Inflation of Type I Error of Log-Rank Test with Inappropriately Generated Censoring Data

Yong-Hoon Kim

The Graduate School
Yonsei University
Department of Biostatistics and Computing

Inflation of Type I Error of Log-Rank Test with Inappropriately Generated Censoring Data

A Master's Thesis
Submitted to the Department of Biostatistics and Computing
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Master of Science

Yong-Hoon Kim

December 2024

This certifies that the master's thesis of ***Yong-Hoon Kim*** is approved.

Chung Mo Nam: Thesis Supervisor

Inkyung Jung: Thesis Committee Member #1

Hyun-Soo Zhang: Thesis Committee Member #2

The Graduate School

Yonsei University

December 2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Inflation of Type I Error of Log-Rank Test
with Inappropriately Generated Censoring Data

When simulating survival data, some types of data generation lead to erroneous results. In, in an appropriate generating method called random generation, the event time and censoring time are separately generated based on assumed distributions. In the case where the censoring proportion is fixed, for example, the Weibull distribution does not have a closed form for calculating the censoring distribution parameter, particularly when the shape parameters differ. This often leads to the use of inappropriate data generation methods to simplify the process. In this study, we aimed to investigate the problems caused by inappropriate data generation through simulations and mathematical validation. Specifically, we evaluated Type I error rates of the log-rank test in a two-sample setting and examined the correlation between event times(T) and censoring times(C).

In inappropriate generating method I, after generating event time based on assumed Exponential and Weibull distributions, censoring indicator is generated using a Bernoulli distribution. In cases where censoring occurs, the censoring time is replaced by the generated event time. Furthermore, in inappropriate generating method II, censoring time is generated based on a $Uniform(0, T_i)\ distribution$, introducing between $T\ and\ C$.

The Type I error of the log-rank test was well controlled in the random generation whether the predefined censoring proportions were equal or not between groups, whereas it was inflated in the inappropriate method when the censoring proportions were unequal. However, an inappropriate method appeared to effectively control the Type I error when the censoring proportions were equal.

This is likely due to the log-rank test, where the dependent censoring between $T\ and\ C$ results in the conditional distribution of $T\ given\ C$ becoming different from the marginal distribution of T which can distort the actual differences between groups, creating the illusion of well-controlled Type I error rates.

Additional simulations were conducted to investigate this issue. One group was generated using the random generating method, while the other group was generated using the inappropriate generating method II. The results showed that even when the censoring proportions were equal between the groups, the Type I error of the log-rank test was inflated. This finding suggests that the increase in Type I errors is not due to unequal censoring proportions, but rather due to the inappropriate data-generating process that induced dependent censoring. Additionally, Spearman correlation between event time and censoring time confirmed that improper data generation introduced dependency.

Key words: Censored Data, Random Generating, Inappropriate Data Generating, Log-Rank Test, Type I error

1. Introduction

In survival analysis, there are various approaches for generating survival data. As different approaches used in previous studies, Alam et al. (2022) generated event times and censoring times separately, assumed the same distribution, and used the minimum as observed time. Additionally, Wan (2016) generated event times assuming a Weibull distribution and created censoring times under two conditions: one where the shape parameter of the censoring time matched the event time, and another where it differed. A numerical root-finding algorithm was employed to determine the scale parameter when using a different shape parameter for the censoring time. This method is commonly used to generate survival data and is referred to as random or independent generation.

Furthermore, event times can be generated by drawing random values from a uniform distribution and applying an inverse transformation to an Exponential distribution. The censoring indicator can be generated based on a Bernoulli distribution using the true event rate as a parameter. The authors then employed the approach of multiplying the generated survival time with a random value from a uniform distribution for censored observations (Kuss et al., 2021).

A recent study investigated the impact of unequal censoring proportions and insufficient follow-up under dependent censoring in clinical trials comparing survival between two groups (Srivastava et al., 2021). In this study, event times were generated assuming Exponential and Weibull distributions, and censoring times were generated from a uniform distribution ranging from zero to the corresponding generated event times. The study found that as the difference in censoring proportions between the treatment groups increased, there was a tendency for Type I error to increase and power to decrease in all tests.

Research on censoring proportions in survival data has been extensively conducted over the years. Beltangady and Frankowski evaluated the performance of log-rank and Wilcoxon type tests to compare two survival distributions in the presence of unequal random censoring for small sample sizes. They concluded that the inequality of censoring proportions affected the power of all tests, and that greater differences in censoring proportions led to lower power estimates (Beltangady and Frankowski, 1989). Wang et al. proposed several improvements of the permutation log-rank test for comparing two survival distributions with different censoring distributions. Their method involves imputing failure and censoring times based on Kaplan-Meier estimates of the survival and censoring distributions. Additionally, they introduced a permutation test specifically designed to address unequal censoring using this imputation approach (Wang et al., 2010).

When the censoring proportion is fixed, survival data are generally generated using the random generation method described in Section 2.2. However, when the shape parameters of the Weibull distribution differ, a closed-form expression for calculating the scale parameter of the censoring distribution does not exist. To address this issue, some studies have employed inappropriate data generation methods. However, improper methods often lead to erroneous results. The objective of this study is to identify the issues arising from such practices through mathematical proofs and simulations across equal and unequal censoring proportions between two groups. This study makes a notable contribution by addressing the issue of inappropriately generating censored data, which commonly occurs in clinical trials, and by comparing incorrect data generation methods across several simulation settings.

The remainder of this paper is organized as follows. Section 1 introduces the study and describes its purpose. In Section 2, we explain independent censoring, the approach for deriving the distribution parameter of the censoring time under this condition, and the log-rank test. Also, in the same section, we summarize the three methods, one using random generating method and two using inappropriate methods, generating survival data. In Section 3, we present the theoretical framework for both dependent and independent data generation of event times and censoring times. In Section 4 sets up different simulation setting and compares the Type I error control of the two-sample log-rank test and the correlation between for each method. Finally, Section 5 concludes the study and discusses the implications of this research.

2. Methods

2.1 Notation

Assume that there are n_i subjects. For $j = 1, \cdots, n_i$, T_{ij} denotes the survival time for the ith subject and in jth group. Also, C_{ij} denotes the censoring time for the ith subject and in jth group. We observe (X_{ij}, δ_{ij}), where the observed event time is denoted $X_{ij} = min(T_{ij}, C_{ij})$ and $\delta_{ij} = I(X_{ij} = C_{ij})$ serves as the censoring indicator.

The hazard function is denoted by $h(t)$, and the survival function is denoted by $S(t)$. The predefined censoring proportion is indicated as p. For the distribution parameters, λ is denoted as the parameter in Exponential distribution or the scale parameter in the Weibull distribution. The shape parameter in the Weibull distribution is denoted as α. Lastly, the Spearman correlation is represented by r_s.

2.2 Parameter of censoring time distribution for fixed censoring proportion

The event time is generated using the Exponential and Weibull distributions, respectively, and each distribution is as follows:

- **Exponential distribution**
  - $h(x) = \lambda$
  - $S(x) = exp(-\lambda x)$
  - $f(x) = \lambda exp(-\lambda x)$

- **Weibull distribution**
  - $h(x) = \lambda\alpha x^{\alpha-1}$
  - $S(x) = exp(-\lambda x^{\alpha})$
  - $f(x) = \alpha\lambda x^{\alpha-1} exp(-\lambda x^{\alpha})$

where $\alpha > 0$ is a shape parameter, and $\lambda > 0$ is a scale parameter. In the case of a Weibull distribution, the condition $\alpha \neq 1$ is required, because if $\alpha = 1$, the distribution becomes equivalent to the Exponential distribution.

The parameters of the censoring time distribution for each distribution can be derived using the joint probability distribution function of the event and the censoring time distribution when the censoring proportion is predefined and fixed. First, when event time is generated from an Exponential distribution, the censoring time parameter can be derived as follows:

- **Exponential distribution**

p (Censoring proportion) $= P(T > C)$

$$= \int_0^\infty \int_c^\infty \lambda_1 e^{-\lambda_1 t} \cdot \lambda_2 e^{-\lambda_2 c} dtdc$$

$$= \int_0^\infty \left[-e^{-\lambda_1 t} \cdot \lambda_2 e^{-\lambda_2 c}\right]_0^\infty dc$$

$$= \left[-\frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda_2)}\right]_0^\infty$$

$$= \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

- **Weibull distribution**

p (Censoring proportion) $= P(T > C)$

$$= \int_0^\infty \int_c^\infty \alpha_1 \lambda_1 t^{\alpha_1 - 1} \exp(-\lambda_1 t) \cdot \alpha_2 \lambda_2 c^{\alpha_2 - 1} \exp(-\lambda_2 c^{\alpha_2})\, dtdc$$

$$= \int_0^\infty \alpha_2 \lambda_2 c^{\alpha_2 - 1} \exp(-\lambda_2 c^{\alpha_2}) \int_c^\infty \alpha_1 \lambda_1 t^{\alpha_1 - 1} \exp(-\lambda_1 t^{\alpha_1})\, dtdc$$

$$= \int_0^\infty \alpha_2 \lambda_2 c^{\alpha_2 - 1} \exp(-\lambda_2 c^{\alpha_2}) \exp\,(-\lambda_1 c^{\alpha_1}) dc$$

Since a closed form no longer exists in the Weibull distribution, we add the condition that the shape parameters, α_1 and α_2, are equal for further mathematical derivations.

$$= \int_0^\infty \alpha_1 \lambda_2 c^{\alpha_1 - 1} \exp(-\lambda_2 c^{\alpha_1} - \lambda_1 c^{\alpha_1})\, dc$$

$$= \left[\frac{\lambda_2}{-(\lambda_1 + \lambda_2)} \exp\,(-(\lambda_1 + \lambda_2) c^{\alpha_1}\right]_0^\infty$$

$$= \frac{\lambda_1}{\lambda_1 + \lambda_2} \quad \& \quad \alpha_1 = \alpha_2$$

As shown in the above equation, the relationship between the parameters of the assumed distributions for the event and censoring times can be derived. Since the censoring proportion is fixed and λ_1 is given, λ_2 can be calculated.

Based on the derived λ_2, the censoring time distribution can be generated. This method of separately deriving the event and censoring time distributions is referred to as independent generating. Typically, the times are derived randomly, and survival data is created by comparing the two times: if the event occurs first, the censoring indicator is set to 1; otherwise, it is set to 0.

However, for example, when the shape parameter of the Weibull distribution differ ($\alpha_1 \neq \alpha_2$), there is no closed-form for calculating the scale parameter λ_2, requiring the use of numerical methods such as the Newton-Raphson algorithm. However, this process is computationally intensive and complex, leading to several cases in previous studies in which inappropriate methods were employed to generate survival data.

To address this issue, some studies have employed inappropriate data generation methods. These methods lead to erroneous results when generating survival data. The aim of this study is to identify and verify the issues arising from such practices through theoretical framework and simulations.

2.3 Data generation

2.3.1 Appropriate generating method

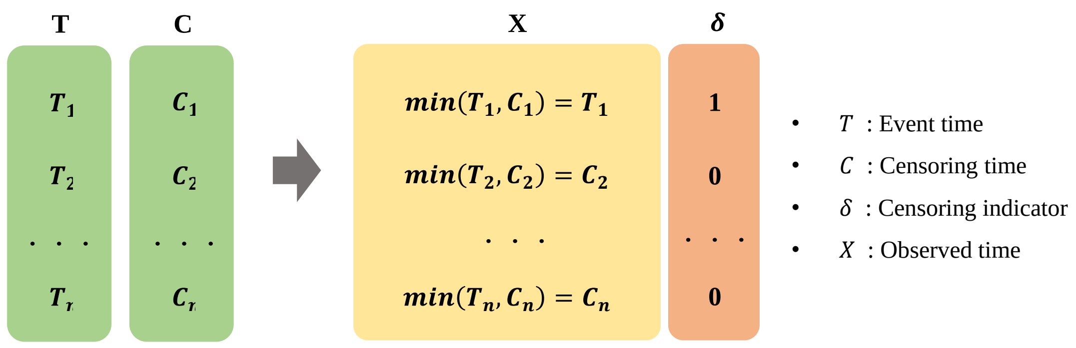
The following methods were explained based on a two-sample setting.

In an appropriate generating method, also called random generation, the event times T_{ij}, $i = 1, \dots, n_j$ and $j = 1, 2$ is generated based on the assumed Exponential and Weibull distribution. Also, the censoring times C_{ij} can be generated based on the distribution assumption of the parameter, as explained through the parameter relationship described in 2.2. The censoring indicator is obtained as $\delta_{ij} = I(T_{ij} \le C_{ij})$, in other words, if an event occurs, $\delta_{ij} = 1$; if censoring occurs, $\delta_{ij} = 0$. Observed time X_{ij} can be expressed as follows:

$$X_{ij} = \begin{cases} T_{ij} & if \ \ \delta_{ij} = 1 \\ C_{ij} & if \ \ \delta_{ij} = 0 \end{cases}$$

Finally, the generated survival data is (X_{ij}, δ_{ij}), $i = 1, \dots, n_j$ and $j = 1, 2$.

Figure 1. Diagram of the random generating method for survival data (X, δ)



2.3.2 Inappropriate generating method I

In inappropriate generating method I, first, the event time, T_{ij}, $i = 1, \dots, n_j$ and $j = 1, 2$ is generated based on the assumed distribution. Next, the censoring indicator is generated using bernoulli distribution, $\delta_{ij} \sim Bernoulli(1-p)$, where p is prespecified censoring proportion. The censoring time C_{ij} is then expressed as follows:

$$C_{ij} = \begin{cases} T_{ij} & if\ \delta_{ij} = 0 \\ Uniform(T_{ij}, C^*_{ij}) & if\ \delta_{ij} = 1 \end{cases}$$

When censoring occurs, $\delta_{ij} = 0$, and the previously generated event time is used as the censoring time. C^*_{ij} is replaced by the percentile values (75%, 85% and 95%) of the censoring time randomly generated in inappropriate method I. However, in this case, if $C^*_{ij} \le T_{ij}$, it becomes difficult to properly generate to uniform distribution. Additionally, even if $\delta_{ij} = 1$, in order to check the correlation between event time and censoring time, there must be no missing values, and all data for each subject must be available. An important point is that this procedure is only required when calculating the correlation. In all other cases, if $\delta_{ij} = 1$, there is no need to generate a censoring time. Therefore, when it is necessary to check the correlation, the survival data is generated as follows with modification.

$$C_{ij} = \begin{cases} T_{ij} & if\ \delta_{ij} = 0 \\ T_{ij} + Uniform(0, C^*_{ij}) & if\ \delta_{ij} = 1 \end{cases}$$

2.3.3 Inappropriate generating method II

In inappropriate generating method I, the censoring time C_{ij} was set to be exactly the pre-generated event time T_{ij}. Instead in inappropriate generating method II, when censoring occurs ($\delta_{ij} = 0$), more randomness can be imposed on C_{ij} by sampling it from a $Unif(0, T_{ij})$ distribution, which is formulated as follows:

$$C_{ij} = \begin{cases} Uniform(0, T_{ij}) & if \ \delta_{ij} = 0 \\ Uniform(T_{ij}, C_{ij}^{*}) & if \ \delta_{ij} = 1 \end{cases}$$

By making the above modifications, it is possible to generate censoring time earlier than the pre-generated event time T_{ij}. Similar to inappropriate method I the case where $\delta_{ij} = 1$ was modified as follows to check the correlation.

$$C_{ij} = \begin{cases} Uniform(0, T_{ij}) & if \ \delta_{ij} = 0 \\ T_{ij} + Uniform(0, C_{ij}^{*}) & if \ \delta_{ij} = 1 \end{cases}$$

2.4 Log-Rank Test

The log-rank test is a representative method for testing the homogeneity of survival functions between two or more groups. The data generated from $K(\geq 2)$ populations may include correct censoring and have $t_1 < t_2 < \cdots < t_D$ distinct observed values. For group $j = 1,2,\cdots,K,$ time point $i = 1,2,\cdots,D,$ d_{ij} is denoted as the number of events in the j^{th} group at time t_i and Y_{ij} is denoted as the number of individuals at risk in the j^{th} group at time t_i. Also, $d_i = \sum_{j=1}^{K} d_{ij}$ and $Y_i = \sum_{j=1}^{K} Y_{ij}$ denote the number of events and the number of individuals at risk at the time point t_i in the combined data, respectively.

We test the following hypotheses:

$H_0 : h_1(t) = h_2(t) = \cdots = h_k(t),$ for all $t \leq \tau$

$H_1 :$ at least one of the $h_i(t)$'s is different for some $t \leq \tau$

Here τ denotes as the largest time at which all of the groups have at least one subject at risk. For example, in a two-sample case, if the null hypothesis is true, then at time t_i, $\frac{d_i}{Y_i} = \frac{d_{1i}}{Y_{1i}} = \frac{d_{2i}}{Y_{2i}}$.

The test statistic of the log-rank test can be expressed as follows:

$$Z_j(\tau) = \sum_{i=1}^{D}\left[\frac{d_{ij}}{Y_{ij}} - \frac{d_i}{Y_i}\right], \qquad j = 1, \dots, K$$

Then rewriting the test statistics $Z_j(\tau)$,

$$Z_j(\tau) = \sum_{i=1}^{D} \frac{1}{Y_{ij}}\left[d_{ij} - Yij\frac{d_i}{Y_i}\right] = \sum_{i=1}^{D} \frac{1}{Y_{ij}}\left[O_{ij} - E_{ij}\right], \qquad j = 1, \dots, K$$

In the case where the number of groups $K = 2$, the test statistic is as follows:

$$Z = \frac{\sum_{i=1}^{D} W(t_i)\left[d_{i1} - Y_{i1}\left(\frac{d_i}{Y_i}\right)\right]}{\sqrt{\sum_{i=1}^{D} W(t_i)^{2}\frac{Y_{1i}}{Y_i}\left(1 - \frac{Y_{1i}}{Y_i}\right)\left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i}}$$

The test statistic Z under the null hypothesis, H_0 approaches $Z \sim N(0,1)$ as $n \to \infty$. When the weight function $W(t_i)$ is 1 for all time points t_i, it represents the standard log-rank test statistics. Depending on the $W(t_i)$, various forms of weighted log-rank test statistics exist (Fleming et al., 1987).

3. Theoretical framework

3.1 Generation of censoring data

Proposition.

When the time-to-event T and censoring time C are dependent, the observed data ($X = min(T, C)$ and $\delta = I(T \leq C)$) alone is insufficient to identify and compare the marginal survival functions between two groups.

Definition.

For the pdf f of the observables (X, δ) with parameters vector θ, θ is identifiable if any given θ uniquely determines the density f of (X, δ), i.e. if $f_{X,\delta,\theta_1} \equiv f_{X,\delta,\theta_2}$, then $\theta_1 = \theta_2$.

Proof.

Let T be the random variable for the event time and C the random variable for the censoring time, and the observed survival time $X = min(T, C)$ and $\delta = I(T \leq C)$. The PDFs of T and C in groups 1 and 2 denoted as $f_1(t), g_1(c)$ and $f_2(t), g_2(c)$, respectively. Also, the CDFs and survival functions in groups 1 and 2 are $F_1(t), S_1(t) = 1 - F_1(t)$ and $F_2(t), S_2(t) = 1 - F_2(t)$, respectively.

The distribution of the observed data (X, δ) is determined by the joint distribution of T and C in each group, $Q_1(t, c) = \Pr(T \leq t, C \leq c)$ and $Q_2(t, c) = \Pr(T \leq t, C \leq c)$.

I. Under independent censoring

$$S_1(t) = \Pr(T_1 > t) = \int_0^\infty \Pr(T_1 > t | C_1 = c) \cdot \Pr(C_1 = c)\, dc$$

$$= \int_0^\infty \Pr(T_1 > t, C_1 = c) / \Pr(C_1 = c) \cdot \Pr(C_1 = c)\, dc$$

$$= \int_0^\infty \Pr(T_1 > t) \cdot \Pr(C_1 = c)\, dc$$

$$= \Pr(T_1 > t) \cdot \int_0^\infty g_1(c) dc = \Pr(T_1 > t) = S_1(t)$$

such that from the observed $Pr(X_1 = c, \delta = 0)$, the marginal survival function $S_1(t)$ can be separated out as $\Pr(T_1 > t, C_1 = c) = S_1(t) \cdot g_1(c)$ due to independence between T and C.

Hence, the observed (X, δ) in both groups 1 and 2 can be used to obtain the marginal survival functions $S_1(t)$ and $S_2(t)$ to test for $H_0 : S_1(t) = S_2(t)$.

II. Under dependent censoring

Let the strength of dependence between T and C be θ $(-1 \leq 0 \leq 1)$.

Now, the observed $\Pr(X_1 = c, \delta = 0) = \Pr(T_1 > t, C_1 = c) = S_1(t|c, \theta) \cdot g_1(c) = S_1^*(t) \cdot g_1(c) \neq S_1(t) \cdot g_1(c)$, where $S_1^*(t)$ is not the marginal survival function of t but a function of t, c, and θ.

Therefore, the marginal survival function $S_1(t)$ can no longer be separated out from the observed (X, δ), i.e., $\Pr(T_1 > t, C_1 = c)$ no longer factorizes such that $S_1(t)$ is isolated. Thus, comparing the observed (X, δ) between the two groups 1 and 2 is not equivalent to comparing the marginal survival functions $S_1(t)$ and $S_2(t)$ and cannot be used to test for $H_0 : S_1(t) = S_2(t)$.

The above presents the proof in a general case. In contrast, the following demonstrates the theoretical framework as applied to the actual generating scenario in the simulation section. The framework of inappropriate generating method II (IG II) is as follows:

$$\text{IG II } = \begin{cases} C \mid T = t, \delta = 0 \sim Unif(0, t) \\ C \mid T = t, \delta = 1 \sim Unif(t, C^*) \end{cases}$$

and censoring proportion is defined by $\Pr(\delta = 0) = p$.

- **Joint CDF**: $Q_{T,C}(t, c) = \Pr(T \leq t, C \leq c)$

$= \Pr(T \leq t, C \leq c, \delta = 0) + \Pr(T \leq t, C \leq c, \delta = 1)$

$= \Pr(T \leq t, C \leq c \mid \delta = 0) \times \Pr(\delta = 0) + \Pr(T \leq t, C \leq c \mid \delta = 1) \times \Pr(\delta = 1)$

$$= p \times \int_0^c \Pr(T \leq t \mid C = w, \delta = 0) \cdot \Pr(C = w \mid \delta = 0) dw$$

$$+ (1 - p) \times \int_0^t \Pr(C \leq c \mid T = z, \delta = 1) \cdot \Pr(T = z \mid \delta = 1) dz$$

$$= p \int_0^c \Pr(T \leq t \mid C = w, \delta = 0)\, g(w) dw + (1 - p) \int_0^t \frac{c - z}{C^* - z} I(z \leq c \leq C^*) f(z)\, dz$$

For the derivation of the formula, we represented it using a joint PDF.

- **Joint PDF**: $q_{(T,C)}(t,c) = \frac{\partial^2}{\partial t \partial c} Q_{T,C}(t,c)$

$$= \frac{\partial}{\partial t}\left[\frac{\partial}{\partial c}\left[p\int_0^c \Pr(T \le t \mid C = w, \delta = 0)\, g(w)dw\right]\right]$$

$$+ \frac{\partial}{\partial c}\left[\frac{\partial}{\partial t}\left[(1-p)\int_0^t \frac{c-z}{C^*-z}\, I(z \le c \le C^*) f(z)\, dz\right]\right]$$

$$= \frac{\partial}{\partial t}[p\Pr(T \le t \mid C = c, \delta = 0) g(c)] + \frac{\partial}{\partial c}\left[(1-p)\frac{-1}{C^*-t} I(t \le c \le C^*) f(t)\right]$$

$$= pg(c)f(T = t \mid C = c, \delta = 0) + (1-p)f(t)\frac{-1}{C^*-t} I(t \le c \le C^*)$$

The joint PDFs for each group 1 and 2 can be expressed as follows:

Group 1: $p_1 g_1(c) f_1(T = t \mid C = c, \delta = 0) + (1-p_1) f_1(t) \frac{-1}{C^*-t} I(t \le c \le C^*)$

Group 2: $p_2 g_2(c) f_2(T = t \mid C = c, \delta = 0) + (1-p_2) f_2(t) \frac{-1}{C^*-t} I(t \le c \le C^*)$

As the survival functions of each group change, $f_1(t)$ becomes $f_1^*(t)$, and $f_2(t)$ becomes $f_2^*(t)$. If $p_1 = p_2$, the joint PDFs structure is the same across groups under equal censoring, but when $p_1 \ne p_2$, with unequal censoring by groups, the distribution of the survival functions is altered. The test statistic such as log-rank differs between groups, leading to an increase in Type I error.

Additionally, when survival data is generated using random generating for group 1 and inappropriate generating for group 2, the survival function $f_1(t)$ of group 1 remains unchanged, whereas the survival function $f_2(t)$ of group 2 changes to $f_2^*(t)$. In this case, when the censoring proportions are equal, i.e., $p_1 = p_2$, the distribution of the survival functions differ, resulting in a violation of homogeneity between the two groups.

4. Simulation

4.1 Simulation setting

Before the simulation settings, event times for all generating methods were generated based on the following approach under distributional assumptions. First, let X be a random variable with a continuous and strictly increasing cumulative distribution function (CDF). Then, we define a random variable Y denoted as $Y = F(X)$, then, Y follows a uniform distribution on the interval $[0, 1]$. In other words, $Y = F(X) \sim Uniform(0,1)$.

Next, for the event time T in the Exponential distribution, $F(T) = 1 - S(T) \overset{d}{=} 1 - exp(-\lambda T) \overset{d}{=} U \sim Uniform(0,1)$. Finally, $exp(-\lambda T) \overset{d}{=} U$, and $T \overset{d}{=} -\log(U)/\lambda$. Similarly, in the Weibull distribution, $exp(-\lambda T^{\alpha}) \overset{d}{=} U$, and $T \overset{d}{=} (-\log(U)/\lambda)^{1/\alpha}$.

In the one-sample case, the correlation between the event time and censoring time was examined. The correlation was measured using the Spearman correlation coefficient (r_s). In the case of the Pearson correlation, the normality assumption is required when both variables are continuous, and it only measures linear relationships. Moreover, because it measures the correlation between actual values, it is highly sensitive to outliers, which can distort the overall correlation. Therefore, Spearman correlation, which is rank-based, non-parametric, and does not require any assumptions, was used. For each simulation setting, 100 iterations were performed and the correlation was calculated as the average of these iterations. Additionally, Kaplan-Meier plots were generated to compare the different methods and prespecified censoring proportions.

In the two-sample case, as in the one-sample case, the correlation between the event time and censoring time was examined. Additionally, under the null hypothesis $H_0: h_1(t) = h_2(t)$, the Type I error of the log-rank test was calculated by performing 1,000 iterations. To make this more intuitive within the table, the %bias measure was also examined. %Bias is an indicator that shows the percentage difference between the calculated value and the reference value, and can be used to determine if there is any bias. The reference value was set at a significance level of 0.05. Letting the calculated Type I error as α, it is calculated as $\%bias = \left(\frac{\alpha - 0.05}{0.05}\right) \times 100$.

Both cases where the sample sizes in the two groups were equal or unequal were examined, considering both general situations and realistic data collection conditions. The detailed simulation settings are shown in Table 1 below.

For the two samples, the simulation setting was modified to use a pre-specified censoring proportion to check the Spearman correlation, unlike the censoring proportion setting used to observe the Type I error rate in Table 1. The reason for this modification is that setting the censoring proportion to zero could result in an infinite value for censoring times, and this adjustment was made to prevent such occurrences. $(p_1, p_2) =$ (10,10), (20,20), (30,30), (40,40), (10,20), (20,30), (30,40), (20,10), (30,20), (40,30), (10,30), (20,40), (30,10), (40,20), (10,40), (40,10).

Table 1. All scenarios for data generation settings

	Sample size	Censoring proportion	Method	Distribution parameter
One Sample	$n \in \{30, 100, 1000\}$	$p \in \{10, 20, 30, 40\}$	Appropriate	$T_i \sim Exp(5.873)\ or\ Weibull(2,2)$ $C_i \sim Exp(\lambda_2)\ or\ Weibull(shape, 2)$
			Inappropriate I	$T_i \sim Exp(5.873)\ or\ Weibull(2,2)$
			Inappropriate II	
Two Sample	Equal sample size $(n_1, n_2) \in$ {(30,30), (100,100), (1000,1000)}	$(p_1, p_2) \in$ {(0,0), (10,10), (20,20), (30,30), (40,40), (0,10), (10,20), (20,30), (30,40), (10,0), (20,10), (30,20), (40,30), (0,20), (10,30), (20,40), (20,0), (30,10), (40,20), (0,30), (10,40), (30,0), (40,10), (0,40), (40,0)}	Appropriate	$T_{ij} \sim Exp(5.873)\ or\ Weibull(2,2)$ $C_{ij} \sim Exp(\lambda_2)\ or\ Weibull(shape, 2)$
			Inappropriate I	$T_{ij} \sim Exp(5.873)\ or\ Weibull(2,2)$
			Inappropriate II	
	Unequal sample size $(n_1, n_2) \in$ {(30,100), (100, 1000)}		Appropriate	$T_{ij} \sim Exp(5.873)\ or\ Weibull(2,2)$ $C_{ij} \sim Exp(\lambda_2)\ or\ Weibull(shape, 2)$
			Inappropriate I	$T_{ij} \sim Exp(5.873)\ or\ Weibull(2,2)$
			Inappropriate II	

4.2 Simulation result

Tables 2 - 5 compared the type I error of the log-rank test and %bias for each method. Tables 8-9 and Appendix Tables 1 - 10 presented the Spearman correlation coefficients for each scenario in both one-sample and two-sample scenarios.

Figure 2 - 3 showed Kaplan-Meier (KM) curves comparing each method under Exponential and Weibull distribution for a one-sample case with $N = 1000$ and $p = 0.3$. As shown in the figure, the log-rank tests conducted using the three different data generation methods under the same simulation settings yielded significant differences (p-value < 0.001). This indicates that the null hypothesis, which states that "all three hazard functions are identical," can be rejected. Based on the KM curves derived from the appropriate method, it was observed that inappropriate method 1 tended to overestimate survival probabilities. This overestimation can be attributed to the fact that censoring times were taken directly from event times, which prolonged the at-risk set denominator and led to inflated survival probability estimates. Meanwhile, under the assumption of a Weibull distribution, the KM curves generated by the appropriate method and inappropriate method II did not show substantial differences. However, under the assumption of an Exponential distribution, the KM curves showed clear differences between the two methods. These findings indicate that the appropriateness of the data generation method can significantly impact the survival analysis results.

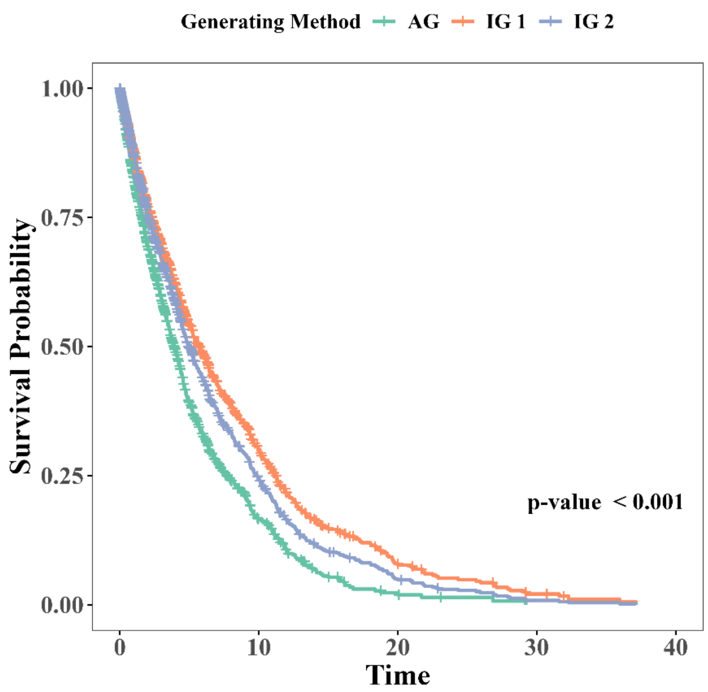
In tables 2 and 4, the Type I error of the log-rank test under each method are shown, based on survival data generated assuming Exponential and Weibull distributions with equal sample sizes. In appropriate generating method, the Type I error remained close to 0.05, irrespective of the simulation settings. On the other hand, in inappropriate generating methods I and II, the larger the sample size and the difference in censoring proportions between groups, the more Type I error inflated from 0.05.

Specifically, in method I, when the sample size was (1000,1000), the %bias approached an enormous value. Even when the difference in censoring proportions between the groups was the same, a higher overall censoring proportion led to a higher Type I error. Also, the Type I error was observed to increase as the extent of unequal censoring between the two groups. In tables 3 and 5, where the sample sizes were unequal, a similar pattern was observed as with equal sample sizes.

However, an inappropriate method appeared to control the Type I error when the censoring proportions were equal. This is likely due to the log-rank test, where improperly generated data can distort the actual differences between groups, creating the illusion of well-controlled Type I error. Additional simulations are conducted to investigate this issue. One group was generated using the random generating method, while the other group was generated using the inappropriate generating method II. The results showed that even when the censoring proportions were equal between the groups, the Type I error of the log-rank test was inflated in tables 6 - 7. Specifically, in Figures 4 – 5, as the sample size increases and the censoring proportion becomes higher, the Type I error increased. This finding suggests that the increase in Type I errors is not simply due to unequal censoring proportions, but rather due to the inappropriate data-generating process.

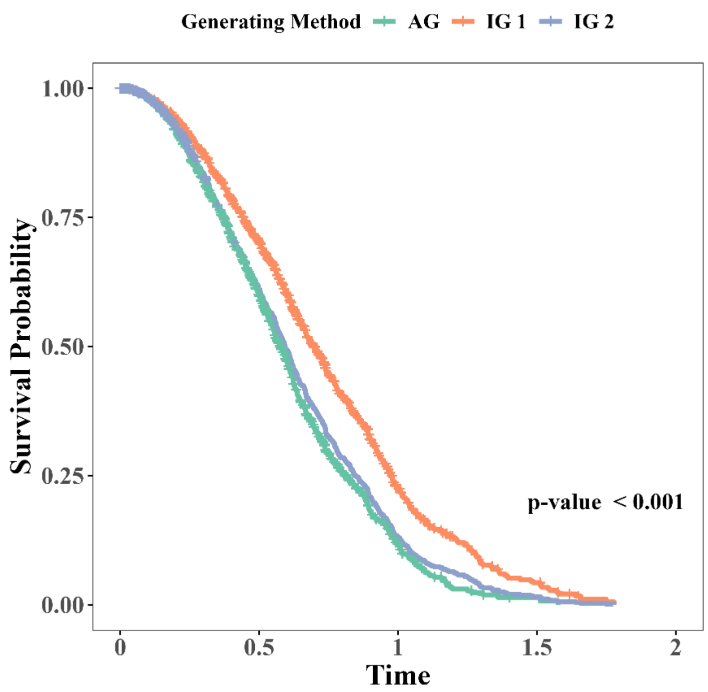
Thus, the Spearman correlation coefficient was examined between event time (T) and censoring time (C) to investigate the properties of inappropriately generated data for each method. (Tables 8 - 9, Appendix tables 1 - 10) In both one-sample and two-sample cases, no correlation was found in random generating. However, in inappropriate method I, a correlation was observed across all settings. Notably, as the censoring proportion increased, the correlation also increased; the lower the percentile replaced upon event occurrence, the higher the correlation value. Additionally, in inappropriate method II, while the correlation coefficient was smaller than that of inappropriate method I across all simulation settings, a correlation was still observed.

Figure 2. Kaplan-Meier curves comparing each method under Exponential distribution for one sample with $N = 1000$ and censoring proportion $p = 0.3$



Note. AG: Appropriate generating method, IG1: Inappropriate generating method I, IG2: Inappropriate generating method II

Figure 3. Kaplan-Meier curves comparing each method under Weibull distribution for one sample with $N = 1000$ and censoring proportion $p = 0.3$



Note. AG: Appropriate generating method, IG1: Inappropriate generating method I, IG2: Inappropriate generating method II

Table 2. Type I error for each method with predefined censoring proportion and sample size where n_1 and n_2 are equal assumed Exponential distribution $T_1, T_2 \sim Exp(5.873)$

p_1	p_2	Appropriate method				Inappropriate method I				Inappropriate method II			
		(30,30)	(100,100)	(1000,1000)	%Bias	(30,30)	(100,100)	(1000,1000)	%Bias	(30,30)	(100,100)	(1000,1000)	%Bias
0	**0**	0.055	0.057	0.053	**6**	0.054	0.051	0.054	**8**	0.054	0.051	0.054	**8**
10	**10**	0.053	0.055	0.06	**20**	0.051	0.058	0.056	**12**	0.062	0.06	0.047	**-6**
20	**20**	0.058	0.052	0.057	**14**	0.056	0.054	0.058	**16**	0.057	0.051	0.056	**12**
30	**30**	0.058	0.055	0.063	**26**	0.06	0.053	0.048	**-4**	0.058	0.057	0.051	**2**
40	**40**	0.055	0.057	0.052	**4**	0.067	0.056	0.052	**4**	0.059	0.056	0.044	**-12**
0	**10**	0.052	0.05	0.058	**16**	0.065	0.11	0.618	**1136**	0.056	0.069	0.19	**280**
10	**20**	0.061	0.055	0.054	**8**	0.073	0.126	0.65	**1200**	0.056	0.084	0.241	**382**
20	**30**	0.062	0.045	0.063	**26**	0.084	0.139	0.71	**1320**	0.066	0.081	0.281	**462**
30	**40**	0.051	0.056	0.054	**8**	0.094	0.141	0.78	**1460**	0.074	0.085	0.358	**616**
10	**0**	0.061	0.049	0.059	**18**	0.073	0.125	0.639	**1178**	0.056	0.081	0.252	**404**
20	**10**	0.053	0.049	0.057	**14**	0.079	0.117	0.689	**1278**	0.057	0.086	0.279	**458**
30	**20**	0.051	0.055	0.059	**18**	0.074	0.127	0.741	**1382**	0.065	0.078	0.344	**588**
40	**30**	0.053	0.057	0.057	**14**	0.075	0.135	0.823	**1546**	0.074	0.081	0.397	**694**
0	**20**	0.057	0.051	0.047	**-6**	0.119	0.308	0.998	**1896**	0.07	0.121	0.684	**1268**

10	**30**	0.055	0.046	0.06	**20**	0.139	0.36	0.997	**1894**	0.076	0.144	0.754	**1408**
20	**40**	0.053	0.052	0.052	**4**	0.153	0.387	0.999	**1898**	0.098	0.167	0.85	**1600**
20	**0**	0.057	0.045	0.06	**20**	0.134	0.321	0.996	**1892**	0.075	0.122	0.713	**1326**
30	**10**	0.047	0.054	0.052	**4**	0.127	0.353	0.999	**1898**	0.089	0.143	0.797	**1494**
40	**20**	0.049	0.051	0.053	**6**	0.141	0.388	0.999	**1898**	0.1	0.168	0.875	**1650**
0	**30**	0.059	0.05	0.06	**20**	0.215	0.611	1	**1900**	0.104	0.237	0.968	**1836**
10	**40**	0.052	0.057	0.052	**4**	0.253	0.686	1	**1900**	0.124	0.285	0.99	**1880**
30	**0**	0.046	0.048	0.056	**12**	0.244	0.625	1	**1900**	0.119	0.236	0.976	**1852**
40	**10**	0.048	0.055	0.059	**18**	0.247	0.673	1	**1900**	0.136	0.284	0.994	**1888**
40	**0**	0.05	0.058	0.051	**2**	0.388	0.864	1	**1900**	0.17	0.429	1	**1900**
0	**40**	0.054	0.058	0.05	**0**	0.384	0.879	1	**1900**	0.154	0.408	1	**1900**

Note. %Bias was calculated based on the Type I error when the sample sizes of the two groups were (1000,1000).

Table 3. Type I error for each method with predefined censoring proportion and sample size where n_1 and n_2 are unequal assumed Exponential distribution $T_1, T_2 \sim Exp(5.873)$

p_1	p_2	Appropriate method			Inappropriate method I			Inappropriate method II		
		(30,100)	(100,1000)	%Bias	(30,100)	(100,1000)	%Bias	(30,100)	(100,1000)	%Bias
0	**0**	0.052	0.052	**4**	0.058	0.057	**14**	0.058	0.057	**14**
10	**10**	0.054	0.053	**6**	0.062	0.056	**12**	0.066	0.055	**10**
20	**20**	0.055	0.047	**-6**	0.059	0.062	**24**	0.047	0.055	**10**
30	**30**	0.056	0.041	**-18**	0.062	0.056	**12**	0.058	0.048	**-4**
40	**40**	0.053	0.048	**-4**	0.062	0.049	**-2**	0.054	0.052	**4**
0	**10**	0.054	0.051	**2**	0.105	0.201	**302**	0.079	0.116	**132**
10	**20**	0.052	0.055	**10**	0.106	0.204	**308**	0.084	0.113	**126**
20	**30**	0.054	0.044	**-12**	0.104	0.232	**364**	0.086	0.12	**140**
30	**40**	0.055	0.043	**-14**	0.106	0.242	**384**	0.085	0.142	**184**
10	**0**	0.053	0.051	**2**	0.066	0.143	**186**	0.058	0.07	**40**
20	**10**	0.054	0.046	**-8**	0.077	0.159	**218**	0.052	0.067	**34**
30	**20**	0.054	0.044	**-12**	0.086	0.169	**238**	0.053	0.081	**62**
40	**30**	0.052	0.043	**-14**	0.085	0.183	**266**	0.058	0.085	**70**
0	**20**	0.05	0.058	**16**	0.208	0.582	**1064**	0.115	0.239	**378**

10	**30**	0.052	0.049	**-2**	0.215	0.635	**1170**	0.119	0.276	**452**
20	**40**	0.048	0.047	**-6**	0.23	0.683	**1266**	0.13	0.322	**544**
20	**0**	0.05	0.044	**-12**	0.167	0.461	**822**	0.066	0.147	**194**
30	**10**	0.055	0.042	**-16**	0.178	0.5	**900**	0.081	0.165	**230**
40	**20**	0.054	0.047	**-6**	0.186	0.562	**1024**	0.089	0.196	**292**
0	**30**	0.049	0.048	**-4**	0.384	0.925	**1750**	0.173	0.471	**842**
10	**40**	0.046	0.048	**-4**	0.42	0.952	**1804**	0.197	0.549	**998**
30	**0**	0.052	0.042	**-16**	0.321	0.809	**1518**	0.112	0.305	**510**
40	**10**	0.057	0.047	**-6**	0.347	0.859	**1618**	0.147	0.381	**662**
40	**0**	0.053	0.046	**-8**	0.495	0.975	**1850**	0.2	0.554	**1008**
0	**40**	0.052	0.056	**12**	0.625	1	**1900**	0.274	0.746	**1392**

Note. %Bias was calculated based on the Type I error when the sample sizes of the two groups were (100,1000).

Table 4. Type I error for each method with predefined censoring proportion and sample size where n_1 and n_2 are equal assumed Weibull distribution $T_1, T_2 \sim Weib(2,2)$

p_1	p_2	Appropriate method				Inappropriate method I				Inappropriate method II			
		(30,30)	(100,100)	(1000,1000)	**%Bias**	(30,30)	(100,100)	(1000,1000)	**%Bias**	(30,30)	(100,100)	(1000,1000)	**%Bias**
0	**0**	0.055	0.057	0.053	**6**	0.054	0.051	0.054	**8**	0.054	0.051	0.054	**8**
10	**10**	0.053	0.055	0.06	**20**	0.051	0.058	0.056	**12**	0.064	0.061	0.047	**-6**
20	**20**	0.058	0.052	0.057	**14**	0.056	0.054	0.058	**16**	0.052	0.053	0.048	**-4**
30	**30**	0.058	0.055	0.063	**26**	0.06	0.053	0.048	**-4**	0.054	0.055	0.051	**2**
40	**40**	0.055	0.057	0.052	**4**	0.067	0.056	0.052	**4**	0.059	0.054	0.045	**-10**
0	**10**	0.052	0.05	0.058	**16**	0.065	0.11	0.618	**1136**	0.051	0.058	0.106	**112**
10	**20**	0.061	0.055	0.054	**8**	0.073	0.126	0.65	**1200**	0.06	0.065	0.143	**186**
20	**30**	0.062	0.045	0.063	**26**	0.084	0.139	0.71	**1320**	0.057	0.068	0.161	**222**
30	**40**	0.051	0.056	0.054	**8**	0.094	0.141	0.78	**1460**	0.071	0.069	0.207	**314**
10	**0**	0.061	0.049	0.059	**18**	0.073	0.125	0.639	**1178**	0.054	0.074	0.132	**164**
20	**10**	0.053	0.049	0.057	**14**	0.079	0.117	0.689	**1278**	0.055	0.079	0.171	**242**
30	**20**	0.051	0.055	0.059	**18**	0.074	0.127	0.741	**1382**	0.058	0.064	0.19	**280**
40	**30**	0.053	0.057	0.057	**14**	0.075	0.135	0.823	**1546**	0.066	0.068	0.23	**360**
0	**20**	0.057	0.051	0.047	**-6**	0.119	0.308	0.998	**1896**	0.058	0.086	0.373	**646**

10	**30**	0.055	0.046	0.06	**20**	0.139	0.36	0.997	**1894**	0.069	0.102	0.449	**798**
20	**40**	0.053	0.052	0.052	**4**	0.153	0.387	0.999	**1898**	0.083	0.104	0.539	**978**
20	**0**	0.057	0.045	0.06	**20**	0.134	0.321	0.996	**1892**	0.062	0.097	0.424	**748**
30	**10**	0.047	0.054	0.052	**4**	0.127	0.353	0.999	**1898**	0.066	0.087	0.509	**918**
40	**20**	0.049	0.051	0.053	**6**	0.141	0.388	0.999	**1898**	0.074	0.106	0.604	**1108**
0	**30**	0.059	0.05	0.06	**20**	0.215	0.611	1	**1900**	0.071	0.138	0.747	**1394**
10	**40**	0.052	0.057	0.052	**4**	0.253	0.686	1	**1900**	0.098	0.157	0.848	**1596**
30	**0**	0.046	0.048	0.056	**12**	0.243	0.625	1	**1900**	0.081	0.144	0.789	**1478**
40	**10**	0.048	0.055	0.059	**18**	0.247	0.673	1	**1900**	0.099	0.167	0.859	**1618**
40	**0**	0.05	0.058	0.051	**2**	0.388	0.864	1	**1900**	0.112	0.233	0.973	**1846**
0	**40**	0.054	0.058	0.05	**0**	0.384	0.879	1	**1900**	0.103	0.23	0.962	**1824**

Note. %Bias was calculated based on the Type I error when the sample sizes of the two groups were (1000,1000).

Table 5. Type I error for each method with predefined censoring proportion and sample size where n_1 and n_2 are unequal assumed Weibull distribution $T_1, T_2 \sim Weib(2,2)$

p_1	p_2	Appropriate method			Inappropriate method I			Inappropriate method II		
		(30,100)	(100,1000)	%Bias	(30,100)	(100,1000)	%Bias	(30,100)	(100,1000)	%Bias
0	**0**	0.052	0.052	**4**	0.058	0.057	**14**	0.058	0.057	**14**
10	**10**	0.054	0.053	**6**	0.062	0.056	**12**	0.064	0.054	**8**
20	**20**	0.055	0.047	**-6**	0.059	0.063	**26**	0.05	0.056	**12**
30	**30**	0.056	0.041	**-18**	0.063	0.057	**14**	0.056	0.053	**6**
40	**40**	0.053	0.048	**-4**	0.063	0.05	**0**	0.06	0.055	**10**
0	**10**	0.054	0.051	**2**	0.105	0.201	**302**	0.067	0.087	**74**
10	**20**	0.052	0.055	**10**	0.106	0.204	**308**	0.075	0.081	**62**
20	**30**	0.054	0.044	**-12**	0.104	0.232	**364**	0.066	0.089	**78**
30	**40**	0.055	0.043	**-14**	0.105	0.245	**390**	0.069	0.104	**108**
10	**0**	0.053	0.051	**2**	0.066	0.143	**186**	0.055	0.055	**10**
20	**10**	0.054	0.046	**-8**	0.078	0.159	**218**	0.053	0.056	**12**
30	**20**	0.054	0.044	**-12**	0.087	0.17	**240**	0.054	0.059	**18**
40	**30**	0.052	0.043	**-14**	0.089	0.182	**264**	0.059	0.07	**40**
0	**20**	0.05	0.058	**16**	0.208	0.582	**1064**	0.092	0.156	**212**

10	**30**	0.052	0.049	**-2**	0.215	0.635	**1170**	0.091	0.177	**254**
20	**40**	0.048	0.047	**-6**	0.231	0.683	**1266**	0.096	0.204	**308**
20	**0**	0.05	0.044	**-12**	0.167	0.461	**822**	0.053	0.089	**78**
30	**10**	0.055	0.042	**-16**	0.183	0.498	**896**	0.058	0.103	**106**
40	**20**	0.054	0.047	**-6**	0.191	0.559	**1018**	0.065	0.114	**128**
0	**30**	0.049	0.048	**-4**	0.384	0.925	**1750**	0.119	0.264	**428**
10	**40**	0.046	0.048	**-4**	0.42	0.952	**1804**	0.136	0.346	**592**
30	**0**	0.052	0.042	**-16**	0.32	0.81	**1520**	0.063	0.152	**204**
40	**10**	0.057	0.047	**-6**	0.349	0.859	**1618**	0.089	0.193	**286**
40	**0**	0.053	0.046	**-8**	0.495	0.974	**1848**	0.105	0.29	**480**
0	**40**	0.052	0.056	**12**	0.625	1	**1900**	0.166	0.469	**838**

Note. %Bias was calculated based on the Type I error when the sample sizes of the two groups were (100,1000).

Table 6. Type 1 error for group 1: appropriate method (random generating), group 2: inappropriate generating method II assumed Exponential distribution $T_1, T_2 \sim Exp(5.873)$

p_1	p_2	$n_1 = n_2$				$n_1 \neq n_2$		
		$(30, 30)$	$(100, 100)$	$(1000, 1000)$	%Bias	$(30, 100)$	$(100, 1000)$	%Bias
0	**0**	0.054	0.051	0.054	**8**	0.058	0.057	**14**
10	**10**	0.048	0.068	0.198	**296**	0.072	0.105	**110**
20	**20**	0.066	0.12	0.672	**1244**	0.106	0.226	**352**
30	**30**	0.103	0.213	0.956	**1812**	0.167	0.422	**744**
40	**40**	0.163	0.378	1	**1900**	0.234	0.654	**1208**

Note. %Bias was calculated based on the type 1 error when the sample sizes of the two groups were (1000,1000) and (100,1000).

Table 7. Type 1 error for group 1: appropriate method (random generating), group 2: inappropriate generating method II assumed Weibull distribution $T_1, T_2 \sim Weib(2,2)$

$\mathbf{p_1}$	$\mathbf{p_2}$	$\boldsymbol{n_1 = n_2}$				$\boldsymbol{n_1 \neq n_2}$		
		$(\mathbf{30}, \mathbf{30})$	$(\mathbf{100}, \mathbf{100})$	$(\mathbf{1000}, \mathbf{1000})$	**%Bias**	$(\mathbf{30}, \mathbf{100})$	$(\mathbf{100}, \mathbf{1000})$	**%Bias**
0	**0**	0.054	0.051	0.054	**8**	0.058	0.057	**14**
10	**10**	0.047	0.058	0.111	**122**	0.061	0.089	**78**
20	**20**	0.046	0.084	0.36	**620**	0.082	0.142	**184**
30	**30**	0.074	0.133	0.729	**1358**	0.118	0.263	**426**
40	**40**	0.107	0.221	0.953	**1806**	0.159	0.442	**784**

Note. %Bias was calculated based on the type 1 error when the sample sizes of the two groups were (1000,1000) and (100,1000).

Figure 4. Type 1 error for group 1: appropriate method (random generating), group 2: inappropriate method II in two sample where n_1 and n_2 are equal

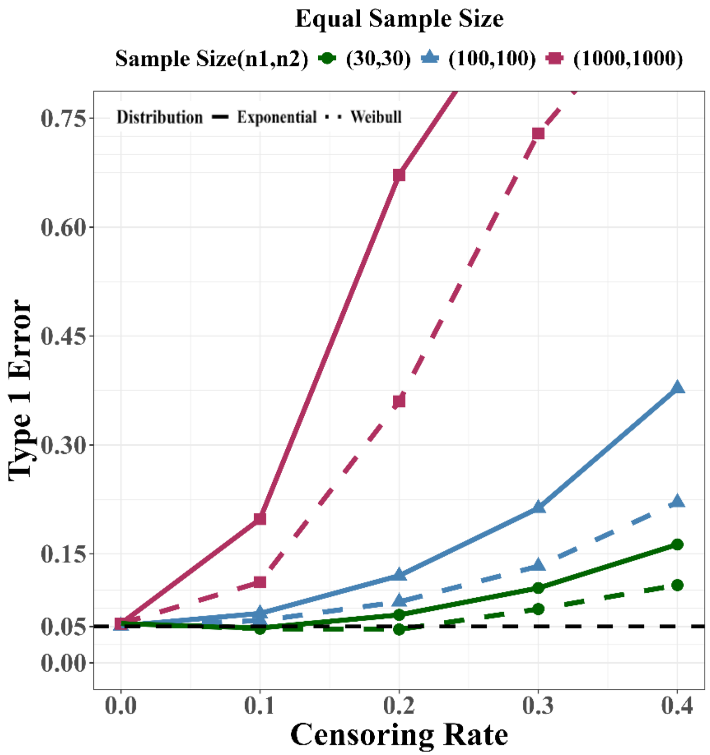


Figure 5. Type 1 error for group 1: appropriate method (random generating), group 2: inappropriate method II in two sample where n_1 and n_2 are unequal

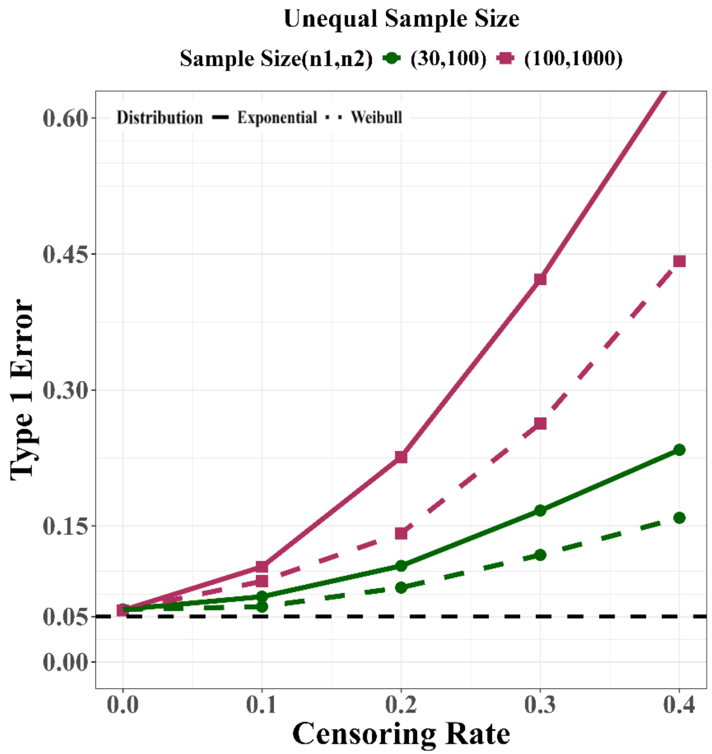


Table 8. Spearman correlation coefficient under each method in one sample assumed Exponential distribution $T \sim Exp(5.873)$

Method		p	n		
			30	100	1000
Appropriate method		**10**	-0.002	-0.001	0.003
		20	0.013	-0.011	-0.004
		30	0.046	-0.006	-0.005
		40	0.013	-0.007	0.001
Inappropriate method I	**C* (75% percentile)**	**10**	0.195	0.204	0.211
		20	0.389	0.4	0.401
		30	0.569	0.571	0.571
		40	0.714	0.716	0.719
	C* (85% percentile)	**10**	0.15	0.157	0.161
		20	0.32	0.321	0.318
		30	0.487	0.479	0.473
		40	0.643	0.629	0.628
	C* (95% percentile)	**10**	0.099	0.106	0.109
		20	0.237	0.231	0.227
		30	0.383	0.363	0.353
		40	0.532	0.5	0.498
Inappropriate method II	**C* (75% percentile)**	**10**	0.223	0.178	0.195
		20	0.339	0.347	0.344
		30	0.432	0.435	0.448
		40	0.516	0.522	0.531
	C* (85% percentile)	**10**	0.181	0.133	0.149
		20	0.278	0.279	0.274
		30	0.366	0.366	0.375
		40	0.459	0.459	0.465
	C* (95% percentile)	**10**	0.135	0.084	0.1
		20	0.205	0.202	0.195
		30	0.287	0.278	0.284
		40	0.38	0.373	0.374

Table 9. Spearman correlation coefficient under each method in one sample assumed Weibull distribution $T \sim Weib(2,2)$

Method		p	n: 30	100	1000
Appropriate method		10	-0.023	-0.014	0.001
		20	-0.005	0.005	-0.001
		30	0.02	0.023	0.003
		40	0.019	0.013	-0.005
Inappropriate method I	C* (75% percentile)	10	0.343	0.35	0.359
		20	0.46	0.477	0.48
		30	0.577	0.576	0.576
		40	0.668	0.665	0.669
	C* (85% percentile)	10	0.304	0.308	0.315
		20	0.416	0.427	0.427
		30	0.527	0.523	0.52
		40	0.624	0.612	0.614
	C* (95% percentile)	10	0.254	0.254	0.26
		20	0.357	0.36	0.36
		30	0.465	0.452	0.446
		40	0.562	0.538	0.539
Inappropriate method II	C* (75% percentile)	10	0.342	0.309	0.327
		20	0.382	0.394	0.394
		30	0.4	0.403	0.419
		40	0.414	0.426	0.436
	C* (85% percentile)	10	0.309	0.271	0.287
		20	0.346	0.354	0.351
		30	0.368	0.366	0.38
		40	0.386	0.395	0.402
	C* (95% percentile)	10	0.265	0.221	0.237
		20	0.296	0.299	0.296
		30	0.32	0.315	0.327
		40	0.347	0.35	0.356

5. Conclusion and discussion

A random generation method is generally used when generating survival data. In this approach, the event and censoring times are generated independently. The distribution parameters for censoring times are then determined based on the assumed two distributions under the condition of a fixed censoring proportion. We define this as the appropriate method. When both time distributions are Exponential, the censoring parameter can be calculated directly. However, when Weibull distributions are assumed, a closed-form expression cannot be derived. In such cases, if the shape parameters are assumed equal, the scale parameter can be determined. Otherwise, the scale parameter is calculated numerically using the root-finding Newton-Raphson algorithm. However, this process is time consuming and complex, leading to many instances in previous studies where survival data were inappropriately generated. Inappropriate data generation can result in misleading results. To address this, we conducted simulations and developed a theoretical framework based on two inappropriate methods.

The simulation demonstrated Type I error under various simulation settings for three different methods: one using random generation and the other two using inappropriate methods. Through this study, we aimed to investigate the inflation in Type I error when survival data were generated inappropriately.

The simulation results showed that when survival data were generated using an appropriate method, the Type I error was well controlled regardless of whether the censoring proportions were equal or unequal. In contrasts, when survival data were generated using an inappropriate method, the Type I error was not controlled in all cases where the censoring proportions were unequal. Furthermore, when one group's data was generated using the random generating method and the other group's data was generated using the inappropriate method II, it was observed that the Type I error was not controlled even when the censoring proportions of the two groups were equal. The Spearman correlation between the event times and censoring times showed that a correlation existed when generating inappropriate methods. The reason for the presence of a correlation in the inappropriate generating method is due to the fact that the data were generated dependently, not randomly.

Therefore, it is essential to appropriately generate survival data. Failure to consider this and using improperly generated data, such as a simulation section, can lead to erroneous results. This can be applied not only to the log-rank test statistic but also to other statistical measures. Moreover, further research is needed to explore new methods that can effectively control the Type I error even under dependently censored survival data.

Appendix

Appendix Table 1. Spearman correlation coefficient for appropriate method in two sample with n_1 and n_2 are equal and unequal assumed Exponential distribution $T_1, T_2 \sim Exp(5.873)$

p_1	p_2	$n_1 = n_2$			$n_1 \neq n_2$	
		(30,30)	(100,100)	(1000,1000)	(30,100)	(100,1000)
10	**10**	-0.023	-0.014	0.007	-0.012	-0.009
20	**20**	-0.006	-0.003	0.004	-0.005	-0.004
30	**30**	0.018	0.013	0.002	0.016	0.008
40	**40**	0.02	-0.009	0.002	0.015	-0.006
10	**20**	0.037	-0.02	-0.003	0.031	-0.011
20	**30**	-0.004	-0.005	-0.005	-0.009	-0.003
30	**40**	0.015	0.006	0.001	0.008	0.002
20	**10**	0.045	-0.002	0	0.034	0
30	**20**	0.013	0.007	0	0.012	0.004
40	**30**	-0.04	0	0	-0.032	0
10	**30**	-0.008	-0.016	-0.003	-0.005	-0.009
20	**40**	0.004	0.007	0.002	-0.002	0.003
30	**10**	-0.009	0.008	-0.003	-0.007	0.003
40	**20**	-0.004	0.018	0.001	-0.006	0.009
10	**40**	-0.002	0.014	-0.002	-0.001	0.009
40	**10**	0.001	-0.005	0.001	-0.001	-0.003

Appendix Table 2. Spearman correlation coefficient for appropriate method in two sample with n_1 and n_2 are equal and unequal assumed Weibull distribution $T_1, T_2 \sim Weib(2,2)$

p_1	p_2	$n_1 = n_2$			$n_1 \neq n_2$	
		(30,30)	(100,100)	(1000,1000)	(30,100)	(100,1000)
10	**10**	-0.024	-0.012	0.008	-0.012	-0.009
20	**20**	-0.004	-0.001	0.003	-0.004	-0.003
30	**30**	0.016	0.013	0.003	0.014	0.007
40	**40**	0.019	-0.008	0.001	0.014	-0.005
10	**20**	0.034	-0.019	-0.004	0.03	-0.01
20	**30**	-0.009	-0.007	-0.005	-0.011	-0.004
30	**40**	0.018	0.005	0.001	0.009	0.002
20	**10**	0.044	-0.001	0	0.032	0.001
30	**20**	0.013	0.007	0	0.012	0.004
40	**30**	-0.041	0	-0.001	-0.032	-0.001
10	**30**	-0.005	-0.014	-0.003	-0.004	-0.008
20	**40**	0.003	0.008	0.002	-0.001	0.003
30	**10**	-0.01	0.007	-0.003	-0.007	0.002
40	**20**	0	0.019	0.002	-0.002	0.01
10	**40**	-0.002	0.014	-0.002	-0.001	0.009
40	**10**	0	-0.004	0.001	-0.002	-0.003

Appendix Table 3. Spearman correlation coefficient for inappropriate method I in two sample with n_1 and n_2 are equal assumed Exponential distribution $T_1, T_2 \sim Exp(5.873)$

p_1	p_2	C* (75% percentile)			C* (85% percentile)			C* (95% percentile)		
		(30,30)	(100,100)	(1000,1000)	(30,30)	(100,100)	(1000,1000)	(30,30)	(100,100)	(1000,1000)
10	**10**	0.194	0.21	0.21	0.149	0.163	0.16	0.098	0.113	0.108
20	**20**	0.389	0.405	0.399	0.319	0.326	0.316	0.236	0.236	0.225
30	**30**	0.57	0.561	0.573	0.489	0.468	0.477	0.385	0.351	0.359
40	**40**	0.714	0.721	0.72	0.642	0.638	0.628	0.532	0.515	0.5
10	**20**	0.3	0.296	0.306	0.239	0.234	0.24	0.175	0.164	0.168
20	**30**	0.482	0.481	0.486	0.405	0.398	0.395	0.312	0.297	0.29
30	**40**	0.643	0.643	0.647	0.567	0.555	0.552	0.463	0.437	0.428
20	**10**	0.301	0.311	0.305	0.243	0.248	0.238	0.178	0.178	0.167
30	**20**	0.495	0.485	0.489	0.423	0.399	0.399	0.33	0.297	0.295
40	**30**	0.652	0.641	0.645	0.578	0.552	0.55	0.477	0.435	0.427
10	**30**	0.403	0.39	0.393	0.342	0.321	0.319	0.267	0.239	0.234
20	**40**	0.562	0.562	0.56	0.489	0.48	0.472	0.394	0.374	0.362
30	**10**	0.384	0.386	0.393	0.322	0.315	0.32	0.247	0.233	0.235
40	**20**	0.573	0.541	0.563	0.505	0.455	0.475	0.412	0.346	0.366
10	**40**	0.449	0.462	0.467	0.39	0.394	0.397	0.312	0.306	0.308
40	**10**	0.474	0.459	0.464	0.417	0.391	0.393	0.34	0.303	0.303

Appendix Table 4. Spearman correlation coefficient for inappropriate method I in two sample with n_1 and n_2 are unequal assumed Exponential distribution $T_1, T_2 \sim Exp(5.873)$

p_1	p_2	C* (75% percentile)		C* (85% percentile)		C* (95% percentile)	
		(30,100)	(100,1000)	(30,100)	(100,1000)	(30,100)	(100,1000)
10	**10**	0.203	0.21	0.156	0.162	0.105	0.111
20	**20**	0.402	0.404	0.328	0.324	0.24	0.234
30	**30**	0.563	0.566	0.476	0.471	0.363	0.353
40	**40**	0.716	0.721	0.637	0.634	0.517	0.508
10	**20**	0.299	0.302	0.236	0.237	0.167	0.166
20	**30**	0.488	0.483	0.405	0.396	0.305	0.292
30	**40**	0.649	0.645	0.567	0.554	0.454	0.432
20	**10**	0.301	0.307	0.244	0.242	0.177	0.172
30	**20**	0.482	0.487	0.405	0.4	0.308	0.297
40	**30**	0.646	0.643	0.565	0.551	0.455	0.431
10	**30**	0.398	0.389	0.329	0.317	0.247	0.233
20	**40**	0.563	0.562	0.486	0.477	0.383	0.367
30	**10**	0.384	0.387	0.321	0.316	0.246	0.233
40	**20**	0.568	0.552	0.494	0.467	0.397	0.357
10	**40**	0.455	0.464	0.391	0.393	0.302	0.304
40	**10**	0.473	0.46	0.415	0.391	0.335	0.302

Appendix Table 5. Spearman correlation coefficient for inappropriate method I in two sample with n_1 and n_2 are equal assumed Weibull distribution $T_1, T_2 \sim Weib(2,2)$

p_1	p_2	C* (75% percentile)			C* (85% percentile)			C* (95% percentile)		
		(30,30)	(100,100)	(1000,1000)	(30,30)	(100,100)	(1000,1000)	(30,30)	(100,100)	(1000,1000)
10	**10**	0.342	0.355	0.358	0.303	0.314	0.314	0.252	0.262	0.259
20	**20**	0.46	0.483	0.479	0.416	0.432	0.426	0.357	0.365	0.359
30	**30**	0.578	0.567	0.578	0.529	0.512	0.522	0.468	0.44	0.45
40	**40**	0.668	0.673	0.669	0.624	0.622	0.614	0.562	0.552	0.54
10	**20**	0.411	0.41	0.42	0.367	0.364	0.371	0.312	0.305	0.31
20	**30**	0.521	0.524	0.528	0.474	0.474	0.473	0.41	0.406	0.403
30	**40**	0.621	0.622	0.624	0.574	0.57	0.568	0.512	0.5	0.494
20	**10**	0.41	0.423	0.419	0.368	0.377	0.37	0.316	0.318	0.309
30	**20**	0.531	0.524	0.531	0.487	0.472	0.477	0.428	0.404	0.407
40	**30**	0.632	0.621	0.622	0.587	0.568	0.566	0.528	0.499	0.493
10	**30**	0.477	0.468	0.47	0.436	0.419	0.419	0.384	0.357	0.356
20	**40**	0.571	0.576	0.574	0.525	0.525	0.52	0.468	0.458	0.449
30	**10**	0.456	0.462	0.47	0.412	0.414	0.419	0.358	0.352	0.356
40	**20**	0.586	0.555	0.577	0.545	0.502	0.523	0.485	0.432	0.453
10	**40**	0.495	0.51	0.516	0.455	0.462	0.467	0.399	0.401	0.403
40	**10**	0.526	0.507	0.512	0.484	0.459	0.463	0.426	0.396	0.399

Appendix Table 6. Spearman correlation coefficient for inappropriate method II in two sample with n_1 and n_2 are unequal assumed Weibull distribution $T_1, T_2 \sim Weib(2,2)$

p_1	p_2	C* (75% percentile)		C* (85% percentile)		C* (95% percentile)	
		(30,100)	(100,1000)	(30,100)	(100,1000)	(30,100)	(100,1000)
10	**10**	0.35	0.356	0.309	0.314	0.257	0.26
20	**20**	0.474	0.483	0.428	0.431	0.366	0.364
30	**30**	0.571	0.571	0.519	0.516	0.45	0.443
40	**40**	0.667	0.672	0.619	0.619	0.55	0.546
10	**20**	0.41	0.415	0.365	0.368	0.307	0.307
20	**30**	0.526	0.525	0.475	0.472	0.409	0.403
30	**40**	0.625	0.622	0.575	0.569	0.508	0.496
20	**10**	0.414	0.421	0.37	0.373	0.314	0.313
30	**20**	0.523	0.528	0.475	0.475	0.411	0.406
40	**30**	0.626	0.622	0.577	0.568	0.512	0.496
10	**30**	0.472	0.466	0.427	0.417	0.368	0.353
20	**40**	0.572	0.575	0.525	0.523	0.461	0.453
30	**10**	0.459	0.465	0.414	0.416	0.356	0.353
40	**20**	0.581	0.567	0.537	0.513	0.474	0.443
10	**40**	0.5	0.511	0.457	0.463	0.396	0.399
40	**10**	0.522	0.509	0.479	0.46	0.422	0.396

Appendix Table 7. Spearman correlation coefficient for inappropriate method II in two sample with n_1 and n_2 are equal assumed Exponential distribution $T_1, T_2 \sim Exp(5.873)$

p_1	p_2	C* (75% percentile)			C* (85% percentile)			C* (95% percentile)		
		(30,30)	(100,100)	(1000,1000)	(30,30)	(100,100)	(1000,1000)	(30,30)	(100,100)	(1000,1000)
10	**10**	0.222	0.204	0.189	0.18	0.16	0.142	0.135	0.111	0.094
20	**20**	0.338	0.339	0.34	0.277	0.272	0.27	0.204	0.194	0.193
30	**30**	0.433	0.453	0.451	0.368	0.383	0.377	0.289	0.292	0.287
40	**40**	0.516	0.531	0.53	0.459	0.468	0.462	0.379	0.379	0.372
10	**20**	0.266	0.268	0.268	0.214	0.212	0.209	0.155	0.148	0.146
20	**30**	0.366	0.393	0.4	0.306	0.326	0.328	0.228	0.245	0.243
30	**40**	0.514	0.477	0.493	0.458	0.41	0.422	0.382	0.322	0.331
20	**10**	0.241	0.26	0.27	0.188	0.203	0.211	0.131	0.139	0.148
30	**20**	0.41	0.39	0.395	0.35	0.322	0.323	0.281	0.238	0.238
40	**30**	0.514	0.486	0.491	0.455	0.418	0.421	0.376	0.33	0.332
10	**30**	0.303	0.328	0.325	0.252	0.271	0.264	0.19	0.203	0.196
20	**40**	0.423	0.427	0.437	0.363	0.36	0.369	0.289	0.276	0.285
30	**10**	0.326	0.319	0.325	0.277	0.261	0.264	0.218	0.194	0.195
40	**20**	0.44	0.418	0.436	0.383	0.353	0.367	0.308	0.273	0.283
10	**40**	0.369	0.358	0.365	0.321	0.303	0.308	0.261	0.236	0.239
40	**10**	0.368	0.362	0.362	0.322	0.307	0.305	0.266	0.239	0.237

Appendix Table 8. Spearman correlation coefficient for inappropriate method II in two sample with n_1 and n_2 are unequal assumed Exponential distribution $T_1, T_2 \sim Exp(5.873)$

p_1	p_2	C* (75% percentile)		C* (85% percentile)		C* (95% percentile)	
		(30,100)	(100,1000)	(30,100)	(100,1000)	(30,100)	(100,1000)
10	**10**	0.212	0.201	0.168	0.156	0.122	0.108
20	**20**	0.344	0.343	0.28	0.275	0.205	0.196
30	**30**	0.435	0.451	0.366	0.38	0.28	0.289
40	**40**	0.521	0.532	0.462	0.466	0.377	0.377
10	**20**	0.265	0.268	0.211	0.211	0.15	0.148
20	**30**	0.377	0.395	0.314	0.326	0.232	0.242
30	**40**	0.508	0.483	0.448	0.415	0.365	0.326
20	**10**	0.252	0.267	0.2	0.209	0.139	0.146
30	**20**	0.393	0.394	0.33	0.324	0.253	0.24
40	**30**	0.507	0.49	0.444	0.421	0.362	0.332
10	**30**	0.311	0.324	0.256	0.265	0.187	0.196
20	**40**	0.433	0.433	0.371	0.365	0.293	0.281
30	**10**	0.322	0.318	0.271	0.259	0.208	0.192
40	**20**	0.439	0.425	0.378	0.359	0.299	0.276
10	**40**	0.361	0.362	0.31	0.305	0.244	0.238
40	**10**	0.357	0.36	0.31	0.303	0.25	0.235

Appendix Table 9. Spearman correlation coefficient for inappropriate method II in two sample with n_1 and n_2 are equal assumed Weibull distribution $T_1, T_2 \sim Weib(2,2)$

p_1	p_2	C* (75% percentile)			C* (85% percentile)			C* (95% percentile)		
		(30,30)	(100,100)	(1000,1000)	(30,30)	(100,100)	(1000,1000)	(30,30)	(100,100)	(1000,1000)
10	**10**	0.341	0.334	0.321	0.308	0.295	0.281	0.265	0.247	0.231
20	**20**	0.382	0.387	0.39	0.345	0.347	0.348	0.295	0.292	0.293
30	**30**	0.402	0.421	0.42	0.369	0.383	0.381	0.322	0.332	0.329
40	**40**	0.414	0.434	0.433	0.386	0.402	0.399	0.346	0.356	0.353
10	**20**	0.351	0.358	0.359	0.314	0.318	0.318	0.266	0.266	0.265
20	**30**	0.375	0.404	0.41	0.339	0.366	0.369	0.29	0.314	0.315
30	**40**	0.446	0.415	0.429	0.418	0.38	0.393	0.379	0.331	0.343
20	**10**	0.327	0.349	0.361	0.289	0.309	0.319	0.244	0.256	0.267
30	**20**	0.411	0.397	0.405	0.379	0.357	0.364	0.335	0.305	0.31
40	**30**	0.453	0.423	0.428	0.423	0.389	0.392	0.379	0.34	0.344
10	**30**	0.354	0.378	0.375	0.318	0.34	0.335	0.272	0.291	0.285
20	**40**	0.391	0.4	0.414	0.359	0.363	0.376	0.314	0.313	0.326
30	**10**	0.374	0.368	0.375	0.34	0.33	0.336	0.297	0.281	0.285
40	**20**	0.409	0.394	0.413	0.377	0.358	0.374	0.333	0.31	0.324
10	**40**	0.385	0.374	0.383	0.355	0.339	0.346	0.313	0.292	0.298
40	**10**	0.385	0.377	0.38	0.355	0.341	0.343	0.315	0.295	0.295

Appendix Table 10. Spearman correlation coefficient for inappropriate method II in two sample with n_1 and n_2 are unequal assumed Weibull distribution $T_1, T_2 \sim Weib(2,2)$

p_1	p_2	C* (75% percentile)		C* (85% percentile)		C* (95% percentile)	
		(30,100)	(100,1000)	(30,100)	(100,1000)	(30,100)	(100,1000)
10	**10**	0.336	0.332	0.3	0.293	0.254	0.243
20	**20**	0.388	0.392	0.351	0.351	0.299	0.296
30	**30**	0.404	0.42	0.369	0.381	0.319	0.33
40	**40**	0.421	0.435	0.391	0.402	0.349	0.356
10	**20**	0.353	0.359	0.314	0.319	0.265	0.266
20	**30**	0.385	0.404	0.348	0.365	0.297	0.312
30	**40**	0.442	0.421	0.411	0.385	0.368	0.336
20	**10**	0.337	0.356	0.298	0.315	0.251	0.263
30	**20**	0.399	0.403	0.363	0.363	0.313	0.31
40	**30**	0.447	0.427	0.414	0.392	0.368	0.343
10	**30**	0.361	0.374	0.323	0.336	0.275	0.285
20	**40**	0.404	0.408	0.371	0.371	0.324	0.321
30	**10**	0.37	0.369	0.335	0.33	0.289	0.28
40	**20**	0.412	0.403	0.377	0.366	0.329	0.317
10	**40**	0.378	0.377	0.346	0.341	0.301	0.295
40	**10**	0.376	0.377	0.344	0.341	0.301	0.293

Bibliography

Srivastava, Deo Kumar, et al. "Impact of unequal censoring and insufficient follow-up on comparing survival outcomes: Applications to clinical studies." Statistical Methods in Medical Research 30.9 (2021): 2057-2074.

Wan, Fei. "Simulating survival data with predefined censoring proportions for proportional hazards models." Statistics in Medicine 36.5 (2017): 838-854.

Kuss, Oliver, and Annika Hoyer. "A proportional risk model for time-to-event analysis in randomized controlled trials." Statistical Methods in Medical Research 30.2 (2021): 411-424.

Alam, Tasneem Fatima, M. Shafiqur Rahman, and Wasimul Bari. "On estimation for accelerated failure time models with small or rare event survival data." BMC Medical Research Methodology 22.1 (2022): 169.

Beltangady, Mohan S., and Ralph F. Frankowski. "Effect of unequal censoring on the size and power of the logrank and Wilcoxon types of tests for survival data." Statistics in Medicine 8.8 (1989): 937-945.

Wang, Rui, Stephen W. Lagakos, and Robert J. Gray. "Testing and interval estimation for two-sample survival comparisons with small sample sizes and unequal censoring." Biostatistics 11.4 (2010): 676-692.

Harrington, David P., and Thomas R. Fleming. "A class of rank test procedures for censored survival data." Biometrika 69.3 (1982): 553-566.

Fleming, Thomas R., David P. Harrington, and Margaret O'sullivan. "Supremum versions of the log-rank and generalized Wilcoxon statistics." Journal of the American statistical Association 82.397 (1987): 312-320.

Klein, John P., and Melvin L. Moeschberger. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media, 2006.

De Winter, Joost CF, Samuel D. Gosling, and Jeff Potter. "Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data." Psychological methods 21.3 (2016): 273.

국 문 요 약

부적절한 중도절단 데이터 생성 시 로그-순위 검정의 제 1종 오류 증가

생존 데이터를 생성할 때, 일부 데이터 생성 방식은 잘못된 결과를 초래할 수 있다. 일반적으로 무작위 생성(random generating)이라 불리는 데이터 생성 방법은 사건발생 시간(T)과 중도절단 시간(C)을 각각 독립적으로 분포를 가정하여 생성한다. 그러나, 예를 들어 중도절단율이 고정된 경우, 와이블 분포의 형상모수가 다를 때 중도절단 분포의 척도모수를 계산하기 위한 닫힌 형태의 식이 존재하지 않는다. 이를 단순화하기 위해 일부 연구에서는 부적절한 데이터 생성을 사용해왔다. 본 연구는 이로 인해 발생하는 문제를 증명과 시뮬레이션을 통해 확인하고자 한다. 구체적으로 두 그룹 간 로그-순위 검정의 1종 오류(Type I Error)와 사건발생 및 중도절단 시간 간 상관관계를 검토하였다.

부적절한 데이터 생성 방법 I은 사건발생 시간을 생성한 후, 베르누이 분포를 사용하여 중도절단 여부를 나타내는 지시함수를 생성한다. 중도절단이 발생하는 경우, 중도절단 시간은 생성된 사건발생 시간으로 대체한다. 부적절한 데이터 생성 방법 II에서는 중도절단 시간을 $Uniform(0, T_i)$ 분포를 기반으로 생성한다.

시뮬레이션 결과, 무작위 생성 방법에서는 사전에 정의된 중도절단율이 두 그룹 간

동일하거나 다른 경우 모두 로그-순위 검정의 1종 오류가 잘 통제되었다. 그러나 부적절한 방법에서는 두 그룹 간 중도절단율이 다른 경우 1종 오류가 증가하였다. 반면, 중도절단율이 동일한 경우 부적절한 방법에서도 1종 오류가 잘 통제되는 것으로 보였다. 이는 T와 C의 종속성으로 인해 C를 조건으로 한 T의 조건부 분포가 T의 주변 분포와 달라지면서, 그룹 간 실제 차이가 왜곡되고 로그-순위 검정에서 1종 오류가 잘 통제되는 것처럼 보이는 결과를 초래한다.

이를 확인하기 위해 추가 시뮬레이션을 진행하였다. 한 그룹은 적절한 방법으로 데이터를 생성하고, 다른 그룹은 부적절한 방법 II로 생성하였다. 그 결과, 두 그룹 간 중도절단율이 동일한 경우에도 1종 오류가 통제되지 않았다. 이는 그룹 간 중도절단율의 차이 때문이 아닌, 부적절한 데이터 생성과정에서 비롯된 문제임을 나타낸다. 또한 사건발생 시간과 중도절단 시간 간의 스피어만 상관계수를 확인한 결과, 부적절한 데이터 생성 방식이 사건발생 시간과 중도절단 시간 간에 종속성이 존재함을 확인하였다.

핵심되는 말: 중도절단 데이터, 무작위 생성, 부적절한 데이터 생성, 로그-순위 검정, 제 1종 오류