

Statistical Powers of Some Tests for Checking Homogeneity of Survival Distributions with Disjointed Ends in the Presence of Censoring

Babalola Bayowa Teniola^{1,2*}, Adeleke Raphael Ayantunji², Halid Omobolaji Yusuf², Olubiyi Adenike Olufunmilola², Ogunsakin Ropo Ebenezer³, Adigun Kehinde Abimbola⁴, Adejuwon Samuel Oluwaseun⁴, Adarabioyo Mumini Idowu⁴, Ogunboyo Ojo Femi⁵, Fadugba Sunday Emmanuel⁶, Egbon Osafu Augustine⁷, Akinyemi Oluwadare², Ogunwale Olukunle Daniel², Faweya Olanrewaju², Kawiso Martin¹

¹Department of Mathematics and Statistics, Kampala International University, Uganda

²Department of Statistics, Ekiti State University, Ado-Ekiti, Nigeria

³Biostatistics, Discipline of Public Health Medicine, Howard College, University of KwaZulu-Natal, South Africa

⁴Department of Mathematical and Physical Science, Afe Babalola University, Ado-Ekiti, Nigeria

⁵Department of Epidemiology and Biostatistics, University of Medical Sciences, Ondo, Nigeria

⁶Department of Mathematics, Ekiti State University, Ado-Ekiti, Nigeria

⁷Institute of Mathematics and Computer Science, University of Sao Paulo, Sao Carlos, Brazil

**Corresponding Author: bayowa.babalola@kiu.ac.ug*

ABSTRACT. This paper considered the comparison of some tests for assessing the overall homogeneity of Kaplan-Meier survival curves under low and high censoring rates when the curves are disjointed towards the end. The performances of these tests were measured by their statistical powers. Monte Carlo simulation study was conducted to evaluate and numerically compare the relative performances of Log-rank, Wilcoxon, Tarone-Ware, Peto-Peto, Modified Peto-Peto, the Fleming-Harrington (1,1), and the Babalola-Adeleke tests. The result obtained shows that the Babalola-Adeleke and Fleming-Harrington (1,1) tests have more

Received: Jun. 20, 2022.

2010 Mathematics Subject Classification. 65C20.

Key words and phrases. survival analysis; Kaplan-Meier; survival curves; statistical power; censoring.

robust performances than the other five popular tests with relatively high power in detecting differences when the censoring rates in the groups are both low and high. The highest overall average powers under low and high censoring rates were produced by Babalola-Adeleke and Fleming-Harrington (1,1) tests respectively. Hence, these two tests are the most suitable tests for diagnosing homogeneity of survival curves under these conditions.

1. Introduction

The rate at which survival analysis is advancing and gaining popularity in every field of study is pretty impressive. The nature of data obtained in the area of Biostatistics has necessitated the growth in the volume of works done in the survival analysis [1-5]. Survival analysis is also of massive use in Engineering and Social sciences fields [6-8]. A very predominant method in Survival analysis is Kaplan-Meier method, which is capable of estimating the survivorship function for different sample sizes. Several scholars have established its huge efficiency in capturing necessary survival details in cohort studies and otherwise. The Kaplan-Meier estimator is a nonparametric method that allows for the incorporation of censoring for the purpose of estimation of probabilities of survival [9-12]. More related and relevant research works have also been reported in the literature.

The log-rank test is arguably the most popular test in testing for homogeneity of survival distribution. However, it may fail to recognize some crucial differences that exist among groups whereby the main difference takes place very early in the study or towards the end of the study [13]. This is because it was proposed in order to give equal weight to all failures among the follow-up [14]. The shortfall of the log-rank test is in the assumption that the hazard ratio of the groups should be proportional along the follow-up period as that is the only condition that makes the test superior to others [15-17]. When this assumption is not met, that is when the hazard ratio is non-constant, the Gehan-Wilcoxon and Tarone-Ware tests can be more powerful than the log-rank test [18,19]. The Peto-Peto test is also efficient when the proportional hazard assumption is violated [10]. The strength of the Fleming-Harrington tests (F-H) is in its flexibility. Unlike the other tests, it allows for the choice of weights and focuses on crossing the hazard ratios of groups [19]. Different combinations of the weight, therefore, yield different tests entirely.

[20] compared the statistical powers of some nonparametric tests and concluded that the Peto-Prentice generalized Wilcoxon statistic performed best under the investigated situation. [15,21] examined the properties of the tests based on linear rank statistics and the effect of unequal censoring

by using various combinations of censoring proportions, respectively. In the paper, Wilcoxon test had the lowest relative power of all tests examined. [22,23] and [24] were interested in the comparison of the Wilcoxon and the Log-rank tests under different scenarios. [25] added more tests, which are the Tarone-Ware, Peto-Peto, and F-H tests to the comparison of the log-rank and Wilcoxon tests when the sample size is quite small. It was concluded in the paper that the choice of weight function has a tremendous impact on the power of the tests under any given situation. The importance of simulations and Monte Carlo methods in modern research were the focus of [26]. [27] proposed a modified one-sample log-rank test, and a sample size formula was derived based on its exact variance to provide a study design that preserves the type I error. [28] discussed the versatile tests for comparing survival curves based on weighted log-rank statistics. [29] proposed a nonparametric test for the comparison of survival curves using the median. [30] examined the tests for comparing survival curves with right-censored data. In the study, the type I error rate of Logrank test was equal or close to the nominal value.

[31] developed a new method and demonstrated that this method outclassed some existing methods and relatively performed better under low and high censoring rates when the Kaplan-Meier survival curves are proportional. It was also ascertained that when there are crossing survival curves, the powers of the tests are relatively low since none of the tests gave statistical power in close of one. Other relevant works on censoring and other methodologies are [1],[32-38].

Thus, this paper considers a typical situation whereby the survival curves of the two groups are similar at the beginning of the study but gradually diverged towards the end. The censoring rates were categorized into two parts (low and high censoring rates). The censoring times among the groups were carefully chosen to fit into the intended survival pattern. All survival times were simulated from an exponential distribution. The outcome of this study will assist researchers as a further guide for their choice of tests when survival curves are disjointed towards the end.

Hence, the novelty of this study would be in comparing the relatively new Babalola-Adeleke test with some of the popular methods for checking homogeneity of Kaplan-Meier survival curves with disjointed ends under both high and low censoring rates. It is expected that the findings of this study would help the users of survival analysis as it will certainly further expose to them performances of the tests under consideration. It will also guide in decision making when confronted with the choosing of the most appropriate test to detect differences in survival curves

with disjointed ends. To the best of our knowledge, this is the first study that would compare Babalola-Adeleke test with others under this particular situation.

2. Methodology

Given that there are two groups, that is, groups 1 and 2, where the survival times were observed and recorded as t_j . The number of observed failures (death) in group 1 and group 2 being m_{1j} and m_{2j} respectively, the number not experiencing the event of interest being $n_{1j} - m_{1j}$ and $n_{2j} - m_{2j}$ for group 1 and group 2 respectively, and the number at risk is $n_j = n_{1j} + n_{2j}$

Table 1. Table used for test of equality of the survivorship function in two groups at observed survival time t_j

Event/Group	1	2	Total
Number of death	m_{1j}	m_{2j}	$m_{1j} + m_{2j}$
Number not dying	$n_{1j} - m_{1j}$	$n_{2j} - m_{2j}$	$n_{1j} + n_{2j} - m_{1j} - m_{2j}$
Number at risk	n_{1j}	n_{2j}	$n_j = n_{1j} + n_{2j}$

The various multiple-group versions of the two-group test statistic is obtained by computing a weighted difference between the observed and the expected numbers of events. Table 2 presents a K groups pattern for the test of equality.

Table 2. Table used for test of equality of the survivorship function in K groups at observed survival time t_j

Event/Group	1	2	...	k	...	K	Total
Number of death	m_{1j}	m_{2j}	...	m_{kj}	...	m_{Kj}	m_j
Number not dying	$n_{1j} - m_{1j}$	$n_{2j} - m_{1j}$...	$n_{kj} - m_{kj}$...	$n_{2j} - m_{Kj}$	$n_j - m_j$
Number at risk	n_{1j}	n_{2j}	...	n_{kj}	...	n_{Kj}	n_j

where, $m_j = m_{1j} + m_{2j} + \dots + m_{Kj}$

$n_j = n_{1j} + n_{2j} + \dots + n_{Kj}$

$$n_j - m_j = n_{1j} + n_{2j} + \dots + n_{Kj} - m_{1j} - m_{2j} - \dots - m_{Kj}$$

Based on the argument above, the test hypothesis considered is:

$$H_0 : S_1(t) = S_2(t)$$

$$H_1 : S_1(t) \neq S_2(t)$$

For the test statistics of the tests, see: [39-42] and [8]. The tests are based on some assumptions namely: censoring is unrelated to prognosis; the survival probabilities are equal for subjects recruited early and late in the study; the events happened at the times specified.

Simulation Study

The use of simulation study for the examination of statistical powers of tests under a variety of situations is a popular concept which is well reported in the literature. Over the years, Monte-Carlo simulations have been employed for testing heterogeneity of survival distributions when the proportional hazard assumption is satisfied and when it is not. Therefore, a Monte Carlo simulation to compare the statistical power of the Log-rank, Wilcoxon, Tarone-Ware, Peto-Peto, Modified Peto-Peto, Fleming-Harrington(1,1), and Babalola-Adeleke tests was conducted. It is a known fact that due to the flexibility of the Fleming-Harrington test, there are several options for its weights. Hence, for the purpose of placing weights of hazard in the middle, Fleming-Harrington (1,1) was selected since every other test either places equal weight across the board or places more weight at the beginning or towards the end. Figure 1 shows the survival curves of two groups that have a similar pattern for some time but have a disjointed end. Therefore, all the simulated datasets followed this pattern.

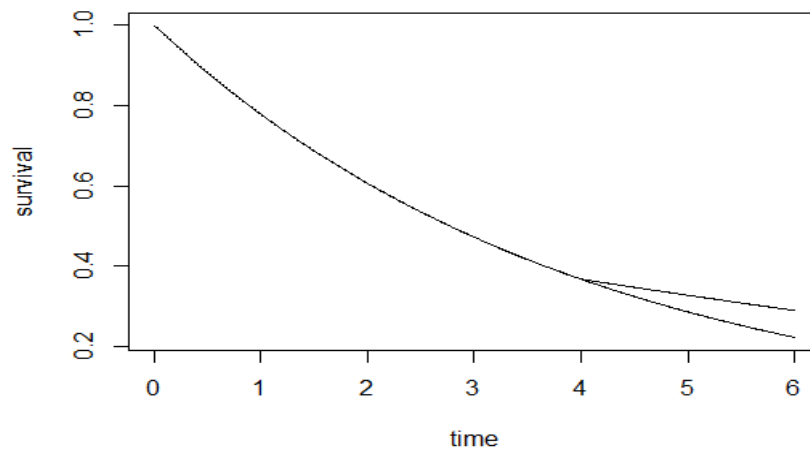


Figure 1. Figure of the Situation for consideration in the simulation study

For each of the combination of the sample sizes, 5000 iterations were simulated in order to obtain statistically viable powers of the aforesaid tests. Since the larger the number of iterations, the better the result. The estimated statistical power was obtained as the proportion of 5000 repeated random samples where the hypothesis of no difference in the survival curves (null hypothesis) at the 0.05 significance level is correctly rejected.

3. Results

Considering the sub-situation with low censoring rates in both groups, the survival times in Group 1 follow an exponential distribution with a mean of 4 (rate 0.25), and in Group 2, the survival times follow an exponential distribution with mean 4(rate 0.25) as well. In order to get disconnected survival curves towards the end, if the survival time in Group 2 is greater than or equal to 4, then the survival time is automatically simulated from an exponential distribution with a mean 40(rate 0.025). In order to have low censoring rates in the two groups, if the survival time is greater than the maximum survival time divided by 1.25 into both groups, then the observation was censored. These yielded an overall average censoring rate of 4.50% and 9.99% in Groups 1 and 2, respectively. Table 3 displays the result of the powers of the seven tests obtained from the simulation conducted for this sub-situation under low censoring rates alongside the censoring rates. The censoring rates in both groups decrease as the sample sizes increase. The same trend is also exhibited in mixed sample sizes. Table 3. Powers of the tests and censoring rates for the Situation (low censoring rates)

Sample size	Log-rank	Wilcoxon	Tarone - Ware	Peto-Peto	Modified Peto-Peto	Fleming-Harrington	Babalola-Adeleke	Censoring (%)	rates
20,20	0.0798	0.0698	0.0732	0.0698	0.0688	0.1016	0.0810	8.3840	11.0760
40,40	0.1824	0.0878	0.1144	0.0884	0.0872	0.1906	0.1884	4.8805	9.7795
50,50	0.2386	0.1082	0.1428	0.1072	0.1062	0.2360	0.2462	4.0228	9.6964
60,60	0.2890	0.1142	0.1554	0.1132	0.1126	0.2706	0.2976	3.4980	9.7923
80,80	0.3914	0.1404	0.2096	0.1390	0.1386	0.3480	0.3998	2.8098	9.5333
100,100	0.4508	0.1610	0.2344	0.1578	0.1578	0.3916	0.4586	2.3252	9.6056
20,50	0.0784	0.0674	0.0668	0.0670	0.0674	0.1082	0.0796	8.5350	9.6608
50,20	0.1880	0.0852	0.1170	0.0836	0.0834	0.1756	0.1966	4.1364	11.0430
50,100	0.2690	0.1104	0.1516	0.1056	0.1054	0.2908	0.2774	4.0152	9.5900
100,50	0.3890	0.1374	0.2016	0.1346	0.1356	0.3218	0.3976	2.3938	9.5516

From Table 3, it is evident that the powers of all the tests increase as the sample size increase as the highest powers recorded for all the tests is obtained at sample size (100,100). The Babalola-Adeleke test has the highest power at the largest equal sample size, with a value of 0.4586. The Babalola-Adeleke test outperforms all the other tests at all sample sizes except when the sample sizes were (20,40), (40,40) and (20,50) for the Fleming-Harrington test. The Peto-Peto and the Modified Peto-Peto produced similar results under this Situation with just small differences in the powers of the two tests across all the sample sizes, which is not statistically significant judging by student t-test. However, the Peto-Peto test still outperforms the Modified Peto-Peto under equal sample sizes. The statistical description of Table 3 is given in Table 4.

Table 4. Descriptive statistics of the power of the tests for the Situation (low censoring rates)

	Log-rank	Wilcoxon	Tarone- Ware	Peto- Peto	Modified Peto-Peto	Fleming- Harrington	Babalola- Adeleke
Mean	0.2556	0.1082	0.1466	0.1066	0.1063	0.2435	0.2622
Standard Error	0.0406	0.0099	0.0178	0.0096	0.0097	0.0312	0.0413
Median	0.2538	0.1093	0.1472	0.1064	0.1058	0.2533	0.2618
Standard Deviation	0.1283	0.0313	0.0563	0.0304	0.0306	0.0988	0.1306
Kurtosis	-1.0388	-0.9181	-0.9519	-0.9273	-0.9537	-1.1006	-1.0237
Skewness	0.0424	0.2763	0.0984	0.2998	0.3179	-0.1156	-0.0025
Range	0.3724	0.0936	0.1676	0.0908	0.0904	0.2900	0.3790
Minimum	0.0784	0.0674	0.0668	0.0670	0.0674	0.1016	0.0796
Maximum	0.4508	0.1610	0.2344	0.1578	0.1578	0.3916	0.4586

Table 4 shows that the Babalola-Adeleke test has the highest mean of 0.2622 as the average power of the method across all the combinations of sample sizes and the standard error of 0.0413. This is followed by the Log-rank test with an average statistical power of 0.2556 with a standard error of 0.0406, while the Modified Peto-Peto test resulted in the lowest average statistical power 0.1063 with standard error 0.0097. The descriptive statistics of the Modified Peto-Peto and Peto-Peto tests are similar. The median powers for the tests arranged in descending order are 0.2618, 0.2538, 0.2533, 0.1472, 0.1093, 0.1064, and 0.1058, which are results of the Babalola-Adeleke test, Log-rank, Fleming-Harrington, Tarone-Ware, Wilcoxon tests, Peto-Peto and Modified Peto-Peto, respectively.

For skewness, the result shows that the power of all the tests is positively skewed except for the Babalola-Adeleke test and Fleming-Harrington test, which indicates that both the mean and the median are less than the mode of the powers of the tests. The negative values of the Kurtosis indicate that the distribution of the powers has lighter tails and a flatter peak than the normal distribution.

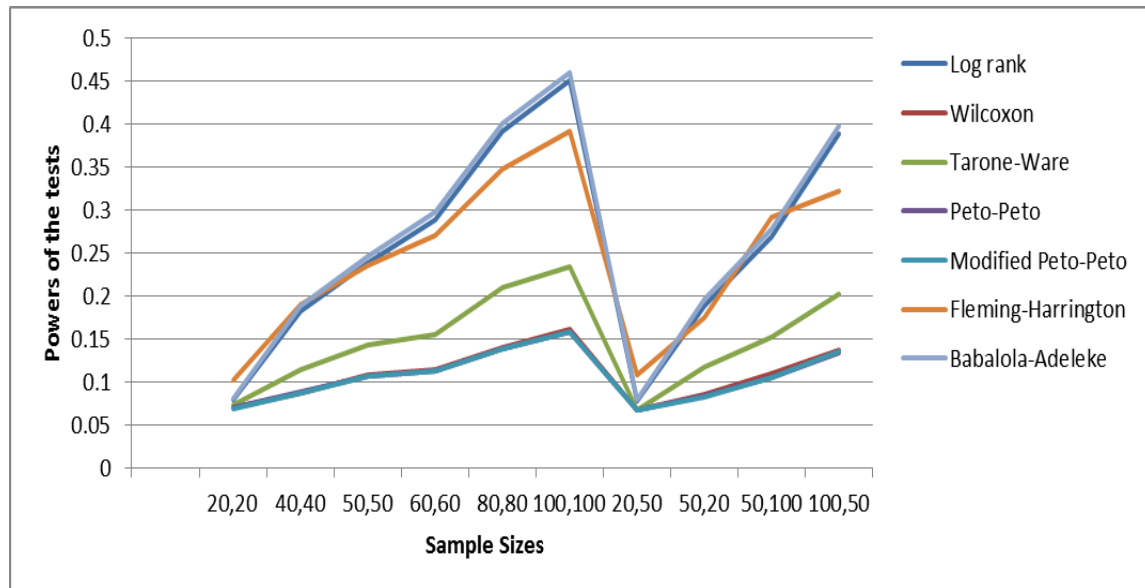


Figure 2. A chart showing the statistical powers of the tests under the Situation with low censoring rates

3.1 The situation with high censoring rates

In the presence of high censoring rates in both groups, the survival times in Group 1 follow an exponential distribution with a mean of 4 (rate 0.25), and in Group 2, the survival times follow an exponential distribution with mean 4(rate 0.25) as well. In order to get disconnected survival curves towards the end, if the survival time in Group 2 is greater than or equal to 4, then the survival time is automatically simulated from an exponential distribution with a mean 40(rate 0.025). Additionally, in order to have high censoring rates in both groups, if the survival time is greater than the minimum survival time plus two. That is, (the minimum survival time in both groups +2), then the observation was censored. These yielded an overall average censoring rate of 59.3096% and 55.6807% in Groups 1 and 2, respectively. These censoring rates are quite high since more than half of the cohorts in both groups censored. The result of the powers of the tests when there are high censoring rates is displayed in Table 5. Unlike the first sub-situation with low censoring rates, the censoring rates in both groups increase with sample size.

Table 5. Powers of the tests and censoring rates for the Situation (High censoring rates)

Sample size	Log-rank	Wilcoxon	Tarone-Ware	Peto-Peto	Modified Peto-Peto	Fleming-Harrington	Babalola-Adeleke	Censoring rates(%)	
20,20	0.0546	0.0526	0.0530	0.0520	0.0526	0.0692	0.0554	57.8950	53.9850
40,40	0.0724	0.068	0.0712	0.0664	0.0664	0.0834	0.0722	59.2980	55.7400
50,50	0.0796	0.0802	0.0806	0.0784	0.0786	0.0918	0.0796	59.3356	56.0164
60,60	0.0964	0.0924	0.0916	0.0902	0.0902	0.1004	0.0964	59.7593	56.1180
80,80	0.1118	0.1056	0.1098	0.1038	0.1038	0.1178	0.1118	59.8665	56.3325
100,100	0.1264	0.1230	0.1250	0.1180	0.1180	0.1378	0.1264	59.9832	56.4456
20,50	0.0576	0.0592	0.0584	0.0576	0.0574	0.0680	0.0576	57.9630	55.7360
50,20	0.0786	0.0704	0.0750	0.0698	0.0700	0.1068	0.0788	59.4056	53.8740
50,100	0.0968	0.0934	0.094	0.0890	0.0892	0.0958	0.0966	59.5496	56.5508
100,50	0.1092	0.1022	0.1048	0.1002	0.1000	0.1192	0.1092	60.0398	56.0084

Generally, the powers of all the tests are low. Even at that, the Fleming-Harrington still outperforms the other tests. As expected, the powers increase as the sample sizes increase. This could indicate that at much larger sample sizes, the powers of the tests could attain higher values than the ones reported.

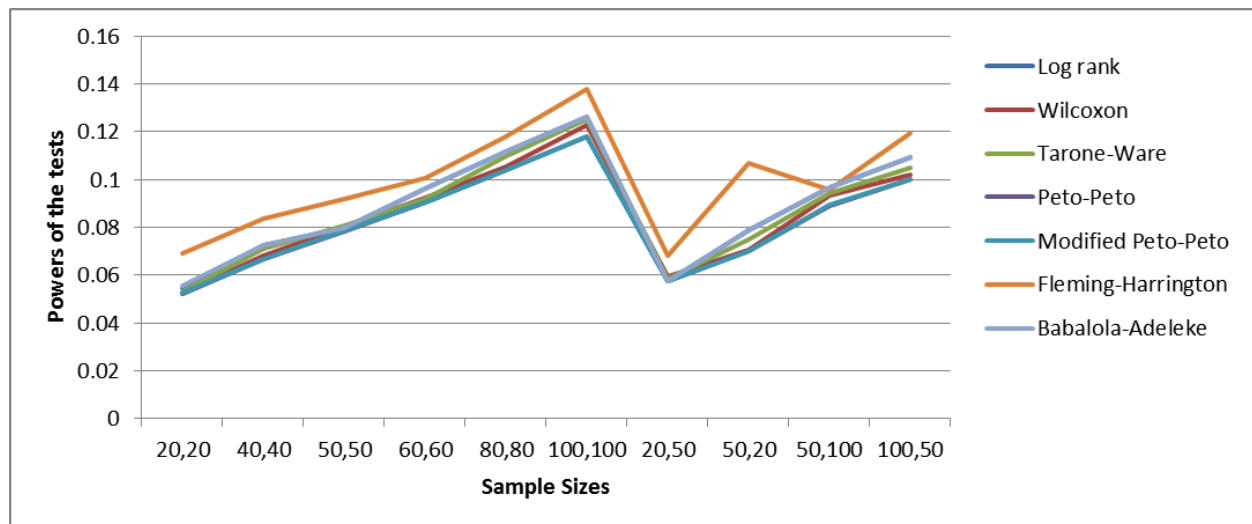


Figure 3. A chart showing the statistical powers of the tests under the Situation with high censoring rates

Figure 3 above further reiterates the outstanding performance of the Fleming-Harrington test under this Situation and censoring rates. It apparently outclasses all the other tests when the sample sizes are the same in the two groups. The value of its power is only in the range of the other tests when the sample size is 50 in the first group and 100 in the second group. In any other sample size, it outperforms all the other tests.

Table 6. Descriptive statistics of the power of the tests for the Situation (High censoring rates)

	Log- rank	Wilcoxon	Tarone- Ware	Peto- Peto	Modified Peto-Peto	Fleming- Harrington	Babalola-A deleke
Mean	0.0883	0.0847	0.0863	0.0825	0.0826	0.0990	0.0884
Standard Error	0.0075	0.0071	0.0073	0.0068	0.0067	0.0071	0.0075
Median	0.0880	0.0863	0.0861	0.0837	0.0839	0.0981	0.0880
Standard Deviation	0.0238	0.0224	0.0230	0.0214	0.0213	0.0223	0.0236
Kurtosis	-1.0034	-0.8489	-0.7644	-0.9598	-0.9645	-0.5284	-1.0139
Skewness	0.0534	0.1778	0.1554	0.1432	0.1486	0.1630	0.0713
Range	0.0718	0.0704	0.0720	0.0660	0.0654	0.0698	0.0710
Minimum	0.0546	0.0526	0.0530	0.0520	0.0526	0.0680	0.0554
Maximum	0.1264	0.1230	0.1250	0.1180	0.1180	0.1378	0.1264

From Table 6, the Fleming-Harrington test has the highest mean of 0.0990 as the average power of the method across all the combinations of sample sizes and the standard error of 0.0071. This is followed by the Babalola-Adeleke test with an average statistical power of 0.0884 with a standard error of 0.0075, while the Peto-Peto test resulted in the lowest average statistical power 0.0825 with standard error 0.0068. As in the case of low censoring rates in this Situation, the descriptive statistics of the Modified Peto-Peto and Peto-Peto tests are similar. However, the Modified Peto-Peto performs better than Peto-Peto under the condition. The median powers for the tests arranged in descending order are 0.0981, 0.0880, 0.0880, 0.0863, 0.0861, 0.0839, and 0.0837, which are results for Fleming-Harrington, Babalola-Adeleke, Log-rank, Wilcoxon, Tarone-Ware, Modified Peto-Peto, and Peto-Peto, respectively. For skewness and Kurtosis, the result shows that the power of all the tests is positively skewed with negative Kurtosis.

3.2 Application of the tests to real-life data

Survival in patients with Acute Myelogenous Leukemia was studied with the interest of knowing the impact of the standard course of chemotherapy extension [43,44]. The variables in the study were

time, which is the survival or censoring time, and event (recurrence of AML cancer) is indicated by the variable "status" 1 = event (recurrence) and 0 = no event (censored). The treatment group was represented by the variable "x", which indicates if maintenance chemotherapy was given (Maintained) or not (Non-maintained).

This is a popular data set with 8.33% patients censored in group 1(maintained) and 36.36% in the second group (non-maintained). The property of this data set is "slightly" similar to the situation under study as the survival curves have a similar pattern from the beginning of the study till about the week 45(though not exactly the same form from the beginning). Then homogeneity of the survival curves can be investigated. This is the closest real-life data we have at our disposal for the situation under study.

The test hypothesis is:

$$H_0 : S_{\text{Maintained}}(t) = S_{\text{Nonmaintained}}(t)$$

$$H_1 : S_{\text{Maintained}}(t) \neq S_{\text{Nonmaintained}}(t)$$

Table 7. Comparison of the results of the different tests using the Acute Myelogenous Leukemia

Method	Log-rank	Wilcoxon	Tarone-Ware	Peto-Peto	Modified Peto-Peto	Fleming-Harrington	Babalola-Adeleke
χ^2 - value	3.3964	2.7233	2.9816	3.5880	3.5670	1.4310	3.6236
p-value	0.0654	0.0988	0.0842	0.0582	0.0590	0.2316	0.0570

Table 7 clearly shows that all the tests validate that the Kaplan-Meier survival curves of those who were maintained and those who were not maintained are not significantly different as none of the p-values is less than 0.05. All the tests yielded very low chi-squared values. This result is consistent with the results earlier reported.

4. Conclusion

Generally, the powers of all the tests are low. Even at that, the Fleming-Harrington still outperforms the other tests. The powers increase as the sample sizes increase. This could indicate that at much larger sample sizes, the powers of the tests could attain higher values than the ones reported. A general comment about this situation, that is when the survival curves are separate towards the end is that, the powers of the tests are also low as expected. This means that it is quite difficult for the different tests to correctly diagnose survival curves because of the similarity of the curves for a larger part of the study (not until towards the end of the study). The low values of the

power are expected, and it has been reported by other researches as well. Generally, across all the sample sizes, the overall average of the power of the entire tests combined is lower when dealing with high censoring rates (0.0874) than when dealing with lower censoring rate (0.1756).

Authors' Contributions

B.T. conceived the idea presented. B.T., O.A. developed the theory and performed the computations. B.T., O.Y., A.O., O.D., O.F, and R.E. verified the analytical methods. B.T., K.A., O., O., S.O., M.I., M., and S.E. wrote the manuscript with input from all authors. R.A. and O.Y. supervised the findings of this paper. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] B.T. Babalola, R.E. Ogunsakin, O.A. Egbon, et al. A Simulation Based Comparative Study of Some Tests for Checking Homogeneity of Non-Crossing Survival Curves Under High Censoring Rates, *J. Appl. Probab. Stat.* 17 (2022), 87-99.
- [2] J.C. Goldsack, A. Coravos, J.P. Bakker, et al. Verification, Analytical Validation, and Clinical Validation (V3): The Foundation of Determining Fit-For-Purpose for Biometric Monitoring Technologies (BioMeTs), *Npj Digit. Med.* 3 (2020), 55. <https://doi.org/10.1038/s41746-020-0260-4>.
- [3] J. Lepš, P. Šmilauer, *Biostatistics with R: An Introductory Guide for Field Biologists*, Cambridge University Press, Cambridge, (2020).
- [4] K. Sumathi, D. Balakrishnan, V. Naveen, et al. Talent Flow Employee Analysis Based Turnover Prediction on Survival Analysis, *Ann. Roman. Soc. Cell Biol.* 25 (2021), 3844-3857.
- [5] T. Saegusa, Z. Zhao, H. Ke, et al. Detecting Survival-Associated Biomarkers From Heterogeneous Populations, *Sci Rep.* 11 (2021), 3203. <https://doi.org/10.1038/s41598-021-82332-y>.
- [6] Z. Cai, Y. Wang, H. Cao, et al. Life Prediction of Self-Locking Nut for Aeroengine Based on Survival Analysis and Bayesian Network, in: *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, IEEE, Singapore, Singapore, 2020: pp. 414–418. <https://doi.org/10.1109/IEEM45057.2020.9309737>.
- [7] S. Nematollahi, S. Nazari, Z. Shayan, et al. Improved Kaplan-Meier Estimator in Survival Analysis Based on Partially Rank-Ordered Set Samples, *Comput. Math. Methods Med.* 2020 (2020), 7827434. <https://doi.org/10.1155/2020/7827434>.

- [8] B.T. Babalola, W.B. Yahya, Effects of Collinearity on Cox Proportional Hazard Model With Time Dependent Coefficients: A Simulation Study, *J. Biostat. Epidemiol.* 5(2020), 172-182.
<https://doi.org/10.18502/jbe.v5i2.2348>.
- [9] E. Ilker, A. Sulaiman, A. Rukayya, The Kaplan Meier Estimate in Survival Analysis, *Biometrics Biostat. Int. J.* 5 (2017), 00128. <https://doi.org/10.15406/bbij.2017.05.00128>.
- [10] D.G. Kleinbaum, M. Klein, *Survival Analysis a Self-Learning Text*, Springer, New York, 44-66, (2005).
- [11] C.J. Pelz, J.P. Klein, *Analysis of Survival Data: A Comparison of Three Major Statistical Packages (SAS, SPSS and BMDP)*. Working paper (Medical College of Wisconsin, Milwaukee). Rep.17: 1-6. (1996).
<https://www.mcw.edu/-/media/MCW/Departments/Biostatistics/tr017.pdf>.
- [12] E.L. Kaplan, P. Meier, Nonparametric Estimation From Incomplete Observations, *J. Amer. Stat. Assoc.* 53 (1958), 457-481.
- [13] J. Klein, J. Rizzo, M.-J. Zhang, N. Keiding, Statistical Methods for The Analysis and Presentation of the Results of Bone Marrow Transplants. Part I: Unadjusted analysis, *Bone Marrow Transplant.* 28 (2001), 909–915. <https://doi.org/10.1038/sj.bmt.1703260>.
- [14] E.T. Lee, J.W. Wang, *Statistical Methods for Survival Data Analysis*, John Wiley & Sons Inc. New Jersey, (2003).
- [15] T.R. Fleming, D.P. Harrington, M. O'sullivan, Supremum Versions of the Log-Rank and Generalized Wilcoxon Statistics, *J. Amer. Stat. Assoc.* 82 (1987), 312–320.
<https://doi.org/10.1080/01621459.1987.10478435>.
- [16] J.W. Lee, Some Versatile Tests Based on the Simultaneous Use of Weighted Log-Rank Statistics, *Biometrics.* 52 (1996), 721-725. <https://doi.org/10.2307/2532911>.
- [17] S. Buyske, R. Fagerstrom, Z. Ying, A Class of Weighted Log-Rank Tests for Survival Data When the Event Is Rare, *J. Amer. Stat. Assoc.* 95 (2000), 249–258. <https://doi.org/10.1080/01621459.2000.10473918>.
- [18] R.E. Tarone, J. Ware, On Distribution-Free Tests for Equality of Survival Distributions, *Biometrika.* 64 (1977), 156–160. <https://doi.org/10.1093/biomet/64.1.156>.
- [19] M.S. Pepe, T.R. Fleming, Weighted Kaplan-Meier Statistics: A Class of Distance Tests for Censored Survival Data, *Biometrics.* 45 (1989), 497-507. <https://doi.org/10.2307/2531492>.
- [20] R.B. Latta, A Monte Carlo Study of Some Two-Sample Rank Tests With Censored Data, *J. Amer. Stat. Assoc.* 76 (1981), 713–719. <https://doi.org/10.1080/01621459.1981.10477710>.
- [21] M.S. Beltangady, R.F. Frankowski, Effect of Unequal Censoring on the Size and Power of the Logrank and Wilcoxon Types of Tests for Survival Data, *Stat. Med.* 8 (1989), 937–945.
<https://doi.org/10.1002/sim.4780080805>.
- [22] E. Letón, P. Zuluaga, Equivalence Between Score and Weighted Tests for Survival Curves, *Commun. Stat. – Theory Methods.* 30 (2001), 591–608. <https://doi.org/10.1081/sta-100002138>.

- [23] E. Letón, P. Zuluaga, Relationships Among Tests for Censored Data, *Biom. J.* 47 (2005), 377–387.
<https://doi.org/10.1002/bimj.200410115>.
- [24] A. Akbar, G.R. Pasha, Properties of Kaplan-Meier Estimator: Group Comparison of Survival Curves, *Eur. J. Sci. Res.* 32 (2009), 391–397.
- [25] T. Jurkiewicz, E. Wycinka, Significance Tests of Differences Between Two Crossing Survival Curves for Small Samples. *Acta Univ. Lodzensis Folia Oecon.* 255 (2011), 114-119. <http://hdl.handle.net/11089/690>.
- [26] P.C. Austin, Generating Survival Times to Simulate Cox Proportional Hazards Models With Time-Varying Covariates, *Stat. Med.* 31 (2012), 3946 – 3958. <https://doi.org/10.1002/sim.5452>.
- [27] J. Wu, A New One-Sample Log-Rank Test, *J. Biometrics Biostat.* 05 (2014), 1000210.
<https://doi.org/10.4172/2155-6180.1000210>.
- [28] T.G. Karrison, Versatile Tests for Comparing Survival Curves Based on Weighted Log-Rank Statistics, *Stata J.* 16 (2016), 678–690. <https://doi.org/10.1177/1536867x1601600308>.
- [29] Z. Chen, G. Zhang, Comparing Survival Curves Based on Medians, *BMC Med. Res. Methodol.* 16 (2016), 33.
<https://doi.org/10.1186/s12874-016-0133-3>.
- [30] P.G. Karadeniz, I. Ercan, Examining Tests for Comparing Survival Curves With Right Censored Data, *Stat. Transition. New Ser.* 18 (2017), 311–328. <https://doi.org/10.21307/stattrans-2016-072>.
- [31] B.T. Babalola, R.A. Adeleke, O.Y. Halid, et al. Statistical Powers of an Alternative Test for Comparison of Survival Distributions With Crossed Survival Curves in the Presence of Censoring: A Simulation Study, *Int. J. Civil Eng. Technol.* 10 (2019), 366-379.
- [32] M. Stevenson, An Introduction to Survival Analysis, EpiCentre, IVABS. Massey Massey University, (2009).
http://www.massey.ac.nz/massey/fms/Colleges/College%20of%20Sciences/Epicenter/docs/ASVCS/Stevenson_survival_analysis_195_721.pdf.
- [33] X. Wang, F. Bai, H. Pang, et al. Bias-adjusted Kaplan–Meier Survival Curves for Marginal Treatment Effect in Observational Studies, *J. Biopharmaceutical Stat.* 29 (2019), 592–605.
<https://doi.org/10.1080/10543406.2019.1633659>.
- [34] R.L.M.C. Martinez, J.D. Naranjo, A Pretest for Choosing Between Logrank And Wilcoxon Tests in the Two-Sample Problem, *METRON.* 68 (2010), 111–125. <https://doi.org/10.1007/bf03263529>.
- [35] J. Xie, C. Liu, Adjusted Kaplan–Meier Estimator and Log-Rank Test With Inverse Probability of Treatment Weighting for Survival Data, *Stat. Med.* 24 (2005), 3089–3110. <https://doi.org/10.1002/sim.2174>.
- [36] A. Winnett, P. Sasieni, Adjusted Nelson–Aalen Estimates With Retrospective Matching, *J. Amer. Stat. Assoc.* 97 (2002), 245–256. <https://doi.org/10.1198/016214502753479383>.
- [37] S. Galimberti, P. Sasieni, M.G. Valsecchi, A Weighted Kaplan-Meier Estimator for Matched Data With Application to the Comparison of Chemotherapy And Bone-Marrow Transplant in Leukaemia, *Stat. Med.* 21 (2002), 3847–3864. <https://doi.org/10.1002/sim.1357>.

- [38] B.T. Babalola, R.A. Adeleke, O.Y. Halid, et al. An Alternative Test for Comparison of Survival Distributions With Proportional Hazard Functions in the Presence of Low Censoring Rates, *J. Appl. Stat. Probab.* 15 (2020), 61-75.
- [39] X. Lin, Q. Xu, A New Method for the Comparison of Survival Distributions, *Pharmaceut. Stat.* 9 (2010), 67–76. <https://doi.org/10.1002/pst.376>.
- [40] J. Shanahan, A New Method for the Comparison of Survival Distributions, Master's Thesis, University of South Carolina, (2013). <https://scholarcommons.sc.edu/etd/555>.
- [41] C. Dardis, Package "survMisc". (2018). <https://cran.r-project.org/web/packages/survMisc/survMisc.pdf>.
- [42] H. Uno, L. Tian, B. Claggett, L.J. Wei, A Versatile Test for Equality of Two Survival Functions Based on Weighted Differences of Kaplan-Meier Curves, *Stat. Med.* 34 (2015), 3680–3695. <https://doi.org/10.1002/sim.6591>.
- [43] R.G. Miller, *Survival Analysis*, John Wiley & Sons, Hoboken, 1981.
- [44] S.H. Embury, L. Elias, P.H. Heller, et al. Remission Maintenance Therapy in Acute Myelogenous Leukaemia, *Western J. Med.* 126 (1977), 267-272.