A SERIES OF TWO-SAMPLE NON-PARAMETRIC TESTS
OF QUANTILE RESIDUAL LIFETIME

by

Yimeng Liu

MS, University of Pittsburgh, 2013
MPH, The Chinese University of Hong Kong, China, 2011
BS, Wuhan University, China, 2006

Submitted to the Graduate Faculty of
the Graduate School of Public Health in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh
2017
UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Yimeng Liu

It was defended on

July 18th 2017

and approved by

Abdus S Wahed, PhD, Associate Professor

Department of Biostatistics

Graduate School of Public Health, University of Pittsburgh

Gong Tang, PhD, Associate Professor

Department of Biostatistics

Graduate School of Public Health, University of Pittsburgh

Gary M Marsh, PhD, Professor

Department of Biostatistics

Graduate School of Public Health, University of Pittsburgh

Yu Cheng, PhD, Associate Professor

Department of Statistics, Psychiatry, and Biostatistics

School of Arts and Sciences, University of Pittsburgh
Dissertation Director: Abdus S Wahed, PhD, Associate Professor

Department of Biostatistics

Graduate School of Public Health, University of Pittsburgh
A SERIES OF TWO-SAMPLE NON-PARAMETRIC TESTS OF QUANTILE RESIDUAL LIFETIME

Yimeng Liu, PhD
University of Pittsburgh, 2017

ABSTRACT

Quantile residual lifetime (QRL) is of significant interest in many clinical studies as an easily interpretable quantity compared to other summary measures of survival distributions. In cancer or other fatal diseases, treatments are often compared based on the distributions or quantiles of the residual lifetime. Thus a common problem of interest is to test the equality of the QRL between two populations.

In the first chapter of this dissertation, we propose two classes of tests to compare two QRLs; one class is based on the difference between two estimated QRLs, and the other is based on the estimating function of the QRL, where the estimated QRL from one sample is plugged into the QRL-estimating-function of the other sample. We outline the asymptotic properties of the proposed test statistics. Simulation studies demonstrate that the proposed tests produced Type I errors closer to the nominal level and are superior to some existing tests based on both Type I error and power. Our proposed statistics are also computationally less intensive and more straightforward compared to tests based on the confidence intervals.

Moreover, for experimental designs, such as for randomized control trials, there is often no interest or need to adjust for the confounding factors due to the nature of the randomization. However, when there is missing data, adding the covariate information can help in improving the efficiency of estimators, e.g., the estimated treatment effect. In the second chapter, we propose an augmented inverse probability weighting estimator (AIPW) for QRL by incorporating the auxil-
inary covariate information in the presence of right-censoring. Simulation studies shows that our proposed estimator has smaller variance compared to the inverse probability estimator (IPW) and the Kaplan-Meier type estimator for the QRL when the auxiliary covariate is associated with the survival outcome. In contrast, there is minimal efficiency gain over our previously proposed test of equality of two QRLs when using the AIPW estimator compared to the IPW or Kaplan-Meier type estimators. We applied the proposed methods to a randomized multicenter Phase III trial for breast cancer patients with positive lymph nodes.

The public health significance of this dissertation is that it provides new methods to estimate and compare the QRLs of the time to event data, that can be used in in epidemiological and clinical studies to estimate and test QRLs more efficiently. This will help make superior inference of the survival outcome in future randomized clinical trials and the observational studies.
# TABLE OF CONTENTS

1.0 INTRODUCTION .............................................. 1
   1.1 Quantile residual lifetime .............................. 1
   1.2 Improving efficiency of the quantile residual lifetime estimator .... 3

2.0 A SERIES OF TWO-SAMPLE NON-PARAMETRIC TEST OF QUANTILE RESIDUAL LIFETIME .......... 6
   2.1 Estimation of the quantile residual lifetime ............... 6
   2.2 Tests on the quantile residual lifetimes among groups .......... 8
      2.2.1 Tests based on the difference between two quantile residual lifetime estimators 8
      2.2.2 Tests based on the estimating function $\hat{u}_k(.)$ .................. 9
   2.3 Large sample properties .................................. 10
      2.3.1 Asymptotic normality of the estimating function .......... 10
      2.3.2 Asymptotic normality of the difference statistic .......... 11
      2.3.3 Asymptotic normality of the PI statistics ................. 12
      2.3.4 Estimation of the asymptotic variance .................. 13
   2.4 Simulation study ......................................... 14
      2.4.1 Type I error ........................................ 14
      2.4.2 Power ............................................. 15
   2.5 Application to NSABP B-30 study .......................... 29
   2.6 Discussion .............................................. 30

3.0 EFFICIENT ESTIMATOR OF THE QUANTILE RESIDUAL LIFETIME .......... 33
   3.1 The inverse probability weighted estimators for the QRL .......... 33
   3.2 The efficient augmented inverse probability weighted estimators for the QRL .... 34
3.3 Simulation studies .................................................. 37
  3.3.1 Efficiency gain of the AIPW estimator ....................... 37
  3.3.2 Efficiency gain of two-sample test of the QRLs with the AIPW estimator . . . 41
3.4 Application in NSABP B-30 study ................................. 45
3.5 Discussion .............................................................. 46
4.0 FUTURE TOPIC ............................................................ 50
APPENDIX A. TYPE I ERROR STUDY WITH DIFFERENT SURVIVAL DISTRIBUTIONS-
  TABLES ................................................................. 51
APPENDIX B. TYPE I ERROR STUDY WITH DIFFERENT SURVIVAL DISTRIBUTIONS-
  FIGURES ................................................................. 62
BIBLIOGRAPHY ............................................................. 72
### LIST OF TABLES

1. Estimated Type I error of median residual life at different $t_0$s for various sample sizes ($n$) and censoring rates ($P_c$) ................................................................. 16
2. Estimated Type I error of 0.25–quantile residual life at different $t_0$s for various sample sizes ($n$) and censoring rates ($P_c$) ................................................................. 17
3. Estimated Type I error of 0.75–quantile residual life at $t_0$s for various sample sizes ($n$) and censoring rates ($P_c$) ................................................................. 18
4. Estimated power of median residual life at $t_0 = 1$ for various $\tau_0$s, sample sizes ($n$) and censoring rates ($P_c$) ................................................................. 23
5. Estimated power of 0.25–quantile residual life at $t_0 = 1$ for various $\tau_0$s, sample sizes ($n$) and censoring rates ($P_c$) ................................................................. 24
6. Estimated power of 0.75–quantile residual life at $t_0 = 1$ for various $\tau_0$s, sample sizes ($n$) and censoring rates ($P_c$) ................................................................. 25
7. Estimated $p$–QRLs in years at different $t_0$s and results of the test of equality of two $p$–QRLs between AC→T and ATC groups by different methods .......................... 31
8. Estimated bias and Monte-Carlo variance of three 0.5–QRL estimators using one auxiliary covariate ................................................................. 40
9. Estimated bias and Monte-Carlo variance of three 0.5–QRL estimators for $t_0 = 1$ using two auxiliary covariates ................................................................. 41
10. Estimated bias and Monte-Carlo variance of three 0.5–QRL estimators for $t_0 = 1$ using one auxiliary covariate under miss-specification .......................... 43
11. Estimated bias, Monte-Carlo variance and power of the two sample plug-in tests with the AIPW and Kaplan-Meier type (KM) estimator for 0.5–QRLs at $t_0 = 1$ .......................... 44
12 Distribution of the auxiliary variables by treatment groups .......................... 47
13 Estimated \( p \)-QRLs in years using the AIPW method at different \( t_0 \)s and results of the test of equality of two \( p \)-QRLs for AC\( \rightarrow \)T and ATC group by methods using the KM-type and AIPW estimators .................................................. 48
14 Parameters set-up for different survival distribution used in the simulation. ...... 52
15 Estimated Type I error for various tests of median residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario i .................. 53
16 Estimated Type I error for various tests of 0.25- residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario i .................. 54
17 Estimated Type I error for various tests of 0.75- residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario i .................. 55
18 Estimated Type I error for various tests of median residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario ii ............. 56
19 Estimated Type I error for various tests of 0.25- residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario ii ............. 57
20 Estimated Type I error for various tests of 0.75- residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario ii ............. 58
21 Estimated Type I error for various tests of median residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario iii ............. 59
22 Estimated Type I error for various tests of 0.25- residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario iii ............. 60
23 Estimated Type I error for various tests of 0.75- residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario iii ............. 61
LIST OF FIGURES

1 Estimated Type I error of median residual life at different $t_0$s for various sample sizes (n) and censoring rates ($P_c$) ............................................. 19

2 Estimated Type I error of 0.25—quantile residual life at different $t_0$s for various sample sizes (n) and censoring rates ($P_c$) ................................. 20

3 Estimated Type I error of 0.75—quantile residual life at different $t_0$s for various sample sizes (n) and censoring rates ($P_c$) ................................. 21

4 Estimated power of median residual life at $t_0 = 1$ for various $\tau_0$, sample sizes (n) and censoring rates ($P_c$) ......................................................... 26

5 Estimated power of 0.25—quantile residual life at $t_0 = 1$ for various $\tau_0$, sample sizes (n) and censoring rates ($P_c$) ......................................................... 27

6 Estimated power of 0.75—quantile residual life at $t_0 = 1$ for various $\tau_0$, sample sizes (n) and censoring rates ($P_c$) ......................................................... 28

7 Estimated $p$—quantile residual lifetimes and the corresponding point wise 95% confidence intervals for breast cancer patients with positive lymph nodes from NSABP data ................................................................. 31

8 Monte-Carlo coverage rates for the estimated 95%CI of the 0.5—QRL for various censoring rates ($P_c$), sample sizes (n) and $\beta_x$ .................................................... 42

9 Estimated Type I error for various tests of median residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario i ...................... 63

10 Estimated Type I error for various tests of 0.25— residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario i ...................... 64
11  Estimated Type I error for various tests of 0.75— residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario i  ............... 65
12  Estimated Type I error for various tests of median residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario ii  ............... 66
13  Estimated Type I error for various tests of 0.25— residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario ii  ............... 67
14  Estimated Type I error for various tests of 0.75— residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario ii  ............... 68
15  Estimated Type I error for various tests of median residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario iii  ............... 69
16  Estimated Type I error for various tests of 0.25— residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario iii  ............... 70
17  Estimated Type I error for various tests of 0.75— residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario iii  ............... 71
1.0 INTRODUCTION

1.1 QUANTILE RESIDUAL LIFETIME

The residual lifetime, defined as the remaining time that a person can survive or is free of certain disease given that he or she already survived to or did not have the disease until a fixed time point, is one of the most desired outcome measurements in epidemiological and clinical studies. In cancer studies, the residual lifetime expectancy is a very appealing and straightforward outcome measurement. Mayer et al. [2015] used the quotient of observed survival and the estimated residual life expectancy to assess the age effect within three treatment groups in a randomized controlled trial of colon cancer patients. Vock et al. [2013] developed a causal-effect model to evaluate the residual life distribution of patients after receiving an organ transplantation. Phadnis et al. [2015] proposed an ensemble survival model to estimate the relative residual life lost due to two types of strokes. Residual lifetime has also been used in breast cancer clinical trials by physicians to advise prospective patients by explaining the distribution of their predicted remaining lifetime conditional upon receiving a secondary treatment given that they already survived 5 years, for example, with tamoxifen [Goss et al., 2003; Coombes et al., 2004].

The residual lifetime distribution is commonly summarized by mean, median, or other quantiles. Because survival data are right censored, quantiles of the residual lifetime (QRL), are more stable summaries than the mean residual lifetime. The estimation of the QRL and its relationship to the distribution of the survival time have been extensively studied in the literature [Schmittlein and Morrison, 1981; Gupta and Langford, 1984; Gelfand and Kottas, 2003; Jeong et al., 2008; Ma and Yin, 2010]. Studies have compared two independent QRLs via non-parametric test of two QRLs or regression-based approaches [Jeong et al., 2008; Franco-Pereira et al., 2012; Lin et al., 2015; Tsai et al., 2015].
In Jeong et al. [2008], a non-parametric test was developed to compare two independent median residual lifetimes at a fixed time point. This method is an extension of Su and Wei [1993]’s non-parametric confidence interval (CI) procedure for the ratios of two median failure times. The CI is obtained by inverting a test statistic constructed using the estimating equations for the corresponding median residual lifetimes. The test statistic is minimized with respect to the unknown median residual lifetime of one group, which is also known as the nuisance parameter. One appealing aspect of this method is that it does not require estimating the density function of the failure time. However, based on the simulation study presented in the paper, this approach is too conservative with its empirical 95% coverage probabilities ranging from 0.96 to 0.98.

Franco-Pereira et al. [2012] developed a non-parametric method to compare two QRLs by constructing confidence bands of the difference between two QRL functions. Lin et al. (2015) proposed a semi-parametric estimator of the conditional QRL based on the Cox proportional hazard model and developed test statistics to compare two QRLs based on the difference and the ratio of the two estimated conditional QRLs. Both methods relied on the bootstrap procedure for inference. Tsai et al. [2015] proposed an alternative confidence interval procedure for the ratio of two median residual lifetimes for left-truncated and right-censored data. The confidence interval of the ratio is derived based on the length of the confidence interval of the median residual lifetime. The estimated asymptotic variance of the QRL estimator is obtained as a function of the upper and lower limit by inverting an asymptotic pivot CI for the QRL, which is estimated using the Lynden-Bell estimator [Tsai et al., 1987].

Therefore, the goal of the first topic in this dissertation is to develop a test of two quantile residual lifetimes that achieves the nominal Type I error and greater power compared to the existing method. In Chapter 2, we proposed a series of non-parametric test statistics for comparing two independent QRLs and discussed the asymptotic properties of these statistics. We also compared the performance of our proposed tests to the existing method by a Monte-Carlo simulation study and applied our proposed tests to a multi-center randomized clinical trial.
1.2 IMPROVING EFFICIENCY OF THE QUANTILE RESIDUAL LIFETIME ESTIMATOR

Estimation of the QRL for right-censored data using the non-parametric approaches had been considered by several authors [Jeong et al., 2008; Zhang et al., 2015]. The most straightforward approach is to solve the estimating equation of the QRL by replacing the survival functions with the Kaplan-Meier (KM) curves [Jeong et al., 2008]. In Chapter 2, we discuss the properties of this estimator and construct our proposed test statistics for comparing two QRLs based on this approach. Because the KM curves are non-parametric consistent estimators of the survival functions, this approach makes no assumption on the distribution of the underlying time to event. Another way to account for the right-censoring is the inverse probability weighted (IPW) method. By considering the right-censoring as a special type of missing data, we can construct a weighted estimating equation for the QRL based on the complete data (the observed event times). The weight equals to the probability of not being censored at the observed event time. If the censoring time is independent of the event time, this probability can be estimated using the KM curve of the censoring distribution at the observed event time. When the censoring probability is consistently estimated, the QRL estimate solved from the weighted estimating equation is also consistent. Wahed [2010] proved that both approaches produce equivalent quantile estimators with similar asymptotic efficiencies in the presence of the right-censoring. Zhang et al. [2015] proposed a smoothed estimator for the QRL by estimating the survival functions using a kernel smoothing method in the estimating equation for right censored data.

Besides the non-parametric approaches described above, semi-parametric approaches that focus on making inference on the QRL conditional on some covariate information are also developed in the literature [Gelfand and Kottas, 2003; Jung et al., 2009; Ma and Wei, 2012; Lin et al., 2015]. Gelfand and Kottas [2003] proposed a Bayesian semi-parametric median residual life model based upon a semi-parametric accelerated failure time regression model. In the Bayesian framework, a Dirichlet process mixture model is specified as the prior of the distribution for the failure time. Jung et al. [2009] developed a time-specific log-linear regression model on the QRL, which is an extension of the work by Ying et al. [1995]. The estimation process in this method does not require estimating the density functions. Ma and Wei [2012] proposed a two-stage procedure to estimate
the conditional QRL given the covariates from the IPW weighted estimating equations based on a generalized linear conditional QRL model. Most of the methods are based on the independent censoring assumption. Lin et al. [2015] proposed a method to estimate the QRL given covariates under the dependent censoring. Their method involves estimating conditional survival functions from an auxiliary Cox proportional hazard model.

In the observational studies, methods that make inference based on some covariates are usually used for the purpose of adjusting for confounders. However, for experimental designs, such as for randomized control trials, there is often no interest or need to adjust for the confounding factors due to the nature of the randomization. However, when there is missing data, adding the covariate information can help improve the efficiency in estimating the parameter of interest, for example, the treatment effect. Robins et al. [1994] proposed a class of augmented inverse probability weighting estimators (AIPW) by assuming that the missing mechanism is missing at random (MAR). The corresponding estimating equation is the same as the IPW estimating equation augmented with an arbitrary function that has a zero expectation at the truth. According to the semi-parametric theory, the efficiency of this estimator is maximized when the arbitrary function is chosen to be the expected value of the IPW function conditional on the observed data. Scharfstein et al. [1999] noted that the AIPW estimators are consistent to the true population means even when one of the outcome regression model and the missing mechanism is misspecified, which is referred to as double robustness. The extensions of the AIPW estimators for various data problems were studied by Wahed and Tsiatis, 2006, Zhang et al., 2008, Cao et al., 2009, Steingrimsson and Strawderman, 2016, Yang and Zhou, 2017 and Lok et al., 2017.

Wahed and Tsiatis [2006] proposed an efficient estimator for related quantities of the survival distribution by adding the information of the auxiliary variables in a two-stage randomized clinical trail design when some individuals are missing by design in the second stage with censored data. Zhang et al. [2008] proposed a class of augmented estimators for the treatment effects using the auxiliary baseline covariate information that achieved better efficiency in the randomized clinical trail. Cao et al. [2009] proposed an alternative doubly robust estimators for population means with incomplete data based on the AIPW estimators. Their estimators have better performance compared to the AIPW when some estimated probabilities of missingness are close to zero. Steingrimsson and Strawderman [2016] proposed a class of augmented estimators for the parameters in
a semi-parametric accelerated failure time model with missing covariates under MAR. Yang and Zhou [2017] developed an augmented IPW estimator for the treatment effect in a Cox regression model for two-stage randomization designs and the simulation study proved improved efficiency compared to the IPW estimator by adding the baseline covariate information. Lok et al. [2017] proposed an augmented IPW estimator for the cumulative incidence function in the presence of competing risk under both independent and dependent censoring mechanisms. The simulation study shows improved efficiency compared to a non-augmented estimator for the cumulative incidence function.

Wahed [2009] proposed several AIPW estimators for the quantiles of the survival distribution with right-censored data in a two-stage randomized clinical trial design. These estimators were shown to have increased efficiency compared to the IPW estimators by adding the information from covariates that are related to the underlying survival distribution. In Chapter 3 of this dissertation, we extend the AIPW estimator to the estimation of the QRL in the presence of the right-censoring. The goal is to develop an AIPW type estimator for the QRL that gains more efficiency by adding auxiliary covariate information.
2.0 A SERIES OF TWO-SAMPLE NON-PARAMETRIC TEST OF QUANTILE RESIDUAL LIFETIME

In this Chapter, we propose two classes of test statistics to compare two independent QRLs. Our methods achieve Type I error closer to the nominal level, and better power compared to Jeong et al. [2008]’s method. In Section 2.1, we define the QRL and describe how to estimate it using the Kaplan-Meier (KM) survival curve. In Section 2.2, we introduce the existing method proposed by Jeong et al. [2008] and propose two new classes of test statistics for testing the equality of two QRLs. In Section 2.3, we outline the asymptotic properties of our proposed tests and discuss various methods to estimate their standard errors. In Section 2.4, we compare our proposed methods with Jeong et al.’s method in a Monte-Carlo simulation study. In Section 2.5, we apply the proposed methods to a randomized breast cancer clinical trial.

2.1 ESTIMATION OF THE QUANTILE RESIDUAL LIFETIME

Assume that the underlying time to event $T$ follows a continuous distribution. We define the $p$–quantile residual lifetime ($p$–QRL), denoted as $\theta_{p,t_0}$, such that $p$ fraction of the population have the event by $\theta_{p,t_0} + t_0$, given that they already survived to time $t_0$. In other words, the $p$–QRL is the $p$–quantile of the remaining lifetime for patients who survived to time $t_0$. As a result, $\theta_{p,t_0}$ satisfies

$$Pr(T - t_0 \leq \theta_{p,t_0} | T \geq t_0) = p.$$ 

Or equivalently,

$$Pr(T > t_0 + \theta_{p,t_0} | T \geq t_0) - (1 - p) = 0.$$
Because $Pr(T > t_0 + \theta_{p,t_0} | T \geq t_0) = S(t_0 + \theta_{p,t_0}) / S(t_0)$, where $S(t)$ is the survival function of the time to event $T$, we can simplify the above equation as

$$u(\theta_{p,t_0}) = S(t_0 + \theta_{p,t_0}) - (1 - p)S(t_0) = 0. \quad (2.1)$$

When $p = 0.5$, the corresponding $p$–QRL is the median residual lifetime meaning that 50% of the subjects in the population will remain alive by $\theta_{0.5,t_0} + t_0$ given that they already survived to $t_0$.

Suppose we follow $n_k$ ($k = 1, 2$) individuals in the $k^{th}$ group until the event occurs or the individual can not be followed any longer due to study termination or drop out, resulting in random right censoring. Define $T_{ki}$ ($i = 1, ..., n_k$) as the failure time for patients in the $k^{th}$ group with the survival function $S_k(t)$. The observed data for the $k^{th}$ group is $\{(X_{ki}, \delta_{ki}), i = 1, ..., n_k\}$, where $X_{ki} = min(T_{ki}, C_{ki})$ is the minimum of the potential failure time $T_{ki}$ and the potential censoring time $C_{ki}$; and $\delta_{ki}$ is the indicator of the observed failure time ($\delta_{ki} = 1$ if $X_{ki} = T_{ki}$; 0 otherwise).

Let us denote the $p$–QRL for the $k^{th}$ population by $\theta_{p,t_0,k}$. The estimation of $\theta_{p,t_0,k}$ relies on the product limit or KM survival curve estimated from the data. More specifically, we replace the unknown survival functions in Equation 2.1 with their KM estimates and solve for $\theta_{p,t_0}$ to obtain an estimate. In other words, $\hat{\theta}_{p,t_0,k}$, the estimated $p$-QRL for the $k^{th}$ group is obtained as a solution to the equation

$$\hat{u}_k(\theta_{p,t_0,k}) = \hat{S}_k(t_0 + \theta_{p,t_0,k}) - (1 - p)\hat{S}_k(t_0) = 0, \quad (2.2)$$

where

$$\hat{S}_k(t) = \prod_{i: \delta_{ki}=1} \frac{n_{ki} - d_{ki}}{n_{ki}}$$

is the KM estimator of $S_k(t)$ from the observed survival data in the $k^{th}$ group, $n_{ki}$ is the number at risk right before the observed death time $T_{ki}$, and $d_{ki}$ is the number of death at time $T_{ki}$. Because the KM survival curves are step functions and hence infinite or no solutions may be possible, we define

$$\hat{\theta}_{p,t_0,k} = \inf \{t : \hat{S}_k(t) \leq (1 - p)\hat{S}_k(t_0)\} - t_0. \quad (2.3)$$

Note that when $t_0 = 0$, the estimate is a $p$–quantile estimate of the random failure time.
Our goal is to make inference about the relationship between two QRLs. More specifically, we would like to test the hypothesis

\[ H_0 : \frac{\tau_{p,t_0}}{\tau_{p,t_0,1}} = \tau_0, \]

where \( \tau_{p,t_0} = \frac{\theta_{p,t_0,2}}{\theta_{p,t_0,1}} \) is the ratio between two population \( p\)–QRLs, and \( \tau_0 \) is a known value.

Jeong et al. (2008) extended Su and Wei (1993)’s procedure of comparing two median failure times to test the equality of two median residual lifetimes. Their test statistic, when adopted to compare two \( p\)–QRLs, can be written as

\[
Q_{p,t_0}(\tau_0) = \inf_{\theta_{p,t_0,1}} W_{t_0}(\tau_0, \theta_{p,t_0,1}),
\]

(2.4)

where \( W_{t_0}(\tau_0, \theta_{p,t_0,1}) = \frac{\hat{u}_2(\theta_{p,t_0,1})}{\sigma_{u_1}^2/n_1} + \frac{\hat{u}_2(\theta_{p,t_0,1}, \tau_0)}{\sigma_{u_2}^2/n_2} \) and \( \sigma_{uk}^2 \) is the estimated variance of \( \sqrt{n_k} \hat{u}_k(\theta_{p,t_0,k}) \), defined in Section 2.3.1. Jeong et al. (2008) argued that \( Q_{p,t_0}(\tau_0) \) asymptotically follows a \( \chi^2_1 \) distribution hence the null hypothesis is rejected when \( Q_{p,t_0}(\tau_0) > \chi^2_{1,1-\alpha} \). We will refer to this test as minimum chi-square test or MCS.

As described previously, MCS approach was shown to be conservative in the sense that the coverage probabilities far exceed the nominal level (Jeong et al., 2008). In the sequel, we propose two alternative approaches to MCS: (i) test based on the difference between two QRL estimators, and (ii) test based on the estimating function \( \hat{u}_k(\cdot) \), defined in Equation 2.2.

### 2.2.1 Tests based on the difference between two quantile residual lifetime estimators

This approach is a straightforward application of Wald test using (2), where one estimates the \( p\)–QRL for both groups, and then constructs a Wald test based on the estimated difference \( \hat{d}_{p,t_0} = \hat{\theta}_{p,t_0,1} - \hat{\theta}_{p,t_0,2}/\tau_0 \), where \( \hat{\theta}_{p,t_0,k} \) is the estimated \( p\)–QRL for the \( k^{th} \) population. We can write the test statistic as

\[
z_{t_0}^{Diff} = \frac{\hat{d}_{p,t_0}}{\sqrt{\hat{\sigma}^2_{Diff}}},
\]

(2.5)

where \( \hat{\sigma}^2_{Diff} \) is an estimate of the asymptotic variance of the difference statistic \( \hat{d}_{p,t_0} \). There are several different approaches to estimate the asymptotic variance and the details will be outlined.
in Section 2.3.4. It can be shown that the statistic $z_{t_0}^{Diff}$ asymptotically follows a standard normal distribution when the sample sizes for both groups are large. The larger the absolute value of the statistic, the more the evidence there is against the null hypothesis. Therefore, we reject $H_0$ when $|z_{t_0}^{Diff}| > z_{1-\alpha}^{*}$, the $(1-\alpha)^{th}$ quantile of the standard normal distribution. We will refer to Equation 2.5 as a difference-based test, or for brevity, DIFF.

### 2.2.2 Tests based on the estimating function $\hat{u}_k(.)$

This approach is based on the estimating function $\hat{u}_k(.)$. Recall that we obtain the estimated $p-$QRL $\hat{\theta}_{p,t_0,k}$ from the estimating equation $\hat{u}_k(\theta_{p,t_0,k})$. In Section 2.3.1, we will show that $\hat{\theta}_{p,t_0,1}$ and $\hat{\theta}_{p,t_0,2}$ are consistent estimators of the population $p-$QRL of Group 1 and Group 2. Hence, under the null hypothesis, if we multiply both $p-$QRL estimators by the corresponding constant ($\tau_0$ or $1/\tau_0$), they become a $p-$QRL estimator of the other group. Therefore, under the null hypothesis, $u_1(\theta_{p,t_0,2}/\tau_0) = 0$ and $u_2(\theta_{p,t_0,1}/\tau_0) = 0$. This motivates us to define another class of test statistics

$$z_{t_0}^{2:1} = \frac{\hat{u}_1(\hat{\theta}_{p,t_0,2}/\tau_0)}{\sqrt{\hat{\sigma}_{2:1}^2}}$$

and

$$z_{t_0}^{1:2} = \frac{\hat{u}_2(\hat{\theta}_{p,t_0,1}/\tau_0)}{\sqrt{\hat{\sigma}_{1:2}^2}}$$

where $\hat{\sigma}_{2:1}^2$ and $\hat{\sigma}_{1:2}^2$ are the estimated asymptotic variances of the functions $\hat{u}_1(\hat{\theta}_{p,k}/\tau_0)$ and $\hat{u}_2(\hat{\theta}_{p,\tau}/\tau_0)$, respectively. Under the null hypothesis, $\hat{u}_1(\hat{\theta}_{p,k}/\tau_0)$ and $\hat{u}_2(\hat{\theta}_{p,\tau}/\tau_0)$ are expected to be close to zero and therefore larger value of $z_{t_0}^{j:k}(j,k = 1 \text{ or } 2 \text{ and } j \neq k)$, will support the alternative hypothesis: $H_a : \tau_{p,t_0} \neq \tau_0$. These statistics will be referred to as plug-in tests or PI.

In Section 2.3.2, we show that $z_{t_0}^{j:k}$ asymptotically follows a standard normal distribution as the sample size becomes large and the ratio between them approaches to a constant. As a result, we can reject the null hypothesis if the statistic is larger than the critical value from the standard normal distribution. Note that if $t_0$ equals to 0 and $\tau_0$ equals to 1, the test becomes a test of equality between two $p-$quantiles of the survival distributions.
2.3 LARGE SAMPLE PROPERTIES

In this section, we provide an outline of the large sample property of the proposed statistics. In particular, we argue that the statistics \( z_{t_0}^{Diff} \) and \( z_{t_0}^{j:k} \) are asymptotically normal, and therefore, tests based on the critical values from the standard normal distribution are appropriate.

2.3.1 Asymptotic normality of the estimating function

For the \( k^{th} \) group, the estimating equation for the \( p-QRL \) is a function of the KM survival curves \( \hat{S}_k(t) \) at \( t_0 + \theta_{p,t_0,k} \) and \( t_0 \), namely,

\[
\hat{u}_k(\theta_{p,t_0,k}) = \hat{S}_k(t_0 + \theta_{p,t_0,k}) - (1 - p) \hat{S}_k(t_0).
\] (2.8)

Because \( \hat{S}_k(t) \) is a uniformly consistent estimator of \( S_k(t) \) (Theorem 3.4.2, Fleming and Harrington, 1991) and \( S_k(t_0 + \theta_{p,t_0,k}) - (1 - p) S_k(t_0) = 0 \) at the true value of \( \theta_{p,t_0,k} = \theta_{p,t_0,k,0} \), \( \hat{u}_k(\theta_{p,t_0,k}) \) converges uniformly to zero. Therefore, under certain conditions, as a solution of the estimating equation \( \hat{u}_k(\theta_{p,t_0,k}) \), \( \hat{\theta}_{p,t_0,k} \) is a consistent estimator of the \( p-QRL \).

By Corollary 3.2.1 of Fleming and Harrington (1991), Jeong et al. (2008) argued that at the true value \( \theta_{p,t_0,k,0} \),

\[
\sqrt{n_k} \hat{u}_k(\theta_{p,t_0,k,0}) = n_k^{-1/2} \sum_{i=1}^{n_k} z_{ki} + o_p(1),
\] (2.9)

where

\[
z_{ki} = -S_k(t_0 + \theta_{p,t_0,k,0}) \int_{t_0}^{t_0 + \theta_{p,t_0,k,0}} \frac{dM_{ki}(s)}{y_k(s)} + (1 - p) S_k(t_0) \int_{0}^{t_0} \frac{dM_{ki}(s)}{y_k(s)}
\]

and \( M_{ki}(t) = N_{ki}(t) - \int_{0}^{t} Y_{ki}(s) d\Lambda_k(s) \) is the usual martingale, defined in Theorem 1.3.2, Fleming and Harrington (1991), \( N_{ki}(t) = I(T_{ki} \leq t, \delta_{ki} = 1) \) is the counting process at time \( t \) for the \( i^{th} \) individual in Group \( k \), \( Y_{ki}(t) = I(X_{ki} \geq t) \), \( \Lambda_k(t) \) is the cumulative hazard function of Group \( k \) at time \( t \), and \( y_k(s) \) is such that \( \sum_{i=1}^{n_k} Y_{ki}(t) \to y_k(s) \). Hence, \( z_{k1}, \ldots, z_{kn_k} \) are \( n_k \) independent random variables with mean zero. As a result, by martingale central limit theorem (Theorem 5.2.3, Fleming and Harrington, 1991), at the true value \( \theta_{p,t_0,k,0} \),

\[
\sqrt{n_k} \hat{u}_k(\theta_{p,t_0,k,0}) \overset{d}{\to} N(0, \sigma_{uk}^2),
\] (2.10)
where $\sigma_{uk}^2 = E(z_{ki}^2)$ is the asymptotic variance of $\sqrt{n_k} \hat{u}_k(\theta_{p,t0,k,0})$, which can be consistently estimated by $\hat{\sigma}_{uk}^2 = \sum_{i=1}^{n_k} \hat{z}_{ki}^2/n_k$, where

$$ \hat{z}_{ki} = -\hat{S}_k(t_0 + \hat{\theta}_{p,t0,k}) \int_0^{t_0 + \hat{\theta}_{p,t0,k}} \frac{d\hat{M}_{ki}(s)}{\bar{Y}_k(s)} + (1 - p)\hat{S}_k(t_0) \int_0^{t_0} \frac{d\hat{M}_{ki}(s)}{\bar{Y}_k(s)}, $$

(2.11)

$\hat{M}_{ki}(t) = N_{ki}(t) - \int_0^t Y_{ki}(s)d\hat{\Lambda}_k(s)$, $\bar{Y}_k(t) = \sum_{i=1}^{n_k} Y_{ki}(t)$ and $\hat{\Lambda}_k(t) = \int_0^t Y_{ki}^{-1}(s)dN_k(s)$ is the Nelson-Aalen estimator of the cumulative hazard function in Group $k$.

### 2.3.2 Asymptotic normality of the difference statistic

First we argue that the $p$–QRL estimator $\hat{\theta}_{p,t0,k}$ asymptotically follows a normal distribution by writing it as a function of $\hat{u}_k(\theta_{p,t0,k,0})$, which is the estimating function at the true value $\theta_{p,t0,k,0}$. Then it follows that $z_{t0}^{Diff}$ is a linear combination of two QRL estimators, both of which asymptotically follow a normal distribution. Our derivation assumes that the true survival functions are continuous and hence differentiable.

Recall that $\hat{\theta}_{p,t0,k}$ is a solution to the equation $\hat{u}_k(\theta_{p,t0,k}) = 0$, and therefore satisfies $\hat{u}_k(\hat{\theta}_{p,t0,k}) = 0$. By applying the Taylor’s expansion of $\hat{u}_k(\hat{\theta}_{p,t0,k})$ at the truth $\theta_{p,t0,k,0}$, and ignoring higher order terms, we have

$$ \sqrt{n_k}(\hat{\theta}_{p,t0,k} - \theta_{p,t0,k,0}) = -\sqrt{n_k} \hat{u}_k(\theta_{p,t0,k,0})/\hat{S}_k(t_0 + \theta_{p,t0,k,0}) + o_p(1). $$

(2.12)

As $n_k \to \infty$, $\hat{S}_k(t_0 + \theta_{p,t0,k,0})$ converges in probability to $S_k'(t_0 + \theta_{p,t0,k,0})$ where $S_k'(t)$ is the first derivative of the survival function $S_k(t)$ at time $t$ and equals $-f_k(t)$, the negative density function of the time to event in Group $k$. Consequently, by Slutsky’s theorem and Equation 2.10, the left side of Equation 2.12 follows an asymptotic normal distribution. More specifically,

$$ \sqrt{n_k}(\hat{\theta}_{p,t0,k} - \theta_{p,t0,k,0}) \xrightarrow{d} N(0, \sigma_k^2), $$

(2.13)

where $\sigma_k^2 = \{f_k(t_0 + \theta_{p,t0,k,0})\}^{-2}\sigma_{uk}^2$. Therefore, the asymptotic variance of $\hat{\theta}_{p,t0,k}$ is given by $\sigma_k^2/n_k$. Thus, under $H_0$, as $n_1 \to \infty$ and $n_1/n_2 \to C$, a constant, we can show that

$$ \sqrt{n_1}\hat{d}_{p,t0} \xrightarrow{d} N(0, \sigma_{Diff}^2), $$

(2.14)

where $\sqrt{n_1}\hat{d}_{p,t0} = \sqrt{n_1}\hat{\theta}_{p,t0,1} - \sqrt{n_1/n_2}\sqrt{\frac{n_2}{\tau_0}}\hat{\theta}_{p,t0,2} + \sigma_{Diff}^2 = \{f_1(t_0 + \theta_{p,t0,1,0})\}^{-2}\sigma_{u1}^2 + C\{f_2(t_0 + \theta_{p,t0,2,0})\}^{-2}\sigma_{u2}^2$. This variance can be estimated by the methods described in Section 2.3.4.
2.3.3 Asymptotic normality of the PI statistics

Let us first consider the statistic $z_{t_0}^{2:1}$ defined in Equation 2.6. The numerator of $z_{t_0}^{2:1}$, by a Taylor expansion, around the true value $\theta_{p,t_0,2,0}$, can be approximated as

$$\hat{u}_1(\hat{\theta}_{p,t_0,2}/\tau_0) = \hat{u}_1(\theta_{p,t_0,2,0}/\tau_0) + (\hat{\theta}_{p,t_0,2} - \theta_{p,t_0,2}) \hat{S}_1'(t_0 + \theta_{p,t_0,2}/\tau_0)(1/\tau_0) + o_p(n_1^{-1/2}).$$

Using Equation 2.12 we can then write

$$\sqrt{n_1} \hat{u}_1\left(\frac{\hat{\theta}_{p,t_0,2}}{\tau_0}\right) = \sqrt{n_1} \hat{u}_1\left(\frac{\theta_{p,t_0,2,0}}{\tau_0}\right) - \sqrt{n_1} \frac{n_2}{n_2} \sqrt{n_2} \hat{u}_2(\theta_{p,t_0,2,0}) \hat{S}_1'(t_0 + \theta_{p,t_0,2,0}/\tau_0) + o_p(1).$$

For large $n_1$ and $n_2$, the above is equivalent to

$$\sqrt{n_1} \hat{u}_1\left(\frac{\hat{\theta}_{p,t_0,2}}{\tau_0}\right) = \sqrt{n_1} \hat{u}_1\left(\frac{\theta_{p,t_0,2,0}}{\tau_0}\right) - \left(\frac{\sqrt{C}}{\tau_0}\right) f_1(t_0 + \theta_{p,t_0,2,0}/\tau_0) \hat{u}_2(\theta_{p,t_0,2,0}) + o_p(1), \quad (2.15)$$

where, as before, $S_1'(t)$ is the derivative of the survival function and replaced with $-f_k(t)$. Because under the null hypothesis, $H_0 : \theta_{p,t_0,2,0}/\tau_0 = \theta_{p,t_0,1,0}$, both $\sqrt{n_1} \hat{u}_1(\theta_{p,t_0,2,0}/\tau_0) = \sqrt{n_1} \hat{u}_1(\theta_{p,t_0,1,0})$, and $\sqrt{n_2} \hat{u}_2(\theta_{p,t_0,2,0})$ converges in distribution to independent normal random variables, by Slutsky’s theorem,

$$\sqrt{n_1} \hat{u}_1\left(\frac{\hat{\theta}_{p,t_0,2}}{\tau_0}\right) \overset{d}{\rightarrow} N(0, \sigma^2_{2:1}), \quad (2.16)$$

where

$$\sigma^2_{2:1} = \sigma^2_{u1}(\theta_{p,t_0,1,0}) + \frac{C}{\tau_0^2} \left\{ \frac{C}{f_2(t_0 + \theta_{0,t_0,2,0})} \right\}^2 \sigma^2_{u2}(\theta_{p,t_0,2,0}) \quad (2.17)$$

The same argument can be used to show that $\sqrt{n_2} \hat{u}_2(\hat{\theta}_{p,t_0,1,0}/\tau_0)$ also follows an asymptotic normal distribution or, more explicitly,

$$\sqrt{n_2} \hat{u}_2\left(\frac{\hat{\theta}_{p,t_0,1}}{\tau_0}\right) \overset{d}{\rightarrow} N(0, \sigma^2_{1:2}), \quad (2.18)$$

where

$$\sigma^2_{1:2} = \sigma^2_{u2}(\theta_{p,t_0,2,0}) + \frac{\tau_0^2}{C} \left\{ \frac{C}{f_1(t_0 + \theta_{0,t_0,1,0})} \right\}^2 \sigma^2_{u1}(\theta_{p,t_0,1,0}). \quad (2.19)$$
2.3.4 Estimation of the asymptotic variance

In Sections 2.3.2 and 2.3.3, we outlined the asymptotic normality of the numerators of our proposed statistics $z_{t_0}^{Diff}$, $z_{t_0}^{2:1}$ and $z_{t_0}^{1:2}$. Here we describe how to estimate their asymptotic variances.

Note that all of the asymptotic variances ($\sigma^2_{Diff}, \sigma^2_{2:1}$ and $\sigma^2_{2:1}$) are functions of the density curves of the failure time, the survival functions and the population QRLs in both groups. Hence, we can easily estimate the variances by replacing all unknown parameters with their consistent estimators. To obtain stable and smooth non-parametric density estimates in the presence of censoring, we used the inverse probability weighted (IPW) Gaussian kernel density. The weight equals to the inverse of the probability that the observation is not being censored at the observed failure time and can be estimated using the KM survival function of the censoring distribution. More specifically, the density estimate for Group $k$ is defined as

$$\hat{f}_{k,IPW}(x) = \frac{\sum_{i=1}^{n_k} K_{hi}(x) \delta_{ki} \hat{S}_{Ck}(T_{ki})}{\sum_{i=1}^{n_k} \delta_{ki} \hat{S}_{Ck}(T_{ki})},$$

(2.20)

where $K_{hi}(x) = \frac{1}{h \sqrt{2\pi}} \exp\left\{-\frac{1}{2} \frac{(x-T_{ki})^2}{h^2}\right\}$, $h$ is the bandwidth, $\hat{S}_{Ck}(T_{ki})$ is the KM survival function of the time to censoring at the observed time to failure $T_{ki}$ for Group $k$ and $\delta_{ki}$ is the indicator function of the observed failure time for the $i^{th}$ individual in Group $k$.

As noted before, the asymptotic variances also contain the unknown QRLs. We took three approaches to construct an estimator for the QRL to be plugged into the asymptotic variance formulas. The first approach is to use the individual estimate described in Equation 2.3. In the remaining sections of the paper, we use "Wald" to denote the variance estimators using this approach. The second and third approaches use a weighted linear combination of both estimates. For example, under the null hypothesis, we can use

$$\hat{\theta}_{p,t_0,w1} = \frac{w_1 \hat{\theta}_{p,t_0,1} + w_2 \hat{\theta}_{p,t_0,2}}{w_1 + w_2},$$

(2.21)

and

$$\hat{\theta}_{p,t_0,w2} = \frac{w_1 \hat{\theta}_{p,t_0,1} \tau_0 + w_2 \hat{\theta}_{p,t_0,2}}{w_1 + w_2},$$

(2.22)

where $w_k$ is the weight for Group $k$. We used two forms of weights, (i) the number of events in the group and (ii) the KM estimate at time $t_0$ for the group. We use "WI" and "WII" to denote the variance estimators using the second and third approach, respectively. We used the same IPW kernel density estimate for all the variance estimators.
2.4 SIMULATION STUDY

To evaluate the performance of the proposed test statistics, we designed simulation studies to estimate their Type I error and statistical power under different scenarios, and compared them with the results of the MCS approach. In our simulation study, we evaluated the Type I error and power of the \( p \)-QRL test statistics for \( p = 0.5 \) (median residual lifetime), 0.25 and 0.75. We compared the performance of the three proposed methods, namely, the DIFF test and two tests based on the PI statistics defined in Section 2.2. These latter tests were defined by using the following two strategies: (i) plug-in the QRL estimate from the group with larger number of events into the \( \hat{u}(\cdot) \) function of the other group and (ii) plug-in the QRL estimate from the group with smaller number of events into the \( \hat{u}(\cdot) \) function of the other group.

2.4.1 Type I error

We generated independent failure times for the \( k^{th} \) group (\( k = 1 \) or 2) from a Weibull survival function

\[
S(t) = \exp\{-\left(\frac{t}{\rho_k}\right)^{\eta_k}\},
\]

where \( \rho_k \) and \( \eta_k \) are the scale and shape parameters, respectively. We generated the failure time for the \( i^{th} \) person in Group \( k \) by

\[
T_{ki} = \frac{1}{\rho_k} \left\{ -\log(1 - U_{ki}) \right\}^{\frac{1}{\eta_k}},
\]

where \( U_{ki} \sim UNIF(0, 1) \).

By setting Equation 2.1 to zero, we can solve \( \theta_{p,t_0,k} \) for Group \( k \), as

\[
\theta_{p,t_0,k} = \frac{1}{\rho_k} \left\{ \log \left( \frac{1}{1-p} \right) + \left( \frac{\rho_k t_0}{\eta_k} \right)^{\frac{1}{\eta_k}} \right\} - t_0,
\]

where \( t_0 = 0, 1, 2, 3, \) and 4. To evaluate the Type I error, we set \( \rho_1 = \rho_2 = 0.2 \) and \( \eta_1 = \eta_2 = 2 \).

Censoring times are generated from a uniform distribution with parameters \( a \) and \( b \). Assuming the failure time \( T \sim S(t) \) and the censoring time \( C \sim UNIF(a, b) \), where \( S(t) \) is the survival function of the failure time, the censoring rate can be calculated as

\[
P_c = Pr(T \geq C) = \int_a^b S(x) \frac{1}{b-a} dx
\]
Different censoring rates can be achieved by varying the parameters in the uniform distribution. We choose three censoring rates ($P_c = 0\%, 20\%$ and $40\%$) to investigate their effect on the performance of all methods. Each Type I error is calculated based on 10,000 simulations with three different sample sizes (50, 100 and 200 per group).

Table 1 shows the estimated Type I error for QRL tests for sample sizes 50 and 200 with 20\% and 40\% observations censored when $p = 0.5$ (median residual life). Tables 2 and 3 shows the corresponding results for the 0.25-QRL and 0.75-QRL tests. In general, the estimated Type I error increased with decreasing sample size, increasing censoring rate, quantile, and time ($t_0$) at which the QRL is being estimated. In most cases, DIFF test had an estimated Type I error larger than the nominal level. The PI tests have estimated Type I errors close to the nominal level of 0.05, regardless of the weighting methods used in the variance estimators (WI and WII). MCS test was the most conservative, yielding estimated Type I errors, ranging from 0.025 to 0.035. In all cases, the empirical Type I errors deviated from the nominal 0.05 level for large values of $t_0$, which is reasonable as the reduced numbers at risk reduce the precision of the test statistic. Figure 1 shows the estimated Type I error over $t_0$ for various tests of equality of median residual life for varying sample sizes and censoring rates. Figures 2 and 3 show the corresponding results for 0.25-QRLs and 0.75-QRLs. Only four of the ten methods from Table 1 are presented in Figure 1 for the sake of clarity. It is clear that the proposed PI statistics maintain Type I errors closer to the nominal level in most of the scenarios, compared to the MCS or the DIFF test. When the sample size is small with a high censoring rate (e.g., $n = 50$, $P_c = 40\%$), none of the tests perform well for large quantiles (e.g., $p = 0.75$). We also performed the simulation studies by generating the failure times from the log-normal and exponential distributions. The results are similar (Appendices A and B).

### 2.4.2 Power

Based on the results of the Type I error simulation, we designed a simulation study to evaluate the power to reject the null hypothesis across various values of $\tau_0$ for the proposed tests. Because the Wald variance estimator did not perform well based on the Type I error results, we excluded them from the power analysis. Thus, powers of all three proposed methods with the WI and WII variance estimators were studied. The simulated failure times in the power study were generated.
Table 1: Estimated Type I error of median residual life at different $t_0$s for various sample sizes (n) and censoring rates ($P_c$)

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th>MCS Wald</th>
<th>DIFF Wald</th>
<th>DIFF WI</th>
<th>DIFF WII</th>
<th>PI Wald</th>
<th>PI WI</th>
<th>PI WII</th>
<th>PI* Wald</th>
<th>PI* WI</th>
<th>PI* WII</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>n = 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.026</td>
<td>0.075</td>
<td>0.065</td>
<td>0.064</td>
<td>0.063</td>
<td>0.057</td>
<td>0.055</td>
<td>0.047</td>
<td>0.050</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.028</td>
<td>0.073</td>
<td>0.062</td>
<td>0.062</td>
<td>0.058</td>
<td>0.049</td>
<td>0.048</td>
<td>0.044</td>
<td>0.050</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.027</td>
<td>0.076</td>
<td>0.058</td>
<td>0.057</td>
<td>0.059</td>
<td>0.048</td>
<td>0.045</td>
<td>0.041</td>
<td>0.049</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.030</td>
<td>0.094</td>
<td>0.060</td>
<td>0.057</td>
<td>0.070</td>
<td>0.050</td>
<td>0.047</td>
<td>0.038</td>
<td>0.047</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.033</td>
<td>0.112</td>
<td>0.064</td>
<td>0.060</td>
<td>0.081</td>
<td>0.056</td>
<td>0.052</td>
<td>0.043</td>
<td>0.051</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.025</td>
<td>0.063</td>
<td>0.046</td>
<td>0.044</td>
<td>0.048</td>
<td>0.043</td>
<td>0.039</td>
<td>0.044</td>
<td>0.042</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.026</td>
<td>0.075</td>
<td>0.054</td>
<td>0.052</td>
<td>0.055</td>
<td>0.049</td>
<td>0.045</td>
<td>0.050</td>
<td>0.049</td>
<td>0.052</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.026</td>
<td>0.086</td>
<td>0.059</td>
<td>0.056</td>
<td>0.062</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.048</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.033</td>
<td>0.099</td>
<td>0.060</td>
<td>0.057</td>
<td>0.074</td>
<td>0.053</td>
<td>0.048</td>
<td>0.050</td>
<td>0.049</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.049</td>
<td>0.117</td>
<td>0.073</td>
<td>0.068</td>
<td>0.098</td>
<td>0.070</td>
<td>0.062</td>
<td>0.079</td>
<td>0.060</td>
<td>0.072</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n = 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.031</td>
<td>0.063</td>
<td>0.059</td>
<td>0.059</td>
<td>0.057</td>
<td>0.054</td>
<td>0.053</td>
<td>0.045</td>
<td>0.048</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.030</td>
<td>0.055</td>
<td>0.048</td>
<td>0.048</td>
<td>0.051</td>
<td>0.046</td>
<td>0.046</td>
<td>0.041</td>
<td>0.047</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.028</td>
<td>0.062</td>
<td>0.053</td>
<td>0.053</td>
<td>0.052</td>
<td>0.046</td>
<td>0.045</td>
<td>0.039</td>
<td>0.046</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.027</td>
<td>0.067</td>
<td>0.054</td>
<td>0.053</td>
<td>0.054</td>
<td>0.046</td>
<td>0.045</td>
<td>0.039</td>
<td>0.047</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.027</td>
<td>0.074</td>
<td>0.057</td>
<td>0.057</td>
<td>0.059</td>
<td>0.048</td>
<td>0.047</td>
<td>0.034</td>
<td>0.045</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.031</td>
<td>0.058</td>
<td>0.052</td>
<td>0.052</td>
<td>0.050</td>
<td>0.048</td>
<td>0.047</td>
<td>0.045</td>
<td>0.050</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.027</td>
<td>0.059</td>
<td>0.052</td>
<td>0.052</td>
<td>0.051</td>
<td>0.047</td>
<td>0.047</td>
<td>0.044</td>
<td>0.049</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.029</td>
<td>0.061</td>
<td>0.052</td>
<td>0.051</td>
<td>0.051</td>
<td>0.046</td>
<td>0.045</td>
<td>0.043</td>
<td>0.049</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.026</td>
<td>0.066</td>
<td>0.053</td>
<td>0.052</td>
<td>0.049</td>
<td>0.043</td>
<td>0.042</td>
<td>0.037</td>
<td>0.047</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.028</td>
<td>0.075</td>
<td>0.055</td>
<td>0.055</td>
<td>0.054</td>
<td>0.048</td>
<td>0.046</td>
<td>0.038</td>
<td>0.044</td>
<td>0.045</td>
<td></td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 2: Estimated Type I error of 0.25—quantile residual life at different $t_0$s for various sample sizes (n) and censoring rates ($P_c$)

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MCS</td>
<td>DIFF</td>
<td>PI</td>
<td>PI*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.022</td>
<td>0.044</td>
<td>0.042</td>
<td>0.043</td>
<td>0.030</td>
<td>0.042</td>
<td>0.041</td>
<td>0.058</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.025</td>
<td>0.049</td>
<td>0.042</td>
<td>0.042</td>
<td>0.034</td>
<td>0.044</td>
<td>0.043</td>
<td>0.052</td>
<td>0.042</td>
</tr>
<tr>
<td>20%</td>
<td>1</td>
<td>0.023</td>
<td>0.063</td>
<td>0.050</td>
<td>0.050</td>
<td>0.035</td>
<td>0.043</td>
<td>0.042</td>
<td>0.050</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.020</td>
<td>0.091</td>
<td>0.063</td>
<td>0.061</td>
<td>0.042</td>
<td>0.050</td>
<td>0.047</td>
<td>0.049</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.022</td>
<td>0.107</td>
<td>0.067</td>
<td>0.064</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.021</td>
<td>0.052</td>
<td>0.051</td>
<td>0.051</td>
<td>0.029</td>
<td>0.044</td>
<td>0.040</td>
<td>0.069</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.022</td>
<td>0.055</td>
<td>0.049</td>
<td>0.049</td>
<td>0.029</td>
<td>0.042</td>
<td>0.038</td>
<td>0.057</td>
<td>0.048</td>
</tr>
<tr>
<td>40%</td>
<td>2</td>
<td>0.019</td>
<td>0.078</td>
<td>0.061</td>
<td>0.060</td>
<td>0.038</td>
<td>0.050</td>
<td>0.045</td>
<td>0.056</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.021</td>
<td>0.096</td>
<td>0.064</td>
<td>0.061</td>
<td>0.042</td>
<td>0.055</td>
<td>0.050</td>
<td>0.051</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.025</td>
<td>0.111</td>
<td>0.069</td>
<td>0.068</td>
<td>0.043</td>
<td>0.054</td>
<td>0.049</td>
<td>0.058</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.029</td>
<td>0.049</td>
<td>0.049</td>
<td>0.049</td>
<td>0.044</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.027</td>
<td>0.051</td>
<td>0.048</td>
<td>0.048</td>
<td>0.038</td>
<td>0.045</td>
<td>0.044</td>
<td>0.052</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.027</td>
<td>0.061</td>
<td>0.052</td>
<td>0.052</td>
<td>0.046</td>
<td>0.050</td>
<td>0.050</td>
<td>0.051</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.026</td>
<td>0.062</td>
<td>0.049</td>
<td>0.049</td>
<td>0.045</td>
<td>0.047</td>
<td>0.046</td>
<td>0.046</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.025</td>
<td>0.072</td>
<td>0.055</td>
<td>0.055</td>
<td>0.049</td>
<td>0.050</td>
<td>0.049</td>
<td>0.046</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.026</td>
<td>0.048</td>
<td>0.045</td>
<td>0.045</td>
<td>0.039</td>
<td>0.046</td>
<td>0.045</td>
<td>0.050</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.024</td>
<td>0.054</td>
<td>0.049</td>
<td>0.049</td>
<td>0.040</td>
<td>0.044</td>
<td>0.043</td>
<td>0.052</td>
<td>0.050</td>
</tr>
<tr>
<td>40%</td>
<td>2</td>
<td>0.027</td>
<td>0.059</td>
<td>0.051</td>
<td>0.051</td>
<td>0.042</td>
<td>0.047</td>
<td>0.046</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.024</td>
<td>0.064</td>
<td>0.052</td>
<td>0.052</td>
<td>0.042</td>
<td>0.044</td>
<td>0.045</td>
<td>0.050</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.021</td>
<td>0.076</td>
<td>0.054</td>
<td>0.054</td>
<td>0.046</td>
<td>0.049</td>
<td>0.048</td>
<td>0.048</td>
<td>0.055</td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 3: Estimated Type I error of 0.75—quantile residual life at \( t_0 \)s for various sample sizes (\( n \)) and censoring rates (\( P_c \))

<table>
<thead>
<tr>
<th>( P_c )</th>
<th>( t_0 )</th>
<th>Method</th>
<th>( n = 50 )</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MCS</td>
<td>DIFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald WI WII</td>
<td>Wald WI WII</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.035</td>
<td>0.110 0.064 0.060 0.107 0.054 0.053 0.040 0.054 0.060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>1</td>
<td>0.037 0.107 0.062 0.060 0.105 0.050 0.048 0.038 0.050 0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.042 0.113 0.061 0.058 0.116 0.053 0.051 0.047 0.049 0.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.062 0.125 0.064 0.061 0.141 0.062 0.059 0.063 0.061 0.068</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.124 0.138 0.074 0.070 0.198 0.084 0.076 0.110 0.072 0.084</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.046 0.113 0.064 0.061 0.117 0.056 0.053 0.066 0.054 0.062</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.058 0.119 0.066 0.063 0.123 0.062 0.059 0.073 0.058 0.069</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
<td>0.078 0.125 0.067 0.065 0.139 0.069 0.063 0.089 0.063 0.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.126 0.132 0.073 0.069 0.180 0.085 0.077 0.125 0.072 0.085</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.218 0.137 0.074 0.072 0.242 0.108 0.095 0.200 0.083 0.101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.026 0.078 0.054 0.052 0.070 0.048 0.047 0.031 0.046 0.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.027 0.078 0.056 0.055 0.071 0.046 0.045 0.032 0.045 0.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
<td>0.027 0.077 0.053 0.051 0.068 0.043 0.041 0.033 0.043 0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.028 0.087 0.054 0.053 0.074 0.047 0.046 0.037 0.045 0.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.036 0.100 0.058 0.057 0.093 0.052 0.050 0.053 0.048 0.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.026 0.078 0.054 0.052 0.070 0.048 0.047 0.031 0.046 0.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.027 0.078 0.056 0.055 0.071 0.046 0.045 0.032 0.045 0.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
<td>0.027 0.077 0.053 0.051 0.068 0.043 0.041 0.033 0.043 0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.028 0.087 0.054 0.053 0.074 0.047 0.046 0.037 0.045 0.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.036 0.100 0.058 0.057 0.093 0.052 0.050 0.053 0.048 0.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual life (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time \( t_0 \), both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 1: Estimated Type I error of median residual life at different $t_0$s for various sample sizes (n) and censoring rates ($P_c$)
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 2: Estimated Type I error of 0.25—quantile residual life at different $t_0$s for various sample sizes (n) and censoring rates ($P_c$)
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 3: Estimated Type I error of 0.75—quantile residual life at different $t_0$s for various sample sizes (n) and censoring rates ($P_c$)
from a parametric proportional hazards model (Cox) with a single covariate \( z \). The model is specified as

\[
S(t; z) = \exp\{-(\rho t)^\eta \exp(\beta z)\},
\]

(2.26)

where \( z \) is the group indicator with \( z = 1 \) for Group 2 and \( z = 0 \) for Group 1. Under this model, the failure time in Group 1 follows a Weibull distribution with the shape parameter \( \rho \) and the scale parameter \( \eta \), while the failure time in Group 2 follows a Weibull distribution with the shape parameter \( \rho^* = \rho \exp(\beta/\eta) \) and the same scale parameter \( \eta \). The hazard ratio between Group 2 and Group 1 equals to \( \exp(\beta) \). Thus, when \( \beta \) deviates from 1 in both directions, the ratio between the two QRLs also deviates away from 1. To evaluate the relationship between the power and the ratio, various values of \( \beta \) were considered along with various censoring rates. The \( \beta \)s used in the study are -0.8, -0.6, -0.3, 0, 0.3 and 0.8, which corresponds to \( \tau_{0.5,t_0,0} = \theta_{0.5,t_0,2,0}/\theta_{0.5,t_0,1,0} = 1.61, 1.43, 1.20, 1.00, 0.83 \) and 0.60 when \( t_0 = 1 \) and \( \tau_{0.5,t_0,0} = 1.73, 1.51, 1.24, 1.00, 0.80 \) and 0.55 when \( t_0 = 2 \). All results are calculated based on 10,000 simulations under three different sample sizes: 50, 100 and 200 per group.

Comparison of power across different methods produced similar conclusions for varying values of \( \beta \), sample size, censoring rate, and \( t_0 \). Therefore we only presented the estimated power against the alternative ratio of the 0.5-QRL when \( t_0 = 1 \) for \( n = 50, 100, \) and 200 under 20\% and 40\% censoring (Table 4 and Figure 4). In general, the power of all tests increased as the true \( \tau \) moved away from null (\( \tau_0 = 1 \)), as the sample size increased, and as the censoring rates decreased. The DIFF statistic tends to have greater power compared to the other methods, while the MCS statistic results in the smallest power among all the methods. For example, when \( \tau_{0.5,1,0} = 1.43, \) \( n = 100, \) and 20\% observations are censored, the estimated power of the MCS method is 0.76, compared to the estimated power of 0.85 for the DIFF test (for both WI and WII) and 0.81 and 0.82 for the PI tests. Similar patterns were observed from the results of the 0.25-QRL (Table 5 Figure 5) and 0.75-QRL (Table 6 Figure 6). In summary, the DIFF test has the largest estimated power, closely followed by the PI tests, outperforming the MCS test. But note that the DIFF test fails to maintain a Type I error close to the nominal level (Section 2.4.1).

To investigate the performance of PI and MCS tests in a simulation scenario that mimics our application data presented in Section 2.5, we performed another simulation study. We generated two hypothetical survival samples with same and different underlying exponential distributions,
Table 4: Estimated power of median residual life at $t_0 = 1$ for various $\tau_0$s, sample sizes (n) and censoring rates ($P_c$)

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$\tau_{0.5,1.0}$</th>
<th>Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MCS</td>
<td>DIFF</td>
<td>PI</td>
<td>PI*</td>
<td>WI</td>
<td>WII</td>
<td>WI</td>
<td>WII</td>
<td>WII</td>
</tr>
<tr>
<td>n=50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>1.61</td>
<td>0.709</td>
<td>0.808</td>
<td>0.808</td>
<td>0.755</td>
<td>0.753</td>
<td>0.759</td>
<td>0.761</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.43</td>
<td>0.438</td>
<td>0.569</td>
<td>0.569</td>
<td>0.512</td>
<td>0.510</td>
<td>0.536</td>
<td>0.538</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>0.123</td>
<td>0.190</td>
<td>0.190</td>
<td>0.166</td>
<td>0.163</td>
<td>0.173</td>
<td>0.176</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.025</td>
<td>0.060</td>
<td>0.059</td>
<td>0.053</td>
<td>0.052</td>
<td>0.050</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>0.116</td>
<td>0.184</td>
<td>0.184</td>
<td>0.167</td>
<td>0.164</td>
<td>0.172</td>
<td>0.175</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.673</td>
<td>0.747</td>
<td>0.746</td>
<td>0.732</td>
<td>0.730</td>
<td>0.751</td>
<td>0.753</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>1.61</td>
<td>0.598</td>
<td>0.755</td>
<td>0.755</td>
<td>0.671</td>
<td>0.661</td>
<td>0.665</td>
<td>0.670</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.43</td>
<td>0.343</td>
<td>0.513</td>
<td>0.514</td>
<td>0.425</td>
<td>0.416</td>
<td>0.454</td>
<td>0.459</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>0.099</td>
<td>0.167</td>
<td>0.164</td>
<td>0.129</td>
<td>0.123</td>
<td>0.151</td>
<td>0.155</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.026</td>
<td>0.054</td>
<td>0.052</td>
<td>0.049</td>
<td>0.045</td>
<td>0.049</td>
<td>0.052</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>0.091</td>
<td>0.144</td>
<td>0.141</td>
<td>0.131</td>
<td>0.127</td>
<td>0.134</td>
<td>0.140</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.531</td>
<td>0.640</td>
<td>0.636</td>
<td>0.625</td>
<td>0.616</td>
<td>0.629</td>
<td>0.635</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n=200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>1.61</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.43</td>
<td>0.970</td>
<td>0.986</td>
<td>0.986</td>
<td>0.983</td>
<td>0.983</td>
<td>0.978</td>
<td>0.978</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>0.442</td>
<td>0.565</td>
<td>0.564</td>
<td>0.530</td>
<td>0.530</td>
<td>0.544</td>
<td>0.544</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.030</td>
<td>0.051</td>
<td>0.051</td>
<td>0.047</td>
<td>0.046</td>
<td>0.048</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>0.445</td>
<td>0.531</td>
<td>0.530</td>
<td>0.517</td>
<td>0.517</td>
<td>0.525</td>
<td>0.525</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>1.61</td>
<td>0.998</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.43</td>
<td>0.926</td>
<td>0.965</td>
<td>0.965</td>
<td>0.957</td>
<td>0.957</td>
<td>0.948</td>
<td>0.948</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>0.343</td>
<td>0.455</td>
<td>0.454</td>
<td>0.420</td>
<td>0.419</td>
<td>0.434</td>
<td>0.434</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.028</td>
<td>0.053</td>
<td>0.053</td>
<td>0.048</td>
<td>0.047</td>
<td>0.047</td>
<td>0.047</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>0.321</td>
<td>0.416</td>
<td>0.415</td>
<td>0.403</td>
<td>0.402</td>
<td>0.405</td>
<td>0.406</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.990</td>
<td>0.996</td>
<td>0.996</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 5: Estimated power of 0.25-quantile residual life at $t_0 = 1$ for various $\tau_0$s, sample sizes (n) and censoring rates ($P_c$)

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$\tau_{0.5,1,0}$</th>
<th>Method</th>
<th>(n=50)</th>
<th>(n=200)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MCS</td>
<td>DIFF</td>
<td>PI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WI</td>
<td>WII</td>
<td>WI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WI</td>
<td>WII</td>
<td>WI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WI</td>
<td>WII</td>
<td>WI</td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.68</td>
<td>0.374</td>
<td>0.577</td>
<td>0.576</td>
<td>0.490</td>
</tr>
<tr>
<td>1.48</td>
<td>0.210</td>
<td>0.365</td>
<td>0.365</td>
<td>0.303</td>
</tr>
<tr>
<td>1.22</td>
<td>0.066</td>
<td>0.125</td>
<td>0.124</td>
<td>0.107</td>
</tr>
<tr>
<td>1.00</td>
<td>0.025</td>
<td>0.047</td>
<td>0.046</td>
<td>0.046</td>
</tr>
<tr>
<td>0.81</td>
<td>0.069</td>
<td>0.109</td>
<td>0.108</td>
<td>0.111</td>
</tr>
<tr>
<td>0.57</td>
<td>0.353</td>
<td>0.481</td>
<td>0.480</td>
<td>0.473</td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.68</td>
<td>0.346</td>
<td>0.538</td>
<td>0.538</td>
<td>0.437</td>
</tr>
<tr>
<td>1.48</td>
<td>0.186</td>
<td>0.338</td>
<td>0.338</td>
<td>0.255</td>
</tr>
<tr>
<td>1.22</td>
<td>0.052</td>
<td>0.116</td>
<td>0.116</td>
<td>0.094</td>
</tr>
<tr>
<td>1.00</td>
<td>0.022</td>
<td>0.049</td>
<td>0.049</td>
<td>0.042</td>
</tr>
<tr>
<td>0.81</td>
<td>0.048</td>
<td>0.112</td>
<td>0.111</td>
<td>0.102</td>
</tr>
<tr>
<td>0.57</td>
<td>0.279</td>
<td>0.425</td>
<td>0.423</td>
<td>0.407</td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.68</td>
<td>0.953</td>
<td>0.984</td>
<td>0.984</td>
<td>0.977</td>
</tr>
<tr>
<td>1.48</td>
<td>0.778</td>
<td>0.865</td>
<td>0.865</td>
<td>0.841</td>
</tr>
<tr>
<td>1.22</td>
<td>0.246</td>
<td>0.320</td>
<td>0.320</td>
<td>0.295</td>
</tr>
<tr>
<td>1.00</td>
<td>0.027</td>
<td>0.047</td>
<td>0.047</td>
<td>0.045</td>
</tr>
<tr>
<td>0.81</td>
<td>0.235</td>
<td>0.308</td>
<td>0.308</td>
<td>0.305</td>
</tr>
<tr>
<td>0.57</td>
<td>0.945</td>
<td>0.970</td>
<td>0.970</td>
<td>0.968</td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.68</td>
<td>0.938</td>
<td>0.973</td>
<td>0.973</td>
<td>0.964</td>
</tr>
<tr>
<td>1.48</td>
<td>0.729</td>
<td>0.833</td>
<td>0.833</td>
<td>0.799</td>
</tr>
<tr>
<td>1.22</td>
<td>0.205</td>
<td>0.303</td>
<td>0.302</td>
<td>0.275</td>
</tr>
<tr>
<td>1.00</td>
<td>0.022</td>
<td>0.055</td>
<td>0.054</td>
<td>0.048</td>
</tr>
<tr>
<td>0.81</td>
<td>0.183</td>
<td>0.266</td>
<td>0.266</td>
<td>0.260</td>
</tr>
<tr>
<td>0.57</td>
<td>0.896</td>
<td>0.934</td>
<td>0.935</td>
<td>0.937</td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 6: Estimated power of 0.75—quantile residual life at $t_0 = 1$ for various $\tau_0$s, sample sizes (n) and censoring rates ($P_c$)

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$\tau_{0.5,1,0}$</th>
<th>Method</th>
<th>MCS</th>
<th>DIFF</th>
<th>PI</th>
<th>PI*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>WI</td>
<td>WII</td>
<td>WI</td>
<td>WII</td>
</tr>
<tr>
<td>n=50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>1.58</td>
<td>0.745</td>
<td>0.779</td>
<td>0.780</td>
<td>0.769</td>
<td>0.762</td>
</tr>
<tr>
<td></td>
<td>1.41</td>
<td>0.501</td>
<td>0.602</td>
<td>0.600</td>
<td>0.561</td>
<td>0.555</td>
</tr>
<tr>
<td></td>
<td>1.19</td>
<td>0.147</td>
<td>0.226</td>
<td>0.224</td>
<td>0.189</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.037</td>
<td>0.063</td>
<td>0.061</td>
<td>0.055</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>0.148</td>
<td>0.211</td>
<td>0.208</td>
<td>0.194</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.62</td>
<td>0.715</td>
<td>0.742</td>
<td>0.731</td>
<td>0.768</td>
<td>0.753</td>
</tr>
<tr>
<td>40%</td>
<td>1.58</td>
<td>0.607</td>
<td>0.612</td>
<td>0.611</td>
<td>0.637</td>
<td>0.620</td>
</tr>
<tr>
<td></td>
<td>1.41</td>
<td>0.402</td>
<td>0.435</td>
<td>0.430</td>
<td>0.439</td>
<td>0.423</td>
</tr>
<tr>
<td></td>
<td>1.19</td>
<td>0.147</td>
<td>0.171</td>
<td>0.167</td>
<td>0.157</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.058</td>
<td>0.066</td>
<td>0.063</td>
<td>0.062</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>0.146</td>
<td>0.173</td>
<td>0.168</td>
<td>0.177</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>0.62</td>
<td>0.590</td>
<td>0.591</td>
<td>0.579</td>
<td>0.640</td>
<td>0.621</td>
</tr>
<tr>
<td>n = 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>1.58</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>1.41</td>
<td>0.982</td>
<td>0.994</td>
<td>0.994</td>
<td>0.993</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>1.19</td>
<td>0.492</td>
<td>0.618</td>
<td>0.619</td>
<td>0.583</td>
<td>0.582</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.028</td>
<td>0.048</td>
<td>0.048</td>
<td>0.041</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>0.472</td>
<td>0.587</td>
<td>0.587</td>
<td>0.564</td>
<td>0.564</td>
</tr>
<tr>
<td></td>
<td>0.62</td>
<td>0.999</td>
<td>0.999</td>
<td>0.998</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>40%</td>
<td>1.58</td>
<td>0.989</td>
<td>0.992</td>
<td>0.992</td>
<td>0.991</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>1.41</td>
<td>0.898</td>
<td>0.950</td>
<td>0.952</td>
<td>0.941</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>1.19</td>
<td>0.338</td>
<td>0.465</td>
<td>0.463</td>
<td>0.427</td>
<td>0.426</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.028</td>
<td>0.052</td>
<td>0.051</td>
<td>0.045</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>0.338</td>
<td>0.446</td>
<td>0.446</td>
<td>0.421</td>
<td>0.419</td>
</tr>
<tr>
<td></td>
<td>0.62</td>
<td>0.987</td>
<td>0.990</td>
<td>0.990</td>
<td>0.993</td>
<td>0.993</td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; WI denotes the variance estimators using the QRL estimate weighted by the number of events; WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 4: Estimated power of median residual life at $t_0 = 1$ for various $\tau_0$, sample sizes (n) and censoring rates ($P_c$)
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 5: Estimated power of 0.25-quantile residual life at $t_0 = 1$ for various $\tau_0$, sample sizes ($n$) and censoring rates ($P_c$)
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 6: Estimated power of $0.75 -$ quantile residual life at $t_0 = 1$ for various $\tau_0$, sample sizes (n) and censoring rates ($P_c$)
70% censoring rate and sample size 1500. These parameters were set to produce a population as close to the NSABP B30 data as possible. Note that in this simulation set (and in the application in Section 2.5), the censoring rate is very high (70%) and hence many of the upper quantiles are not estimable. We calculated the estimated Type I error and power for the tests of equality of the 0.1–QRLs (at years 1 & 5) and 0.2–QRLs at year 1 based on 5000 simulations. Across all the scenarios, only PI tests maintained correct Type I error. The Type I error of the DIFF test is largely inflated to as high as 0.75, which we believe is due to the high censoring rate. Among the tests with Type I error below the nominal level, the power of the proposed PI tests were higher compared to the MCS test (data not shown).

2.5 APPLICATION TO NSABP B-30 STUDY

We applied the proposed methods to analyze the data from the B-30 study conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP), currently part of the NRG Oncology. The NSABP B-30 study is a randomized multicenter phase III trial designed to compare the effect of three adjuvant chemotherapy regimens in breast cancer patients with positive lymph nodes [Swain et al., 2010]. The main interests were to examine whether the regimen with four cycles of doxorubicin, docetaxel and cyclophosphamide (ATC) was more effective than the regimen with four cycles of doxorubicin and cyclophosphamide followed by four cycles of docetaxel (AC→T), and whether the four cycles of doxorubicin and docetaxel (D-D) was as effective as the ATC regimen. For this application, we compared the 0.1 and 0.2–quantile residual lifetimes of the overall survival at various time points between the ATC and the AC→T groups.

Of the 3567 patients randomized to the AC→T or the ATC regimen between 1994 and 2004, 3501 (1745 in the AC→T group and 1756 in the ATC group) had supporting documents of follow-up and were included in the final analysis. A total of 797 deaths (361 in the AC→T and 436 in the ATC group) were observed by Dec.31, 2012. The median and maximum follow up times were 9.5 and 13.7 years, respectively. The estimated 10-year overall survival was 75.8% and 71.7% for the AC→T and ATC group, respectively. The time of the adjuvant chemotherapy ranged from 3 to 6 months. After chemotherapy, all patients with estrogen-receptor positive breast cancer received
hormone therapy at least for five years. Therefore, we chose \( t_0 \) to be years 1 and 5 in the analysis to study the residual survival after phases of the therapies. Figure 7 shows the estimated QRLs at years 1 and 5 for both groups. The patterns of the 0.1-QRL across different \( t_0 \) are slightly different between the two groups. For example, at year 1, the estimated 0.1-QRL of the AC→T group is 3.64 (95% CI: 3.49-3.79) compared to 3.93 (95% CI: 3.77-4.10) years in the ATC group, implying 90% of the patients are expected to survive at least 3.64 additional years when treated with AC→T and 3.93 additional years when treated with ATC if they survived through the first year. At year 5, the 0.1-QRLs were 4.98 (95% CI: 4.80-5.15) and 3.32 (95% CI: 3.14-3.49) years for AC→T and ATC respectively. The 0.2-QRLs at year 5 are not estimable due to the high survival, however, at year 1, the estimated 0.2-QRLs are 9.03 (95% CI: 8.80-9.27) years and 7.49 (95% CI: 7.25-7.73) years for the AC→T and ATC groups, respectively.

Table 7 shows the estimated ratio and the p-value of the test of the equality of QRLs between the ATC and the AC→T group, using the PI and MCS methods. Because the p-values using the two weighted variance estimates are similar, we only showed the results of the PI tests with the WI variance estimator. In general, the MCS test had the largest p-value consistently across all the quantiles and time points. For example, at 1 year, the p-value of the MCS test was 0.52 compared to the PI tests (p=0.16 to 0.26). These results are consistent with the simulation study described in Section 2.4 where MCS test was more conservative than the proposed PI tests, making it harder to reject the null hypothesis.

In summary, the AC→T regimen significantly improved the residual life among the post-surgery breast cancer patients as compared to the ATC regimen.

### 2.6 DISCUSSION

In this Chapter, we proposed several test statistics to compare two independent QRLs for data subject to right censoring. Our methods are based on the difference between two QRL estimators (DIFF test) and the plug-in approach (PI tests). The asymptotic properties of the proposed statistics are also outlined. In the simulation studies, we evaluated these methods under various scenarios with various distributions of the failure time in each group, different values of the ratio between two population QRLs, and different censoring rates.
Figure 7: Estimated $p-$quantile residual lifetimes and the corresponding point wise 95% confidence intervals for breast cancer patients with positive lymph nodes from NSABP data

Table 7: Estimated $p-$QRLs in years at different $t_0$s and results of the test of equality of two $p-$QRLs between AC$\rightarrow$T and ATC groups by different methods

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Estimated $p-$QRL (95%CI)</th>
<th>$p-$value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AC$\rightarrow$T</td>
<td>ATC</td>
</tr>
<tr>
<td>$t_0 = 1$ year, $p = 0.1$</td>
<td>3.64(3.49-3.79)</td>
<td>3.93(3.77-4.10)</td>
</tr>
<tr>
<td>$t_0 = 1$ year, $p = 0.2$</td>
<td>9.03(8.80-9.27)</td>
<td>7.49(7.25-7.73)</td>
</tr>
<tr>
<td>$t_0 = 5$ year, $p = 0.1$</td>
<td>4.98(4.80-5.15)</td>
<td>3.32(3.14-3.49)</td>
</tr>
</tbody>
</table>

$p-$QRL = $p-$quantile residual lifetimes; MCS = minimum chi-square statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated QRL from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both PI and PI* tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22.
Our results show that PI tests maintain the Type I error better across all situations compared to the DIFF test and the existing minimum chi-square statistic approach (MCS). DIFF test had an inflated Type I error when the censoring rate was higher than 40%. Variance under the null hypothesis was estimated by separately weighting the estimated quantiles using the number of events or the KM survival function. Both forms of weights resulted in very similar test statistics with comparable performances. Moreover, the empirical power of the proposed tests was greater than the power of the MCS test across all scenarios. Although the DIFF test had greater power compared to the other tests, it failed to maintain the nominal Type I error and hence is not recommended. One limitation we observed in the simulation study is that when the true \( \tau \) is too far away from the null hypothesized value 1, the power of all methods decrease for large quantiles and \( t_0 \) in small sample (\( n<200 \)). This may be due to the unreliability of the KM curve estimate in the tails. (data not shown)

Unlike MCS test, our methods require estimating the density functions in the variance estimation. However, we obtain stable estimates of the variance functions by using the IPW Gaussian Kernel density estimate instead of the empirical density estimator.

Besides the improvement in Type I error and power, our proposed tests are straightforward and less computationally intensive, with the closed-form variance estimators available, compared to the confidence interval based approaches (e.g. MCS). Our approaches, however, only allow comparison between two QRLs. Future work will focus on the extension of these methods to the comparison of QRLs across more than two independent groups.
3.0 EFFICIENT ESTIMATOR OF THE QUANTILE RESIDUAL LIFETIME

In the second Chapter of this dissertation, we have proposed various tests for comparing the equality of the quantiles of residual life and compared them to existing methods. Note that these methods did not use any covariate information. We anticipate that by incorporating information from patient covariates, as has been discussed in Section 1.2 of Chapter 1, the efficiency in estimating the QRLs can be improved. In Section 3.1, we introduce inverse probability weighted estimator (IPW) for QRL. In Section 3.2, we outline how to add the covariate information and obtain the efficient augmented inverse probability weighted estimator (AIPW) for QRL. We also discuss the variance estimator for the AIPW estimator. In Section 3.3, we conduct Monte-Carlo studies to evaluate the efficiency gain of the AIPW estimator from the Kaplan-Meier (KM) type estimator and the IPW estimator. We also conduct simulation studies to compare the power of using the AIPW estimator to the KM-type estimator in the two-sample plug-in tests proposed in the Chapter 2. In Section 3.4, we apply the proposed plug-in tests with the AIPW estimator to the same randomized breast cancer clinical trial that we used in the Section 2.5 of Chapter 2.

3.1 THE INVERSE PROBABILITY WEIGHTED ESTIMATORS FOR THE QRL

Based on Equation 2.1, a naive estimator \( \hat{\theta}^{naive} \) for the QRL can be obtained by solving the equation

\[
\sum_{i=1}^{n} \delta_i [I(T_i \leq t_0 + \theta) - (1 - p)I(T_i \leq t_0) - p] = 0. \tag{3.1}
\]

For simplicity, we drop \( p \) and \( t_0 \) from the notation. This estimator is biased because it assumes same weight for each observed failure time (complete cases), that is, it does not account for the fact
that patients have varying likelihood of being completely observed. To get an unbiased estimator
for the QRL, we define \( \hat{\theta}^{IPW} \) from the weighted estimating equation

\[
\sum_{i=1}^{n} \frac{\delta_i[I(T_i \leq t_0 + \theta) - (1 - p)I(T_i \leq t_0) - p]}{\pi_i} = 0,
\] (3.2)

and solve for \( \theta \) to obtain \( \hat{\theta}^{IPW} \), where \( \pi_i = P(T_i \leq C_i) \) is the probability of not being censored at the observed failure time. We refer to \( \hat{\theta}^{IPW} \) as the inverse-probability-of-censoring-weighted (IPW) QRL estimator. The IPW estimator can be shown to be consistent to the true QRL when \( \pi_i \)s are known or correctly estimated. However, \( \pi_i \) is generally unknown and hence needs to be estimated in order to solve the estimating equation 3.2. Assuming that the right-censoring is independent of the failure time, \( \pi_i \) can be consistently estimated by the KM curve of the censoring distribution at the observed failure time, which we denote by \( S_C(T_i) \). The assumption of independent censoring can be relaxed such that conditional on baseline covariates, the probability of being completely observed does not depend on the censoring. In this case, we can estimate \( \pi_i \) from some conditional models based on the covariate information. The asymptotic normality of \( \hat{\theta}^{IPW} \) can be derived from theories of influence functions and M-estimation and its asymptotic variance can be estimated using the robust sandwich estimators.

### 3.2 THE EFFICIENT AUGMENTED INVERSE PROBABILITY WEIGHTED
ESTIMATORS FOR THE QRL

Suppose that one would like to estimate \( \beta \) from full data \( Z_i \) using the estimating equation \( \sum_{i=1}^{n} m(Z_i, \beta) = 0 \). Because of the presence of the missing data (indicated by \( \Delta_i = 0 \)), however, the full data is observed (indicated by \( \Delta_i = 1 \)) only for a subset of individuals in the sample. According to the theory described in Robins et al. [1994], under MAR, an efficient semi-parametric AIPW estimator can be obtained from the equation

\[
\sum_{i=1}^{n} \left\{ \frac{\Delta_i}{\pi(\Delta_i = 1, Z_i, \psi_0)} m(Z_i, \beta) + \frac{\Delta_i - \pi(\Delta_i = 1, Z_i, \psi_0)}{\pi(\Delta_i = 1, Z_i, \psi_0)} L_i \right\} = 0,
\] (3.3)
where \( \pi(\Delta_i = 1, Z_i, \psi_0) \) is the probability of being observed for the \( i^{th} \) subject, \( \psi_0 \) is the nuisance parameters for the missing mechanisms, \( L_i \) is an arbitrary function that depends on the observed data for the \( i^{th} \) subject. The first part of this estimating equation is the IPW estimating equation from the complete data and the second part is the augmented term which adds the predicted information from the missing data that has expectation zero under the truth. Because the AIPW estimators use both information from the complete and missing data, they have better efficiency compared to the IPW estimators.

Tsiatis [2007] gave details on the AIPW estimators for models with survival distribution that is subject to right-censoring. By considering censoring as a special type of monotone missing, an efficient estimator for the parameter of interest \( \beta \) in the survival distribution can be obtained from the equation

\[
\sum_{i=1}^{n} \left[ \frac{\delta_i}{S_C(X_i, G_i(X_i))} m(Z_i, \beta) + \int_{0}^{\infty} \frac{dM_i(r)}{S_C(r, G_i(r))} L_i(G_i(r)) \right] = 0, \tag{3.4}
\]

where \( X_i \) denotes the observed failure or censoring time as defined in Section 2.1, \( G_i(t) \) denotes the observed information upon time \( t \), \( S_C(t, G(t)) \) denotes the conditional survival function of the censoring distribution at time \( t \) given the observed information up to but not including time \( t \), \( G_i(t) \), for the \( i^{th} \) individual, \( M_i(t) = S_C(t, G_i(t)) - N_i(t) = I(C_i \leq t, \delta_i = 0) \), \( Y_i(t) = I(X_i \geq t) \), \( \Lambda_C(t) = \int_{0}^{t} Y^{-1}(s)d\lambda_C(s) \) is the cumulative hazard function of the censoring distribution, \( \lambda_C(t) \) is the hazard function of the censoring distribution at time \( t \), \( Y(t) = \sum_{i=1}^{n} Y_i(t) \) and \( N_C(t) = \sum_{i=1}^{n} N_i(t) \). By Theorem 10.4 in Tsiatis [2007], the most efficient AIPW estimator for \( \beta \) would then be obtained by choosing

\[
L_i(G_i(r)) = E\{m(Z_i, \beta)|T_i \geq r, G_i(r)\}. \tag{3.5}
\]

Following the same notation with the data set-up as in Section 2.1 and by replacing \( m(Z_i, \beta, t_0) \) in Equation 3.5 with the individual estimating equation for QRL when there is no censoring, we can obtain an efficient AIPW estimator for QRL by solving the equation

\[
\sum_{i=1}^{n} \left[ \frac{\delta_i}{S_C(T_i, G_i(T_i))} m(T_i, \theta, t_0) + \int_{0}^{\infty} \frac{dM_i(r)}{S_C(r, G_i(r))} L_i(G_i(r)) \right] = 0, \tag{3.6}
\]
where \( m(T_i, \theta, t_0) = I(T_i \leq t_0 + \theta) − (1 − p)I(T_i \leq t_0) − p \) and \( L_i\{G_i(r)\} = E\{m(T_i, \theta, t_0)|T_i \geq r, G_i(r)\} \). When censoring does not depend on any covariate history, Equation 3.6 is simplified to

\[
\sum_{i=1}^{n} \left[ \frac{\delta_i}{S_C(T_i)} m(T_i, \theta, t_0) + \int_{0}^{\infty} \frac{dM_C(r)}{S_C(r)} L_i\{G_i(r)\} \right] = 0. \tag{3.7}
\]

A two-step estimating process can be used to solve Equation 3.7 for \( \theta \). In the first step, we plug-in an initial QRL estimate, such as the IPW QRL estimate, for \( \theta \) in the augmented part of the equation. Then, except for \( \theta \), we replace all the other unknown functions with their consistent estimators in the first part of the equation. Therefore, the plug-in estimating equation in the first step can be wrote as

\[
\sum_{i=1}^{n} \left[ \frac{\delta_i}{S_C(T_i)} m(T_i, \theta, t_0) + \int_{0}^{\infty} \frac{d\hat{M}_C(r)}{S_C(r)} \hat{L}_i\{G_i(r), \hat{\theta}^{(0)}\} \right] = 0, \tag{3.8}
\]

where \( \hat{M}_C(t) = N_C(t) − \int_{0}^{t} Y_i(s)d\hat{\Lambda}_C(s), \hat{\Lambda}_C(t) = \int_{0}^{t} Y^{-1}(s)dN_C(s) \) is the Nelson-Aalen estimator of the cumulative hazard function for the censoring distribution and \( \hat{\theta}^{(0)} \) denotes initial estimate for \( \theta \). The key problem is how to estimate \( \hat{L}_i\{G_i(t), \theta\} = E\{m(T_i, \theta, t_0)|T_i \geq t, G_i(t)\} \) for a given \( \theta \). One way is to fit a conditional regression model of the survival distribution using all the data that has an observed event time or censoring time equal to or greater than time \( t \). In the model, we add the observed covariates that are prognostic of the survival distribution as auxiliary variables to help predict the conditional survival functions, which are denoted by \( S\{t_0 + \theta|T \geq t, G_i(t)\} \) and \( S\{t_0|T \geq t, G_i(t)\} \). By plugging in the estimated conditional survival functions, we can obtain the estimator of the conditional expectation in Equation 3.7 as

\[
\hat{L}\{G_i(t), \hat{\theta}^{(0)}\} = \hat{S}\{t_0 + \hat{\theta}^{(0)}|T \geq t, G_i(t)\} − (1 − p)\hat{S}\{t_0|T \geq t, G_i(t)\}. \tag{3.9}
\]

In the second step, we solve \( \hat{\theta}^{(1)} \) from Equation 3.8. Repeat steps 1 and 2 until \( \hat{\theta} \) converges. We denote this estimator as \( \hat{\theta}^{AIPW} \).

Following Wahed [2009], the influence function of the AIPW estimator can be written as

\[
\Psi_i = f_{T_i}^{-1}(t_0 + \theta) \left[ \frac{\delta_i}{S_C\{T_i, G_i(T_i)\}} m(T_i, \theta, t_0) + \int_{0}^{\infty} \frac{dM_C\{r, G_i(r)\}}{S_C\{r, G_i(r)\}} L_i\{G_i(r)\} \right]. \tag{3.10}
\]
Therefore, a natural estimator for the asymptotic variance of the AIPW estimator can be de-
rived as
\[
\hat{\text{Var}}(\hat{\theta}_{\text{AIPW}}) = n^{-2} \sum_{i=1}^{n} \hat{\Psi}_i^2, \tag{3.11}
\]
where \(\hat{\Psi}_i = \hat{f}_{T_i}^{-1}(t_0 + \hat{\theta}_{\text{AIPW}}) \left[ \frac{\delta_i}{S_{C\{T_i,G_i(t_i)\}}} \hat{m}(T_i, \theta, t_0) + \int_0^\infty \frac{dM_{G_i(r,G_i(r))}}{S_{C\{r,G_i(r)\}}} \hat{L}_i\{G_i(r), \hat{\theta}_{\text{AIPW}}\} \right] \) and \(\hat{m}(T_i, \theta, t_0) = I(T_i \leq t_0 + \hat{\theta}_{\text{AIPW}}) - (1 - p)I(T_i \leq t_0) - p \) and \(\hat{L}_i\{G_i(t), \hat{\theta}_{\text{AIPW}}\} = \hat{S}\{t_0 + \hat{\theta}_{\text{AIPW}}|T \geq t, G_i(t)\} - (1 - p)\hat{S}\{t_0|T \geq t, G_i(t)\}\).

3.3 SIMULATION STUDIES

We conducted a Monte-Carlo (MC) simulation study to evaluate the efficiency gain of the AIPW estimator by comparing its Monte-Carlo variance with the KM-type estimator (Equation 2.3) and the IPW estimator. We also conducted another Monte-Carlo simulation study to evaluate the potential power gain of applying the AIPW estimator to the previously proposed two-sample tests of QRLs (Equations 2.6 and 2.7) over the results using the original KM-type estimator. In the simulation study, we generated the failure times from the Weibull distributions and used the Cox PH regression to estimate the conditional mean of the augmented term \(L\{G_i(t), \hat{\theta}^{(0)}\}\).

3.3.1 Efficiency gain of the AIPW estimator

In the efficiency gain study, the simulated failure time was generated from a parametric proportional hazards model (Cox) with covariates \(x\), where \(x = \{x_1, x_2, ..., x_k\}^T\) is a \(k\) by 1 vector. The model is specified as
\[
S(t; x) = \exp\{- (\rho t)^\eta \exp(\beta x)\}, \tag{3.12}
\]
where \(\beta\) is the \(k\) vector of the covariate effects for \(x\). Under this model, the failure time follows a Weibull distribution with the shape parameter \(\rho \ast \exp(\beta x)\) and the scale parameter \(\eta\). For each patient, the independent failure time is generated by
\[ T = \frac{1}{\rho \exp(\beta x)} \left\{ -\log(1 - U) \right\}^{\frac{1}{\gamma}}, \]

where \( U \sim UNIF(0, 1) \). Note that in the simulated model, the failure time depends on the covariates \( x \), which will be used as the auxiliary variables in estimating the QRL using the AIPW method.

For simplicity, independent censoring times are generated from a uniform distribution separately and do not depend on the covariates \( x \) in Equation 3.12. To evaluate the relationship between the efficiency gain and the parameters of the survival model, we conducted the simulation studies with different sample sizes, effect sizes (\( \beta \)) and censoring rates.

Tables 8 and 9 show the estimated bias and Monte-Carlo variance of the KM-type estimator, IPW estimator and the AIPW estimators for the 0.5-QRL and the ratio between the variance of each estimator and the AIPW estimator. In Table 8, only a binary covariate \( (x \sim BIN(1, 0.5)) \) is used, while in Table 9 both binary and continuous covariates are used \( (x_1 \sim BIN(1, 0.5); x_2) \). The estimated bias for all the estimators are very small, ranging from 0-1.4\%, indicating that all three estimators are consistent to the true QRL. In general, the efficiency gain of the AIPW estimator over the other two estimators increases as the censoring rate, the sample size and the absolute effect size \( \beta \) increases. For example, in Scenario I, when the censoring rate is 60\%, the Monte-Carlo variance of the AIPW estimator for 0.5-QRL at \( t_0 = 0.3 \) is 0.0267, while the variance of the KM-type and IPW estimators are 0.0302 and 0.0300, respectively. Thus, the efficiency gain of the AIPW estimator over the KM-type and IPW estimators are 13\% and 12\%. When the censoring rate decreases, the efficiency of the AIPW estimator over other estimators decreases. For example, when the censoring rate decreases to 31\%, the efficiency gain is 6\% over both the KM-type and IPW estimators. When there is no censoring, the efficiency gain of the AIPW estimator is almost zero over the other two estimators. A similar pattern is observed for the scenarios with two auxiliary variables (Table 9). In Scenario II, the AIPW estimator achieves the largest efficiency gain (26\% for both estimators) at the largest considered sample size (n=500) with other parameters being fixed. In Scenario III, the AIPW estimator achieves the largest efficiency gain (17\% for both estimators) when \( \beta_x = 3 \), compared to no efficiency gain when \( \beta_x = 0 \). This implies that when the covariate is highly correlated with the survival time, using the auxiliary variables will help gain efficiency.
We also calculated the 95% coverage rate for the confidence interval using the variance estimator of the AIPW estimator and compared it to the 95% coverage rate of the KM-type estimator. Figure 8 shows the estimated 95% coverage rates for three estimators for various censoring percentages, sample sizes and $\beta_x$s in Scenario I to III. The coverage rates are much lower than 95% when censoring rate is higher than 40% for the KM-type estimator. When the censoring rate is higher than 50%, the coverage rates for the IPW estimator are also low. However, the AIPW estimator maintains a closer coverage rate to 95% regardless of the censoring rates. For example, in Scenario I, when the censoring rate is 15%, the coverage rates of the AIPW, KM type and IPW estimators are 0.94, 0.94 and 0.95, respectively. However, when the censoring rates increase to 50%, the coverage rates of the AIPW estimator is 96%, while the coverage rates of the KM type and IPW estimators decrease to 0.85 and 0.90. The average length of the 95%CI for the AIPW estimators in our simulation studies ranged from 0.3-1.7.

To evaluate the efficiency gain of the AIPW estimator under the miss-specification of the conditional prediction model, we conducted another Monte-Carlo study by generating the failure time from a log-normal accelerate failure (AFT) model

$$T_i = \exp(\beta_0 + \beta_x x_i + \sigma z_i),$$

where $Z_i$ follows a standard normal distribution, $x_i \sim BIN(1, 0.5)$ is the covariate related to the survival distribution, $\beta_0$ is the shape parameters of the log-scale failure time when $x_i = 0$ and $\beta_x$ is coefficient effect for $x$, $\sigma$ is the scale parameter of the log-scale failure time, respectively. The log-scale failure time follows a normal distribution with mean $\beta_0 + \beta_x x$ and variance $\sigma^2$.

In this case, the assumption of the proportional hazards is no longer hold for the log-normal AFT model. Therefore, the condition model used to predict the information for censored observation are miss-specified.

Table 10 shows the estimated bias and Monte-Carlo variance of the KM-type, IPW and AIPW estimators for the 0.5-QRL and the ratio between the variance of each estimator and the AIPW estimator when conditional model is miss-specified. For all scenarios in Table 10, only a binary covariate ($x \sim BIN(0, 1)$) is used. The results show all three estimators are consistent to the true $\theta_0$, with the estimated bias ranging from $0 - 0.3\%$. Moreover, when the conditional model used to predict the information for the censored observations is miss-specified, using AIPW estimator still
Table 8: Estimated bias and Monte-Carlo variance of three 0.5—QRL estimators using one auxiliary covariate

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Changing Parameter</th>
<th>$\theta_0$</th>
<th>KM-type</th>
<th>IPW</th>
<th>AIPW</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bias</td>
<td>Var</td>
<td>Bias</td>
<td>Var</td>
</tr>
<tr>
<td>I:</td>
<td>$P_c$</td>
<td>60%</td>
<td>1.73</td>
<td>0.6</td>
<td>0.0302</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51%</td>
<td>1.73</td>
<td>0.6</td>
<td>0.0251</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>$\beta_x = 2,$</td>
<td>41%</td>
<td>1.73</td>
<td>0.6</td>
<td>0.0203</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>$t_0 = 0.3,$</td>
<td>31%</td>
<td>1.73</td>
<td>0.6</td>
<td>0.0177</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>n=500</td>
<td>15%</td>
<td>1.73</td>
<td>0.6</td>
<td>0.0168</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0%</td>
<td>1.73</td>
<td>0.6</td>
<td>0.0146</td>
<td>0.6</td>
</tr>
<tr>
<td>II:</td>
<td>n=500</td>
<td>500</td>
<td>1.73</td>
<td>0.9</td>
<td>0.0310</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>400</td>
<td>1.73</td>
<td>1.2</td>
<td>0.0391</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>1.74</td>
<td>1.0</td>
<td>0.0498</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.73</td>
<td>1.0</td>
<td>0.0665</td>
<td>1.0</td>
</tr>
<tr>
<td>III:</td>
<td></td>
<td>3</td>
<td>1.47</td>
<td>1.4</td>
<td>0.0246</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>2.08</td>
<td>0.5</td>
<td>0.0183</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2.89</td>
<td>0.3</td>
<td>0.0211</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>3.87</td>
<td>0.3</td>
<td>0.0290</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1</td>
<td>4.96</td>
<td>0.0</td>
<td>0.0550</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2</td>
<td>6.09</td>
<td>0.3</td>
<td>0.1349</td>
<td>0.3</td>
</tr>
</tbody>
</table>

All results are based on 3000 simulations; The observed survival time is generated from $S(t; x) = \exp\{-0.2t^2 \exp(\beta_x x)\}$ where $x \sim BIN(1, 0.5)$; Bias is calculated as (mean estimate - $\theta_0$)/$\theta_0 \times 100$; Ratio is calculated as the variance of each estimator divided by the variance of the AIPW estimator.
Table 9: Estimated bias and Monte-Carlo variance of three 0.5–QRL estimators for \( t_0 = 1 \) using two auxiliary covariates

<table>
<thead>
<tr>
<th>Scenario</th>
<th>( P_c )</th>
<th>( \theta_0 )</th>
<th>KM Bias</th>
<th>KM Var</th>
<th>IPW Bias</th>
<th>IPW Var</th>
<th>AIPW Bias</th>
<th>AIPW Var</th>
<th>Ratio KM</th>
<th>Ratio IPW</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
<td>56%</td>
<td>1.98</td>
<td>0.9</td>
<td>0.0364</td>
<td>0.9</td>
<td>0.0363</td>
<td>0.9</td>
<td>0.0329</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>IV</td>
<td>41%</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0223</td>
<td>0.1</td>
<td>0.0223</td>
<td>0.1</td>
<td>0.0212</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>IV</td>
<td>22%</td>
<td>0.6</td>
<td>0.6</td>
<td>0.0190</td>
<td>0.6</td>
<td>0.0190</td>
<td>0.6</td>
<td>0.0182</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>V</td>
<td>56%</td>
<td>1.68</td>
<td>0.8</td>
<td>0.0276</td>
<td>0.8</td>
<td>0.0275</td>
<td>0.7</td>
<td>0.0250</td>
<td>1.11</td>
<td>1.10</td>
</tr>
</tbody>
</table>

IV: \( \beta_{x_1} = 1.6, \beta_{x_2} = 0.06; \) V: \( \beta_{x_1} = 0.5, \beta_{x_2} = 0.08; \) All results are based on 3000 simulations; The observed survival time is generated from \( S(t;x) = \exp(-0.2t \cdot \exp(\beta_{x_1} x_1 + \beta_{x_2} x_2)) \), \( x_1 \sim BIN(1,0.5), x_2=I(z = k) * UNIF(a_k,b_k) \), \( a_1 = 9, a_2 = 26, a_3 = 35, a_4 = 55, b_1 = 25, b_2 = 34, b_3 = 54 \) and \( b_4 = 64; \) \( z \sim Multinomial(n = 1, p_1 = 0.16, p_2 = 0.2, p_3 = 0.43, p_4 = 0.21) \); Bias is calculated as \((\text{mean estimate} - \theta_0)/\theta_0*100; \) Ratio is calculated as the variance of each estimator divided by the variance of the AIPW estimator.

Gain efficiency from the KM-type and the IPW estimators. For example, for Scenario VI, when the censoring rate is 54\% and the sample size is 300, the Monte-Carlo variance of the AIPW estimator for 0.5-QRL at \( t_0 = 1 \) is 0.0567, while the variance of the KM-type and IPW estimators are 0.0628 and 0.0627, respectively. Thus, the efficiency gain of the AIPW estimator over the KM-type and IPW estimators are both 11\%. When the censoring rate decreases to 31\%, the efficiency gains of the AIPW estimator over other estimators decrease to 4\% and 3\%.

3.3.2 Efficiency gain of two-sample test of the QRLs with the AIPW estimator

The previous simulation shows the AIPW estimator has significant efficiency gain in estimating the QRL over the KM-type and IPW estimators. Therefore, we conducted another Monte-Carlo simulation study to evaluate the efficiency gain of applying the AIPW estimator in the previously proposed plug-in (PI) two-sample QRL test over the same test using the KM-type estimator. The new test statistics is obtained by replacing all the KM-type estimators with the AIPW estimators in Equations 2.6 and 2.7. In all the Scenarios for the two-sample test, the failure times are generated through the survival function \( S(t;z;x) = \exp\{-0.2t \cdot \exp(\beta_{trt} z + \beta_x x)\} \), where \( z \sim BIN(0,1) \) is the group indicator and \( x \sim BIN(0,1) \) is a binary covariate used as the auxiliary variable in the AIPW estimation.
All results are based on 3000 simulations; The observed survival time is generated from $S(t; x) = \exp\{-0.2t\exp(\beta_x x)\}$ where $x \sim BIN(1, 0.5)$; Scenario I: $\beta_x = 2$, $t_0 = 0.3$, $n = 500$; Scenario II: $\beta_x = 2$, $t_0 = 1$, $P_c = 56\%$; Scenario III: $t_0 = 1$, $P_c = 56\%$, $n = 500$

Figure 8: Monte-Carlo coverage rates for the estimated 95% CI of the 0.5–QRL for various censoring rates ($P_c$), sample sizes ($n$) and $\beta_x$. 

42
Table 10: Estimated bias and Monte-Carlo variance of three 0.5-QRL estimators for $t_0 = 1$ using one auxiliary covariate under miss-specification

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$P_c$</th>
<th>$n$</th>
<th>$\theta_0$</th>
<th>KM-type</th>
<th>IPW</th>
<th>AIPW</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bias</td>
<td>Var</td>
<td>Bias</td>
<td>Var</td>
</tr>
<tr>
<td>VI</td>
<td>54%</td>
<td>300</td>
<td>0.0</td>
<td>0.0628</td>
<td>0.0</td>
<td>0.0627</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.3</td>
<td>0.0365</td>
<td>0.3</td>
<td>0.0364</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>31%</td>
<td>300</td>
<td>0.0</td>
<td>0.0468</td>
<td>0.0</td>
<td>0.0467</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.0</td>
<td>0.0274</td>
<td>0.0</td>
<td>0.0274</td>
<td>0.0</td>
</tr>
<tr>
<td>VII</td>
<td>56%</td>
<td>300</td>
<td>0.0</td>
<td>0.0300</td>
<td>0.0</td>
<td>0.0299</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>0.0</td>
<td>0.0180</td>
<td>0.0</td>
<td>0.0180</td>
<td>0.0</td>
</tr>
</tbody>
</table>

VI: $\beta_0 = 1$, $\beta_x = 0.8$ and $\sigma = 0.47$; VII: $\beta_0 = 0.59$, $\beta_x = 0.8$ and $\sigma = 0.47$; All results are based on 3000 simulations; The observed survival time is generated from $T = \exp(\beta_0 + \beta_xx + \sigma z)$, where $x \sim \text{BIN}(1, 0.5)$, $z \sim \mathcal{N}(0, 1)$, and $\sigma = P_c = $ censoring percentage; Bias is calculated as $(\text{mean estimate} - \theta_0)/\theta_0 \times 100$; Ratio is calculated as the variance of each estimator divided by the variance of the AIPW estimator.

The results of the simulation studies show that using the AIPW estimator in the PI tests still achieves the correct Type I error close to 0.05, ranging from 0.03-0.06 (Data now shown). Table 11 shows the Monte-Carlo power for the PI tests of the 0.5-QRL at $t_0 = 1$ using the AIPW estimator and KM-type estimator in four different scenarios. In all four scenarios, the true model of the failure time includes the treatment effect and a binary effect. Therefore, the AIPW estimator is obtained using the binary covariate as the auxiliary variable.

In general, although the Monte-Carlo variance of the AIPW estimators are smaller than the variance of the KM-type estimators, using the AIPW estimator only gains minimal power in the PI tests compared with the KM-type estimators. For example, in Scenario VIII, the estimated variance of the AIPW estimators based on 3000 simulations for $z = 0$ and 1 are 0.0442 and 0.0236, gaining 13% and 11% efficiency over the KM-type estimators (variances equal to 0.0498 and 0.0263 for $z=0$ and 1, respectively). However, the estimated power based on 3000 simulations for the two PI tests with the AIPW estimator are 0.48 and 0.60, compared to 0.48 and 0.59 for the tests with the KM-type estimators.
Table 11: Estimated bias, Monte-Carlo variance and power of the two sample plug-in tests with the AIPW and Kaplan-Meier type (KM) estimator for $0.5 - \text{QRLs at } t_0 = 1$

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Method</th>
<th>Power</th>
<th>QRL estimation</th>
<th>Ratio(QRL)</th>
<th>Ratio(Power)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$z=0$</td>
<td>$z=1$</td>
<td>$z=0$</td>
</tr>
<tr>
<td>VIII</td>
<td>KM</td>
<td>0.48</td>
<td>1.3</td>
<td>0.0498</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>AIPW</td>
<td>0.48</td>
<td>1.1</td>
<td>0.0442</td>
<td>1.0</td>
</tr>
<tr>
<td>IX</td>
<td>KM</td>
<td>0.97</td>
<td>0.5</td>
<td>0.0416</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>AIPW</td>
<td>0.97</td>
<td>0.5</td>
<td>0.0404</td>
<td>0.7</td>
</tr>
<tr>
<td>X</td>
<td>KM</td>
<td>0.74</td>
<td>0.0</td>
<td>0.0368</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>AIPW</td>
<td>0.73</td>
<td>-0.1</td>
<td>0.0327</td>
<td>0.7</td>
</tr>
<tr>
<td>XI</td>
<td>KM</td>
<td>0.99</td>
<td>0.0</td>
<td>0.0298</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>AIPW</td>
<td>0.99</td>
<td>-0.1</td>
<td>0.0293</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Power results are based on testing the $H_0: \tau = 1$. All results are based on 3000 simulations with $n=300$ and $S(t; z; x) = \exp\{-(0.2t)^2 \exp(\beta_{trt} z + \beta_xx)\};$ Scenario VIII: $\beta_{trt} = 0.7$, $\beta_x = 2$, $P_c = 50\%$ & $\tau_0 = 0.69$; Scenario IX: $\beta_{trt} = 0.7$, $\beta_x = 1$, $P_c = 60\%$ & $\tau_0 = 0.63$; Scenario X: $\beta_{trt} = -0.7$, $\beta_x = 2$, $P_c = 43\%$ & $\tau_0 = 1.50$; Scenario XI: $\beta_{trt} = -0.7$, $\beta_x = 1$, $P_c = 43\%$ & $\tau_0 = 1.54$; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated QRL from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both PI and PI* tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; Ratio for QRL is calculated as the variance of the KM estimator divided by the variance of the AIPW estimator; Ratio for Power is calculated as the power of the PI test using the AIPW estimator divided by the power of same PI test using the KM type estimator.
3.4 APPLICATION IN NSABP B-30 STUDY

To demonstrate the use of our proposed AIPW estimator, we re-analyzed the data-set from the NSABP B-30 study in Chapter 2 (Section 2.5) by applying the AIPW QRL estimator and the two-sample PI tests with the AIPW estimator.

To obtain the AIPW estimator, we chose six covariates that are potentially related to the survival outcome of the breast cancer patients as the auxiliary variables, including two demographic variables (BMI and Age in years), estrogen receptor value (ER), menopausal status (yes or no), mastectomy type (lumpectomy or mastectomy) and maximum pathologic tumor size in centimeter. For simplicity, patients with any of the covariates missing are removed from the analysis. Among 3501 patients with supporting documents of follow-up, 1683 (53 with missing tumor size and 9 with missing menopausal status are removed) in the AC→T group and 1709 (34 with missing tumor size and 13 with missing menopausal status are removed) in the ATC group have all six covariates observed and included in the analysis.

Table 12 shows the distribution of the auxiliary variables by treatment groups. In general, all six covariates distributed equally between the breast cancer patients assigned to the AC→T and ATC group. For example, the percentage of the patients with negative or borderline estrogen receptor value are 18.4% for both groups. Moreover, 45.5% and 46% of the patients are in menopausal status in the AC→T and ATC group, respectively. The maximum pathologic tumor size among 19.1% of the patients in both groups are 4 centimeter or larger. The equal distributions of the covariates between the treatment groups indicate that the randomization is performed well.

Similar to the data analysis presented in Chapter 2, we obtained the AIPW estimates of the 0.1-QRL at year 1 and the 0.2-QRLs at year 1 and year 5 for the AC→T and ATC group. We also tested the equality of the two QRLs using the proposed two-sample PI tests with the AIPW estimators. Table 13 shows the estimated QRLs with the 95%CIs and the ratio between the two QRL estimates. In general, the AIPW estimates are very similar to the KM estimates (Data now shown). At year 1, the estimated 0.1-QRL of the AC→T group is 3.67 (95%CI: 3.46-3.87) compared to 3.93 (95%CI: 3.66-4.20) years in the ATC group, implying 90% of the patients are expected to survive at least 3.67 additional years when treated with AC→T and 3.93 additional years when treated with ATC if they survived through the first year. At year 5, the 0.1-QRLs were 4.58 (95%CI: 4.34-4.83) and
3.32 (95% CI: 3.02-3.61) years for AC→T and ATC respectively. At year 1, the estimated 0.2-QRLs are 8.82 (95% CI: 8.43-9.22) years and 7.47 (95% CI: 6.94-8.00) years for the AC→T and ATC groups, respectively.

The last four columns of Table 13 shows the p-values for the two-sample PI tests based on the KM-type QRL estimators and the proposed AIPW QRL estimators. Except for the PI* test of 0.2-QRL at year 5, the p-values are very similar between the PI tests with the KM-type and AIPW estimators. For example, the p-value of the PI test for the 0.1-QRLs at year 1 is 0.192 based on the KM-type estimator, while the p-value is 0.176 for the test based on the AIPW QRL estimator. Although the p-value of the PI* test for 0.2-QRL at year 1 is 0.001 based on the KM-type estimators, compared to 0.015 for the PI* test based on the AIPW estimators, the results of the both tests are still consistently significant. In general, using the AIPW estimators in the PI tests gave consistent results with the PI tests based on the KM-type estimators in comparing the QRLs of patients who received four cycles of doxorubicin and cyclophosphamide followed by four cycles of docetaxel and the patients who received four cycles of doxorubicin, docetaxel and cyclophosphamide at the same time.

3.5 DISCUSSION

In this Chapter, we proposed an AIPW type estimator to estimate the QRLs by adding the information of the auxiliary variables. The AIPW method aims to use the conditional model with the auxiliary covariates to predict the survival outcome for patients with observed time to censoring. The AIPW estimating equation was constructed as a combination of the IPW estimating equation from the observed survival outcomes and the conditional prediction information from the observed censoring outcomes. It was expected that adding the conditional prediction from the covariates can increase the efficiency in estimating the QRLs when the auxiliary covariates are truly related to the survival outcome.

In the simulation studies, we evaluated the efficiency of the AIPW QRL estimator and its 95% CI of the confidence interval based on the proposed variance estimator under different censoring rates, sample sizes and covariate effects. Compared to the KM-type and IPW estimators,
Table 12: Distribution of the auxiliary variables by treatment groups

<table>
<thead>
<tr>
<th>Auxiliary variables</th>
<th>Categories</th>
<th>AC→T (n=1683)</th>
<th>ATC (n=1709)</th>
<th>Total (n=3392)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estrogen receptor</td>
<td>Negative or Borderline</td>
<td>310 (18.4)</td>
<td>315 (18.4)</td>
<td>625 (18.4)</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>900 (53.5)</td>
<td>912 (53.4)</td>
<td>1812 (53.4)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>107 (6.4)</td>
<td>108 (6.3)</td>
<td>215 (6.3)</td>
</tr>
<tr>
<td></td>
<td>Positive or Borderline</td>
<td>366 (21.7)</td>
<td>374 (21.9)</td>
<td>740 (21.8)</td>
</tr>
<tr>
<td>BMI</td>
<td>&lt;18.5</td>
<td>13 (0.8)</td>
<td>25 (1.5)</td>
<td>38 (1.1)</td>
</tr>
<tr>
<td></td>
<td>18.5-24.9</td>
<td>551 (32.7)</td>
<td>593 (34.7)</td>
<td>1144 (33.7)</td>
</tr>
<tr>
<td></td>
<td>25-29.9</td>
<td>544 (32.3)</td>
<td>521 (30.5)</td>
<td>1065 (31.4)</td>
</tr>
<tr>
<td></td>
<td>30+</td>
<td>575 (34.2)</td>
<td>570 (33.4)</td>
<td>1145 (33.8)</td>
</tr>
<tr>
<td>Age (year)</td>
<td>&lt;40</td>
<td>206 (12.2)</td>
<td>187 (10.9)</td>
<td>393 (11.6)</td>
</tr>
<tr>
<td></td>
<td>40-49</td>
<td>563 (33.5)</td>
<td>571 (33.4)</td>
<td>1134 (33.4)</td>
</tr>
<tr>
<td></td>
<td>50-59</td>
<td>566 (33.6)</td>
<td>585 (34.2)</td>
<td>1151 (33.9)</td>
</tr>
<tr>
<td></td>
<td>60+</td>
<td>348 (20.7)</td>
<td>366 (21.4)</td>
<td>714 (21)</td>
</tr>
<tr>
<td>Menopausal</td>
<td>Yes</td>
<td>765 (45.5)</td>
<td>786 (46)</td>
<td>1551 (45.7)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>918 (54.5)</td>
<td>923 (54)</td>
<td>1841 (54.3)</td>
</tr>
<tr>
<td>Type of mastectomy</td>
<td>Lumpectomy</td>
<td>830 (49.3)</td>
<td>840 (49.2)</td>
<td>1670 (49.2)</td>
</tr>
<tr>
<td></td>
<td>Mastectomy</td>
<td>853 (50.7)</td>
<td>869 (50.8)</td>
<td>1722 (50.8)</td>
</tr>
<tr>
<td>Maximum pathologic</td>
<td>&lt;2cm</td>
<td>573 (34)</td>
<td>591 (34.6)</td>
<td>1164 (34.3)</td>
</tr>
<tr>
<td></td>
<td>2-2.59cm</td>
<td>465 (27.6)</td>
<td>461 (27)</td>
<td>926 (27.3)</td>
</tr>
<tr>
<td></td>
<td>2.6-3.9cm</td>
<td>323 (19.2)</td>
<td>331 (19.4)</td>
<td>654 (19.3)</td>
</tr>
<tr>
<td></td>
<td>4cm+</td>
<td>322 (19.1)</td>
<td>326 (19.1)</td>
<td>648 (19.1)</td>
</tr>
</tbody>
</table>
Table 13: Estimated \( p \)-QRLs in years using the AIPW method at different \( t_0 \)s and results of the test of equality of two \( p \)-QRLs for AC→T and ATC group by methods using the KM-type and AIPW estimators

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Estimated ( p )-QRL (95%CI)</th>
<th>Ratio</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_0 = 1 ) year, ( p = 0.1 )</td>
<td>( 3.67 ) (3.46-3.87) ( 3.93 ) (3.66-4.20)</td>
<td>1.07</td>
<td>0.321 0.192 0.321 0.176</td>
</tr>
<tr>
<td>( t_0 = 1 ) year, ( p = 0.2 )</td>
<td>( 8.82 ) (8.43-9.22) ( 7.47 ) (6.94-8.00)</td>
<td>0.85</td>
<td>0.009 0.001 0.008 0.015</td>
</tr>
<tr>
<td>( t_0 = 5 ) year, ( p = 0.1 )</td>
<td>( 4.58 ) (4.34-4.83) ( 3.32 ) (3.02-3.61)</td>
<td>0.72</td>
<td>0.001 &lt;0.001 &lt;0.001 0.001</td>
</tr>
</tbody>
</table>

\( p \)-QRLs = \( p \)-quantile residual lifetimes; The Ratio of the \( p \)-QRL is calculated as the AIPW estimator in the ATC group divided by the AIPW estimator in the AC→T group; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated QRL from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both PI and PI* tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22.

the efficiency gain of the AIPW estimators ranged from 2%-26%. From the simulation studies, we also observed that the efficiency gain of the AIPW estimator over the KM-type and IPW estimators increases as the censoring rate, sample size and the absolute effect size increase. This observation is reasonable based on the assumption of the AIPW methods that more information will be added into the estimating equation when there are more censoring observations (larger censoring rate or bigger sample size). If the absolute effect sizes of the auxiliary covariates are bigger, which means the covariates are highly significant predictors of the survival outcome, we would also expect the conditional modeling using the auxiliary covariates will add more information to estimate the QRL, this results in more efficiency gain compared to the estimators without using the covariate information. When there is no censoring or the auxiliary covariate is not related to the survival outcome (\( \beta_x = 0 \)), there is no efficiency gain in using the AIPW estimator compared to the KM-type and the IPW estimators.

We also plugged the AIPW estimator into the two-sample PI tests proposed in the first chapter and compared the Type I error and the power of the PI tests using the AIPW estimators with those using the KM-type estimators in the simulation studies. Based on the simulation results, using the AIPW estimator in the PI tests has correct Type I error that is close to the nominal level. However,
the efficiency gain observed for estimating the QRL did not translate into similar increase in the power for the two-sample PI tests over the tests based on the KM-type estimator.

We note that the 95% CI of the KM-type QRL estimator has much lower coverage rates ranging from 75% to 90% when the censoring rate is above 30%. However, the 95% CI coverage rate of the AIPW estimators are generally close to 95% across all the Scenarios, ranging from 93% to 97% in most Scenarios. Thus, when the censoring rate is high (above 30%), we recommend to use the AIPW estimator to make the inference for the QRL. However, the proposed two-sample PI tests still works no matter which estimator is being used.
4.0 FUTURE TOPIC

In the future, we are going to extend our proposed plug-in tests to compare the QRLs of more than two groups. Furthermore, we also plan to develop methods modeling the QRLs as a function of the time and apply the AIPW approach to its estimating process.
APPENDIX A

TYPE I ERROR STUDY WITH DIFFERENT SURVIVAL DISTRIBUTIONS-TABLES

For the simulation study in Chapter 2 (Section 2.4.1), we also evaluated the Type I errors for the 0.5, 0.25 and 0.75-QRLs with different survival distributions and compared the results among the proposed methods (DIFF and PI tests) and the MCS approach. We considered three additional scenarios, which we generated the failure time of both groups follow (i) two independent log-normal distributions; (ii) two independent exponential distributions and (iii) independent Weibull and log-normal distributions.

For the Log-normal distribution, failure times are generated from

$$T_{ki} = \exp(\mu_k + \sigma_k Z_{ki}),$$

where $Z_{ki}$ follows a standard normal distribution, $\mu_k$ and $\sigma_k$ are the shape and scale parameters of the log-scale failure time in Group $k$, respectively. The log-scale failure time follows a normal distribution with mean $\mu_k$ and variance $\sigma_k^2$.

For the exponential distribution, failure times are generated from

$$T_{ki} = \frac{1}{\lambda}\{-\log(1 - U_{ki})\},$$

where $U_{ki} \sim UNIF(0, 1)$.

Table 14 shows the parameters of the survival distributions used to evaluate the type I error. Each type I error is calculated based on 10,000 simulations with three different sample sizes (50, 100 and 200 per group).
Table 14: Parameters set-up for different survival distribution used in the simulation.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>$T_1, T_2 \sim \text{Log-normal}(\mu = 1.39, \sigma = 0.47)$</td>
</tr>
<tr>
<td></td>
<td>$t_0 = 0, 1, 2, 3, 4$</td>
</tr>
<tr>
<td></td>
<td>$\tau_{0.5,t_0} = 1$</td>
</tr>
<tr>
<td>ii</td>
<td>$T_1 \sim \text{Exponential}(\lambda = 2)$</td>
</tr>
<tr>
<td></td>
<td>$T_2 \sim \text{Exponential}(\lambda = 1.64)$</td>
</tr>
<tr>
<td></td>
<td>$t_0 = 0, 0.3, 0.6, 1$</td>
</tr>
<tr>
<td></td>
<td>$\tau_{0.5,t_0} = 0.8187$</td>
</tr>
<tr>
<td>iii</td>
<td>$T_1 \sim \text{Weibull}(\rho = 0.27, \eta = 2)$</td>
</tr>
<tr>
<td></td>
<td>$T_2 \sim \text{Log-normal}(\mu = 1.39, \sigma = 0.47)$</td>
</tr>
<tr>
<td></td>
<td>$t_0 = 0, 1, 2, 3, 4$</td>
</tr>
<tr>
<td></td>
<td>$\tau_{0.5,t_0} = 1.30, 1.34, 1.30, 1.32, 1.43$</td>
</tr>
</tbody>
</table>

Tables 15, 16 & 17 show the estimated Type I error for the QRL tests for sample sizes 50 and 200 with 20% and 40% observations censored for Scenario i. Tables 18, 19 & 20 show the corresponding results for the scenario ii and Tables 21, 22 & 23 show the corresponding results for Scenario iii.
Table 15: Estimated Type I error for various tests of median residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario i

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th>DIFF</th>
<th>PI</th>
<th>PI*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MCS</td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 50$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.025</td>
<td>0.078</td>
<td>0.062</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.026</td>
<td>0.078</td>
<td>0.061</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.024</td>
<td>0.085</td>
<td>0.063</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.025</td>
<td>0.107</td>
<td>0.065</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.032</td>
<td>0.129</td>
<td>0.062</td>
<td>0.057</td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.026</td>
<td>0.075</td>
<td>0.055</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.029</td>
<td>0.081</td>
<td>0.058</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.028</td>
<td>0.082</td>
<td>0.055</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.029</td>
<td>0.107</td>
<td>0.062</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.051</td>
<td>0.135</td>
<td>0.066</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.028</td>
<td>0.068</td>
<td>0.060</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.030</td>
<td>0.065</td>
<td>0.057</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.032</td>
<td>0.066</td>
<td>0.054</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.029</td>
<td>0.074</td>
<td>0.058</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.025</td>
<td>0.091</td>
<td>0.065</td>
<td>0.064</td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.031</td>
<td>0.063</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.029</td>
<td>0.064</td>
<td>0.053</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.026</td>
<td>0.067</td>
<td>0.055</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.027</td>
<td>0.074</td>
<td>0.056</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.025</td>
<td>0.082</td>
<td>0.056</td>
<td>0.055</td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 16: Estimated Type I error for various tests of 0.25-- residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario i

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th>[n = 50]</th>
<th>[n = 200]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>[MCS]</td>
<td>[DIFF]</td>
<td>[PI]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
</tr>
<tr>
<td>0</td>
<td>0.022</td>
<td>0.038</td>
<td>0.037</td>
<td>0.037</td>
</tr>
<tr>
<td>0</td>
<td>0.024</td>
<td>0.042</td>
<td>0.041</td>
<td>0.040</td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.026</td>
<td>0.054</td>
<td>0.044</td>
</tr>
<tr>
<td>2</td>
<td>0.023</td>
<td>0.084</td>
<td>0.059</td>
<td>0.058</td>
</tr>
<tr>
<td>3</td>
<td>0.026</td>
<td>0.116</td>
<td>0.069</td>
<td>0.067</td>
</tr>
<tr>
<td>4</td>
<td>0.017</td>
<td>0.048</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>0</td>
<td>0.020</td>
<td>0.047</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.024</td>
<td>0.061</td>
<td>0.049</td>
</tr>
<tr>
<td>2</td>
<td>0.022</td>
<td>0.094</td>
<td>0.063</td>
<td>0.061</td>
</tr>
<tr>
<td>3</td>
<td>0.023</td>
<td>0.127</td>
<td>0.071</td>
<td>0.068</td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 17: Estimated Type I error for various tests of \(0.75\)—residual life at different \(t_0\) for various sample sizes \((n)\) and censoring rates \((P_c)\) for Scenario \(i\)

<table>
<thead>
<tr>
<th>(P_c)</th>
<th>(t_0)</th>
<th>(n = 50)</th>
<th>(n = 200)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MCS DIFF PI PI*</td>
<td>MCS DIFF PI PI*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald WI WII Wald WI WII Wald WI WII Wald WI WII</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.031 0.122 0.055 0.052 0.106 0.044 0.043 0.025 0.045 0.050</td>
<td>0.027 0.090 0.056 0.054 0.077 0.045 0.045 0.019 0.042 0.043</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.032 0.131 0.061 0.056 0.111 0.048 0.047 0.022 0.047 0.051</td>
<td>0.026 0.085 0.055 0.053 0.071 0.041 0.041 0.020 0.041 0.041</td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
<td>0.031 0.132 0.060 0.057 0.118 0.051 0.048 0.028 0.048 0.052</td>
<td>0.026 0.095 0.061 0.060 0.079 0.048 0.048 0.018 0.044 0.045</td>
</tr>
<tr>
<td>20%</td>
<td>3</td>
<td>0.049 0.153 0.056 0.054 0.143 0.053 0.049 0.048 0.054 0.060</td>
<td>0.026 0.102 0.056 0.055 0.083 0.045 0.044 0.019 0.047 0.047</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.107 0.153 0.058 0.055 0.189 0.067 0.060 0.099 0.062 0.075</td>
<td>0.030 0.113 0.053 0.051 0.090 0.045 0.044 0.029 0.042 0.043</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.039 0.136 0.063 0.059 0.120 0.054 0.051 0.052 0.052 0.060</td>
<td>0.029 0.090 0.061 0.059 0.077 0.048 0.048 0.031 0.047 0.049</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.044 0.140 0.067 0.064 0.125 0.060 0.056 0.059 0.054 0.062</td>
<td>0.030 0.083 0.055 0.053 0.072 0.042 0.042 0.025 0.040 0.042</td>
</tr>
<tr>
<td>40%</td>
<td>2</td>
<td>0.056 0.142 0.065 0.061 0.130 0.058 0.055 0.064 0.055 0.063</td>
<td>0.031 0.097 0.058 0.055 0.080 0.046 0.045 0.031 0.043 0.045</td>
</tr>
<tr>
<td>40%</td>
<td>3</td>
<td>0.088 0.147 0.066 0.061 0.157 0.067 0.061 0.098 0.063 0.073</td>
<td>0.034 0.112 0.048 0.047 0.093 0.043 0.043 0.024 0.048 0.040 0.042</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.204 0.150 0.063 0.058 0.224 0.088 0.076 0.180 0.074 0.090</td>
<td>0.034 0.112 0.048 0.047 0.093 0.043 0.043 0.024 0.048 0.040 0.042</td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time \(t_0\), both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 18: Estimated Type I error for various tests of median residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario ii

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th>MCS</th>
<th>DIFF</th>
<th>PI</th>
<th>PI*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
</tr>
</tbody>
</table>

---

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

56
Table 19: Estimated Type I error for various tests of 0.25—residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario ii

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th>MCS</th>
<th>DIFF</th>
<th>PI</th>
<th>PI*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
</tr>
<tr>
<td>n=50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.024</td>
<td>0.054</td>
<td>0.047</td>
<td>0.047</td>
<td>0.033</td>
<td>0.040</td>
</tr>
<tr>
<td>1</td>
<td>0.024</td>
<td>0.101</td>
<td>0.081</td>
<td>0.081</td>
<td>0.037</td>
<td>0.042</td>
</tr>
<tr>
<td>2</td>
<td>0.023</td>
<td>0.129</td>
<td>0.101</td>
<td>0.098</td>
<td>0.041</td>
<td>0.042</td>
</tr>
<tr>
<td>3</td>
<td>0.025</td>
<td>0.138</td>
<td>0.103</td>
<td>0.101</td>
<td>0.044</td>
<td>0.039</td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.021</td>
<td>0.090</td>
<td>0.075</td>
<td>0.074</td>
<td>0.034</td>
<td>0.045</td>
</tr>
<tr>
<td>1</td>
<td>0.020</td>
<td>0.134</td>
<td>0.099</td>
<td>0.094</td>
<td>0.037</td>
<td>0.041</td>
</tr>
<tr>
<td>2</td>
<td>0.022</td>
<td>0.152</td>
<td>0.104</td>
<td>0.099</td>
<td>0.039</td>
<td>0.040</td>
</tr>
<tr>
<td>3</td>
<td>0.020</td>
<td>0.173</td>
<td>0.100</td>
<td>0.093</td>
<td>0.045</td>
<td>0.042</td>
</tr>
</tbody>
</table>

| n=200 |       |        |      |      |     |      |    |     |      |    |     |
| 20%   |       |        |      |      |     |      |    |     |      |    |     |
| 0     | 0.032 | 0.063  | 0.059| 0.059| 0.040| 0.041| 0.041| 0.059| 0.060| 0.060|
| 1     | 0.027 | 0.087  | 0.083| 0.083| 0.040| 0.042| 0.041| 0.049| 0.052| 0.053|
| 2     | 0.029 | 0.093  | 0.089| 0.089| 0.041| 0.042| 0.042| 0.055| 0.057| 0.057|
| 3     | 0.027 | 0.097  | 0.089| 0.089| 0.041| 0.042| 0.040| 0.052| 0.057| 0.057|
| 40%   |       |        |      |      |     |      |    |     |      |    |     |
| 0     | 0.030 | 0.081  | 0.075| 0.074| 0.040| 0.042| 0.041| 0.053| 0.056| 0.056|
| 1     | 0.029 | 0.097  | 0.085| 0.084| 0.041| 0.041| 0.040| 0.047| 0.053| 0.053|
| 2     | 0.026 | 0.106  | 0.090| 0.089| 0.043| 0.044| 0.043| 0.051| 0.058| 0.058|
| 3     | 0.025 | 0.098  | 0.080| 0.079| 0.037| 0.040| 0.038| 0.041| 0.050| 0.051|

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 20: Estimated Type I error for various tests of 0.75 – residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario ii

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MCS</td>
<td>DIFF</td>
<td>PI</td>
<td>PI*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
<td>WI</td>
</tr>
<tr>
<td>n = 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.030</td>
<td>0.144</td>
<td>0.069</td>
<td>0.066</td>
<td>0.106</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.043</td>
<td>0.139</td>
<td>0.061</td>
<td>0.058</td>
<td>0.118</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.068</td>
<td>0.121</td>
<td>0.046</td>
<td>0.043</td>
<td>0.130</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.160</td>
<td>0.090</td>
<td>0.030</td>
<td>0.030</td>
<td>0.148</td>
<td>0.038</td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.050</td>
<td>0.141</td>
<td>0.043</td>
<td>0.041</td>
<td>0.112</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.136</td>
<td>0.120</td>
<td>0.039</td>
<td>0.038</td>
<td>0.132</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.213</td>
<td>0.104</td>
<td>0.037</td>
<td>0.037</td>
<td>0.166</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.345</td>
<td>0.098</td>
<td>0.036</td>
<td>0.035</td>
<td>0.230</td>
<td>0.083</td>
</tr>
<tr>
<td>n = 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.026</td>
<td>0.093</td>
<td>0.076</td>
<td>0.075</td>
<td>0.052</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.025</td>
<td>0.097</td>
<td>0.072</td>
<td>0.070</td>
<td>0.060</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.027</td>
<td>0.082</td>
<td>0.055</td>
<td>0.055</td>
<td>0.059</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.064</td>
<td>0.055</td>
<td>0.026</td>
<td>0.027</td>
<td>0.058</td>
<td>0.029</td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.024</td>
<td>0.105</td>
<td>0.056</td>
<td>0.053</td>
<td>0.051</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.030</td>
<td>0.097</td>
<td>0.050</td>
<td>0.048</td>
<td>0.059</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.047</td>
<td>0.070</td>
<td>0.032</td>
<td>0.033</td>
<td>0.055</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.146</td>
<td>0.043</td>
<td>0.024</td>
<td>0.024</td>
<td>0.063</td>
<td>0.035</td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 21: Estimated Type I error for various tests of median residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario iii

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MCS</td>
<td>DIFF</td>
<td>PI</td>
<td>PI*</td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
<td>WI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.030</td>
<td>0.051</td>
<td>0.049</td>
<td>0.048</td>
<td>0.053</td>
<td>0.057</td>
<td>0.058</td>
<td>0.057</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.027</td>
<td>0.052</td>
<td>0.045</td>
<td>0.042</td>
<td>0.053</td>
<td>0.052</td>
<td>0.054</td>
<td>0.055</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.029</td>
<td>0.048</td>
<td>0.046</td>
<td>0.041</td>
<td>0.055</td>
<td>0.049</td>
<td>0.053</td>
<td>0.067</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.031</td>
<td>0.061</td>
<td>0.051</td>
<td>0.040</td>
<td>0.056</td>
<td>0.055</td>
<td>0.060</td>
<td>0.082</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.046</td>
<td>0.078</td>
<td>0.063</td>
<td>0.045</td>
<td>0.081</td>
<td>0.058</td>
<td>0.074</td>
<td>0.100</td>
<td>0.066</td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.028</td>
<td>0.047</td>
<td>0.045</td>
<td>0.042</td>
<td>0.051</td>
<td>0.047</td>
<td>0.049</td>
<td>0.061</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.032</td>
<td>0.053</td>
<td>0.049</td>
<td>0.044</td>
<td>0.048</td>
<td>0.050</td>
<td>0.054</td>
<td>0.065</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.032</td>
<td>0.054</td>
<td>0.051</td>
<td>0.041</td>
<td>0.050</td>
<td>0.048</td>
<td>0.055</td>
<td>0.070</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.040</td>
<td>0.065</td>
<td>0.056</td>
<td>0.042</td>
<td>0.061</td>
<td>0.056</td>
<td>0.067</td>
<td>0.090</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.074</td>
<td>0.103</td>
<td>0.077</td>
<td>0.046</td>
<td>0.104</td>
<td>0.064</td>
<td>0.087</td>
<td>0.121</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>n = 200</td>
<td>0.030</td>
<td>0.045</td>
<td>0.042</td>
<td>0.041</td>
<td>0.053</td>
<td>0.054</td>
<td>0.055</td>
<td>0.050</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.029</td>
<td>0.041</td>
<td>0.046</td>
<td>0.045</td>
<td>0.054</td>
<td>0.049</td>
<td>0.049</td>
<td>0.048</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.029</td>
<td>0.040</td>
<td>0.040</td>
<td>0.039</td>
<td>0.051</td>
<td>0.053</td>
<td>0.054</td>
<td>0.047</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.033</td>
<td>0.037</td>
<td>0.041</td>
<td>0.038</td>
<td>0.053</td>
<td>0.050</td>
<td>0.052</td>
<td>0.052</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.027</td>
<td>0.040</td>
<td>0.043</td>
<td>0.039</td>
<td>0.057</td>
<td>0.050</td>
<td>0.055</td>
<td>0.058</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.030</td>
<td>0.044</td>
<td>0.043</td>
<td>0.041</td>
<td>0.052</td>
<td>0.051</td>
<td>0.052</td>
<td>0.050</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.030</td>
<td>0.044</td>
<td>0.045</td>
<td>0.043</td>
<td>0.052</td>
<td>0.048</td>
<td>0.049</td>
<td>0.052</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.031</td>
<td>0.044</td>
<td>0.043</td>
<td>0.040</td>
<td>0.053</td>
<td>0.047</td>
<td>0.049</td>
<td>0.051</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.028</td>
<td>0.043</td>
<td>0.045</td>
<td>0.040</td>
<td>0.055</td>
<td>0.052</td>
<td>0.056</td>
<td>0.053</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.032</td>
<td>0.047</td>
<td>0.047</td>
<td>0.037</td>
<td>0.059</td>
<td>0.050</td>
<td>0.057</td>
<td>0.064</td>
<td>0.055</td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 22: Estimated Type I error for various tests of 0.25 -- residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario iii

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MCS</td>
<td>DIFF</td>
<td>PI</td>
<td>PI*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.031</td>
<td>0.047</td>
<td>0.040</td>
<td>0.041</td>
<td>0.043</td>
<td>0.048</td>
<td>0.045</td>
<td>0.053</td>
<td>0.051</td>
<td>0.052</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>1</td>
<td>0.028</td>
<td>0.043</td>
<td>0.040</td>
<td>0.038</td>
<td>0.041</td>
<td>0.048</td>
<td>0.044</td>
<td>0.056</td>
<td>0.053</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
<td>0.028</td>
<td>0.051</td>
<td>0.045</td>
<td>0.045</td>
<td>0.043</td>
<td>0.051</td>
<td>0.046</td>
<td>0.055</td>
<td>0.056</td>
<td>0.060</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>3</td>
<td>0.028</td>
<td>0.067</td>
<td>0.057</td>
<td>0.059</td>
<td>0.040</td>
<td>0.049</td>
<td>0.045</td>
<td>0.054</td>
<td>0.057</td>
<td>0.062</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>4</td>
<td>0.029</td>
<td>0.092</td>
<td>0.069</td>
<td>0.072</td>
<td>0.040</td>
<td>0.049</td>
<td>0.045</td>
<td>0.055</td>
<td>0.064</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>0</td>
<td>0.025</td>
<td>0.048</td>
<td>0.044</td>
<td>0.044</td>
<td>0.038</td>
<td>0.045</td>
<td>0.041</td>
<td>0.056</td>
<td>0.054</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>1</td>
<td>0.025</td>
<td>0.053</td>
<td>0.046</td>
<td>0.046</td>
<td>0.054</td>
<td>0.047</td>
<td>0.055</td>
<td>0.053</td>
<td>0.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>2</td>
<td>0.026</td>
<td>0.059</td>
<td>0.049</td>
<td>0.048</td>
<td>0.041</td>
<td>0.053</td>
<td>0.045</td>
<td>0.048</td>
<td>0.050</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>3</td>
<td>0.026</td>
<td>0.082</td>
<td>0.060</td>
<td>0.058</td>
<td>0.038</td>
<td>0.051</td>
<td>0.044</td>
<td>0.052</td>
<td>0.058</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td>4</td>
<td>0.039</td>
<td>0.125</td>
<td>0.079</td>
<td>0.077</td>
<td>0.058</td>
<td>0.063</td>
<td>0.050</td>
<td>0.076</td>
<td>0.074</td>
<td>0.088</td>
<td></td>
</tr>
</tbody>
</table>

| 40%   | 0     | 0.032  | 0.044     | 0.041     | 0.041     | 0.047     | 0.047     | 0.047     | 0.047     | 0.049     | 0.049     |           |
| 40%   | 1     | 0.029  | 0.039     | 0.039     | 0.039     | 0.045     | 0.048     | 0.046     | 0.046     | 0.047     | 0.047     |           |
| 40%   | 2     | 0.028  | 0.051     | 0.051     | 0.050     | 0.052     | 0.053     | 0.051     | 0.050     | 0.052     | 0.054     |           |
| 40%   | 3     | 0.026  | 0.051     | 0.047     | 0.047     | 0.042     | 0.046     | 0.044     | 0.050     | 0.050     | 0.051     |           |
| 40%   | 4     | 0.025  | 0.057     | 0.053     | 0.053     | 0.042     | 0.047     | 0.045     | 0.052     | 0.053     | 0.056     |           |

| 20%   | 0     | 0.030  | 0.046     | 0.044     | 0.044     | 0.047     | 0.049     | 0.047     | 0.051     | 0.052     | 0.052     |           |
| 20%   | 1     | 0.028  | 0.048     | 0.044     | 0.044     | 0.044     | 0.047     | 0.045     | 0.051     | 0.052     | 0.053     |           |
| 20%   | 2     | 0.026  | 0.047     | 0.044     | 0.044     | 0.043     | 0.046     | 0.043     | 0.048     | 0.048     | 0.049     |           |
| 20%   | 3     | 0.024  | 0.049     | 0.048     | 0.048     | 0.041     | 0.045     | 0.042     | 0.046     | 0.049     | 0.051     |           |
| 20%   | 4     | 0.023  | 0.063     | 0.058     | 0.059     | 0.041     | 0.046     | 0.043     | 0.051     | 0.055     | 0.057     |           |

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
Table 23: Estimated Type I error for various tests of $0.75 -$ residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario iii

<table>
<thead>
<tr>
<th>$P_c$</th>
<th>$t_0$</th>
<th>Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MCS</td>
<td>DIFF</td>
<td>PI</td>
<td>PI*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Wald</td>
<td>WI</td>
<td>WII</td>
</tr>
<tr>
<td>n = 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.035</td>
<td>0.088 0.058 0.056 0.073 0.049 0.049 0.060 0.055 0.059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.035</td>
<td>0.088 0.060 0.058 0.073 0.047 0.044 0.057 0.052 0.057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
<td>0.040 0.096 0.062 0.060 0.081 0.051 0.046 0.062 0.052 0.061</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.051</td>
<td>0.116 0.067 0.066 0.101 0.064 0.054 0.086 0.056 0.075</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.141</td>
<td>0.136 0.074 0.069 0.156 0.092 0.065 0.162 0.070 0.102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.045</td>
<td>0.097 0.061 0.057 0.084 0.056 0.052 0.079 0.051 0.060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.046</td>
<td>0.101 0.059 0.055 0.084 0.054 0.048 0.075 0.050 0.060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.055</td>
<td>0.116 0.064 0.060 0.105 0.066 0.055 0.095 0.057 0.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.105</td>
<td>0.137 0.074 0.068 0.141 0.085 0.061 0.132 0.069 0.093</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.255</td>
<td>0.156 0.089 0.074 0.222 0.131 0.082 0.241 0.084 0.133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.030</td>
<td>0.054 0.054 0.053 0.042 0.042 0.042 0.051 0.048 0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.026</td>
<td>0.060 0.057 0.057 0.047 0.047 0.046 0.056 0.053 0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
<td>0.030 0.055 0.054 0.053 0.043 0.044 0.041 0.057 0.051 0.054</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.025</td>
<td>0.065 0.058 0.058 0.049 0.048 0.043 0.060 0.050 0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.033</td>
<td>0.082 0.059 0.060 0.061 0.050 0.040 0.071 0.048 0.060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.027</td>
<td>0.059 0.055 0.056 0.046 0.043 0.043 0.057 0.052 0.054</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.027</td>
<td>0.055 0.051 0.051 0.044 0.042 0.040 0.058 0.048 0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.026</td>
<td>0.059 0.052 0.051 0.046 0.044 0.039 0.053 0.047 0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.032</td>
<td>0.068 0.054 0.054 0.056 0.049 0.040 0.063 0.045 0.054</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.043</td>
<td>0.094 0.055 0.056 0.075 0.054 0.041 0.084 0.047 0.069</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Wald denotes the variance estimators using the individual QRL estimates defined in Equation 2.3, WI denotes the variance estimators using the QRL estimate weighted by the number of events, WII denotes the variance estimators using the QRL estimate weighted by the KM function at time $t_0$, both of the estimators are defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.
APPENDIX B

TYPE I ERROR STUDY WITH DIFFERENT SURVIVAL DISTRIBUTIONS-FIGURES

Figures 9, 10 & 11 show the estimated Type I errors over $t_0$ for the QRL tests for Scenario i. Figures 12, 13 & 14 show the same results for the scenario ii. Figures 15, 16 & 17 show the same results for Scenario iii.
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 9: Estimated Type I error for various tests of median residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario i
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 10: Estimated Type I error for various tests of 0.25— residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario i
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 11: Estimated Type I error for various tests of 0.75—residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario i
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 12: Estimated Type I error for various tests of median residual life at different \( t_0 \) for various sample sizes (n) and censoring rates (\( P_c \)) for Scenario ii
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 13: Estimated Type I error for various tests of $0.25$—residual life at different $t_0$ for various sample sizes ($n$) and censoring rates ($P_c$) for Scenario ii
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 14: Estimated Type I error for various tests of $0.75 - $ residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario ii
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 15: Estimated Type I error for various tests of median residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario iii.
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 16: Estimated Type I error for various tests of 0.25– residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario iii
MCS = minimum chi-square statistic test; DIFF = difference statistic test; PI = plug-in tests; PI test is defined as in Equations 2.6 and 2.7, except that the estimated quantile residual lifetime (QRL) from the sample with larger number of event is plugged into the estimating function of the other group, and vice-versa for PI*; Both DIFF and PI tests used the variance estimators with the QRL estimates weighted by the number of events, defined in Equations 2.21 and 2.22; All results are based on 10,000 simulations.

Figure 17: Estimated Type I error for various tests of 0.75—residual life at different $t_0$ for various sample sizes (n) and censoring rates ($P_c$) for Scenario iii


Tsung-Hsien Tsai, Wei-Yann Tsai, Yunchan Chi, and Sheng-Mao Chang. Confidence intervals for the ratio of two median residual lifetimes with left-truncated and right-censored data. *Biometrics*, 2015.


