAM	$CMFT$	$Ratios$	$p-value$
0.0	3.11	2.51	1.24	0.66
	(1.49 - 10.3)	(1.19-4.45)	(0.42 - 1.72)	
0.5	2.61	2.01	1.30	0.66
	(0.99-9.82)	(0.69-3.95)	(0.3-2.08)	
1.0	2.11	1.57	1.34	0.92
	(0.49-9.32)	(0.86-5.15)	(0.44 - 5.0)	

Table 13 provides the estimates of median residual lifes for the two arms at recurrence using the other method. The point estimates of median residual life since recurrence differed between the two methods (Table 12 vs. Table 13). However the estimate of the ratio of median residual lifes using two methods was similar, as are the p-values, inferring that there were no significant differences across the two arms in terms of median residual life since recurrence, half-year or one year after recurrence.

In Figure 6, we plotted estimated median residual life function for breast cancer patients with recurrence over time by treatment arm. Generally the median residual life is longer for the tamoxifen group compared to the CMFT group at almost all time points.

Table 13: NSABP B-20 Data Analysis : Estimated median residual lifetimes along with 95% confidence intervals (using $\hat{Q}^*$) by treatment. The confidence interval for the ratio, and the p-value for test is computed using W^*.

	$\hat{Q}(0.5, t_0)$			
t_0	TAM	$CMFT$	$Ratio$	$p-value$
0.0	2.36	2.04	1.16	0.56
	(1.42 - 4.34)	(1.06-3.66)	(0.2 - 2.44)	
0.5	1.86	1.54	1.21	0.56
	(0.92-3.84)	(0.56-3.16)	(0.12 - 3.18)	
1.0	1.36	1.09	1.24	0.96
	(0.42-3.34)	(0.19-3.29)	(0.14 - 5.0)	

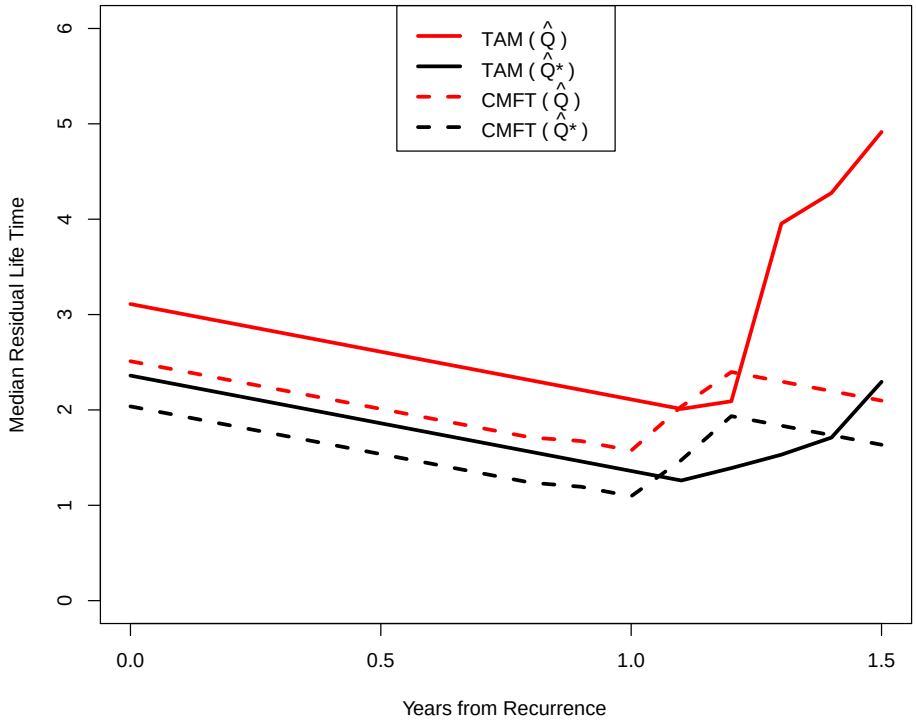

Figure 6: Estimated median residual life function plotted over time for NSABP B-20 data by treatment arm.

3.5 DISCUSSION

In the presence of length-biased data which are left truncated, regular Kaplan-Meier curves are not applicable (Klein and Moeschberger, chapter 4, section 6, [24]) and hence a test based on the Kaplan-Meier curve can be misleading. In this chapter we have proposed two test statistics for testing the equality of quantiles of residual life functions for length-biased right-censored survival data. One test statistic W is based on the truncation product-limit estimator of survival distribution, and the other statistic W^* based on an estimator of the survival distribution (Huang and Qin, 2011) that takes into account the fact that the marginal distribution of the truncation time is the same as that of the residual survival time (time since recruitment to terminating event). The two tests both follow an asymptotic χ^2_1 distribution. Our simulation studies demonstrated that for reasonable sample sizes, both tests maintain Type I error under various censoring proportions ranging as high as 65%. When sample sizes are small, e.g., $n <= 100$ per group, the test statistics do not perform well with Type I error being much smaller than the nominal level. Moreover, when the sample sizes are unequal, both tests require larger sample sizes per group than when the sample sizes are equal. Generally, the power for the W^* test is larger than that for the W test.

Length-biased data assume that the initial event occurs at a stable rate, which means that the truncation time given the overall survival is uniformly distributed. A consequence of this result is that the marginal distribution of the truncation time (A) is the same as that of the residual survival time (V). Asgharian et al. (2006) suggested checking this stationarity assumption by comparing the estimated survival distribution of A and V graphically. We presented the survival curves for A and V for the NSABP B-20 data in Figure 7. These estimates are obtained using regular Kaplan-Meier method. Unfortunately the two survival curves are not similar for this data set. This may be due to the fact that the recurrence (initial event) rates may not be stable over time in this population and hence it is likely that the stationarity assumption is violated.

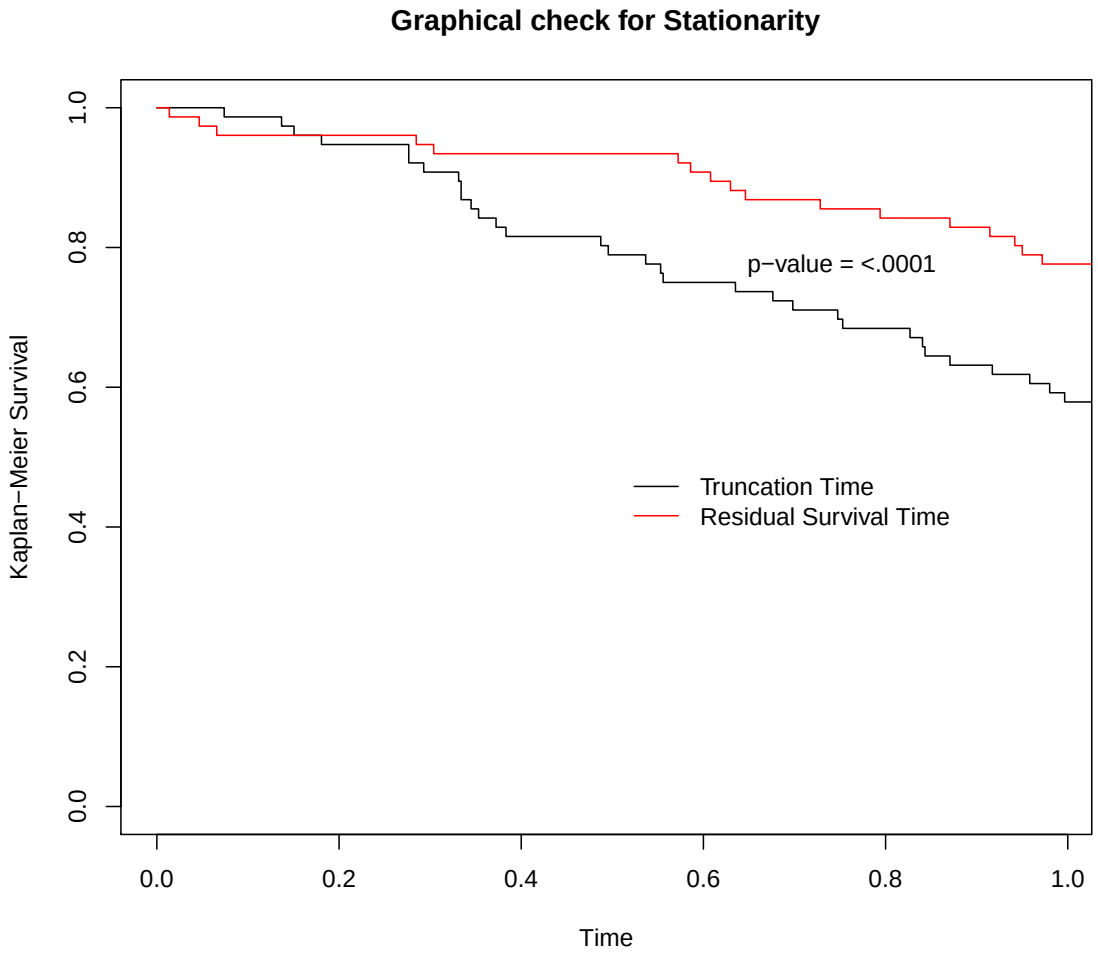

Figure 7: Graphical check for stationarity assumption. The red curve is the survival distribution of the forward survival (time since recruitment) to the terminating event or censoring; the black curve is the survival distribution of the truncation time or backward survival time.

4.0 CONCLUSION AND DISCUSSION

In a study of the natural history of a disease (e.g., cancer), it is often of interest to estimate the survival from the onset of a non-terminal initial event (e.g. recurrence) to a terminal event (death). For example, in a study of hepatitis B, it may be of interest to estimate the survival distribution of a population of patients who acquired hepatitis B; or among patients with hepatitis B, the interest may be to estimate the median time between the occurrence of hepatocellular carcinoma (HCC) and death. To estimate such quantities, one needs to draw a sample of individuals from the respective population. One possible way to obtain such a sample is to start with a random sample of individuals without the initial event (disease, recurrence, HCC) from the population and then follow them to the occurrence of the initial event. Individuals with the initial event are then followed until the occurrence of the terminal event. This process is known as incidence sampling. Incidence sampling renders several practical inconveniences for a researcher as it might take a very large sample to start with to ensure a reasonable number of individuals experience the initial event. Subsequently, it might take a long follow-up to ensure a reasonable number of these individuals to experience the terminal event in order to draw valid conclusions.

An alternative sampling scheme is prevalence sampling, where individuals who have already experienced the initial event but not the terminal event are recruited into the sample at a specific time and are followed until the terminal event. Prevalence sampling reduces the need for a large starting cohort, and possibility, the long follow-up time required in incidence sampling. However, prevalent sample is generally biased. The bias arises due to the fact that people who experienced the terminal event earlier than the recruitment date would not be included in the sample. In other words, individuals with shorter survival times

(time between initial and terminal event) would not be recruited into the sample, and hence the survival based on prevalence sample will generally be overestimated. This bias is called length bias. One important characteristic of length-biased data is that the initial event occurs at a constant rate, which usually is violated if there is an epidemic or if there is an improvement in the diagnosis of the disease.

Survival data from prevalence sampling is naturally left-truncated. Since the follow-up time is often limited, and drop-out may occur at any time during follow-up, the data may also be right-censored. In the last several decades, many authors have addressed the issues of left-truncation and right censoring in survival analysis (Wang et al., 1986[37]; Tsai et al., 1987[29]; and references therein). There is a version of the product-limit estimator (called the truncation product limit estimator [[22], [37]]) that provide an unbiased estimate of the survival function in the presence of left-truncation and independent right-censoring. However, this estimator does not take into account the uniform distribution of the left truncation times (or stationarity), and hence is inefficient. If the incidence rate remains constant over time (Wang, 1991)[35], i.e., if the stationarity assumption holds, the truncation time will follow a Uniform distribution and an estimator that incorporates the truncation time distribution is generally more efficient (Vardi, 1982[31], 1985[32], 1989[32], Gill et al., 1988[15], Vardi and Zhang (1992)[33], Asgharian et al., 2002[2], and Asgharian and Wolfson, 2005[3]) than truncation product limit estimator. Vardi 's (1982[31], 1985[32]) nonparametric maximum likelihood estimator (NPMLE) is not easy to implement while Huang and Qin 's (Huang and Qin, 2011[17]) new non-parametric estimator has a closed-form expression and is almost as efficient as the NPMLE.

In this dissertation, we provided methods for inference about quantile residual life function from right-censored length-biased data based on prevalence sampling. We used TPL and Huang and Qin survival curves to derive estimators of the quantiles of residual life distribution. Since it is not easy to obtain a stable variance estimator, we used a Brookmeyer-type approach (1982, [6]) to construct point-wise confidence intervals for quantiles of residual life. Simulation under various conditions indicated that the proposed confidence intervals maintain nominal coverage and the Huang and Qin approach provides narrower confidence intervals. We also proposed two tests based on the two approaches of estimation. Both tests

maintain Type I error. However, the Huang and Qin approach is generally more powerful than the TPL approach. We demonstrated our methods by estimating the quantiles of residual life functions following recurrence in breast cancer patients.

In this dissertation we used non-parametric methods to estimate the survival curves and draw inference about the quantiles of the residual life. Future research might assume parametric models for the survival distributions and then can use likelihood-based approaches to derive estimators and statistical tests. It would be interesting to see how such estimators or tests compare to the proposed non-parametric counterparts. One other extension would be to formulate regression-based approaches to adjust for covariates while comparing quantiles of residual life across groups.

There is important public health significance to the proposed work. As mentioned earlier, public health professional (PHPs) and clinicians are often interested in studying the time between the occurrence of an initial event and a terminating event, e.g., the time between recurrence and death, the time between the onset of depression and the remission. This helps PHPs develop policies for better treatment of individuals. For example, after recurrence of breast cancer, a women can make a better informed decision about treatment choices if she knows what her median residual life would be under different treatment options. Our methods provide accurate and efficient estimators of the median and quantiles in general of the residual life at a specific time point and hence can serve the purpose.

BIBLIOGRAPHY

[1] Andersen, P. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer-Verlag GmbH.

[2] Asgharian, M., C. E. M'Lan, and D. B. Wolfson (2002). Length-biased sampling with right censoring: an unconditional approach. *Journal of the American Statistical Association 97*(457), 201–209.

[3] Asgharian, M. and D. B. Wolfson (2005). Asymptotic behavior of the unconditional npmle of the length-biased survivor function from right censored prevalent cohort data. *The Annals of Statistics 33*(5), 2109–2131.

[4] Asgharian, M., D. B. Wolfson, and X. Zhang (2006). Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in medicine 25*(10), 1751–1767.

[5] Berger, R. L., D. D. Boos, and F. M. Guess (1988). Tests and confidence sets for comparing two mean residual life functions. *Biometrics*, 103–115.

[6] Brookmeyer, R. and J. Crowley (1982). A confidence interval for the median survival time. *Biometrics*, 29–41.

[7] Chen, Y. Q. and S. Cheng (2005). Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika 92*(1), 19–29.

[8] Cheng Wang, M. (1987). Product limit estimates: A generalized maximum likelihood study. *Communications in Statistics-Theory and Methods 16*(11), 3117–3132.

[9] Chiang, C. L. (1960). A stochastic study of the life table and its applications: I. probability distributions of the biometric functions. *Biometrics 16*(4), 618–635.

[10] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 187–220.

[11] Fisher, B., J. Dignam, B. Emir, J. Bryant, A. DeCillis, N. Wolmark, D. L. Wickerham, N. V. Dimitrov, N. Abramson, J. N. Atkins, et al. (1997). Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *Journal of the National Cancer Institute 89*(22), 1673–1682.

[12] Fisher, B., J.-H. Jeong, J. Bryant, S. Anderson, J. Dignam, E. R. Fisher, and N. Wolmark (2004). Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from national surgical adjuvant breast and bowel project randomised clinical trials. *The Lancet 364*(9437), 858–868.

[13] Fligner, M. A. and S. W. Rust (1982). A modification of mood's median test for the generalized behrens-fisher problem. *Biometrika 69*(1), 221–226.

[14] Gelfand, A. E. and A. Kottas (2003). Bayesian semiparametric regression for median residual life. *Scandinavian Journal of Statistics 30*(4), 651–665.

[15] Gill, R. D., Y. Vardi, and J. A. Wellner (1988). Large sample theory of empirical distributions in biased sampling models. *The Annals of Statistics*, 1069–1112.

[16] Hall, W. and J. A. Wellner (1981). Mean residual life. *Statistics and Related Topics 169*, 184.

[17] Huang, C.-Y. and J. Qin (2011). Nonparametric estimation for length-biased and right-censored data. *Biometrika 98*(1), 177–186.

[18] Jeong, J.-H., S.-H. Jung, and J. P. Costantino (2008). Nonparametric inference on median residual life function. *Biometrics 64*(1), 157–163.

[19] Jung, S.-H., J.-H. Jeong, and H. Bandos (2009). Regression on quantile residual life. *Biometrics 65*(4), 1203–1212.

[20] Kaplan, E. L. and P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association 53*(282), 457–481.

[21] Lagakos, S., L. Barraj, and V. De Gruttola (1988). Nonparametric analysis of truncated survival data, with application to aids. *Biometrika 75*(3), 515–523.

[22] Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3cr quasars. *Monthly Notices of the Royal Astronomical Society 155*, 95.

[23] Maguluri, G. and C.-H. Zhang (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 477–489.

[24] Moeschberger, M. L. and J. Klein (2003). *Survival analysis: Techniques for censored and truncated data: Statistics for Biology and Health*. Springer.

[25] Oakes, D. and T. Dasu (1990). A note on residual life. *Biometrika 77*(2), 409–410.

[26] Padgett, W. and D. T. McNichols (1984). Nonparametric density estimation from censored data. *Communications in Statistics-Theory and Methods 13*(13), 1581–1611.

[27] Schmittlein, D. C. and D. G. Morrison (1981). The median residual lifetime: A characterization theorem and an application. *Operations Research 29*(2), 392–399.

[28] Su, J. Q. and L. Wei (1993). Nonparametric estimation for the difference or ratio of median failure times. *Biometrics*, 603–607.

[29] Tsai, W.-Y., N. P. Jewell, and M.-C. Wang (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika 74*(4), 883–886.

[30] Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 290–295.

[31] Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics 10*(2), 616–620.

[32] Vardi, Y. (1985). Empirical distributions in selection bias models. *The Annals of Statistics*, 178–203.

[33] Vardi, Y. and C.-H. Zhang (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *The Annals of Statistics*, 1022–1039.

[34] Wang, J.-L. and T. P. Hettmansperger (1990). Two-sample inference for median survival times based on one-sample procedures for censored survival data. *Journal of the American Statistical Association 85*(410), 529–536.

[35] Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association 86*(413), 130–143.

[36] Wang, M.-C., R. Brookmeyer, and N. P. Jewell (1993). Statistical models for prevalent cohort data. *Biometrics*, 1–11.

[37] Wang, M.-C., N. P. Jewell, and W.-Y. Tsai (1986). Asymptotic properties of the product limit estimate under random truncation. *the Annals of Statistics*, 1597–1605.

[38] Wei, L. (1992). The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in Medicine 11*(14-15), 1871–1879.

[39] Wolfson, C., D. B. Wolfson, M. Asgharian, C. E. M'Lan, T. Østbye, K. Rockwood, and D. Hogan (2001). A reevaluation of the duration of survival after the onset of dementia. *New England Journal of Medicine 344*(15), 1111–1116.

[40] Yang, G. L. (1978). Estimation of a biometric function. *The Annals of Statistics 6*(1), 112–116.

[41] Ying, Z., S. Jung, and L. Wei (1995). Survival analysis with median regression models. *Journal of the American Statistical Association 90*(429), 178–184.

[42] Zhao, Y. and Y. Q. Chen (2004). A note on empirical likelihood inference of residual life regression. *Department of Mathematics & Statistics Faculty Publications*, 2.