A Slope Test Strategy for Censored Data in Survival Analysis

Yumin Zhang

Jinan University

Zheng Chen

Southern Medical University

Yawen Hou (✉ thouyw@jnu.edu.cn)

Jinan University

Research Article

Keywords: Survival curve test, Slope strategy, Censoring rates

Posted Date: February 16th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1287998/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A slope test strategy for censored data in survival analysis

Yumin Zhang¹, Zheng Chen², Yawen Hou^{1,*}

¹Department of Statistics, College of Economics, Jinan University, No.601, West Huangpu Avenue , Tianhe District, Guangzhou 510632, China.

²Department of Biostatistics, School of Public Health (Guangdong Provincial Key Laboratory of Tropical Disease Research), Southern Medical University. China.

**e-mail:* thouyw@jnu.edu.cn

Abstract

Background: Censored data are nearly ubiquitous in survival analysis.

If the proportional hazard assumption is invalid, traditional test methods, such as family of log-rank test, lose power rapidly. This paper gives a new slope strategy to improve power of test. The aim of this study is to establish effective tests for censoring data which are familiar in biomedical research and medical practice.

Methods: Four tests based on slope are proposed in this paper, and monte carlo simulations with various survival curves, sample size and censoring rates scenarios were performed to evaluate test power. Four common tests, log-rank test, Gehan-Wilcoxon test and Fleming-Harrington with different parameters, were compared with four new slope tests.

Results: The slope tests are more effective when the censoring rate is equal to or greater than 30%, especially for scenarios involving early hazard function differences. Two clinical cases also indicate that the proposed slope methods can detect differences for real data.

Conclusions: Slope tests may provide a supplement of computing method to survival curve test, it will be more useful considering higher censoring rates.

KEY WORDS: Survival curve test; Slope strategy; Censoring rates

1 Introduction

The survival rate is an important component of survival analysis in many fields, such as medicine and industrial reliability analysis. A concern shared by doctors, patients, manufacturers and consumers alike regards the presence of differences between survival curves, which can play a decisive role in supporting the launch of new clinical treatment strategies or new products. The widely used log-rank test^[1] has been shown to have optimal test power under the premise that the proportional hazard is valid^[2]. However, in practical clinical or industrial research, this precondition is often violated^[3-5], and hazard functions easily crossover. In this case, the difference based on the hazard rate will be offset before and after the intersection point^[6, 7], and the power of the log-rank test will suffer greatly^[8]. In addition, the log-rank test and linear rank test of this class will perform poorly in cases with small samples and potential imbalance in the censoring distribution^[9].

The construction of test statistics can be generally divided into two groups, one based on hazard functions and another on survival functions. The former includes the common log-rank^[10] and Wilcoxon^[11] tests, and the latter is a class of weighted Kaplan-Meier (WKM) statistics^[12]. In recent years, many nonparametric test methods based on Kaplan-Meier (KM) estimators have been proposed, which mainly compare the differences in the areas under the estimated survival curves^[8]. The restricted mean survival time (RMST) is estimated as the area under the survival curve from time 0 to a specified time point^[13-17]. However, when two survival curves cross, the difference in the RMST may be offset before and after the crossing point^[18]. Constructing test

statistics to avoid the cross-influence of survival curves could be a way to solve the above problems.

This paper proposes four test statistics based on changes in the measured survival curve. Intuitively, the slope, as an indicator measuring the change in the survival curve, is not affected by the presence of survival curve crossover. The paper is structured as follows. Based on slope differences, four tests are constructed in Section 2. In Section 3, we conduct a series of Monte Carlo simulations to examine the performance of these tests when the survival curves are proportional, the survival curves are crossing, and the hazard functions are different. Finally, two empirical examples from clinical medicine are combined for illustrative purposes. Finally, we summarize our findings and present a general discussion.

2 Methodology

To test whether two survival functions $S_1(t)$ and $S_2(t)$ are equal, we can make the following hypothesis test:

$$H_0 : S_1(t) = S_2(t) \text{ for all } t, \quad H_1 : S_1(t) \neq S_2(t)$$

Let S_{gi} be the survival function of group g after the i events have occurred, $g = 1, 2, i = 1, \dots, N_g$, N_g denotes the total number of events in group g .

$\hat{S}_{gi}$ is the KM estimator of the survival function of group g after only i events. $\hat{S}_{gi}$ can be expressed as follows:

$$\hat{S}_{gi} = \hat{S}_g(t_i) = \prod_{i: X_{gk} \leq t_i} \left(1 - \frac{\Delta N_g(X_{gk})}{Y_g(X_{gk})}\right)$$

where X_{gk} denotes the time when event k happened in group g ; $\Delta N_g(X_{gk})$ is the number of events that happened at time point X_{gk} ; and $Y_g(X_{gk})$ represents the

number of survival data points in the risk of group g at the time point.

For the survival data of group g , the slope constructed at the initial point $(0, S_0)$ and the event occurrence point (t_i, S_i) can be obtained by drawing the survival curve.

$$l_{gi} = (S_{gi} - 1)/t_{gi}$$

The estimator of l_{gi} is

$$\hat{l}_{gi} = (\hat{S}_{gi} - 1) / t_{gi}$$

where t_{gi} is the time at which event i occurred in group g .

2.1 Slope Difference test

2.1.1 Method 1

The goal of this method is to compare the slope between the two survival curves at each time point of the event. To reduce the influence of extreme slope values, the test statistic is defined as the median of the absolute slope difference, denoted by:

$$\Delta_1 = \underset{i \in (0, n]}{\text{median}} |l_{1i} - l_{2i}|$$

Assuming that a total of n events occurred, $n = \min(N_1, N_2)$, then the estimator of Δ_1 is:

$$\hat{\Delta}_1 = \underset{i \in (0, n]}{\text{median}} |\hat{l}_{1i} - \hat{l}_{2i}|$$

2.1.2 Method 2

A series of slope values can be obtained from the interval between the occurrence time and initial time for each group of events. We select the median of these slopes as the representatives for the survival curves:

$$\Delta_2 = \underset{i|t_i \in (0, t_{N_1})}{\text{median}}(l_{1i}) - \underset{j|t_j \in (0, t_{N_2})}{\text{median}}(l_{2j})$$

Δ_2 can be estimated as follows:

$$\hat{\Delta}_2 = \text{median}_{i|t_i \in (0, t_{N_1})}(\hat{l}_{1i}) - \text{median}_{j|t_j \in (0, t_{N_2})}(\hat{l}_{2j})$$

2.1.3 Method 3

The previous two tests contain absolute values, which may increase the sensitivity of the tests. We can define a test that does not include absolute values:

$$\Delta_3 = \text{median}_i(l_{1i} - l_{2i})$$

Δ_3 can be estimated as follows:

$$\hat{\Delta}_3 = \text{median}_i(\hat{l}_{1i} - \hat{l}_{2i})$$

2.1.4 Method 4

In many studies, the treatment effect is often effective early, and later effects are not obvious^[19]. Based on WKM statistics, we choose the weight as the reciprocal of time, increasing the sensitivity of the test to early survival differences. The form of this statistic is analogous to the slope:

$$\Delta_4 = \int_0^\tau \frac{1}{t_i} |S_{1i} - S_{2i}| dt_i$$

where $\tau = \min(T_{N_1}, T_{N_2})$. T_{N_g} , ($g = 1, 2$) is the time that the last event occurred for group g . Δ_4 can be estimated as follows:

$$\hat{\Delta}_4 = \int_0^\tau \frac{1}{t_i} |\hat{S}_{1i} - \hat{S}_{2i}| dt_i$$

2.2 Slope difference permutation test

Bootstrapping and permutation tests are often used to obtain the P value when the distribution of statistics is difficult to derive. The greatest virtue of permutation tests are their accuracy^[20]. The proposed slope difference permutation test (SDPT) is based

on permutation theory^[21]; it generates a P value directly from the permutation resample data rather than from the distribution of the test statistic under the null hypothesis. The SDPT procedure is as follows:

(1) $\{(T_{gn}, \delta_{gn}) : n = 1, \dots, N_g\}, (g=1, 2)$ denote the two initial sets of sample data. T_{gn} is the survival time of individual n in group g , and δ_{gn} is the event indicator. $\delta_{gn} = 1$ denotes the occurrence of an event; $\delta_{gn} = 0$ denotes that the data were censored. Using the censored data, we can obtain Δ .

(2) $\{(T_n, \delta_n) : n = 1, \dots, N\}$ represent mixed-sample data. Two columns of random permutation samples can be obtained by random sampling of mixed samples without replacement, represented by $\{(T_{gn}^*, \delta_{gn}^*) : n = 1, \dots, N_g\}, (g = 1, 2)$. Sample sizes are N_1 and N_2 , respectively. Δ^* is used to represent the slope difference test based on the permuted sample.

(3) Repeating step (2) B times, we obtain B values of Δ^* . For a given significance level α , the P value is defined as the probability of $\Delta^* > \Delta$. When $p < \alpha$, we reject the null hypothesis.

3 Simulations

Classical test methods have been widely used in practice. These tests are no longer applicable in some specific cases, but in most cases, they can still control the type I errors and maintain high power. Among nonparametric tests, the log-rank test is often used for intergroup comparisons of survival data and has excellent performance when the proportional hazards assumption is valid. The Gehan-Wilcoxon test can be regarded as a weighted log-rank test, in which greater emphasis is placed on the early

differences between survival curves. The Fleming-Harrington test is more flexible in terms of weight: with a sufficient weight, the test can focus on the differences between survival curves in a period of interest. Therefore, we selected the log-rank (LG), Gehan-Wilcoxon (GW), and Fleming-Harrington (FH(0,1), FH(1,1)) tests as four classic test methods for comparison with the methods proposed in this paper.

3.1 Simulation design

To thoroughly investigate the test power of the statistics in different scenarios, the simulation is divided into three categories. In the first class of simulations, we assume that the survival data have originated from the same survival distribution and that the hazard functions are either proportional or not proportional and do not cross. Second, we consider survival curves that cross at different times. Finally, we consider hazard functions with differences in the earlier stage.

For conditions where both groups have the same sample size, we use (20,20), (50,50), and (100,100) samples for comparison^[18]. The survival data may also consist of a disproportionate number of samples between the experimental and control groups; in that case, we set unequal sample sizes [(20,30), (40,80), (20,50) and (50,100)]for each group in the simulation^[7]. The significance level is set to $\alpha = 0.05$, and the number of iterations is 1000. For each scenario, the simulation procedure is as follows.

3.1.1 Simulation of survival time

- (1) Generate random variables $U_i, U_i \sim U[0,1]$;
- (2) Obtain the cumulative hazard function $H_0(t)$ and then the inverse function $H_0^{-1}(t)$

depending on the distribution of survival time.

(3) Determine the survival time, $T_i = H_0^{-1}[-\log(U)]$.

3.1.2 Censoring time simulation

Survival data are often censored; for the simulations, we set three censoring rates: 20%, 30% and 50%. Both equal and unequal censor rates are considered for the groups^[22].

(1) We assume that for a censoring rate $F\%$, the censoring time C_i obeys a uniform distribution;

(2) The survival time is generated from an exponential or a Weibull distribution;

(3) The status is set to $\delta = 1$ when there was no censoring. Here, we consider the data to be right censored; we assume that patients enter the study uniformly over a certain time interval and that censoring time C is generated from uniform distributions $U(0, a)$ and $U(0, b)$. The observed survival time is denoted by $T = \min(X, C)$, $\delta = I(X \leq C)$, where X is the actual survival time. We adjust the average censoring rate by the values a and b .

3.1.3 Empirical power

(1) According to the distribution of survival time and the censoring rates, the simulation data are obtained as $\{(T_{gn}, \delta_{gn}): g=1, \dots, N_g\}, (g=1, 2)$. For test method Q , the simulation data are tested to obtain the P value, P_{Q1} ;

(2) Step (1) is repeated an additional 999 times, and we obtain 1000 P values:

$P_{Q1}, P_{Q2}, \dots, P_{Q1000}$;

(3) Power is calculated as $\sum_i 1(p_{Qi} < 0.05) / 1000$.

The specific settings are as follows.

3.2 Survival curve design

3.2.1 Non-crossing curves

Set the following three hazard functions: $h_0 = 1$, $h_1 = 2$, $h_2 = 1.8 + 0.5t$. We consider the following three scenarios for simulating survival data:

- (1) $\{h_0, h_0\}$: the same hazard functions are used to calculate the type I error;
- (2) $\{h_0, h_1\}$: the hazard functions are proportional;
- (3) $\{h_0, h_2\}$: the hazard functions are not proportional and do not cross.

3.2.2 Crossing curves

Set the following survival functions: $S_3 \sim W(2.5, 30)$, $S_7 \sim W(1.5, 5)$. $W(\eta, \theta)$ denotes the Weibull distribution, whose survival function satisfies the form:

$S(t) = \exp\{-t / \theta\}^\eta$. Set the following three hazard functions

$$h_4 = 0.125I(t \leq 1) + 0.01I(t > 1)$$

$$h_5 = 1/12$$

$$h_6 = 1/4I(t \leq 2) + 1/35I(t > 2)$$

$$h_8 = 0.5I(t \leq 1.5) + 0.1I(t > 1.5)$$

- (1) $\{S_3, S_4\}$: two survival curves with an early ($S(t) > 0.6$) crossing;
- (2) $\{S_5, S_6\}$: two survival curves with a middle ($0.4 < S(t) < 0.6$) crossing;
- (3) $\{S_7, S_8\}$: two survival curves with a late ($S(t) < 0.4$) crossing.

3.2.3 Non-crossing and early difference curves

We consider the differences in hazard functions in different periods. The specific settings are as follows. Set two hazard functions with early ($0 \leq t < 0.6$) differences:

$$h_9 = 0.75I(0 \leq t < 0.4) + 3I(0.4 \leq t < 0.6) + 1I(0.6 \leq t)$$

$$h_{10} = 3I(0 \leq t < 0.4) + 0.75I(0.4 \leq t < 0.6) + 1I(0.6 \leq t)$$

According to the hazard functions and survival functions we set above, the survival curves under each scenario are shown in Fig. S1 of Supplemental file.

3.3 Simulation results

(1) Scenario 1.1

When the survival data came from the same distribution, the significance level was set to, and the probability of a type I error in each test fluctuated around approximately 0.05, as shown in Fig. 1. M-1 to M-4 represent Method 1 to Method 4 proposed earlier in this paper.

Fig. 1. Power of test procedures for scenario 1.1(Type I error rates).

(2) Scenario 1.2

When the hazard functions are proportional, the log-rank test achieves good performance; this has the highest power regardless of the sample proportion when

there is no censoring. However, classical tests undergo significant loss of statistical testing power as the censoring rate increases. Compared with classical tests, the SDPTs still have robust power in the case of censoring. When both groups have either 50 or 100 samples, censoring has little influence on the SDP test.

When there is no censoring, the test power of each method tends to consistently increase toward 1 as the sample size increases. However, when the censoring rate is greater than 30%, M-2 and M-3 have obvious advantages, with M-2 being more robust. See the details in Fig. S2 of Supplemental file.

(3) Scenario 1.3

When the hazard functions are not proportional and do not cross, the performance of each test is similar to that in scenario 1.2. The log-rank test has the best performance when there is no censoring. However, classical tests lose power as the censoring rate increases; when the censoring rate reaches 50%, these tests have difficulty assessing differences. In this situation, the powers of the four SDPTs are all superior to those of the classical tests. Among the SDPTs, M-2 has the best performance and maintains robust test power under different simulated conditions. The details are shown in Fig. S3 of Supplemental file.

(4) Scenario 2.1

When the survival distributions cross in the early stage, the differences are mainly reflected in the middle and later stages, and the power of the log-rank test does not decrease substantially. Compared with that of the log-rank test, the power of GW is lower. The FH(0,1) test, which focuses on later differences, performs better than

FH(1,1). Early changes in the survival curve have a great influence on the slope. Since the formulas of M-2 and M-3 do not involve absolute values, positive and negative changes in the slope differences cancel out, so these two methods cannot detect the differences in this scenario. Among the remaining two methods, M-4 not only outperforms M-1 but also maintains high test power under different censoring rates and different sample proportions. The details are shown in Fig. S4 of Supplemental file.

(5) Scenario 2.2

As shown in Fig. S5 of Supplemental file, the power of the log-rank and GW tests substantially decreased when the survival curves crossed at the middle stage; on the other hand, the FH(0,1) and FH(1,1) tests are more applicable in this scenario. Among the SDPTs, when there is no censoring, M-4 performs better overall but still not as well as the FH(0,1) test. When samples are censored, unlike the classical tests, the powers of M-1, M-2 and M-3 increase with increasing censoring rate, with M-2 increasing most.

In this scenario, FH(0,1) is more applicable when there is no censoring. When the censoring rate is greater than or equal to 30%, M-2 and M-3 are better than the FH(0,1), the best classic method, with M-2 being more effective.

(6) Scenario 2.3

When the survival distributions crossover of in the later period, as shown in Fig. S6 of Supplemental file, the power of the FH(0,1) test is the highest when there is no censoring. However, similar to scenario 2.2, the power of this test decreases

substantially as the censoring rate increases. In contrast, the power of M-2 increases with increasing censoring rate.

(7) Scenario 3.1

As shown in Fig. 2, the GW test gives higher weight to previous survival data, so it performs best among the classical tests. However, the power of the SDPTs is higher than that of the GW test in all cases. The SDPTs are more suitable for scenarios in which the survival distributions differ in the early stage. Among them, M-2 performed best.

Fig. 2. Power of test procedures for scenario 3.1.

In summary, the SDPTs can maintain robust test power in cases involving censoring while controlling the type I error, an excellent ability that traditional tests do not possess. In addition, the SDPTs performed best with both large and small sample sizes in cases where the survival curve differed at the early stage. To present

the simulation results clearly, we summarize the optimal test for each case in Tab. 1.

Table 1. Optimal tests for different scenarios.

Scenario	Censoring rate(%)	Small sample size		Large sample size	
		Proportional	Nonproportional	Proportional	Nonproportional
1.2	0	M-3	LG	LG	LG
	20	M-2	M-2	M-3	M-2
	30	M-2	M-2	M-3	M-2
	50	M-2	M-2	M-3	M-2
1.3	0	LG	LG	LG	LG
	20	M-2	M-2	LG	LG
	30	M-2	M-2	M-2	M-2
	50	M-2	M-2	M-2	M-2
2.1	0	FH(0,1)	FH(1,1)	LG	LG
	20	FH(0,1)	FH(0,1)	FH(1,1)	M-4
	30	M-4	FH(0,1)	M-4	M-4
	50	M-4	M-4	M-4	M-4
2.2	0	FH(0,1)	FH(0,1)	FH(0,1)	FH(0,1)
	20	M-3	M-2	FH(0,1)	FH(0,1)
	30	M-2	M-2	M-3	M-2
	50	M-2	M-2	M-3	M-2
2.3	0	FH(0,1)	FH(0,1)	FH(0,1)	FH(0,1)
	20	M-2	M-3	M-4	M-3
	30	M-2	M-2	M-2	M-2
	50	M-2	M-2	M-2	M-2
3.1	0	M-2	M-2	M-2	M-2
	20	M-2	M-2	M-2	M-2
	30	M-3	M-2	M-2	M-2
	50	M-2	M-2	M-2	M-2

4 Applications

4.1 Example 1

The SDPTs were applied to a reconstructed dataset of diabetic patients^[23], which included survival data for 194 patients at high risk for diabetes. The original data is not publicly available, thus we generated data with similar characteristics using a simulation method. This study assessed the long-term prognostic value of myocardial perfusion imaging (MPI) in a high-risk cohort of patients with diabetes mellitus.

Fig. 3. The Kaplan-Meier survival curves with normal vs. abnormal MPI.

Abnormal MPI was detected in 125 patients, and the other patients had normal MPI. The Kaplan–Meier survival curves are presented in Fig. 3. Up to 4 years after MPI, the normal MPI patients had a better prognosis than the abnormal MPI patients. Although early differences can be seen directly from Fig. 3, the P value of the log-rank test was 0.6. When the limit of the test was defined as the 4-year follow-up period, the P value of the log-rank test was 0.036, resulting in a rejection of the null hypothesis. However, this approach was highly susceptible to subjective influence. The P values of the GW test, FH(1,1) test and FH(0,1) test were 0.3, 0.495, and 0.545.

The survival curves crossed, and there was a substantial difference in the early stage. In this scenario, it is more appropriate to use the M-1 and M-2 tests mentioned in this paper. The P value of M-1 was 0.006, and that of M-2 was 0.101. The results from M-3 and M-4 suggested that there was no significant difference between the two survival curves.

4.2 Example 2

Data were derived from clinical datasets of patients treated for gastric cancer. Stablein and Koutrouvelis compared the efficacy of two treatments^[24]; in the study, 45 patients were randomly assigned to each group, and the censoring rates were 4.44% and 13.33%, respectively. The Kaplan–Meier survival curves are presented in Fig. 4. We can conclude that the efficacies of the two treatments were different in the early stage.

For this survival dataset, the P values of the log-rank test, Gehan-Wilcoxon test, FH(0,1) test, and FH(1,1) test were 0.6, 0.05, 0.152 and 0.906, respectively. The SDP tests are suitable for this scenario, as the survival functions are different in the early stage. The P values of M-1, M-2, M-3 and M-4 were 0.008, 0.001, 0.004 and 0.024, respectively. As the selected significance level was 0.05, all SDP tests concluded that there was a significant difference in the efficacies of the two treatments.

Fig. 4. Kaplan-Meier survival curves for chemotherapy vs. chemotherapy+radiotherapy.

5 Discussion

The proposed statistical tests proposed in this paper are based on Kaplan-Meier estimators. Most previous WKM tests were compared based on the difference in the areas under the

survival curves. These tests are more robust than the log-rank test when the hazard functions are not proportional. However, differences in survival curves are also reflected in their slopes, from whose perspective four tests were constructed in this paper. Unlike the area, the differences between slopes are not affected by the crossing of the curves.

When the hazards are proportional and there is no censoring in the survival data, the log-rank test is undoubtedly the best choice. However, a major feature of clinical data is the ease with which it can be censored. In the simulation study, we found that as the censoring rate increased, the powers of the classical tests greatly decreased, while the performance of the SDPTs was more robust. We also simulated scenarios in which the hazard functions differed only at the early, middle or later stage. The simulation results showed that when the hazard functions differed in the early stage, the SDPTs showed high level of power and performed very robustly in all cases. However, the SDPTs were no longer suitable for the middle- and late-change scenarios.

In this paper, four tests based on slope differences were proposed. According to the differences in the morphologies, sample sizes and censoring rates of the survival curves, the optimal test was selected among the four methods. This paper only considered right-censored data, and the processing excludes it from the statistical calculations. Other possible censoring methods need to be further considered. The research in this paper only focuses on a single risk dataset. Competitive risk datasets are also common in clinical medicine and should also be considered in subsequent studies.

References

- [1] Harrington D, and Fleming T. (1982) A class of rank test procedures for censored survival data. *Biometrika*, 69, 553-566.
- [2] Prentice RL. (1978) Linear rank tests with right censored data. *Biometrika*, 65, 167-179.
- [3] Weeda VB, Murawski M, McCabe AJ, et al. (2013) Fibrolamellar variant of hepatocellular carcinoma does not have a better survival than conventional hepatocellular carcinoma – Results and treatment recommendations from the Childhood Liver Tumour Strategy Group (SIOPEL) experience. *Eur J Cancer* 49(12): 2698-2704.
- [4] Smith MJF, Ridgway PF, Catton CN, et al. (2014) Combined management of retroperitoneal sarcoma with dose intensification radiotherapy and resection: Long-term results of a prospective trial. *Radiotherapy and Oncology* 110(1): 165-171.
- [5] Hsieh J-J, Chen H-Y. (2017) A testing strategy for two crossing survival curves. *Communications in Statistics - Simulation and Computation*, 46, 6685-6696.
- [6] Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 32, 2380-2385.
- [7] Li H, Han D, Hou Y, Chen H, and Chen Z. (2015) Statistical inference methods for two crossing survival curves: a comparison of methods. *PloS One*, 10(1): e0116774.
- [8] Lin X, Xu Q. A new method for the comparison of survival distributions. *Pharm Stat*, 9, 67-76.
- [9] Wang R, Lagakos SW, Gray RJ. (2010) Testing and interval estimation for two-sample survival comparisons with small sample sizes and unequal censoring. *Biostatistics*, 2010, 11,

676-692.

[10] Cox D. (1972) Regression Models and Life-Tables. Journal of the Royal Statistics Society, Series B, 34, 187-220.

[11] Peto R, and Peto J. (1972) Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society, Series A, 135, 185-198.

[12] Pepe MS, and Fleming T R. (1989) Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data. Biometrics, 45, 497-507.

[13] Calkins KL, Canan CE, Moore RD, et al. (2018) An application of restricted mean survival time in a competing risks setting: comparing time to ART initiation by injection drug use. BMC Medical Research Methodology, 18, 27.

[14] Lyu J, Hou Y, Chen Z. (2020) The use of restricted mean time lost under competing risks data. BMC Medical Research Methodology, 20, 1-11.

[15] Royston P, Parmar MK. (2013) Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol, 13, 152.

[16] Zhang C, Wu Y, Yin G. (2020) Restricted mean survival time for interval-censored data. Stat Med, 39, 3879-3895.

[17] Zhao L, Claggett B, Tian L, et al. (2016) On the restricted mean survival time curve in survival analysis. Biometrics, 72, 215-221.

[18] Huang X, Lyu J, Hou Y, and Chen Z. (2020) A nonparametric statistical method for two crossing survival curves. Communications in Statistics - Simulation and Computation, 1-10.

[19] Wu L, Gilbert PB. Flexible weighted log-rank tests optimal for detecting early and/or late

survival differences. *Biometrics*, 58, 997-1004.

[20] Efron B, Tibshirani RJ. (1994) *An Introduction to the Bootstrap*. CRC press.

[21] Galimberti S, Valsecchi MG. (2013) Multivariate permutation test to compare survival curves for matched data. *BMC Medical Research Methodology*, 13-16

[22] Chen K, Jin Z, Ying Z. (2002) Semiparametric analysis of transformation models with censored data. *Biometrika*, 89, 659-668.

[23] Boiten HJ, van Domburg RT, Valkema R, et al. (2016) Dobutamine stress myocardial perfusion imaging: 8-year outcomes in patients with diabetes mellitus. *European heart journal Cardiovascular Imaging*, 2016, 17, 871-876.

[24] Stablein DM, Koutrouvelis IA. (1985) A two-sample test sensitive to crossing hazards in uncensored and singly censored data. *Biometrics*, 41, 643-652.

Ethics approval and consent to participate:

The first example data was reconstructed using a simulation method^[23], the second example data in this study is available publicly. No new clinical data was gathered or used. There is thus no need for the ethical approval and consent to participate.

Consent for publication:

Not Applicable.

Availability of data and materials:

The first example data was reconstructed using a simulation method^[23]. The data of two treatments in gastric cancer clinical study that support the findings of this study is available in the article “A two-sample test sensitive to crossing hazards in uncensored and singly censored data” that public by Journal of Biometrics.

Competing interests:

All authors declare no potential conflicts of interest.

Funding:

This work was supported by the National Natural Science Foundation of China (81903411, 82173622, 81673268), the Natural Science Foundation of Guangdong Province (2018A030313849), and the Guangdong Basic and Applied Basic Research Foundation (2019A1515011506).

Authors' contributions:

YH and ZC were involved in the idea of this study, YH and YZ originated the

methodology, YZ prepared the simulation and application results, including figures and tables.

YH, YZ and ZC jointly contributed to drafting the article. All authors have read and approved the final manuscript.

Acknowledgements:

Not applicable.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalfileforAslopeteststrategyforcensoreddatainsurvivalanalysis20220212.docx](#)