On Comparing Survival Curves with Right-Censored Data According to the Events Occur at the Beginning, in the Middle and at the End of Study Period

Pinar Gunel Karadeniz[1,*] and Ilker Ercan[2]

[1]*Department of Biostatistics Gaziantep, SANKO University, Faculty of Medicine, 27090, Turkey*

[2]*Department of Biostatistics Gorukle Campus Bursa, Uludag University, Faculty of Medicine, Turkey*

Abstract: In clinical practice the event of interest does not always occur equally across the study time period. Depending on the disease being investigated, the event that is of interest can occur intensively in different periods of the follow-up time. In such cases, choosing the correct survival comparison test has importance. This study aims to examine and discuss the results of survival comparison tests under some certain circumstances. A simulation study was conducted. We discussed the result of different tests such as Logrank, Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, Modified Peto-Peto tests and tests belonging to Fleming-Harrington test family with (p, q) values; (1, 0), (0.5, 0.5), (1, 1), (0, 1) ve (0.5, 2) by means of Type I error rate that obtained from simulation study, when the event of interest occurred intensively at the beginning of the study, in the middle of the study and at the end of the study time period. As a result of simulation study, Type I error rate of tests is generally lower or higher than the nominal value. In the light of the results, it is proposed to re-examine the tests for cases where events are observed intensively at the beginning, middle and late periods, to carry out new simulation studies and to develop new tests if necessary.

Keywords: Survival analysis, survival curves, comparison of survival curves, right censored observations.

1. INTRODUCTION

In medicine, studies, especially clinical studies, mostly investigate whether a new drug, a new treatment, a new procedure is better than the current methods. In these studies, in addition to the short-term results of the method, the results such as long-term mortality and morbidity are also very important. For instance, in a study of chronic kidney disease, a researcher might want to investigate the survival probability in chronic kidney disease patients with and without dialysis. The event of interest in such a study has two options: survival or death. In such cases, researchers care about patient's survival time under a particular treatment or certain conditions. The data obtained in this type of study is called survival data. The methods applied for survival data belong to survival analysis topic. Survival analysis is performed to determine the survival probability, to compare survival rates of different groups, or to examine the effect of treatment types and other factors on the survival time until the occurrence of the event of interest (death, disease, recurrence etc.) within a certain follow-up period [1].

In survival analysis, survival function which is expressed as a percentage of individuals who are alive up to or more than a specific time t, or a survival curve which is the plot of this function are used [1].

It is fairly easy to estimate the survival function when all the individuals in a study are observed until the end of the study period. However, most clinical trials are completed within a certain time period, and throughout this period it is not always possible to observe all individuals until the event of interest occurs. In such cases, censored data comes to exist. Some special methods have been developed to estimate the survival function in the presence of censored data. One of these methods is the "Life Table" method, and the other is "Kaplan-Meier Product Limit" method [1-2].

While it is important to estimate the survival probability of a specific group, it is quite common in clinical research to compare survival curves of two or more groups. Although there are many tests which are developed to compare survival curves of the groups in the presence of censored data, in the literature, in practice the most frequently used test is the logrank test. The logrank test was proposed by Mantel and Haenszel (1959) and was developed by Mantel (1966) [3-4]. Then, Gehan (1965) presented a generalized version of the Wilcoxon rank sum test for censored data. Peto and Peto (1972) and Tarone and Ware (1977) proposed specific tests for censored data. Fleming and Harrington (1981) and Harrington and Fleming (1982) developed weighted logrank statistics [5-9].

The logrank test gives equal weights across all time periods [10-11]. On the other hand, in clinical practice the event of interest does not always occur equally

*Address correspondence to this author at the SANKO Üniversitesi, Incilipinar St. Gazimuhtar Pasa Boulevard 36, 27090. Sehitkamil Gaziantep Turkey; Tel: 00905056708243; E-mail: pgkaradeniz@sanko.edu.tr

Table 1:   Survival Comparison Tests and their Test Statistics with Explanations

Test	Test Statistic	Explanation
Logrank	$$\dfrac{\left(\sum_j \left(d_{ij} - d_j \dfrac{r_{ij}}{r_j}\right)\right)^2}{\sum_{j=1}^{k} \dfrac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}}$$	Equal weights across all times.
Gehan Generalized Wilcoxon	$$\dfrac{\left(\sum_j r_j \left(d_{ij} - d_j \dfrac{r_{ij}}{r_j}\right)\right)^2}{\sum_{j=1}^{k} r_j^2 \dfrac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}}$$	Places very heavy weight on hazards at the beginning of the study.
Tarone-Ware	$$\dfrac{\left(\sum_j \sqrt{r_j} \left(d_{ij} - d_j \dfrac{r_{ij}}{r_j}\right)\right)^2}{\sum_{j=1}^{k} r_j \dfrac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}}$$	Places heavy weight on hazards at the beginning of the study.
Peto-Peto	$$\dfrac{\left(\sum_j \tilde{S}(t_j) \left(d_{ij} - d_j \dfrac{r_{ij}}{r_j}\right)\right)^2}{\sum_{j=1}^{k} \tilde{S}(t_j)^2 \dfrac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}}$$	Places a little more weight on hazards at the beginning of the study.
Modified Peto-Peto	$$\dfrac{\left(\sum_j \tilde{S}(t_j) r_j /(r_j + 1) \left(d_{ij} - d_j \dfrac{r_{ij}}{r_j}\right)\right)^2}{\sum_{j=1}^{k} [\tilde{S}(t_j) r_j /(r_j + 1)]^2 \dfrac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}}$$	Places a little more weight on hazards at the beginning of the study.
Fleming-Harrington (p, q) **F-H (1,0)** **F-H (0.5,0.5)** **F-H (1,1)** **F-H (0,1)** **F-H (0.5,2)**	$$\dfrac{\left(\sum_j \hat{S}(t_{j-1})^p [1 - \hat{S}(t_{j-1})]^q \left(d_{ij} - d_j \dfrac{r_{ij}}{r_j}\right)\right)^2}{\sum_{j=1}^{k} [\hat{S}(t_{j-1})^p [1 - \hat{S}(t_{j-1})]^q]^2 \dfrac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}}$$	Places weight on hazards at the beginning of the study. Places weight on hazards in the middle of the study. Places weight on hazards in the middle of the study. Places weight on hazards at the end of the study. Places weight on hazards at the end of the study.

t_j, survival time.
d_{ij}, number of individuals who experience the event in group i at time j.
d_j, total number of individuals in both groups who experience the event.
r_{ij}, number of individuals at risk in group i at time j.
r_j, total number of individuals at risk at time j.
r_{1j}, number of individuals at risk in group 1.
r_{2j}, number of individuals at risk in group 2.

$\hat{S}(x) = \prod_{t_j \leq t} \left(1 - \dfrac{d_j}{r_j}\right)$, Kaplan-Meier estimator of survival function.

$\hat{S}(x) = \prod_{t_j < t} \left(1 - \dfrac{d_j}{r_j + 1}\right)$, a modification of Kaplan-Meier estimator.

across the study time period. Depending on the disease being investigated, the event that is of interest can occur intensively in different periods of the follow-up time. In such cases, choosing the correct test has importance.

In literature, there are studies which examine the performance of survival comparison tests. In these studies, simulations have been carried out and generally comparisons are made in terms of power, and some studies have also used the Type I error rate

for comparison [5, 11-13]. The nominal value for Type I error rate was considered as 0.05.

In our previous study, we examined the survival comparison tests in regard to Type I error rates with right-censored data in some defined particular cases with events spread equally during the follow-up time [14]. In present study, by using the same defined cases, we conducted a simulation study and examined and discussed the different test results when the event of interest occurred intensively at the beginning of the study, in the middle of the study and at the end of the study time period.

2. MATERIAL AND METHODS

In this study, in order to examine survival comparison tests, a simulation study with 500 replicates was conducted and type I error rates were obtained and assessed.

The survival comparison tests we used are summarized briefly in Table **1** with their test statistics and short explanations [6-7, 13, 15-17].

The simulation study was conducted for four different sample sizes, 10, 30, 50 and 100, and the event of interest was placed intensively at the beginning of the study, at the end of the study and in the middle of the study time period. While generating the survival data, simulation studies in the literature were reviewed. For exponential distribution the scale parameter was selected as $\beta= 0.5, 1, 1.5$; for Weibull distribution the shape parameter was $\alpha= 1, 2, 3$ and the scale parameter was $\beta= 1.5, 2.5, 3.5$; for lognormal distribution the shape parameter was $\sigma= 1, 2, 3$ and the scale parameter was $m= 0$; for inverse Gaussian distribution the location parameter was $\mu= 0.5$ and the scale parameter was $\lambda= 1, 2, 3$.

The status variable was generated with the same sample sizes 10, 30, 50 and 100 and;

- in order to generate the status variable as it would be when the event of interest was intensively placed at the beginning of the study time period, one-third of the sample size had 0.75 binomial probability, while two-thirds of it had 0.25 binomial probability.

- in order to generate the status variable as it would be when the event of interest was intensively placed in the middle of the study time period, the first one-third of the sample size had

0.25 binomial probability, the middle one-third of the sample size had 0.75 binomial probability and the remaining one-third of the sample size had 0.25 binomial probability.

- in order to generate the status variable as it would be when the event of interest was intensively placed at the end of the study time period, the first two-thirds of the sample size had 0.25 binomial probability, while the remaining one-third of it had 0.75 binomial probability.

Data were generated via R software version 3.0.3, and for the survival curve comparisons NCSS version 07.1.5 was used. The WinAutomation program was used for replications [18-20].

3. RESULTS

Type I error rates according to the simulation study for the sample sizes of n=10, 30, 50 and 100 are given in Tables **2**, **3**, **4**, **5**, **6**, **7**.

Graphs of Type I error rates according to different sample sizes, different distributions and different tests used were given in Figure **1**.

4. DISCUSSION

In the literature, it is suggested that the logrank test should be preferred if it is desired to give equal weight to events occurring throughout the entire study period in a study [21]. However, for this test, the hazard functions of the groups are assumed to be proportional across the time period of interest [10, 22-23]. As long as the hazard ratios of only two groups are constant, logrank is a strong test. However, when the assumption of proportional hazards is violated, that is, when the hazard functions of the groups overlap, the logrank test loses power and the Gehan-Wilcoxon and Tarone-Ware tests become stronger [7, 24]. Likewise, the Peto-Peto test is more effective when the proportional hazard assumption is not provided. When the Fleming-Harrington test family is considered, they are the ones that provide the most flexibility in adjusting weights [25]. Fleming-Harrington tests are focused on tests that are sensitive to conditions of overlapping hazard functions [24].

In addition to the overlapping of hazard ratios issue, depending on the disease being investigated, the event of interest does not appear evenly throughout the follow-up period, and it can occur intensely during different periods of the study period. In these cases,

Table 2: Type I Error Rates of Tests for Exponential Distribution with Different Parameters

Tests	Exponential (0.5)				Exponential (1)				Exponential (1.5)			
	When the Event of Interest Occurs at the Beginning of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0260	0.0120	0.0120	0.0180	0.0160	0.0140	0.0240	0.0040	0.0100	0.0060	0.0060	0.0080
Gehan-Wilcoxon	0.0100	0.0000	0.0100	0.0100	0.0120	0.0160	0.0120	0.0020	0.0040	0.0080	0.0060	0.0080
Tarone-Ware	0.0160	0.0040	0.0100	0.0120	0.0100	0.0140	0.0100	0.0060	0.0060	0.0080	0.0080	0.0080
Peto-Peto	0.0200	0.0020	0.0100	0.0120	0.0120	0.0160	0.0140	0.0080	0.0060	0.0060	0.0080	0.0100
Mod. Peto-Peto	0.0200	0.0020	0.0100	0.0120	0.0120	0.0160	0.0140	0.0080	0.0060	0.0060	0.0080	0.0100
F-H (1, 0)	0.0240	0.0020	0.0100	0.0120	0.0120	0.0160	0.0140	0.0080	0.0080	0.0060	0.0080	0.0100
F-H (0.5, 0.5)	0.0360	0.0240	0.0200	0.0260	0.0300	0.0240	0.0280	0.0200	0.0220	0.0180	0.0140	0.0300
F-H (1, 1)	0.0480	0.0340	0.0240	0.0380	0.0360	0.0400	0.0320	0.0320	0.0300	0.0240	0.0260	0.0360
F-H (0, 1)	0.0500	0.0420	0.0360	0.0380	0.0380	0.0440	0.0500	0.0420	0.0260	0.0240	0.0320	0.0420
F-H (0.5, 2)	0.0540	0.0600	0.0560	0.0420	0.0600	0.0480	0.0580	0.0460	0.0280	0.0340	0.0420	0.0380
Tests	When the Event of Interest Occurs in the Middle of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0340	0.0200	0.0200	0.0360	0.0380	0.0240	0.0280	0.0320	0.0420	0.0360	0.0480	0.0360
Gehan-Wilcoxon	0.0540	0.0540	0.0520	0.0800	0.0580	0.0660	0.0660	0.0780	0.0620	0.0520	0.0920	0.0720
Tarone-Ware	0.0520	0.0340	0.0420	0.0620	0.0460	0.0460	0.0420	0.0540	0.0520	0.0420	0.0760	0.0520
Peto-Peto	0.0520	0.0440	0.0460	0.0700	0.0480	0.0480	0.0440	0.0640	0.0520	0.0440	0.0820	0.0600
Mod. Peto-Peto	0.0540	0.0480	0.0460	0.0700	0.0520	0.0500	0.0440	0.0640	0.0540	0.0440	0.0840	0.0600
F-H (1, 0)	0.0500	0.0440	0.0460	0.0720	0.0480	0.0500	0.0460	0.0620	0.0500	0.0420	0.0800	0.0580
F-H (0.5, 0.5)	0.0200	0.0120	0.0100	0.0120	0.0200	0.0100	0.0160	0.0160	0.0400	0.0240	0.0180	0.0100
F-H (1, 1)	0.0200	0.0140	0.0100	0.0100	0.0240	0.0100	0.0180	0.0120	0.0280	0.0180	0.0120	0.0060
F-H (0, 1)	0.0180	0.0180	0.0180	0.0200	0.0260	0.0180	0.0180	0.0200	0.0280	0.0240	0.0200	0.0120
F-H (0.5, 2)	0.0180	0.0220	0.0260	0.0300	0.0260	0.0300	0.0200	0.0380	0.0240	0.0340	0.0360	0.0300
Tests	When the Event of Interest Occurs at the End of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.1160	0.0880	0.0940	0.0840	0.1140	0.0720	0.0840	0.0940	0.1220	0.0860	0.1000	0.0960
Gehan-Wilcoxon	0.0780	0.0540	0.0920	0.0700	0.0840	0.0840	0.0720	0.0800	0.0880	0.0760	0.0800	0.0960
Tarone-Ware	0.1020	0.0660	0.0920	0.0860	0.1000	0.0920	0.0780	0.0860	0.1040	0.0880	0.0940	0.1100
Peto-Peto	0.1020	0.0680	0.0980	0.1020	0.1040	0.0820	0.0920	0.1000	0.1080	0.0980	0.1080	0.1140
Mod. Peto-Peto	0.0980	0.0680	0.1000	0.1000	0.0980	0.0920	0.0900	0.1000	0.0960	0.0980	0.1060	0.1100
F-H (1, 0)	0.1040	0.0700	0.0980	0.1020	0.1040	0.0860	0.0900	0.1000	0.1100	0.0980	0.1080	0.1140
F-H (0.5, 0.5)	0.1380	0.1080	0.0960	0.0860	0.1340	0.0880	0.0840	0.0940	0.1380	0.0920	0.0980	0.0900
F-H (1, 1)	0.1420	0.1040	0.0840	0.0720	0.1320	0.0880	0.0800	0.0880	0.1380	0.0920	0.0960	0.0800
F-H (0, 1)	0.1420	0.1140	0.0640	0.0660	0.1300	0.0900	0.0600	0.0620	0.1500	0.0920	0.0860	0.0620
F-H (0.5, 2)	0.1460	0.1080	0.0580	0.0540	0.1220	0.0940	0.0700	0.0560	0.1540	0.0780	0.0820	0.0500

Table 3: Type I Error Rates of Tests for Weibull Distribution with Different Parameters (1, 1.5; 1, 2.5; 1, 3.5)

	Weibull (1, 1.5)				Weibull (1, 2.5)				Weibull (1, 3.5)			
Tests	When the Event of Interest Occurs at the Beginning of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0180	0.0220	0.0200	0.0180	0.0080	0.0120	0.0160	0.0120	0.0280	0.0120	0.0160	0.0200
Gehan-Wilcoxon	0.0120	0.0120	0.0180	0.0100	0.0080	0.0060	0.0140	0.0100	0.0140	0.0080	0.0080	0.0080
Tarone-Ware	0.0100	0.0180	0.0160	0.0160	0.0080	0.0080	0.0140	0.0100	0.0180	0.0060	0.0100	0.0120
Peto-Peto	0.0120	0.0180	0.0180	0.0140	0.0100	0.0080	0.0160	0.0120	0.0200	0.0060	0.0120	0.0180
Mod. Peto-Peto	0.0120	0.0180	0.0180	0.0140	0.0100	0.0060	0.0140	0.0120	0.0180	0.0060	0.0120	0.0180
F-H (1, 0)	0.0120	0.0180	0.0180	0.0140	0.0080	0.0100	0.0160	0.0120	0.0220	0.0040	0.0120	0.0180
F-H (0.5, 0.5)	0.0280	0.0420	0.0160	0.0320	0.0120	0.0180	0.0180	0.0240	0.0320	0.0400	0.0160	0.0340
F-H (1, 1)	0.0300	0.0520	0.0240	0.0400	0.0100	0.0260	0.0320	0.0320	0.0300	0.0480	0.0200	0.0380
F-H (0, 1)	0.0340	0.0520	0.0280	0.0480	0.0240	0.0240	0.0500	0.0380	0.0360	0.0700	0.0260	0.0560
F-H (0.5, 2)	0.0360	0.0660	0.0400	0.0620	0.0220	0.0280	0.0580	0.0400	0.0400	0.0760	0.0320	0.0700
Tests	When the Event of Interest Occurs in the Middle of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0260	0.0320	0.0300	0.0280	0.0300	0.0300	0.0400	0.0260	0.0540	0.0280	0.0380	0.0440
Gehan-Wilcoxon	0.0580	0.0540	0.0520	0.0760	0.0460	0.0760	0.0820	0.0580	0.0680	0.0580	0.0660	0.0580
Tarone-Ware	0.0400	0.0460	0.0320	0.0520	0.0380	0.0460	0.0600	0.0460	0.0640	0.0360	0.0520	0.0520
Peto-Peto	0.0380	0.0520	0.0380	0.0600	0.0380	0.0460	0.0720	0.0480	0.0600	0.0420	0.0580	0.0580
Mod. Peto-Peto	0.0440	0.0520	0.0360	0.0620	0.0380	0.0480	0.0700	0.0460	0.0640	0.0440	0.0580	0.0580
F-H (1, 0)	0.0380	0.0520	0.0380	0.0600	0.0360	0.0460	0.0720	0.0480	0.0600	0.0400	0.0560	0.0580
F-H (0.5, 0.5)	0.0280	0.0140	0.0160	0.0140	0.0180	0.0100	0.0180	0.0140	0.0280	0.0060	0.0120	0.0200
F-H (1, 1)	0.0260	0.0160	0.0120	0.0120	0.0140	0.0140	0.0100	0.0140	0.0200	0.0100	0.0180	0.0200
F-H (0, 1)	0.0260	0.0180	0.0260	0.0180	0.0200	0.0200	0.0200	0.0240	0.0220	0.0160	0.0200	0.0240
F-H (0.5, 2)	0.0220	0.0220	0.0320	0.0300	0.0220	0.0280	0.0300	0.0320	0.0240	0.0220	0.0340	0.0420
Tests	When the Event of Interest Occurs at the End of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.2800	0.1160	0.1000	0.0920	0.1060	0.1180	0.1180	0.0780	0.0500	0.0860	0.1060	0.1000
Gehan-Wilcoxon	0.1800	0.1020	0.0780	0.0740	0.0720	0.0820	0.0900	0.0940	0.0400	0.0820	0.0960	0.0740
Tarone-Ware	0.2200	0.1120	0.0840	0.0880	0.0860	0.1000	0.1100	0.0860	0.0420	0.0900	0.1120	0.0900
Peto-Peto	0.2450	0.1160	0.0900	0.1000	0.0820	0.1100	0.1160	0.0960	0.0420	0.0980	0.1140	0.1040
Mod. Peto-Peto	0.2300	0.1180	0.0960	0.1020	0.0860	0.1080	0.1160	0.0960	0.0420	0.0980	0.1120	0.1060
F-H (1, 0)	0.2450	0.1180	0.0920	0.1000	0.0820	0.1100	0.1160	0.0960	0.0420	0.0940	0.1140	0.1040
F-H (0.5, 0.5)	0.3150	0.1080	0.1040	0.0960	0.1200	0.1140	0.1060	0.0680	0.0560	0.0940	0.1020	0.1180
F-H (1, 1)	0.3100	0.1020	0.0920	0.0900	0.1340	0.1100	0.0920	0.0680	0.0460	0.0860	0.0880	0.1180
F-H (0, 1)	0.3450	0.0980	0.0960	0.0940	0.1380	0.1140	0.0840	0.0620	0.0460	0.0880	0.0760	0.0860
F-H (0.5, 2)	0.3250	0.0960	0.0820	0.0860	0.1300	0.1060	0.0780	0.0620	0.0520	0.0820	0.0660	0.0740

Table 4: Type I Error Rates of Tests for Weibull Distribution with Different Parameters (2, 1.5; 2, 2.5; 2, 3.5)

Tests	Weibull (2, 1.5)				Weibull (2, 2.5)				Weibull (2, 3.5)			
	When the Event of Interest Occurs at the Beginning of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0120	0.0180	0.0160	0.0140	0.0160	0.0360	0.0120	0.0160	0.0260	0.0220	0.0160	0.0080
Gehan-Wilcoxon	0.0140	0.0080	0.0080	0.0100	0.0100	0.0160	0.0120	0.0200	0.0060	0.0180	0.0120	0.0080
Tarone-Ware	0.0140	0.0080	0.0120	0.0100	0.0160	0.0200	0.0100	0.0080	0.0140	0.0240	0.0100	0.0080
Peto-Peto	0.0140	0.0100	0.0100	0.0120	0.0160	0.0220	0.0120	0.0080	0.0160	0.0260	0.0120	0.0100
Mod. Peto-Peto	0.0140	0.0080	0.0100	0.0100	0.0140	0.0220	0.0120	0.0060	0.0160	0.0260	0.0120	0.0100
F-H (1, 0)	0.0140	0.0100	0.0100	0.0120	0.0180	0.0220	0.0120	0.0080	0.0160	0.0260	0.0120	0.0100
F-H (0.5, 0.5)	0.0240	0.0200	0.0240	0.0180	0.0280	0.0520	0.0160	0.0160	0.0340	0.0260	0.0220	0.0080
F-H (1, 1)	0.0340	0.0200	0.0360	0.0260	0.0400	0.0620	0.0180	0.0320	0.0440	0.0420	0.0360	0.0200
F-H (0, 1)	0.0360	0.0300	0.0540	0.0320	0.0380	0.0640	0.0320	0.0380	0.0460	0.0540	0.0440	0.0260
F-H (0.5, 2)	0.0420	0.0440	0.0600	0.0400	0.0320	0.0800	0.0340	0.0440	0.0420	0.0580	0.0560	0.0400
Tests	When the Event of Interest Occurs in the Middle of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0380	0.0360	0.0240	0.0340	0.0440	0.0360	0.0240	0.0280	0.0360	0.0320	0.0220	0.0320
Gehan-Wilcoxon	0.0520	0.0780	0.0700	0.0780	0.0640	0.0700	0.0820	0.0660	0.0460	0.0840	0.0540	0.0640
Tarone-Ware	0.0440	0.0580	0.0520	0.0620	0.0540	0.0540	0.0460	0.0480	0.0420	0.0540	0.0320	0.0440
Peto-Peto	0.0480	0.0640	0.0580	0.0660	0.0540	0.0560	0.0500	0.0520	0.0440	0.0640	0.0360	0.0440
Mod. Peto-Peto	0.0460	0.0620	0.0580	0.0660	0.0540	0.0560	0.0500	0.0540	0.0440	0.0640	0.0360	0.0460
F-H (1, 0)	0.0460	0.0620	0.0540	0.0680	0.0540	0.0580	0.0480	0.0540	0.0440	0.0640	0.0360	0.0440
F-H (0.5, 0.5)	0.0280	0.0160	0.0100	0.0080	0.0320	0.0260	0.0160	0.0060	0.0320	0.0180	0.0100	0.0160
F-H (1, 1)	0.0240	0.0100	0.0120	0.0080	0.0260	0.0220	0.0120	0.0040	0.0300	0.0160	0.0080	0.0140
F-H (0, 1)	0.0180	0.0180	0.0300	0.0180	0.0240	0.0260	0.0140	0.0140	0.0360	0.0200	0.0180	0.0240
F-H (0.5, 2)	0.0240	0.0280	0.0360	0.0280	0.0200	0.0320	0.0260	0.0200	0.0340	0.0320	0.0200	0.0360
Tests	When the Event of Interest Occurs at the End of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.1040	0.1080	0.0960	0.0980	0.0900	0.1140	0.1300	0.0900	0.1240	0.1040	0.0940	0.0960
Gehan-Wilcoxon	0.0780	0.0700	0.1020	0.0840	0.0480	0.0860	0.0920	0.0780	0.0760	0.0960	0.0880	0.0780
Tarone-Ware	0.0900	0.0840	0.1000	0.0980	0.0740	0.1180	0.1040	0.0860	0.0880	0.1100	0.1080	0.0960
Peto-Peto	0.0940	0.0920	0.1040	0.1080	0.0720	0.1220	0.1120	0.0960	0.0920	0.1160	0.1160	0.1080
Mod. Peto-Peto	0.0900	0.0900	0.0980	0.1100	0.0620	0.1240	0.1120	0.0980	0.0880	0.1160	0.1160	0.1100
F-H (1, 0)	0.0940	0.0920	0.1040	0.1080	0.0720	0.1220	0.1120	0.0980	0.0940	0.1180	0.1160	0.1080
F-H (0.5, 0.5)	0.1100	0.1100	0.1000	0.0920	0.1040	0.1240	0.1200	0.0860	0.1480	0.1080	0.0880	0.1000
F-H (1, 1)	0.1280	0.1080	0.0840	0.0900	0.1200	0.1200	0.1100	0.0800	0.1600	0.0980	0.0820	0.1060
F-H (0, 1)	0.1320	0.1220	0.0780	0.0580	0.1300	0.1280	0.0920	0.0740	0.1660	0.0940	0.0760	0.0540
F-H (0.5, 2)	0.1320	0.1100	0.0660	0.0540	0.1240	0.1100	0.0840	0.0640	0.1780	0.0900	0.0660	0.0560

Table 5:   Type I Error Rates of Tests for Weibull Distribution with Different Parameters (3, 1.5; 3, 2.5; 3, 3.5)

| Tests	Weibull (3, 1.5)				Weibull (3, 2.5)				Weibull (3, 3.5)			
	When the Event of Interest Occurs at the Beginning of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0260	0.0100	0.0200	0.0100	0.0120	0.0200	0.0140	0.0100	0.0100	0.0280	0.0180	0.0120
Gehan-Wilcoxon	0.0100	0.0100	0.0140	0.0080	0.0080	0.0160	0.0040	0.0040	0.0060	0.0160	0.0120	0.0060
Tarone-Ware	0.0160	0.0120	0.0160	0.0040	0.0100	0.0140	0.0080	0.0060	0.0040	0.0180	0.0160	0.0040
Peto-Peto	0.0160	0.0120	0.0160	0.0060	0.0080	0.0160	0.0100	0.0080	0.0040	0.0240	0.0180	0.0060
Mod. Peto-Peto	0.0140	0.0120	0.0160	0.0060	0.0080	0.0160	0.0100	0.0080	0.0060	0.0240	0.0180	0.0060
F-H (1, 0)	0.0180	0.0100	0.0200	0.0060	0.0080	0.0160	0.0100	0.0060	0.0040	0.0240	0.0180	0.0060
F-H (0.5, 0.5)	0.0340	0.0140	0.0180	0.0120	0.0220	0.0220	0.0280	0.0140	0.0240	0.0420	0.0160	0.0160
F-H (1, 1)	0.0380	0.0220	0.0220	0.0200	0.0240	0.0280	0.0400	0.0220	0.0300	0.0480	0.0220	0.0140
F-H (0, 1)	0.0400	0.0280	0.0340	0.0260	0.0300	0.0420	0.0500	0.0400	0.0360	0.0480	0.0360	0.0200
F-H (0.5, 2)	0.0360	0.0420	0.0360	0.0380	0.0380	0.0540	0.0640	0.0500	0.0380	0.0420	0.0420	0.0340
Tests	When the Event of Interest Occurs in the Middle of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0400	0.0400	0.0300	0.0380	0.0320	0.0160	0.0220	0.0440	0.0360	0.0260	0.0340	0.0380
Gehan-Wilcoxon	0.0600	0.0700	0.0620	0.0820	0.0500	0.0540	0.0560	0.0960	0.0560	0.0640	0.0660	0.0720
Tarone-Ware	0.0480	0.0540	0.0380	0.0640	0.0400	0.0280	0.0320	0.0600	0.0420	0.0420	0.0480	0.0640
Peto-Peto	0.0500	0.0620	0.0400	0.0660	0.0360	0.0260	0.0380	0.0700	0.0420	0.0440	0.0540	0.0680
Mod. Peto-Peto	0.0580	0.0620	0.0440	0.0660	0.0440	0.0260	0.0380	0.0700	0.0480	0.0480	0.0560	0.0680
F-H (1, 0)	0.0500	0.0580	0.0420	0.0680	0.0340	0.0280	0.0380	0.0700	0.0420	0.0460	0.0560	0.0680
F-H (0.5, 0.5)	0.0240	0.0280	0.0260	0.0220	0.0220	0.0080	0.0180	0.0220	0.0300	0.0140	0.0220	0.0180
F-H (1, 1)	0.0180	0.0160	0.0280	0.0220	0.0220	0.0080	0.0180	0.0140	0.0260	0.0100	0.0120	0.0200
F-H (0, 1)	0.0160	0.0160	0.0300	0.0380	0.0220	0.0240	0.0240	0.0140	0.0200	0.0100	0.0140	0.0300
F-H (0.5, 2)	0.0160	0.0220	0.0380	0.0400	0.0200	0.0280	0.0280	0.0180	0.0160	0.0100	0.0220	0.0400
Tests	When the Event of Interest Occurs at the End of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.1020	0.0960	0.0840	0.0920	0.1120	0.0380	0.1040	0.0960	0.0960	0.0940	0.1000	0.0820
Gehan-Wilcoxon	0.0760	0.0740	0.0800	0.0640	0.0780	0.0280	0.0900	0.0840	0.0660	0.0800	0.0880	0.0840
Tarone-Ware	0.0940	0.0840	0.0920	0.0700	0.0920	0.0320	0.1080	0.0900	0.0780	0.0820	0.0960	0.0980
Peto-Peto	0.0960	0.0800	0.0960	0.0780	0.0940	0.0320	0.1140	0.1000	0.0800	0.0920	0.1140	0.1060
Mod. Peto-Peto	0.0900	0.0800	0.0980	0.0780	0.0920	0.0300	0.1120	0.1000	0.0780	0.0880	0.1140	0.1060
F-H (1, 0)	0.0960	0.0820	0.0960	0.0800	0.0980	0.0320	0.1160	0.1000	0.0800	0.0940	0.1100	0.1060
F-H (0.5, 0.5)	0.1320	0.1040	0.1080	0.1080	0.1240	0.0400	0.1080	0.0880	0.1420	0.0940	0.1080	0.0960
F-H (1, 1)	0.1380	0.1080	0.0960	0.0960	0.1420	0.0340	0.0980	0.0900	0.1600	0.0940	0.0940	0.0940
F-H (0, 1)	0.1400	0.1020	0.0840	0.0760	0.1540	0.0420	0.0800	0.0740	0.1680	0.0920	0.0900	0.0880
F-H (0.5, 2)	0.1340	0.1020	0.0760	0.0600	0.1340	0.0360	0.0600	0.0720	0.1780	0.0840	0.0840	0.0800

Table 6:   Type I Error Rates of Tests for Lognormal Distribution with Different Parameters

Tests	Lognormal (0, 1)				Lognormal (0, 2)				Lognormal (0, 3)			
	When the Event of Interest Occurs at the Beginning of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0160	0.0200	0.0140	0.0140	0.0180	0.0200	0.0200	0.0120	0.0200	0.0180	0.0140	0.0180
Gehan-Wilcoxon	0.0120	0.0100	0.0140	0.0100	0.0100	0.0100	0.0120	0.0080	0.0100	0.0140	0.0060	0.0140
Tarone-Ware	0.0180	0.0120	0.0140	0.0080	0.0120	0.0180	0.0140	0.0100	0.0120	0.0140	0.0040	0.0140
Peto-Peto	0.0160	0.0120	0.0160	0.0100	0.0140	0.0200	0.0160	0.0100	0.0140	0.0180	0.0040	0.0140
Mod. Peto-Peto	0.0180	0.0120	0.0160	0.0100	0.0120	0.0180	0.0160	0.0100	0.0120	0.0160	0.0040	0.0160
F-H (1, 0)	0.0160	0.0120	0.0140	0.0100	0.0140	0.0200	0.0160	0.0100	0.0140	0.0200	0.0060	0.0140
F-H (0.5, 0.5)	0.0320	0.0240	0.0240	0.0160	0.0220	0.0280	0.0220	0.0260	0.0220	0.0240	0.0160	0.0280
F-H (1, 1)	0.0320	0.0380	0.0280	0.0260	0.0260	0.0360	0.0260	0.0340	0.0280	0.0320	0.0240	0.0420
F-H (0, 1)	0.0380	0.0520	0.0380	0.0280	0.0340	0.0460	0.0420	0.0400	0.0360	0.0360	0.0380	0.0460
F-H (0.5, 2)	0.0400	0.0660	0.0440	0.0480	0.0280	0.0560	0.0600	0.0520	0.0360	0.0500	0.0500	0.0440
Tests	When the Event of Interest Occurs in the Middle of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0120	0.0280	0.0340	0.0200	0.0300	0.0300	0.0300	0.0400	0.0520	0.0280	0.0380	0.0240
Gehan-Wilcoxon	0.0300	0.0600	0.0720	0.0620	0.0400	0.0600	0.0640	0.0740	0.0760	0.0760	0.0820	0.0560
Tarone-Ware	0.0240	0.0380	0.0520	0.0380	0.0380	0.0400	0.0520	0.0540	0.0620	0.0520	0.0560	0.0360
Peto-Peto	0.0240	0.0480	0.0560	0.0480	0.0380	0.0440	0.0620	0.0560	0.0680	0.0560	0.0600	0.0400
Mod. Peto-Peto	0.0240	0.0460	0.0560	0.0480	0.0360	0.0440	0.0620	0.0600	0.0680	0.0560	0.0600	0.0420
F-H (1, 0)	0.0180	0.0460	0.0540	0.0440	0.0360	0.0440	0.0620	0.0560	0.0680	0.0560	0.0600	0.0380
F-H (0.5, 0.5)	0.0220	0.0100	0.0180	0.0120	0.0180	0.0300	0.0180	0.0160	0.0260	0.0120	0.0180	0.0140
F-H (1, 1)	0.0200	0.0060	0.0160	0.0180	0.0180	0.0200	0.0100	0.0160	0.0240	0.0120	0.0160	0.0120
F-H (0, 1)	0.0200	0.0160	0.0160	0.0240	0.0220	0.0300	0.0200	0.0260	0.0280	0.0280	0.0260	0.0140
F-H (0.5, 2)	0.0160	0.0240	0.0280	0.0400	0.0300	0.0240	0.0400	0.0280	0.0260	0.0320	0.0300	0.0280
Tests	When the Event of Interest Occurs at the End of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.1460	0.0980	0.0860	0.1080	0.0980	0.0840	0.1000	0.1260	0.1100	0.0800	0.1120	0.1260
Gehan-Wilcoxon	0.1080	0.0820	0.0920	0.0960	0.0660	0.0860	0.0840	0.1020	0.0820	0.0780	0.1320	0.0920
Tarone-Ware	0.1220	0.0960	0.0980	0.1280	0.0860	0.1000	0.1040	0.1220	0.0960	0.0920	0.1340	0.1100
Peto-Peto	0.1260	0.1040	0.1060	0.1320	0.0880	0.1020	0.1220	0.1260	0.1000	0.0840	0.1380	0.1100
Mod. Peto-Peto	0.1300	0.1040	0.1060	0.1320	0.0800	0.1000	0.1220	0.1280	0.0940	0.0900	0.1380	0.1120
F-H (1, 0)	0.1280	0.1040	0.1040	0.1320	0.0860	0.0980	0.1240	0.1260	0.0980	0.0840	0.1400	0.1100
F-H (0.5, 0.5)	0.1580	0.0980	0.0880	0.0880	0.1100	0.0860	0.0980	0.1120	0.1420	0.0820	0.1100	0.1260
F-H (1, 1)	0.1740	0.0880	0.0920	0.0860	0.1200	0.0840	0.0940	0.1080	0.1420	0.0820	0.0960	0.1180
F-H (0, 1)	0.1720	0.1080	0.0820	0.0720	0.1240	0.0900	0.0840	0.0860	0.1380	0.0840	0.0780	0.0860
F-H (0.5, 2)	0.1620	0.1100	0.0720	0.0680	0.1280	0.0940	0.0760	0.0640	0.1460	0.0780	0.0700	0.0860

Table 7:   Type I Error Rates of Tests for Inverse Gaussian Distribution with Different Parameters

Tests	Inverse Gaussian (0.5, 1)				Inverse Gaussian (0.5, 2)				Inverse Gaussian (0.5, 3)			
	When the Event of Interest Occurs at the Beginning of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0160	0.0160	0.0080	0.0140	0.0220	0.0100	0.0180	0.0240	0.0180	0.0140	0.0060	0.0160
Gehan-Wilcoxon	0.0160	0.0180	0.0080	0.0140	0.0200	0.0120	0.0120	0.0160	0.0120	0.0140	0.0100	0.0060
Tarone-Ware	0.0180	0.0140	0.0060	0.0140	0.0220	0.0120	0.0120	0.0180	0.0140	0.0140	0.0080	0.0060
Peto-Peto	0.0180	0.0180	0.0060	0.0140	0.0240	0.0120	0.0180	0.0180	0.0180	0.0140	0.0080	0.0100
Mod. Peto-Peto	0.0160	0.0140	0.0060	0.0140	0.0240	0.0120	0.0160	0.0180	0.0160	0.0140	0.0080	0.0120
F-H (1, 0)	0.0180	0.0180	0.0040	0.0140	0.0240	0.0120	0.0180	0.0180	0.0180	0.0140	0.0080	0.0100
F-H (0.5, 0.5)	0.0260	0.0140	0.0140	0.0300	0.0300	0.0180	0.0140	0.0280	0.0100	0.0140	0.0100	0.0240
F-H (1, 1)	0.0340	0.0200	0.0240	0.0480	0.0360	0.0300	0.0220	0.0340	0.0160	0.0240	0.0100	0.0320
F-H (0, 1)	0.0480	0.0240	0.0360	0.0540	0.0420	0.0360	0.0400	0.0440	0.0280	0.0340	0.0200	0.0340
F-H (0.5, 2)	0.0500	0.0320	0.0460	0.0680	0.0440	0.0480	0.0600	0.0420	0.0340	0.0420	0.0260	0.0420
Tests	When the Event of Interest Occurs in the Middle of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.0360	0.0400	0.0260	0.0380	0.0340	0.0260	0.0200	0.0280	0.0280	0.0380	0.0360	0.0340
Gehan-Wilcoxon	0.0580	0.0700	0.0680	0.0700	0.0440	0.0620	0.0660	0.0440	0.0540	0.0560	0.0660	0.0880
Tarone-Ware	0.0420	0.0680	0.0400	0.0460	0.0400	0.0420	0.0400	0.0380	0.0420	0.0460	0.0480	0.0660
Peto-Peto	0.0460	0.0700	0.0440	0.0600	0.0420	0.0440	0.0460	0.0380	0.0480	0.0460	0.0480	0.0740
Mod. Peto-Peto	0.0460	0.0700	0.0460	0.0600	0.0460	0.0460	0.0480	0.0380	0.0520	0.0440	0.0500	0.0760
F-H (1, 0)	0.0420	0.0720	0.0440	0.0580	0.0440	0.0420	0.0480	0.0380	0.0480	0.0460	0.0480	0.0720
F-H (0.5, 0.5)	0.0220	0.0360	0.0240	0.0140	0.0220	0.0160	0.0140	0.0140	0.0160	0.0220	0.0100	0.0100
F-H (1, 1)	0.0200	0.0300	0.0180	0.0140	0.0160	0.0120	0.0140	0.0180	0.0180	0.0180	0.0100	0.0060
F-H (0, 1)	0.0200	0.0280	0.0280	0.0200	0.0180	0.0200	0.0160	0.0140	0.0140	0.0340	0.0140	0.0120
F-H (0.5, 2)	0.0200	0.0220	0.0380	0.0280	0.0200	0.0160	0.0360	0.0240	0.0120	0.0360	0.0180	0.0300
Tests	When the Event of Interest Occurs at the End of the Study Period											
	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100	n=10	n=30	n=50	n=100
Logrank	0.1140	0.1020	0.0920	0.1080	0.1140	0.1100	0.0840	0.1120	0.1100	0.1020	0.0840	0.1100
Gehan-Wilcoxon	0.0740	0.0840	0.0820	0.0880	0.0920	0.0880	0.0820	0.0920	0.0740	0.0840	0.0800	0.0980
Tarone-Ware	0.0980	0.0920	0.0880	0.0920	0.0980	0.1100	0.0940	0.0960	0.0920	0.0960	0.0940	0.1020
Peto-Peto	0.0960	0.1000	0.0980	0.1040	0.1080	0.1120	0.0940	0.1100	0.0940	0.1100	0.1020	0.1180
Mod. Peto-Peto	0.0960	0.0940	0.0960	0.1060	0.1060	0.1140	0.0960	0.1120	0.0960	0.1040	0.1040	0.1180
F-H (1, 0)	0.0920	0.1000	0.1000	0.1040	0.1080	0.1160	0.0960	0.1140	0.0960	0.1080	0.1020	0.1180
F-H (0.5, 0.5)	0.1400	0.0980	0.0940	0.1080	0.1380	0.1040	0.0840	0.1160	0.1420	0.1000	0.0840	0.1060
F-H (1, 1)	0.1540	0.0920	0.0880	0.1180	0.1480	0.1020	0.0840	0.1040	0.1540	0.0980	0.0720	0.1080
F-H (0, 1)	0.1520	0.0820	0.0660	0.0800	0.1640	0.0900	0.0980	0.0740	0.1460	0.0940	0.0720	0.0900
F-H (0.5, 2)	0.1400	0.0860	0.1350	0.0740	0.1740	0.0820	0.0880	0.0680	0.1400	0.0840	0.0600	0.0740
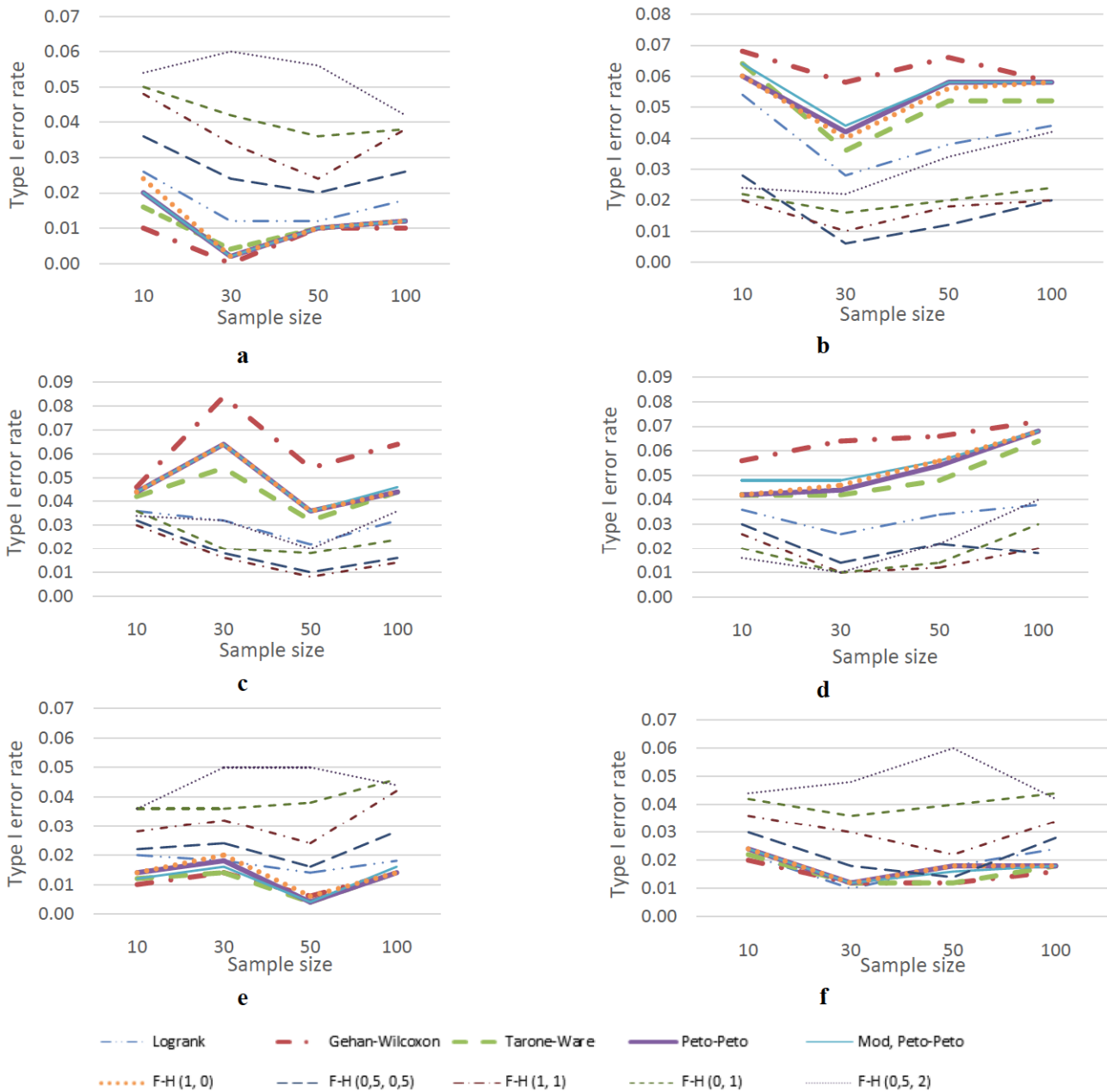
Figure 1: Graphs of type I error rates according to different sample sizes, different distributions and different tests (**a**: Exponential (0.5), event of interest at the beginning, **b**: Weibull (1, 3.5), event of interest in the middle, **c**: Weibull (2, 3.5), event of interest in the middle, **d**: Weibull (3, 3.5), event of interest in the middle, **e**: Lognormal (0, 3), event of interest at the beginning, **f**: Inverse Gaussian (0.5, 2), event of interest at the beginning).

when it is desired to give more weight to events that occur in early or late periods, then the logrank test will not be appropriate anymore [26]. For this reason, a simulation study has been carried out to take into consideration situations where the event of interest is intensively located at the beginning, middle or end of the follow-up period.

In some cases, for example, when a treatment is expected to reduce the risk, the risk of late onset is insignificant and can be ignored. In such cases the

need for the use of tests which give priority to the events taking place at the beginning arises and the Gehan-Wilcoxon and Tarone-Ware tests can be used [24-26]. Likewise, the Peto-Peto test and the Fleming-Harrington (1,0) test of the Fleming-Harrington group focus on the initial events. For values of (p, q) in the Fleming-Harrington test family; while the tests with (0.5, 0.5) and (1, 1) focus on the events that occur in the middle; the tests with (0,1) and (0.5, 2) focus on events occurring late in the follow-up period.

When the simulation results are examined, the type I error rate was found to be lower than the nominal value in almost all of the distributions and tests considered when we worked on conditions where the event of interest was intense at the beginning of the study. Since the Fleming-Harrington test family focuses on events occurring in the middle and at the end of the follow-up time, it may be normal to obtain such results for this test group [9-10]. In our study, for tests known to give weight to the initial events such as Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, Modified Peto-Peto and Fleming-Harrington (1,0), the type I error rate was found to be lower than the nominal value, in contrast to the studies of Harrington and Fleming (1982) and Lee (1996) [9-10].

According to the simulation results for the condition that the event of interest was placed in the middle of the follow-up period intensively; type I error rates for tests except Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, Modified Peto-Peto and Fleming-Harrington (1,0) tend to be lower than the nominal value. The type I error rates are lower than the nominal value in the Fleming-Harrington tests, which are known to give weight to the events in the middle of the follow-up period, when the p and q values are equal [10, 27].

When we examine the results of the simulation under the condition that the event of interest was intensively present at the end of the follow-up period; type I error rates were found to be higher than the nominal value in almost all distributions and in all tests. In Lee's study, it is stated that Fleming-Harrington tests for p < q give more weight to late events [10]. In our simulation study, the results supporting this information were found only for Fleming-Harrington (0.5, 2) test and exponential distribution, and for Fleming-Harrington (0,1) and Fleming-Harrington (0.5, 2) tests and Weibull distribution with (2, 1.5) and (2, 3.5) parameters.

When the event of interest was seen intensely at the beginning, the type I error rate tended to be lower than the nominal value in all distributions; if the event of interest was present at the end period the type I error rate tended to be greater than the nominal value in all distributions. Type I error rates, which are closest to the nominal value, were found for the Weibull distribution when the deaths were seen mostly in the middle of the follow-up period. This may be explained by the fact that the Weibull distribution is a flexible distribution that is compatible with many survival data (Figure **1b**, **c**, **d**).

In addition, when the event of interest was intensely present in the middle of the follow-up period the type I error rate was generally higher than the nominal value in the Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, Modified Peto-Peto and Fleming-Harrington (1,0) tests which focus mainly on the events occurring in the beginning, and it was lower in other tests. The cases in which the event of interest occurred in different parts of the follow-up period, and the tests in which the nominal value near Type I error rate was obtained in all distributions from which survival data are derived, are, respectively; the Fleming-Harrington (.5,2) test, Tarone-Ware test, Peto-Peto test and Fleming-Harrington (1,0) test which is equivalent to the Peto-Peto test. However, the most commonly used logrank test in the literature did not perform very well in all cases, especially if the deaths are not evenly distributed throughout the follow-up period. This is a result that emphasizes the careful use of the logrank test, which is often used to compare survival curves in survival studies.

5. CONCLUSIONS

In the light of these results, it is proposed to re-examine the tests for cases where events are observed intensively at the beginning, middle and late periods, to carry out new simulation studies and to develop new tests if necessary.

REFERENCES

[1]   Fisher LD, Belle GV. Biostatistics, a methodology for the health sciences, John Wiley & Sons Inc, New York, page 1993; 786-807.

[2]   Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 1958; 53(282): 457-481.
https://doi.org/10.1080/01621459.1958.10501452

[3]   Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of National Cancer Institute 1959; 22(4): 719-748.

[4]   Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Reports, 1966; 50(3): 163-170.

[5]   Gehan EA. A generalized Wilcoxon test for comparing arbitrarily single-censored samples. Biometrika 1965; 52: 203-223.
https://doi.org/10.1093/biomet/52.1-2.203

[6]   Peto R, Peto J. Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society 1972; 135(2): 185-207.
https://doi.org/10.2307/2344317

[7]   Tarone RE, Ware J. On distribution-free tests for equality of survival distributions. Biometrika 1977; 64(1): 156-160.
https://doi.org/10.1093/biomet/64.1.156

[8]   Fleming TR, Harrington DP. A class of hypothesis tests for one and two samples censored survival data. Communications in Statistics-Theory and Methods 1981; 10(8): 763-794.
https://doi.org/10.1080/03610928108828073

[9]   Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. Biometrika 1982; 69(3): 553-566.
https://doi.org/10.1093/biomet/69.3.553

[10]   Lee JW. Some versatile tests based on the simultaneous use of weighted log-rank statistics. Biometrics 1996; 52(2): 721-725.
https://doi.org/10.2307/2532911

[11]   Martinez RLMC, Naranjo JD. A pretest for choosing between logrank and Wilcoxon tests in the two-sample problem. Metron: International Journal of Statistics 2010; 68(2): 111-125.
https://doi.org/10.1007/BF03263529

[12]   Graves TS, Pazdan JL. A permutation test analogue to Tarone's test for trend in survival analysis. Journal of Statistical Computation and Simulation 1995; 53(1-2): 79-89.
https://doi.org/10.1080/00949659508811697

[13]   Leton E, Zuluaga P. Relationships among tests for censored data. Biometrical Journal 2005; 47(3): 377-387.
https://doi.org/10.1002/bimj.200410115

[14]   Karadeniz PG, Ercan I. Examining Tests for Comparing of Survival Curves with Right Censored Data. Statistics in Transition 2017; 18(2): 311-328.
https://doi.org/10.21307/stattrans-2016-072

[15]   Latta RB. A monte carlo study of some two-sample rank tests with censored data. Journal of American Statistical Association 1981; 76(375): 713-719.
https://doi.org/10.1080/01621459.1981.10477710

[16]   Hintze JL. NCSS user guide V tabulation, item analysis, proportions, diagnostic tests, and survival / reliability, Published by NCSS, Kaysville, Utah, 2007.

[17]   Oller R, Gomez G. A generalized Fleming and Harrington's class of tests for interval-censored data. The Canadian Journal of Statistics 2012; 40(3): 501-516.
https://doi.org/10.1002/cjs.11139

[18]   R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

[19]   Hintze, J. (2007). NCSS 2007. NCSS, LLC. Kaysville, Utah, USA, 2007. www.ncss.com

[20]   WinAutomation Software Solutions, Softotomotive Ltd, Athens, 2014.

[21]   Lee ET, Wang JW. Statistical Methods for Survival Data Analysis. New Jersey: John Wiley & Sons Inc, 2003.

[22]   Fleming TR, Harrington DP, O'Sullivan M. Supremum versions of the log-rank and generalized Wilcoxon statistics. Journal of the American Statistical Association 1987; 82(397): 312-320.
https://doi.org/10.1080/01621459.1987.10478435

[23]   Buyske S, Fagerstrom R, Ying Z. A class of weighted log-rank tests for survival data when the event is rare. Journal of the American Statistical Association 2000; 95(449): 249-258.
https://doi.org/10.1080/01621459.2000.10473918

[24]   Pepe MS, Fleming TR. Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data. Biometrics 1989; 45(2): 497-507.
https://doi.org/10.2307/2531492

[25]   Kleinbaum DG, Klein M. Survival Analysis a Self-Learning Text. New York: Springer, 2005.

[26]   Klein JP, Rizzo JD, Zhang MJ, Keiding N. Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part I: Unadjusted analysis. Bone Marrow Transplantation 2001; 28(10): 909-915.
https://doi.org/10.1038/sj.bmt.1703260

[27]   Gomez G, Calle ML, Oller R, Langohr K. Tutorial on methods for interval-censored data and their implementation in R. Statistical Modelling 2009; 9(4): 259-297.
https://doi.org/10.1177/1471082X0900900402