Journal of Statistical Research *https://doi.org/10.47302/jsr.2021550107*
2021, Vol. 55, No. 1, pp. 75-100 *ISSN 0256 - 422 X*

A FLEXIBLE NON-PARAMETRIC PROCEDURE FOR TESTING CUMULATIVE HAZARDS WITH APPLICATION TO ONCOLOGY STUDIES

YIMING ZHANG

Department of Statistics, University of Connecticut, Storrs, CT 06226, USA
Email: yiming.3.zhang@uconn.edu

ABIDEMI K. ADENIJI

M-Estimator LLC, Boston, MA 20115, USA
Email: adeniji@m-estimator.com

MING-HUI CHEN*

Department of Statistics, University of Connecticut, Storrs, CT 06226, USA
Email: ming-hui.chen@uconn.edu

SUMMARY

The hazard function is the probability of an event per unit of time for ever smaller time intervals. It has applications to a number of industries including drug development, engineering, finance, insurance and commerce to name a few. We focus on clinical trials, but more specifically, within the therapeutic area of oncology. Here, the hazard function is an important measure that can quantify the changes to the risk of mortality or cancer over time. It is a common and important tool for clinical trial practitioners. In this paper, we develop new non-parametric procedures for testing cumulative hazard functions. From the asymptotic properties of the Kaplan-Meier estimators, we propose procedures that construct test statistics for different tests of hypotheses, including testing if a cumulative hazard function follows a partially known-form hazard, and testing the proportional hazards assumption between two independent samples. Our testing approaches are very flexible since they allow us to choose the testing period and to specify any partially known-form distribution. In addition, the approximate asymptotic distributions of the test statistics are derived under both the null hypothesis and the alternative hypothesis, respectively. Extensive simulation studies show that the proposed procedures enjoy a reasonable Type-I error control and good statistical power under different censoring scenarios. The proposed methodology is further applied to examine the gender-specific mortality hazard rates for young adults with acute myeloid leukemia using the SEER database.

Keywords and phrases: cumulative hazard function, non-parametric test, time-to-event data, oncology study

* Corresponding author

1 Introduction

The hazard function is important to survival analysis because it describes the chance of an event of interest (for example, death or the recurrence of disease) in the next instant in time. Testing a hazard function with a known-form hazard is primarily achieved by testing the observed data via a corresponding distribution. For instance, to test a constant hazard, the exponentiality test is commonly used in practice. Existing statistical methodologies to test exponentiality for time-to-event data are mostly goodness-of-fit tests considering censored data. One of the most commonly used tests was proposed by Hollander and Proschan (1979) to compare the empirical distribution with a completely specified distribution. Hollander and Pena (1992) and Li and Doss (1993) proposed the Pearson-type Chi-square goodness-of-fit test. In recent literature, Han et al. (2017) used a piecewise exponential approach to identify violations to exponentiality with a better control of Type-I error under various censoring mechanisms. Smuts et al. (2019) proposed tests for exponentiality based on a conditional second-moment characterization approach.

The comparison of two hazard functions, the hazard ratio (HR), is a common measure in survival analysis. However, the HR between two samples only makes sense if the two hazard functions are proportional to each other. In addition, the proportional hazards (PH) assumption is important to the famous Cox's regression model (Cox, 1972). Therefore, the PH test is widely discussed in the literature. Existing methods include Wei (1984), Gill and Schumacher (1987), Dabrowska et al. (1989), Grambsch and Therneau (1994), Deshpande and Sengupta (1995), Sahoo and Sengupta (2016), and Xue et al. (2020), etc. Besides, a hypothesis test for an increasing hazard ratio is developed by Sahoo and Sengupta (2017).

In oncology studies, the progression of disease from diagnosis is not consistent, therefore, a constant hazard function over time is not always expected. However, there's interest on testing whether the hazard function follows a particular form within a certain period. For instance, as illustrated in Figure 1, the data generated from a piecewise exponential distribution with change-points at $t = 2$ and $t = 20$ have a constant hazard function during $t \in [2, 20]$. However, the aforementioned goodness-of-fit tests cannot be applied to test the constant hazard function between $t = 2$ and $t = 20$. Similarly, clinicians may also be interested in knowing whether the PH assumption holds within a certain period, rather than the full spectrum of time. For instance, some immunotherapies have been shown to have delayed clinical effects as compared to cytotoxic chemotherapy (Small et al., 2006; Hodi et al., 2010; Chen, 2013). Therefore, the hazards between the treatment and control groups are non-proportional, however the PH assumption may still hold after the delayed period, allowing for an evaluation of the clinical benefit by the HR. To the best of our knowledge, few methodologies in the literature address testing hazard functions within a prespecified period. To address this challenge, we propose flexible non-parametric procedures and develop the corresponding test statistics for testing (i) a partially specified known-form hazard, and (ii) the proportional hazards assumption on a prespecified period.

In this paper, we propose the flexible non-parametric procedures for testing hazard functions based on the asymptotic properties of the Kaplan-Meier (KM) estimator (Kaplan and Meier, 1958). We introduce a test statistic to evaluate whether a particular hazard function follows a known-form hazard on a selected period, this test statistic is a formulation of the KM estimators at the prespecified

knots of the testing period. An F approximation is applied to correct the Type-I inflation problem. The proposed approach is very flexible in practice since the form of the hazard function is only needed to be partially specified, and it is not restricted to exponentiality, it can be specified as the forms from other distributions. We then extend the testing procedure to a two-sample proportional hazards testing problem under a prespecified testing period. Extensive simulation studies show that the proposed methods enjoy a reasonable Type-I error control and good power in both the one-sample and two-sample testing problems. The proposed methodology is further applied to examine the gender-specific mortality hazards for young adults with acute myeloid leukemia (AML) using the Surveillance, Epidemiology, and End Results (SEER) registry system.

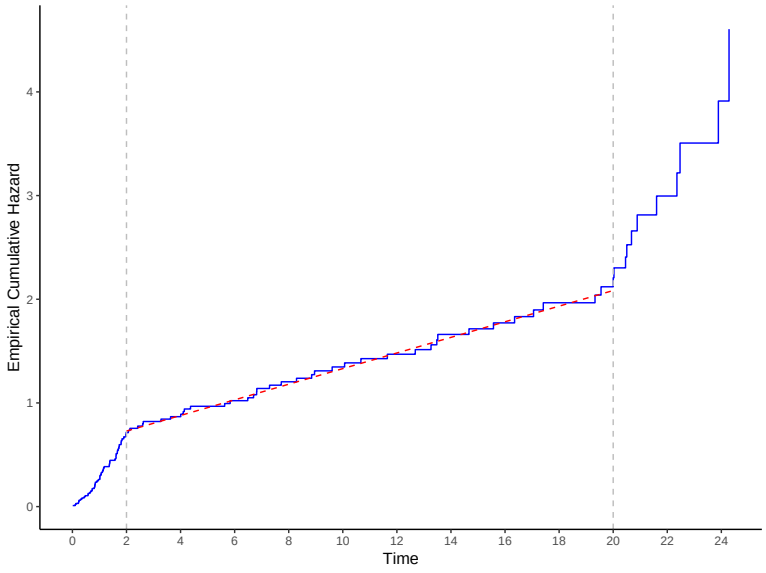

Figure 1: Testing constant hazard in a selected period of time.

The rest of the paper is organized as follows. Section 2 introduces the proposed non-parametric testing procedures for both the one-sample and two-sample problems, including the testing procedures and the distributions of the test statistics under both the null and alternative hypotheses. Section 3 presents simulation studies to examine the empirical performance of the proposed testing procedures on the Type-I error rate control and statistical power. In Section 4, we apply the proposed methodology to the AML data in young adults from the SEER database. The last chapter is a discussion on the proposed methods, it is provided in Section 5.

2 Proposed Method

2.1 One-sample problem: testing a known-form hazard

2.1.1 Hypothesis specification

Suppose there are n independent subjects in a time-to-event dataset. Our main goal is to test whether the hazard function of the sample follows a partially known form within a prespecified period, such that $h(t) = \lambda h_0(t)$ when $t \in [t_l, t_u]$, where λ is an unknown parameter, and t_l and t_u are the lower and upper bounds of the testing period. The form of $h_0(t)$ is determined by the testing problem, for instance, $h_0(t) = 1$ for testing a constant hazard. Then, the hypothesis is equivalently written as for $t \in [t_l, t_u]$,

$$\begin{aligned} &H_0 : H(t) = \lambda H_0(t) - \lambda H_0(t_l) + H(t_l) \text{ versus} \\ &H_A : H(t) \neq \lambda H_0(t) - \lambda H_0(t_l) + H(t_l), \end{aligned} \tag{2.1}$$

where $H(t) = \int_0^t h(u)du$ denote the cumulative hazard function of the sample and $H_0(t) = \int_0^t h_0(u)du$ is the partially known cumulative hazard form. The null hypothesis assumes that the cumulative hazard function follows a specific form $\lambda H_0(t) - \lambda H_0(t_l) + H(t_l)$ between t_l and t_u. If $t_l = 0$ and $t_u = \infty$, the hypothesis (2.1) reduces to

$$H_0 : H(t) = \lambda H_0(t) \text{ versus } H_A : H(t) \neq \lambda H_0(t). \tag{2.2}$$

The hypotheses (2.1) and (2.2) allow for a flexible choice of the cumulative hazard function in the null hypothesis by specifying any $H_0(t)$. For instance, we can choose $H_0(t) = t$ to test a constant hazard, or $H_0(t) = t^\alpha$ to test if $H(t)$ has the same form as a Weibull distribution with a known shape parameter α.

In pursuit of testing the hypothesis (2.1), we start with a partition of the domain of $t \in [t_l, t_u]$, where the prespecified knots are denoted as $t_l = t_0 < t_1 < t_2 < \ldots, t_{k-1} < t_k = t_u$. The survival rates of the knots are denoted as $\boldsymbol{\theta} = (S_{t_0}, S_{t_1}, S_{t_2}, \ldots, S_{t_k})^T$. Consider a function of $\boldsymbol{\theta}$, given by

$$g(\boldsymbol{\theta}) = \left(\frac{\Delta H(t_1)}{\Delta H_0(t_1)}, \frac{\Delta H(t_2)}{\Delta H_0(t_2)}, \ldots, \frac{\Delta H(t_k)}{\Delta H_0(t_k)} \right)^T, \tag{2.3}$$

where $\Delta H(t_i) = H(t_i) - H(t_{i-1}) = -\log S_{t_i} - (-\log S_{t_{i-1}})$, and $\Delta H_0(t_i) = H_0(t_i) - H_0(t_{i-1})$ for $i = 1, 2, \ldots, k$. With respect to each element in $g(\boldsymbol{\theta})$, the numerator is the difference of the true cumulative hazards ($H(t)$'s) between two adjacent knots, while the denominator is the difference of the $H_0(t)$'s specified in the null hypothesis between two adjacent knots. Under the null hypothesis (2.1), all the elements of $g(\boldsymbol{\theta})$ are consistent to the unknown parameter λ. Thus, we re-specify the hypothesis as

$$H_0 : \boldsymbol{c}g(\boldsymbol{\theta}) = \mathbf{0} \text{ versus } H_A : \boldsymbol{c}g(\boldsymbol{\theta}) \neq \mathbf{0}, \tag{2.4}$$

where $\boldsymbol{c}$ is a $(k-1)\times k$ contrast matrix given by

$$\boldsymbol{c} = \begin{bmatrix} 1 & & & & -1 \\ & 1 & & \mathbf{0} & -1 \\ & & 1 & & -1 \\ & \mathbf{0} & & \ddots & \vdots \\ & & & 1 & -1 \end{bmatrix}. \tag{2.5}$$

If the null hypothesis in (2.1) is false, we should reject the null hypothesis in (2.4) as well.

2.1.2 Testing procedures

We construct the testing statistic based on the KM estimator and its asymptotic properties which are broadly discussed in Breslow and Crowley (1974), Wang et al. (1986), and Tsai et al. (1987).

Theorem 1. *Let $\hat{\boldsymbol{\theta}} = (\hat{S}_{t_0}, \hat{S}_{t_1}, \hat{S}_{t_2}, \ldots, \hat{S}_{t_k})^T$ denote the Kaplan-Meier estimators of $\boldsymbol{\theta}$. Under the null hypothesis (2.4), we have*

$$T_1 = n(\boldsymbol{c}g(\hat{\boldsymbol{\theta}}))^T(\boldsymbol{c}\nabla g(\hat{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}\nabla g(\hat{\boldsymbol{\theta}})\boldsymbol{c}^T)^{-1}\boldsymbol{c}g(\hat{\boldsymbol{\theta}}) \xrightarrow{D} \chi^2_{k-1} \tag{2.6}$$

as $n\to\infty$, where $\boldsymbol{\Sigma}$ is a $(k+1)\times(k+1)$ covariance matrix and $\nabla g(\hat{\boldsymbol{\theta}})$ denotes the gradient of $g(\hat{\boldsymbol{\theta}})$ with respect to $\hat{\boldsymbol{\theta}}$.

Proof. Following Breslow and Crowley (1974), the asymptotic properties of the KM estimators imply that $\sqrt{n}(\hat{S}(t)-S(t))$ converges in distribution to a Gaussian process with expectation 0 and a covariance function. Thus, the vector of the KM estimators, $\hat{\boldsymbol{\theta}}$, follows an asymptotic multivariate normal distribution such that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0},\boldsymbol{\Sigma}) \quad \text{as } n\to\infty. \tag{2.7}$$

The asymptotic distribution of $g(\hat{\boldsymbol{\theta}})$ can be obtained by the Delta method (Casella and Berger, 2001) as

$$\sqrt{n}\big(g(\hat{\boldsymbol{\theta}})-g(\boldsymbol{\theta})\big) \xrightarrow{D} N\big(\mathbf{0},\nabla g(\hat{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}\nabla g(\hat{\boldsymbol{\theta}})\big) \quad \text{as } n\to\infty. \tag{2.8}$$

The gradient of $g(\hat{\boldsymbol{\theta}})$ with respect to $\hat{\boldsymbol{\theta}}$ is a $(k+1)\times k$ matrix given by

$$\nabla g(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \frac{1}{\Delta H_0(t_1)\hat{S}_{t_0}} & 0 & 0 & 0 & \ldots & 0 & 0 \\ \frac{-1}{\Delta H_0(t_1)\hat{S}_{t_1}} & \frac{1}{\Delta H_0(t_2)\hat{S}_{t_1}} & 0 & 0 & \ldots & 0 & 0 \\ 0 & \frac{-1}{\Delta H_0(t_2)\hat{S}_{t_2}} & \frac{1}{\Delta H_0(t_3)\hat{S}_{t_2}} & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \frac{-1}{\Delta H_0(t_{k-1})\hat{S}_{t_{k-1}}} & \frac{1}{\Delta H_0(t_k)\hat{S}_{t_{k-1}}} \\ 0 & 0 & 0 & 0 & \ldots & 0 & \frac{1}{\Delta H_0(t_k)\hat{S}_{t_k}} \end{bmatrix}.$$

Thus, under the null hypothesis (2.4) and $\boldsymbol{c}g(\boldsymbol{\theta}) = \mathbf{0}$, (2.6) directly follows from (2.8). □

We then substitute the covariance matrix $\boldsymbol{\Sigma}$ by its estimator $\widehat{\boldsymbol{\Sigma}}$ to obtain an approximation of the test statistic T_1 under the null hypothesis, which is given by

$$T_1^* = n(\boldsymbol{c}g(\hat{\boldsymbol{\theta}}))^T(\boldsymbol{c}\nabla g(\hat{\boldsymbol{\theta}})^T\widehat{\boldsymbol{\Sigma}}\nabla g(\hat{\boldsymbol{\theta}})\boldsymbol{c}^T)^{-1}\boldsymbol{c}g(\hat{\boldsymbol{\theta}}) \overset{\text{approx}}{\sim} \chi^2_{k-1}. \tag{2.9}$$

In practice, the elements in the estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}$ can be obtained by the Greenwood formula (Breslow and Crowley, 1974; Wang et al., 1986), denoted as $\widehat{\boldsymbol{\Sigma}} = n\tilde{\Sigma}$. The diagonal elements in $\tilde{\Sigma}$ are given by the Greenwood formula such that $\widehat{\text{Var}}[\hat{S}_{t_i}] = \hat{S}^2_{t_i}\sum_{t^{(h)}\le t_i}\frac{d_h}{n_h(n_h-d_h)}$, where d_h and n_h denote the number of death and the number of patients at risk at $t^{(h)}$, respectively. The off-diagonal elements in $\tilde{\Sigma}$ are obtained by $\widehat{\text{Cov}}(\hat{S}_{t_j}, \hat{S}_{t_i}) = \frac{\hat{S}_{t_i}}{\hat{S}_{t_j}}\widehat{\text{Var}}(\hat{S}_{t_j})$, where $t_j < t_i$ and $i, j = 0, 1, 2, \ldots, k$.

Remark 1. According to Mushfiqur Rashid et al. (2000), the Chi-square critical value, derived from the asymptotic theory, allows too many rejections under the null hypothesis and inflates the significance level. Thus, to circumvent this problem, we consider an alternative testing statistic given by

$$T_2^* = \frac{T_1^*}{n} \times \frac{n-k+2}{k-1} \overset{\text{approx}}{\sim} F_{k-1,n-k+2} \tag{2.10}$$

under the null hypothesis. The proof of (2.10) directly follows with the result from Muirhead (1982). We recommend to use T_2^* in practice as we find that the F correction does correct the alpha inflation problem in our numerical studies.

Remark 2. When $t_0 = 0$ or there are no events between 0 and t_0, the estimate of $\hat{S}_{t_0}$ is fixed because $\hat{S}_{t_0} = 1$ and $-log(\hat{S}_{t_0}) = 0$, which make $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ have one fewer element. The $k \times k$ matrix of the gradient of $g(\hat{\boldsymbol{\theta}})$ with respect to $\hat{\boldsymbol{\theta}} = (\hat{S}_{t_1}, \hat{S}_{t_2}, \ldots, \hat{S}_{t_k})^T$ is given by

$$\nabla g(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \frac{-1}{\Delta H_0(t_1)\hat{S}_{t_1}} & \frac{1}{\Delta H_0(t_2)\hat{S}_{t_1}} & 0 & 0 & \ldots & 0 & 0 \\ 0 & \frac{-1}{\Delta H_0(t_2)\hat{S}_{t_2}} & \frac{1}{\Delta H_0(t_3)\hat{S}_{t_2}} & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \frac{-1}{\Delta H_0(t_{k-1})\hat{S}_{t_{k-1}}} & \frac{1}{\Delta H_0(t_k)\hat{S}_{t_{k-1}}} \\ 0 & 0 & 0 & 0 & \ldots & 0 & \frac{1}{\Delta H_0(t_k)\hat{S}_{t_k}} \end{bmatrix}.$$

Remark 3. If there are no events in $[t_{i-1}, t_i)$ where $i > 1$, the inverse of the covariance matrix might have a singularity problem. To correct this problem, we drop the knot at t_{i-1} and combine $[t_{i-1}, t_i)$ with $[t_{i-2}, t_{i-1})$. Before conducting the hypothesis test, we repeat this procedure until all the intervals have at least one event.

2.1.3 Power of the test

The distribution of the test statistic under a given alternative hypothesis can be derived in a similar procedure. Suppose that the hazard function under the alternative hypothesis is a known distribution, that is,

$$H_A : \boldsymbol{c}g(\boldsymbol{\theta}) = \boldsymbol{\mu}, \tag{2.11}$$

where $\boldsymbol{\mu}$ is a known vector. Given the fixed knots $t_0, t_1, t_2, \ldots, t_k$, the distribution of the test statistic under the alternative hypothesis is given as follows.

Corollary 2.1. Given $\boldsymbol{c}g(\boldsymbol{\theta}) = \boldsymbol{\mu}$, the distribution of T_1^* is an approximated non-central Chi-square distribution given by

$$T_1^* = n(\boldsymbol{c}g(\hat{\boldsymbol{\theta}}))^T(\boldsymbol{c}\nabla g(\hat{\boldsymbol{\theta}})^T\widehat{\boldsymbol{\Sigma}}\nabla g(\hat{\boldsymbol{\theta}})\boldsymbol{c}^T)^{-1}\boldsymbol{c}g(\hat{\boldsymbol{\theta}}) \overset{\text{approx}}{\sim} \chi^2_{k-1,\nu}, \tag{2.12}$$

where $\nu = \frac{n}{2}\boldsymbol{\mu}^T(\boldsymbol{c}\nabla g(\hat{\boldsymbol{\theta}})^T\widehat{\boldsymbol{\Sigma}}\nabla g(\hat{\boldsymbol{\theta}})\boldsymbol{c}^T)^{-1}\boldsymbol{\mu}$ is the noncentrality parameter.

Similar to (2.10), the distribution of T_2^* under the alternative hypothesis follows an approximated non-central F distribution with degree of freedom $k-1$ and $n-k+2$, and a non-central parameter ν. Therefore, given $\widehat{\boldsymbol{\Sigma}}$, $g(\hat{\boldsymbol{\theta}})$, and $\nabla g(\hat{\boldsymbol{\theta}})$ estimated from the pilot studies, the power of the test can be calculated by $1 - P(F^*_{k-1,n-k+2,1-\alpha})$, where P is the cumulative distribution function of the aforementioned non-central F distribution, and $F^*_{k-1,n-k+2,1-\alpha}$ denotes the F critical value at a $1-\alpha$ confidence level under the null hypothesis.

2.2 Two-sample problem: testing proportional hazards

2.2.1 Hypothesis specification

Suppose that there are n_1 and n_2 independent subjects in the two independent groups, respectively. In order to test if the hazard functions of the two groups are proportional to each other in a prespecified testing period $[t_l, t_u]$, the null hypothesis is specified as for $t \in [t_l, t_u]$,

$$H_0 : \frac{H_1(t) - H_1(t_l)}{H_2(t) - H_2(t_l)} = \lambda \text{ versus } H_A : \frac{H_1(t) - H_1(t_l)}{H_2(t) - H_2(t_l)} \neq \lambda, \tag{2.13}$$

where t_l and t_u are the lower and upper bounds of the testing period, $H_1(t)$ and $H_2(t)$ denote the cumulative hazard functions of group 1 and group 2, respectively, and λ is the constant unknown hazard ratio between these two independent samples. When $t_l = 0$ and $t_u = \infty$, the hypothesis (2.13) reduces to

$$H_0 : \frac{H_1(t)}{H_2(t)} = \lambda \text{ versus } H_A : \frac{H_1(t)}{H_2(t)} \neq \lambda. \tag{2.14}$$

Similar to the testing procedure for the one-sample problem, a partition of the testing period is prespecified with $k+1$ knots distributed on the domain of t, denoted as $t_l = t_0 < t_1 < t_2 < \ldots, t_{k-1} < t_k = t_u$. The survival rates at the knots in the two groups are denoted by $\boldsymbol{\theta}_1 =$

$(S_{1t_0}, S_{1t_1}, S_{1t_2}, \ldots, S_{1t_k})^T$ and $\boldsymbol{\theta}_2 = (S_{2t_0}, S_{2t_1}, S_{2t_2}, \ldots, S_{2t_k})^T$, respectively, and $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$. Consider a function of $\boldsymbol{\theta}$, given by

$$g(\boldsymbol{\theta}) = \left(\frac{\Delta H_1(t_1)}{\Delta H(t_1)}, \frac{\Delta H_1(t_2)}{\Delta H(t_2)}, \ldots, \frac{\Delta H_1(t_k)}{\Delta H(t_k)} \right)^T, \tag{2.15}$$

where $\Delta H(t_i) = \Delta H_1(t_i) + \Delta H_2(t_i)$, and $\Delta H_1(t_i) = H_1(t_i) - H_1(t_{i-1}) = -\log S_{1t_i} - (-\log S_{1t_{i-1}})$ and $\Delta H_2(t_i) = H_2(t_i) - H_2(t_{i-1}) = -\log S_{2t_i} - (-\log S_{2t_{i-1}})$ are the differences of the cumulative hazard functions between the two adjacent knots in group 1 and group 2, respectively. If the proportional hazards assumption holds, then all of elements $g(\boldsymbol{\theta})$ will be consistent to $\lambda/(1+\lambda)$. Thus, in similar fashion to the one-sample problem, we test the following hypothesis

$$H_0 : \boldsymbol{c}g(\boldsymbol{\theta}) = \mathbf{0} \text{ versus } H_A : \boldsymbol{c}g(\boldsymbol{\theta}) \neq \mathbf{0}, \tag{2.16}$$

where $\boldsymbol{c}$ is a $(k-1) \times k$ contrast matrix given in (2.5).

2.2.2 Testing procedures

Theorem 2. *Let $\hat{\boldsymbol{\theta}} = (\hat{S}_{1t_0}, \hat{S}_{1t_1}, \hat{S}_{1t_2}, \ldots, \hat{S}_{1t_k}, \hat{S}_{2t_0}, \hat{S}_{2,t_1}, \hat{S}_{2,t_2}, \ldots, \hat{S}_{2,t_k})^T$ denote the Kaplan-Meier estimators of $\boldsymbol{\theta}$, $N = \frac{n_1 n_2}{n_1+n_2}$, $\frac{n_1}{n_1+n_2} \to m_1$ and $\frac{n_2}{n_1+n_2} \to m_2$ as $n_1 \to \infty$ and $n_2 \to \infty$. Assume that*

$$0 < m_1 < 1 \text{ and } 0 < m_2 < 1. \tag{2.17}$$

Then, under the null hypothesis (2.16), we have

$$T_3 = N(\boldsymbol{c}g(\hat{\boldsymbol{\theta}}))^T (\boldsymbol{c}\nabla g(\hat{\boldsymbol{\theta}})^T \boldsymbol{\Sigma} \nabla g(\hat{\boldsymbol{\theta}}) \boldsymbol{c}^T)^{-1} \boldsymbol{c}g(\hat{\boldsymbol{\theta}}) \xrightarrow{D} \chi^2_{k-1} \tag{2.18}$$

as $n_1 \to \infty$ and $n_2 \to \infty$. $\nabla g(\hat{\boldsymbol{\theta}})$ denotes the gradient of $g(\hat{\boldsymbol{\theta}})$ with respect to $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}$ is a $(2k+2) \times (2k+2)$ covariance matrix given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} m_2 \boldsymbol{\Sigma}_1 & \mathbf{0}_{k+1} \\ \mathbf{0}_{k+1} & m_1 \boldsymbol{\Sigma}_2 \end{bmatrix}, \tag{2.19}$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the asymptotic covariance matrix of $\sqrt{n_1}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)$ and $\sqrt{n_2}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)$, respectively, and $\mathbf{0}_{k+1}$ is a $(k+1) \times (k+1)$ zero matrix.

Proof. Following the asymptotic properties of the KM estimators (2.7), we have

$$\begin{aligned} \sqrt{n_1}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) &\xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_1) \quad \text{as } n_1 \to \infty, \\ \sqrt{n_2}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) &\xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_2) \quad \text{as } n_2 \to \infty. \end{aligned}$$

With the assumption (2.17), we have

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2.20}$$

where $\boldsymbol{\Sigma}$ is given in (2.19). By applying the Delta method, the asymptotic distribution of $g(\hat{\boldsymbol{\theta}})$ is written as

$$\sqrt{N}(g(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \xrightarrow{D} N(\mathbf{0}, \nabla g(\hat{\boldsymbol{\theta}})^T \boldsymbol{\Sigma} \nabla g(\hat{\boldsymbol{\theta}})) \quad \text{as } n_1 \to \infty \text{ and } n_2 \to \infty, \tag{2.21}$$

where $\nabla g(\hat{\boldsymbol{\theta}})$ is the gradient of $g(\hat{\boldsymbol{\theta}})$ with respect to $\hat{\boldsymbol{\theta}}$ with dimension $(2k+2) \times k$, which is given by

$$\nabla g(\hat{\boldsymbol{\theta}}) = \begin{bmatrix}
\frac{\Delta H_2(t_1)}{\hat{S}_{1t_0}\Delta H(t_1)^2} & 0 & 0 & 0 & \dots & 0 & 0 \\
\frac{-\Delta H_2(t_1)}{\hat{S}_{1t_1}\Delta H(t_1)^2} & \frac{\Delta H_2(t_2)}{\hat{S}_{1t_1}\Delta H(t_2)^2} & 0 & 0 & \dots & 0 & 0 \\
0 & \frac{-\Delta H_2(t_2)}{\hat{S}_{1t_2}\Delta H(t_2)^2} & \frac{\Delta H_2(t_3)}{\hat{S}_{1t_2}\Delta H(t_3)^2} & 0 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \frac{-\Delta H_2(t_{k-1})}{\hat{S}_{1t_{k-1}}\Delta H(t_{k-1})^2} & \frac{\Delta H_2(t_k)}{\hat{S}_{1t_{k-1}}\Delta H(t_k)^2} \\
0 & 0 & 0 & 0 & \dots & 0 & \frac{-\Delta H_2(t_k)}{\hat{S}_{1t_k}\Delta H(t_k)^2} \\
\frac{-\Delta H_1(t_1)}{\hat{S}_{2t_0}\Delta H(t_1)^2} & 0 & 0 & 0 & \dots & 0 & 0 \\
\frac{\Delta H_1(t_1)}{\hat{S}_{2t_1}\Delta H(t_1)^2} & \frac{-\Delta H_1(t_2)}{\hat{S}_{2t_1}\Delta H(t_2)^2} & 0 & 0 & \dots & 0 & 0 \\
0 & \frac{\Delta H_1(t_2)}{\hat{S}_{2t_2}\Delta H(t_2)^2} & \frac{-\Delta H_1(t_3)}{\hat{S}_{2t_2}\Delta H(t_3)^2} & 0 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \frac{\Delta H_1(t_{k-1})}{\hat{S}_{2t_{k-1}}\Delta H(t_{k-1})^2} & \frac{-\Delta H_1(t_k)}{\hat{S}_{2t_{k-1}}\Delta H(t_k)^2} \\
0 & 0 & 0 & 0 & \dots & 0 & \frac{\Delta H_1(t_k)}{\hat{S}_{2t_k}\Delta H(t_k)^2}
\end{bmatrix}.$$

Then, under the null hypothesis $\boldsymbol{c}g(\boldsymbol{\theta}) = \mathbf{0}$, the asymptotic chi-square distribution of T_3 directly follows with (2.21) under the null hypothesis $\boldsymbol{c}g(\boldsymbol{\theta}) = \mathbf{0}$. □

Theorem 3. *The test statistic T_3 is invariant if $H_1(t)$ and $H_2(t)$ in (2.13) are switched.*

Proof. In the test statistic T_3, it is obvious to see $\nabla g(\hat{\boldsymbol{\theta}})^T \boldsymbol{\Sigma} \nabla g(\hat{\boldsymbol{\theta}})$ is invariant if we switch $H_1(t)$ and $H_2(t)$ in (2.13). In addition, let

$$g_1(\boldsymbol{\theta}) = \left(\frac{\Delta H_1(t_1)}{\Delta H(t_1)}, \frac{\Delta H_1(t_2)}{\Delta H(t_2)}, \dots, \frac{\Delta H_1(t_k)}{\Delta H(t_k)} \right)^T,$$

$$g_2(\boldsymbol{\theta}) = \left(\frac{\Delta H_2(t_1)}{\Delta H(t_1)}, \frac{\Delta H_2(t_2)}{\Delta H(t_2)}, \dots, \frac{\Delta H_2(t_k)}{\Delta H(t_k)} \right)^T,$$

and let $\mathbf{1}_k$ denote a $k \times 1$ vector with all elements equal to 1. Then, we have

$$\begin{aligned} g_1(\boldsymbol{\theta}) &= \mathbf{1}_k - g_2(\boldsymbol{\theta}) \\ \boldsymbol{c}g_1(\boldsymbol{\theta}) &= \boldsymbol{c}\mathbf{1}_k - \boldsymbol{c}g_2(\boldsymbol{\theta}) = -\boldsymbol{c}g_2(\boldsymbol{\theta}). \end{aligned}$$

Therefore, $T_3 = N(\boldsymbol{c}g(\hat{\boldsymbol{\theta}}))^T(\boldsymbol{c}\nabla g(\hat{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}\nabla g(\hat{\boldsymbol{\theta}})\boldsymbol{c}^T)^{-1}\boldsymbol{c}g(\hat{\boldsymbol{\theta}})$ is invariant if $H_1(t)$ and $H_2(t)$ in (2.13) are switched. □

By replacing $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ with their estimates $\widehat{\boldsymbol{\Sigma}}_1$ and $\widehat{\boldsymbol{\Sigma}}_2$, we approximate the test statistic T_3 by T_3^* and we obtain

$$T_3^* = N(\boldsymbol{c}g(\hat{\boldsymbol{\theta}}))^T(\boldsymbol{c}\nabla g(\hat{\boldsymbol{\theta}})^T\widehat{\boldsymbol{\Sigma}}\nabla g(\hat{\boldsymbol{\theta}})\boldsymbol{c}^T)^{-1}\boldsymbol{c}g(\hat{\boldsymbol{\theta}}) \overset{\text{approx}}{\sim} \chi^2_{k-1} \tag{2.22}$$

under the null hypothesis. In practice, $\widehat{\boldsymbol{\Sigma}}_1$ and $\widehat{\boldsymbol{\Sigma}}_2$ can be derived from the Greenwood formula. Like the one-sample problem, we adopt an F approximation to correct the Type-I error inflation that's present when testing is based on the asymptotic Chi-square distribution. Let $N^* = [N]$ denote the integer part of N, the test statistic based on an approximated F distribution under the null hypothesis is given by

$$T_4^* = \frac{T_3^*}{N^*} \times \frac{N^* - k + 2}{k-1} \overset{\text{approx}}{\sim} F_{k-1,N^*-k+2}. \tag{2.23}$$

Remark 4. The gradient matrix $\nabla g(\hat{\boldsymbol{\theta}})$ needs to be adjusted when $t_0 = 0$ since the KM estimates of $\hat{S}_{1t_0}$ and $\hat{S}_{2t_0}$ are fixed to be 0. The $2k \times k$ matrix of the gradient of $g(\hat{\boldsymbol{\theta}})$ with respect to $\hat{\boldsymbol{\theta}} = (\hat{S}_{1t_1}, \hat{S}_{1t_2}, \ldots, \hat{S}_{1t_k}, \hat{S}_{2,t_1}, \hat{S}_{2,t_2}, \ldots, \hat{S}_{2,t_k})^T$ is given by

$$\nabla g(\hat{\boldsymbol{\theta}}) = \begin{bmatrix}
\frac{-\Delta H_2(t_1)}{\hat{S}_{1t_1}\Delta H(t_1)^2} & \frac{\Delta H_2(t_2)}{\hat{S}_{1t_1}\Delta H(t_2)^2} & 0 & 0 & \ldots & 0 & 0 \\
0 & \frac{-\Delta H_2(t_2)}{\hat{S}_{1t_2}\Delta H(t_2)^2} & \frac{\Delta H_2(t_3)}{\hat{S}_{1t_2}\Delta H(t_3)^2} & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \frac{-\Delta H_2(t_{k-1})}{\hat{S}_{1t_{k-1}}\Delta H(t_{k-1})^2} & \frac{\Delta H_2(t_k)}{\hat{S}_{1t_{k-1}}\Delta H(t_k)^2} \\
0 & 0 & 0 & 0 & \ldots & 0 & \frac{-\Delta H_2(t_k)}{\hat{S}_{1t_k}\Delta H(t_k)^2} \\
\frac{\Delta H_1(t_1)}{\hat{S}_{2t_1}\Delta H(t_1)^2} & \frac{-\Delta H_1(t_2)}{\hat{S}_{2t_1}\Delta H(t_2)^2} & 0 & 0 & \ldots & 0 & 0 \\
0 & \frac{\Delta H_1(t_2)}{\hat{S}_{2t_2}\Delta H(t_2)^2} & \frac{-\Delta H_1(t_3)}{\hat{S}_{2t_2}\Delta H(t_3)^2} & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \frac{\Delta H_1(t_{k-1})}{\hat{S}_{2t_{k-1}}\Delta H(t_{k-1})^2} & \frac{-\Delta H_1(t_k)}{\hat{S}_{2t_{k-1}}\Delta H(t_k)^2} \\
0 & 0 & 0 & 0 & \ldots & 0 & \frac{\Delta H_1(t_k)}{\hat{S}_{2t_k}\Delta H(t_k)^2}
\end{bmatrix}.$$

Remark 5. If there are no events in either of the groups in $[t_{i-1}, t_i)$, the inverse of $\widehat{\boldsymbol{\Sigma}}$ may be singular. To correct this problem, we drop the knot at t_{i-1} and combine $[t_{i-1}, t_i)$ with $[t_{i-2}, t_{i-1})$. Before conducting the hypothesis test, we repeat this procedure until all the intervals have at least one event in each group.

The distribution of the test statistic under a given alternative hypothesis can be derived like the one-sample test we discussed earlier. The distribution of T_4^* under the alternative hypothesis (2.11) is an approximated non-central F distribution with degrees of freedom $k-1$ and $n-k+2$

and a noncentrality parameter $\nu = \frac{N^*}{2}\boldsymbol{\mu}^T(\boldsymbol{c}\nabla g(\hat{\boldsymbol{\theta}})^T\widehat{\boldsymbol{\Sigma}}\nabla g(\hat{\boldsymbol{\theta}})\boldsymbol{c}^T)^{-1}\boldsymbol{\mu}$. Therefore, the power of the hypothesis test (2.23) is given by $1 - P(F^*_{k-1,N^*-k+2,1-\alpha})$, where P is the cumulative distribution function of the non-central F distribution with noncentrality parameter ν, and $F^*_{k-1,N^*-k+2,1-\alpha}$ denotes the F critical value at a $1-\alpha$ confidence level under the null hypothesis.

3 Simulation Studies

3.1 Simulation study 1: testing a partially known-form hazard

To implement the proposed procedure for testing a partially known-form hazard function, we need to prespecify the following parameters: the known part of the cumulative hazard $H_0(t)$, the testing period $[t_l, t_u]$, the number of knots k, and the partition of the testing period. The testing period and $H_0(t)$ are determined by the hypothesis testing problem, while k and $t_1, \ldots, t_{k-1}$ can be specified in different ways. In this simulation study, we investigate the empirical performance of the proposed test under different approaches to determine k and $t_1, \ldots, t_{k-1}$.

3.1.1 Type-I error simulation

Let α and γ denote the shape and scale parameter of a Weibull distribution, respectively. We generate data from three different distributions: an exponential distribution (constant hazard), a Weibull distribution with $\alpha > 1$ (increasing hazard), and a Weibull distribution with $\alpha < 1$ (decreasing hazard). The specifications of these parameters in these three scenarios are given in Table 1. The sample sizes (n) are 100, 200, 300, 500, and 1,000. To evaluate the performance of the proposed test under different censoring cases, we simulate data with no censoring, 20% random censoring, and 40% random censoring. For each of these three censoring cases, 50,000 random samples are generated, and then the proposed test is performed for each of the random samples at the Type-I error rate of 0.05. The empirical Type-I error rate is computed by the proportion of rejected null hypotheses among all the simulation samples.

The performances of the proposed test with different k's and locations of the knots are evaluated in each simulation scenario. We select k to be 3, 5, 7, 10, and 15. Two partitioning approaches (PA) to divide the testing period are considered. The partition approaches are (i) PA1: partitioning the testing period evenly, and (ii) PA2: partitioning the testing period based on the quantile of event times of the true distributions. In practice, PA1 is the simplest way to determine the locations of knots without any preliminary information of the true hazard functions to be tested. However, if we have prior knowledge of the true hazard functions from a pilot study, PA2 may be a better approach since it can prevent the collapse of intervals due to the lack of events. We perform the test using T_2^* instead of T_1^* since the F correction works well in preventing the Type-I error inflation issue.

The empirical Type-I error rates of the proposed test using the simulated data sets with no censoring are reported in Table 2. The empirical Type-I error rates are reasonably controlled when the selected value of k is small. We observe a slight inflation in the Type-I error when the number of knots are larger under PA1. The Type-I error inflation issue is much milder if we partition the testing period using PA2. Thus, PA2 is preferred if we have prior information on the distribution of the data

Table 1: Parameter settings of the Type-I error simulation in Simulation Study 1.

Hazard form	Distribution	α	γ	testing period	$H_0(t)$
Constant hazard	Exponential	1	3	[0, 6.91]	t
Increasing hazard	Weibull	1.5	3	[0, 5.23]	$t^{1.5}$
Decreasing hazard	Weibull	0.75	3	[0, 9.12]	$t^{0.75}$

being tested. Figure 2 displays the comparison between the empirical distribution of the test statistic T_2^*, and its theoretical distribution under the null hypothesis, which confirms that the F distribution is a good approximation for T_2^*. Tables 3 and 4 present the Type-I error rates of the proposed test under 20% and 40% censoring rates, respectively. Inflation of the empirical Type-I error rates is observed as the censoring rate increases, but the tests using a smaller k and PA2 have better resistance to the random censoring.

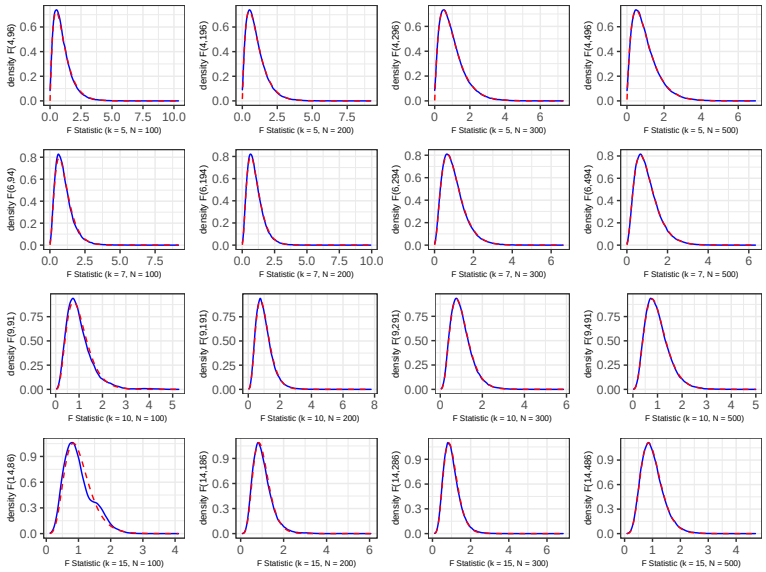

Figure 2: Comparison between the empirical and theoretical distributions of T_2^* under the null hypothesis. Blue solid lines denote the empirical distributions and red dashed lines denote the theoretical F distributions.

3.1.2 Power simulation

The objective of this simulation study is to evaluate the empirical performance on statistical power of our proposed one-sample test under different settings. We generate random samples from Weibull

Table 2: Empirical Type-I error rates with no censoring in Simulation Study 1.

	PA1					PA2				
$n\backslash k$	3	5	7	10	15	3	5	7	10	15
Exponential										
100	0.0501	0.0650	0.0683	0.0573	0.0305	0.0431	0.0443	0.0477	0.0618	0.0576
200	0.0480	0.0578	0.0698	0.0884	0.0873	0.0453	0.0465	0.0462	0.0537	0.0622
300	0.0491	0.0518	0.0629	0.0752	0.1036	0.0451	0.0470	0.0500	0.0521	0.0571
500	0.0499	0.0534	0.0575	0.0655	0.0799	0.0476	0.0496	0.0484	0.0506	0.0546
1000	0.0501	0.0519	0.0540	0.0559	0.0640	0.0496	0.0503	0.0499	0.0496	0.0518
Weibull $\alpha = 1.5$										
100	0.0446	0.0498	0.0587	0.0607	0.0391	0.0431	0.0443	0.0477	0.0618	0.0576
200	0.0462	0.0483	0.0524	0.0639	0.0802	0.0453	0.0465	0.0462	0.0537	0.0622
300	0.0458	0.0471	0.0524	0.0600	0.0724	0.0451	0.0470	0.0500	0.0521	0.0571
500	0.0478	0.0506	0.0522	0.0545	0.0603	0.0476	0.0496	0.0484	0.0506	0.0546
1000	0.0484	0.0486	0.0497	0.0514	0.0529	0.0496	0.0503	0.0499	0.0496	0.0518
Weibull $\alpha = 0.75$										
100	0.0593	0.0783	0.0718	0.0583	0.0315	0.0431	0.0443	0.0477	0.0618	0.0576
200	0.0524	0.0697	0.0876	0.1044	0.0945	0.0453	0.0465	0.0462	0.0537	0.0622
300	0.0521	0.0626	0.0746	0.0993	0.1203	0.0451	0.0470	0.0500	0.0521	0.0571
500	0.0522	0.0571	0.0643	0.0792	0.1086	0.0476	0.0496	0.0484	0.0506	0.0546
1000	0.0498	0.0521	0.0571	0.0626	0.0759	0.0496	0.0503	0.0499	0.0496	0.0518

PA1: partitioning the testing period evenly;

PA2: partitioning the testing period based on the quantile of events of the true distributions.

Table 3: Empirical Type-I error rates with 20% censoring rate in Simulation Study 1.

	PA1					PA2				
$n\backslash k$	3	5	7	10	15	3	5	7	10	15
Exponential										
100	0.0558	0.0650	0.0586	0.0436	0.0230	0.0396	0.0456	0.0524	0.0666	0.0262
200	0.0533	0.0698	0.0844	0.0864	0.0770	0.0440	0.0463	0.0496	0.0576	0.0778
300	0.0519	0.0627	0.0769	0.0961	0.1004	0.0464	0.0494	0.0490	0.0556	0.0658
500	0.0494	0.0580	0.0665	0.0828	0.1134	0.0464	0.0488	0.0494	0.0515	0.0579
1000	0.0495	0.0545	0.0577	0.0664	0.0825	0.0490	0.0496	0.0515	0.0512	0.0524
Weibull $\alpha = 1.5$										
100	0.0436	0.0581	0.0570	0.0501	0.0277	0.0394	0.0456	0.0522	0.0666	0.0262
200	0.0478	0.0533	0.0632	0.0758	0.0795	0.0443	0.0461	0.0496	0.0577	0.0778
300	0.0468	0.0531	0.0585	0.0708	0.0905	0.0466	0.0493	0.0493	0.0556	0.0659
500	0.0484	0.0515	0.0543	0.0610	0.0759	0.0466	0.0491	0.0494	0.0516	0.0581
1000	0.0493	0.0503	0.0527	0.0540	0.0609	0.0487	0.0492	0.0515	0.0514	0.0524
Weibull $\alpha = 0.75$										
100	0.0662	0.0667	0.0571	0.0446	0.0236	0.0394	0.0454	0.0521	0.0671	0.0264
200	0.0607	0.0861	0.0912	0.0879	0.0751	0.0447	0.0457	0.0502	0.0572	0.0775
300	0.0556	0.0733	0.0981	0.1118	0.1066	0.0466	0.0492	0.0497	0.0552	0.0661
500	0.0532	0.0670	0.0804	0.1067	0.1322	0.0467	0.0491	0.0495	0.0518	0.0576
1000	0.0515	0.0584	0.0637	0.0780	0.1036	0.0487	0.0492	0.0513	0.0512	0.0525

PA1: partitioning the testing period evenly;

PA2: partitioning the testing period based on the quantile of events of the true distributions.

Table 4: Empirical Type-I error rates with 40% censoring rate in Simulation Study 1.

	PA1					PA2				
$n\backslash k$	3	5	7	10	15	3	5	7	10	15
Exponential										
100	0.0725	0.0499	0.0414	0.0311	0.0160	0.0502	0.0609	0.0635	0.0482	0.0101
200	0.0710	0.0759	0.0697	0.0629	0.0537	0.0461	0.0551	0.0643	0.0769	0.0844
300	0.0605	0.0851	0.0853	0.0841	0.0776	0.0458	0.0521	0.0587	0.0703	0.0911
500	0.0577	0.0744	0.0940	0.1007	0.1059	0.0483	0.0521	0.0542	0.0612	0.0775
1000	0.0528	0.0617	0.0732	0.0959	0.1209	0.0485	0.0498	0.0527	0.0553	0.0613
Weibull $\alpha = 1.5$										
100	0.0651	0.0527	0.0454	0.0356	0.0164	0.0504	0.0614	0.0635	0.0477	0.0100
200	0.0553	0.0732	0.0694	0.0659	0.0595	0.0469	0.0552	0.0645	0.0773	0.0839
300	0.0506	0.0671	0.0797	0.0796	0.0772	0.0454	0.0525	0.0587	0.0711	0.0921
500	0.0514	0.0608	0.0695	0.0872	0.0939	0.0485	0.0524	0.0542	0.0616	0.0770
1000	0.0515	0.0546	0.0596	0.0697	0.0904	0.0484	0.0494	0.0531	0.0552	0.0613
Weibull $\alpha = 0.75$										
100	0.0524	0.0497	0.0389	0.0311	0.0171	0.0511	0.0616	0.0635	0.0471	0.0099
200	0.0804	0.0758	0.0685	0.0611	0.0516	0.0468	0.0553	0.0645	0.0778	0.0846
300	0.0746	0.0849	0.0840	0.0849	0.0742	0.0456	0.0526	0.0586	0.0715	0.0919
500	0.0652	0.0925	0.0994	0.1085	0.1061	0.0491	0.0523	0.0546	0.0622	0.0778
1000	0.0572	0.0724	0.0913	0.1208	0.1362	0.0481	0.0493	0.0525	0.0559	0.0620

PA1: partitioning the testing period evenly;

PA2: partitioning the testing period based on the quantile of events of the true distribution.

distributions with scale parameter $\gamma = 3$ and different shape parameters $\alpha \in [0.5, 2]$. Then we conduct the proposed test at the Type-I error rate of 0.05 under the three null hypotheses: H_{10} : $H(t) = \lambda t$, $H_{20} : H(t) = \lambda t^{1.5}$, and $H_{30} : H(t) = \lambda t^{0.75}$. The empirical power estimate of each simulation scenario is calculated by the proportion of rejected null hypotheses among the 10,000 simulation samples. The sample size (n) is 300, and the testing periods are set as 0 to the 90% quantile of the true distributions. No censoring is considered in this simulation. Similar to the settings considered in Section 3.1.1, five choices of k and two approaches to determine the locations of knots are considered in this simulation.

Figure 3 shows the power curves of the proposed test for different numbers of knots and different approaches to partition time. A higher power is achieved when the true shape parameter is deviate from the shape parameter specified in the null hypothesis. The tests using PA1 and PA2 have comparable power when the Type-I error rates are controlled at 0.05. When partitioning the testing period using PA2, a higher power is observed with a smaller k.

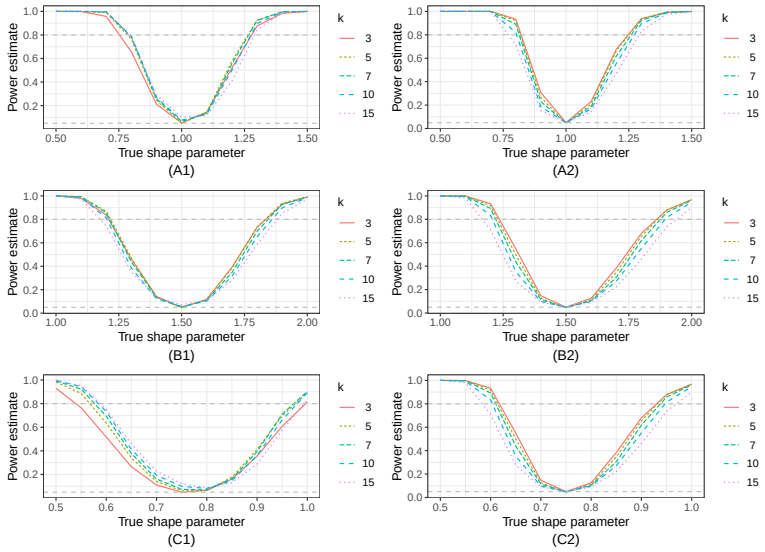

Figure 3: Empirical power curves for (A1) testing H_{10} with PA1; (A2) testing H_{10} with PA2; (B1) testing H_{20} with PA1; (B2) testing H_{20} with PA2; (C1) testing H_{30} with PA1; (C2) testing H_{30} with PA2.

3.2 Simulation study 2: testing the proportional hazard assumption

Similar to the one-sample test, we prespecify the following parameters: the testing period $[0, t_u]$, the number of knots k, and the location of each knot in the testing period to test the proportional hazard assumption. When we are interested in testing if the proportional hazards assumption holds for the entire time in practice, t_u may be chosen as the 95% empirical quantile of the events in both two groups. In this simulation study, we investigate the performance of the proposed procedure for

testing the proportional hazards with various numbers of knots and different approaches to partition the testing period.

3.2.1 Type-I error simulation

In this study, we generate two groups of data from the following three scenarios: exponential distributions with $\gamma_1 = 3$ and $\gamma_2 = 4$; Weibull distributions with $\alpha = 1.5$, $\gamma_1 = 4$ and $\gamma_2 = 5$; Weibull distributions with $\alpha = 0.75$, $\gamma_1 = 1.5$ and $\gamma_2 = 2.5$. In each scenario, the two distributions meet the proportional hazards assumption. The sample sizes per group in the simulation are $n = 100, 200, 300, 500,$ and $1,000$. Fifty thousand simulation runs were generated. Three censoring scenarios are considered in each simulation scenario, which are: no censoring, 20% random censoring, and 40% random censoring. The values of k for the test are 3, 5, 7, 10, and 15. Two approaches to partition the testing period are considered in this simulation study, namely, PA1: partitioning the testing period evenly, and PA2: partitioning the testing period based on the empirical quantile of events in both groups. Similar to Simulation Study 1, we select T_4^* instead of T_3^* to prevent the Type-I error inflation issue using the F correction.

Tables 5 - 7 present the Type-I error rates of the proposed test under the scenarios of no censoring, 20% random censoring, and 40% random censoring, respectively. Like Simulation Study 1, Type-I error inflation is observed in the tests with large k and PA1 (determining the location of knots). The issue gets more severe as the censoring rate increases. However, partitioning time using PA2 controls the Type-I error rate well in all cases of k. It still performs well with an increasing proportion of censored data and has better resistance to a high censoring rate compared with PA1. Thus, PA2 is recommended to partition the time axis in practice.

3.2.2 Power simulation

We study the characteristic of the power of the proposed PH test with different prespecified settings in this simulation study. For each random sample, one group is generated from an exponential distribution with the scale parameter $\lambda_1 = 3$, while the other group is generated from a Weibull distribution with the same scale parameter $\lambda_2 = 3$ and a different shape parameter $\alpha_2 \in [0.5, 2]$. The proportional hazards assumption is violated when α_2 is deviate from 1. The proposed PH testing procedure is then performed in each random sample at the Type-I error rate of 0.05. The empirical power is estimated as the proportion of rejections of the null hypothesis among all the random samples. The sample size (n) is 200. Ten thousand random samples are generated for each simulation scenario. In addition, no censoring is considered in this power simulation. Similar to the Type-I error simulation, five different numbers of k and two approaches to determine the location of knots are considered.

Figure 4 displays the empirical power curves of the proposed PH test with different k's and different approaches to partition the testing period. The proposed test reaches a higher power as the shape parameter of group 2 gets far away from 1. When the Type-I error is well controlled at 0.05 in both PA1 and PA2 ($k = 3$), the testing power using PA2 is higher than that using PA1 to partition the time axis. Among the tests with PA2, a higher power is observed when a smaller number of knots

Table 5: Empirical Type-I error rates with no censoring in Simulation Study 2.

	PA1					PA2				
$n\backslash k$	3	5	7	10	15	3	5	7	10	15
Exponential										
100	0.0601	0.0736	0.0768	0.0687	0.0480	0.0443	0.0431	0.0429	0.0424	0.0465
200	0.0615	0.0708	0.0793	0.0902	0.0929	0.0506	0.0474	0.0482	0.0466	0.0463
300	0.0567	0.0652	0.0718	0.0842	0.0984	0.0494	0.0489	0.0490	0.0475	0.0472
500	0.0534	0.0582	0.0619	0.0724	0.0867	0.0500	0.0494	0.0495	0.0473	0.0477
1000	0.0531	0.0539	0.0564	0.0592	0.0694	0.0505	0.0502	0.0486	0.0511	0.0506
Weibull $\alpha = 1.5$										
100	0.0464	0.0534	0.0554	0.0575	0.0398	0.0418	0.0408	0.0411	0.0408	0.0427
200	0.0528	0.0570	0.0622	0.0666	0.0742	0.0495	0.0492	0.0478	0.0458	0.0460
300	0.0527	0.0560	0.0571	0.0625	0.0716	0.0493	0.0499	0.0490	0.0489	0.0461
500	0.0528	0.0525	0.0552	0.0570	0.0627	0.0506	0.0509	0.0487	0.0481	0.0465
1000	0.0512	0.0510	0.0518	0.0550	0.0554	0.0510	0.0509	0.0498	0.0508	0.0503
Weibull $\alpha = 0.75$										
100	0.0620	0.0757	0.0808	0.0736	0.0552	0.0400	0.0383	0.0395	0.0400	0.0410
200	0.0689	0.0847	0.1009	0.1099	0.1097	0.0482	0.0477	0.0457	0.0446	0.0456
300	0.0623	0.0764	0.0890	0.1038	0.1232	0.0490	0.0496	0.0479	0.0483	0.0437
500	0.0577	0.0663	0.0768	0.0910	0.1122	0.0499	0.0507	0.0496	0.0499	0.0472
1000	0.0537	0.0574	0.0626	0.0713	0.0868	0.0516	0.0495	0.0501	0.0508	0.0495

PA1: partitioning the testing period evenly;

PA2: partitioning the testing period based on the empirical quantile of the events.

Table 6: Empirical Type-I error rates with 20% censoring rate in Simulation Study 2.

	PA1					PA2				
$n\backslash k$	3	5	7	10	15	3	5	7	10	15
Exponential										
100	0.0688	0.0817	0.0841	0.0776	0.0541	0.0481	0.0496	0.0520	0.0592	0.0572
200	0.0653	0.0762	0.0880	0.1016	0.0992	0.0518	0.0515	0.0511	0.0545	0.0626
300	0.0586	0.0689	0.0800	0.0954	0.1114	0.0511	0.0506	0.0517	0.0533	0.0572
500	0.0560	0.0609	0.0681	0.0836	0.1019	0.0491	0.0519	0.0500	0.0520	0.0539
1000	0.0526	0.0571	0.0585	0.0653	0.0769	0.0517	0.0517	0.0524	0.0520	0.0533
Weibull $\alpha = 1.5$										
100	0.0532	0.0632	0.0702	0.0666	0.0481	0.0466	0.0501	0.0512	0.0597	0.0543
200	0.0563	0.0630	0.0700	0.0784	0.0876	0.0521	0.0509	0.0500	0.0543	0.0611
300	0.0544	0.0605	0.0652	0.0727	0.0851	0.0518	0.0509	0.0525	0.0522	0.0556
500	0.0531	0.0547	0.0579	0.0649	0.0775	0.0497	0.0520	0.0521	0.0510	0.0558
1000	0.0518	0.0545	0.0531	0.0588	0.0631	0.0510	0.0501	0.0525	0.0511	0.0525
Weibull $\alpha = 0.75$										
100	0.0692	0.0858	0.0875	0.0821	0.0628	0.0446	0.0474	0.0505	0.0577	0.0511
200	0.0731	0.0932	0.1064	0.1180	0.1182	0.0516	0.0511	0.0512	0.0539	0.0616
300	0.0676	0.0845	0.1010	0.1183	0.1340	0.0505	0.0522	0.0508	0.0524	0.0547
500	0.0597	0.0728	0.0852	0.1031	0.1304	0.0506	0.0521	0.0522	0.0524	0.0531
1000	0.0559	0.0609	0.0662	0.0795	0.0981	0.0513	0.0502	0.0508	0.0508	0.0526

PA1: partitioning the testing period evenly;

PA2: partitioning the testing period based on the empirical quantile of the events.

Table 7: Empirical Type-I error rates with 40% censoring rate in Simulation Study 2.

	PA1					PA2				
$n\backslash k$	3	5	7	10	15	3	5	7	10	15
Exponential										
100	0.0799	0.0946	0.0933	0.0890	0.0703	0.0546	0.0603	0.0721	0.0932	0.0611
200	0.0700	0.0881	0.0999	0.1101	0.1076	0.0544	0.0556	0.0614	0.0713	0.0940
300	0.0621	0.0762	0.0915	0.1084	0.1209	0.0532	0.0545	0.0563	0.0622	0.0745
500	0.0584	0.0665	0.0788	0.0952	0.1190	0.0506	0.0507	0.0562	0.0570	0.0629
1000	0.0529	0.0573	0.0624	0.0729	0.0922	0.0511	0.0502	0.0524	0.0520	0.0551
Weibull $\alpha = 1.5$										
100	0.0640	0.0782	0.0843	0.0839	0.0598	0.0522	0.0588	0.0712	0.0896	0.0577
200	0.0599	0.0730	0.0806	0.0962	0.1029	0.0542	0.0550	0.0600	0.0716	0.0949
300	0.0560	0.0631	0.0722	0.0861	0.1036	0.0527	0.0534	0.0561	0.0612	0.0735
500	0.0528	0.0586	0.0630	0.0738	0.0905	0.0521	0.0516	0.0545	0.0568	0.0623
1000	0.0507	0.0532	0.0569	0.0593	0.0705	0.0506	0.0511	0.0518	0.0534	0.0558
Weibull $\alpha = 0.75$										
100	0.0804	0.0972	0.0991	0.0921	0.0783	0.0519	0.0567	0.0690	0.0866	0.0549
200	0.0826	0.1033	0.1175	0.1240	0.1254	0.0537	0.0560	0.0602	0.0694	0.0909
300	0.0725	0.0946	0.1116	0.1299	0.1415	0.0538	0.0538	0.0562	0.0626	0.0743
500	0.0643	0.0819	0.0960	0.1198	0.1480	0.0505	0.0506	0.0533	0.0571	0.0604
1000	0.0544	0.0620	0.0738	0.0860	0.1137	0.0516	0.0497	0.0528	0.0527	0.0561

PA1: partitioning the testing period evenly;

PA2: partitioning the testing period based on the empirical quantile of the events.

are chosen.

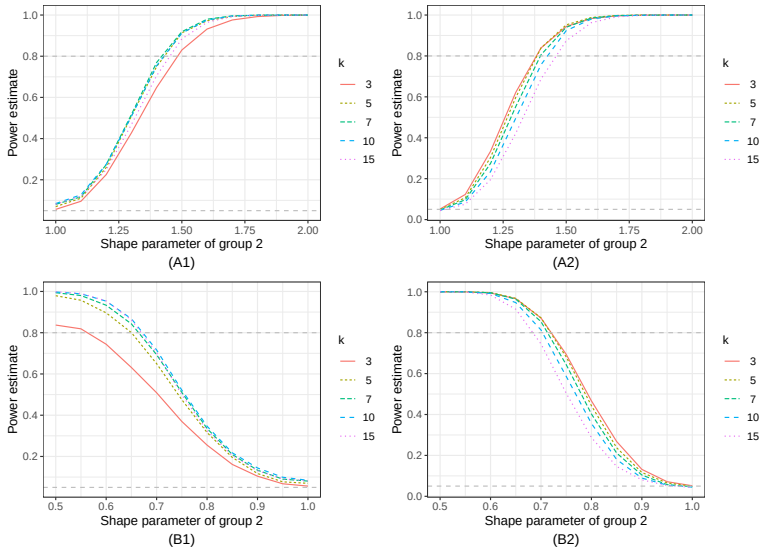

Figure 4: Empirical power curves for (A1) (B1) PH Tests with PA1; (A2) (B2) PH Tests with PA2.

4 Application to the Gender Difference in Morality Hazard of Young Adults with AML

Acute myeloid leukemia (AML) is a heterogeneous disease with a varied incidence and mortality rates across different age groups (Xie et al., 2003). Age at diagnosis is one of the most critical factors that impacts disease progression of AML. AML is less common (Deschler and Lübbert, 2006) and has a lower risk of death in children and young adults (Appelbaum et al., 2006) as compared to older adults. Research in the field has studied the impact of age at diagnosis on the risk of death due to AML, but young adults receive less attention than both younger pediatric and older patient populations (Schmidt, 2006; Wennström et al., 2016). Besides, gender is another host factor that can influence the risk of mortality. Males have a significantly higher mortality risk than females, and several gene mutations related to the high-risk of AML have been reported to be associated with males (Quesada et al., 2019; Yazarloo et al., 2013). Hossain and Xie (2015) investigated the sex variation of AML survival in childhood and young adults and identified that males substantially have a higher risk of death than females in the age group of 20-24 years old at diagnosis. In this application, we apply our proposed methodology to study the sex-specific hazard functions of mortality in young adults (20-24 years old) with AML.

We extract 496 young adult patients with an AML diagnosis between the ages of 20 and 24 from the Surveillance, Epidemiology, and End Results (SEER) program registry system from years

1990 to 2017. The majority of patients are female (50.2%) and Caucasian (75.0%). Two hundred and twenty seven events (deaths) were observed under a maximum 10-year follow-up time. The censoring rate is 54.2%. Figure 5 displays the Kaplan-Meier curves and empirical cumulative hazard curves, respectively, for male and female patients. A change-point of the cumulative hazard function at 30 months is observed for both the gender groups. The log-rank test shows a significant difference in survival rates over time between the two gender groups ($p = 0.006$).

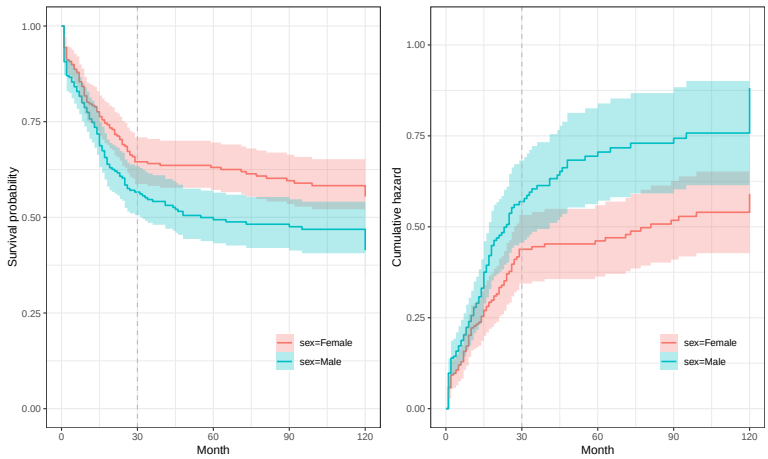

Figure 5: Kaplan Meier curves (left) and empirical cumulative hazard curves (right) of the death due to AML in different young adult gender groups.

In order to describe the difference by a hazard ratio, we first verify the PH assumption between the two gender groups using the proposed two-sample testing procedure. By choosing $k = 10$ and partitioning the time axis based on the empirical quantile of the events in both groups, the proposed method gives a p-value of 0.822. Therefore, we do not reject the null hypothesis of the PH assumption. That is, a hazard ratio is valid to describe the difference in mortality risk between the gender groups. Applying Cox's regression, males are estimated to have 1.45 (95% CI = [1.11, 1.87]) times the risk of mortality than females in young adults after the diagnosis of AML. The PH test for a Cox regression model fit (Grambsch and Therneau, 1994) demonstrates the proportional hazards ($p = 0.67$) which is consistent to the result of the proposed PH test.

Since a change point of the cumulative hazard function was observed at 30 months in both gender groups, we conduct the proposed one-sample test to test a constant hazard of death in two testing period: 0-30 months and 30-90 months. We prespecify $k = 5$ and $H_0(t) = t$, and partition the time axis evenly on each testing period. The null hypothesis of the constant hazard is not rejected in the male group at 30-90 months ($p = 0.230$), the female group at 0-30 months ($p = 0.193$), and the female group at 30-90 months ($p = 0.649$), but the null hypothesis is rejected in male group at 0-30 months ($p < 0.001$). Thus, the group of young adult females with AML might have a piecewise constant hazard function after diagnosis, where a higher constant risk within the first 30 months and a lower constant risk between 30 and 90 months. Specifying different null hypotheses of a Weibull

hazard with a shape parameter α ($H_0(t) = t^\alpha$) to males at 0-30 months, we cannot reject the null hypothesis when $\alpha = 0.7$ ($p = 0.167$) and $\alpha = 0.6$ ($p = 0.186$). It indicates a potential decreasing hazard of mortality in young adult males within the first 30 months after the diagnosis of AML. If we test a constant hazard in 5-25 months for young adult males, the null hypothesis is not rejected ($p = 0.501$) using the proposed test.

5 Discussion

In this paper, we develop the non-parametric procedures for the hypothesis testing problems for cumulative hazard functions in censored time-to-event data. These procedures can be used (i) to test if the cumulative hazard function in one sample follows a partially known-form hazard on a selected period; and (ii) to test the proportional hazards assumption between two independent samples. The proposed approach is very flexible in practice as it does not restrict the testing period to the entire domain of time. Thus, the testing period can be prespecified based on the scientific question of interest. Extensive simulation studies are carried out to examine the performance of different approaches to choose the number of knots and the partition of time. The simulation results indicate that the proposed methods enjoy a reasonable Type-I error control and a good power in both one-sample and two-sample testing problems. The proposed methodology is applied to the SEER database of young adults (20-24 years old at diagnosis) with AML to investigate the gender difference in mortality risk and the sex-specific mortality hazard functions after AML diagnosis.

In practice, choosing an appropriate number of knots and partitioning the time axis properly are essential when applying the proposed methods. Based on our empirical studies, both better Type-I error control and higher empirical power of the tests are achieved by selecting a smaller number of knots. Slightly Type-I error inflation is observed when choosing a large number of knots in a scenario with small to moderate sample size. However, too sparse knots are difficult to capture the pattern of the true cumulative hazard function. Thus, a moderate number of knots (e.g. 5-10) is recommended in practice. For the partition of the testing period, avoiding the collapse of interval due to the lack of events prevents the Type-I error inflation issue and achieves higher power. Therefore, we suggest to partition the testing period based on the prior knowledge when testing a known-form hazard and empirical quantile when testing the PH assumption.

The proposed methods are potentially useful in oncology studies. Although it is seldom to observe a constant failure rate of the primary efficacy endpoint on the entire time in cancer clinical trials, the constant hazard might still be true in a certain period of time. Our method may aid physicians to understand the characteristics of disease progression by testing if the cumulative hazard function follows a known-form hazard in a certain period. Testing the proportionality between two hazards is also important in oncology studies. For instance, the widely used log-rank test (Peto and Peto, 1972) for testing the difference between the treatment and control groups is powerful when the PH holds between the two independent groups. Compared with other PH tests, our proposed procedure is able to test the PH assumption under a certain period. Our proposed method may help clinicians to report the effect size of treatment versus placebo using a hazard ratio in a certain period even if the PH assumption is not true for the entire time.

Acknowledgements

We thank the guest editors for the invitation. We would also like to thank the editor and an anonymous referee for their useful comments and suggestions, which have led to an improved version of the paper.

References

Appelbaum, F. R., Gundacker, H., Head, D. R., Slovak, M. L., Willman, C. L., Godwin, J. E., Anderson, J. E., and Petersdorf, S. H. (2006), "Age and acute myeloid leukemia," *Blood*, 107, 3481–3485.

Breslow, N. and Crowley, J. (1974), "A large sample study of the life table and product limit estimates under random censorship," *The Annals of Statistics*, 437–453.

Casella, G. and Berger, R. L. (2001), *Statistical inference*, Cengage Learning.

Chen, T.-T. (2013), "Statistical issues and challenges in immuno-oncology," *Journal for ImmunoTherapy of Cancer*, 1, 1–9.

Cox, D. R. (1972), "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, 34, 187–202.

Dabrowska, D. M., Doksum, K. A., and Song, J.-K. (1989), "Graphical comparison of cumulative hazards for two populations," *Biometrika*, 76, 763–773.

Deschler, B. and Lübbert, M. (2006), "Acute myeloid leukemia: epidemiology and etiology," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 107, 2099–2107.

Deshpande, J. V. and Sengupta, D. (1995), "Testing the hypothesis of proportional hazards in two populations," *Biometrika*, 82, 251–261.

Gill, R. and Schumacher, M. (1987), "A simple test of the proportional hazards assumption," *Biometrika*, 74, 289–300.

Grambsch, P. M. and Therneau, T. M. (1994), "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, 81, 515–526.

Han, G., Schell, M. J., Zhang, H., Zelterman, D., Pusztai, L., Adelson, K., and Hatzis, C. (2017), "Testing violations of the exponential assumption in cancer clinical trials with survival endpoints," *Biometrics*, 73, 687–695.

Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J. C., et al. (2010), "Improved survival with ipilimumab in patients with metastatic melanoma," *New England Journal of Medicine*, 363, 711–723.

Hollander, M. and Pena, E. A. (1992), "A chi-squared goodness-of-fit test for randomly censored data," *Journal of the American Statistical Association*, 87, 458–463.

Hollander, M. and Proschan, F. (1979), "Testing to determine the underlying distribution using randomly censored data," *Biometrics*, 393–401.

Hossain, M. J. and Xie, L. (2015), "Sex disparity in childhood and young adult acute myeloid leukemia (AML) survival: Evidence from US population data," *Cancer Epidemiology*, 39, 892–900.

Kaplan, E. L. and Meier, P. (1958), "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, 53, 457–481.

Li, G. and Doss, H. (1993), "Generalized Pearson-Fisher chi-square goodness-of-fit tests, with applications to models with life history data," *The Annals of Statistics*, 772–797.

Muirhead, R. J. (1982), *Aspects of multivariate statistical theory*, vol. 197, John Wiley & Sons.

Mushfiqur Rashid, M., Chen, M.-H., and Ganter, S. L. (2000), "Nonparametric analysis of a multigroup incompletely ranked item response data," *Journal of Nonparametric Statistics*, 12, 245–264.

Peto, R. and Peto, J. (1972), "Asymptotically efficient rank invariant test procedures," *Journal of the Royal Statistical Society: Series A (General)*, 135, 185–198.

Quesada, A. E., Routbort, M. J., DiNardo, C. D., Bueso-Ramos, C. E., Kanagal-Shamanna, R., Khoury, J. D., Thakral, B., Zuo, Z., Yin, C. C., Loghavi, S., et al. (2019), "DDX41 mutations in myeloid neoplasms are associated with male gender, TP53 mutations and high-risk disease," *American Journal of Hematology*, 94, 757–766.

Sahoo, S. and Sengupta, D. (2016), "On graphical tests for proportionality of hazards in two samples," *Statistics in Medicine*, 35, 942–956.

— (2017), "Testing the hypothesis of increasing hazard ratio in two samples," *Computational Statistics & Data Analysis*, 114, 119–129.

Schmidt, C. (2006), "Lack of progress in teen and young adult cancers concerns researchers, prompts study," *Journal of the National Cancer Institute*, 98, 1760–1763.

Small, E. J., Schellhammer, P. F., Higano, C. S., Redfern, C. H., Nemunaitis, J. J., Valone, F. H., Verjee, S. S., Jones, L. A., and Hershberg, R. M. (2006), "Placebo-controlled phase III trial of immunologic therapy with sipuleucel-T (APC8015) in patients with metastatic, asymptomatic hormone refractory prostate cancer," *Journal of Clinical Oncology*, 24, 3089–3094.

Smuts, M., Allison, J., and Santana, L. (2019), "New goodness-of-fit tests for exponentiality based on a conditional moment characterisation," *ORiON*, 35, 145–160.

Tsai, W.-Y., Jewell, N. P., and Wang, M.-C. (1987), "A note on the product-limit estimator under right censoring and left truncation," *Biometrika*, 74, 883–886.

Wang, M.-C., Jewell, N. P., and Tsai, W.-Y. (1986), "Asymptotic properties of the product limit estimate under random truncation," *The Annals of Statistics*, 1597–1605.

Wei, L. (1984), "Testing goodness of fit for proportional hazards model with censored observations," *Journal of the American Statistical Association*, 79, 649–652.

Wennström, L., Edslev, P. W., Abrahamsson, J., Nørgaard, J. M., Fløisand, Y., Forestier, E., Gustafsson, G., Heldrup, J., Hovi, L., Jahnukainen, K., et al. (2016), "Acute myeloid leukemia in adolescents and young adults treated in pediatric and adult departments in the Nordic countries," *Pediatric Blood & Cancer*, 63, 83–92.

Xie, Y., Davies, S. M., Xiang, Y., Robison, L. L., and Ross, J. A. (2003), "Trends in leukemia incidence and survival in the United States (1973–1998)," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 97, 2229–2235.

Xue, Y., Wang, H., Yan, J., and Schifano, E. D. (2020), "An online updating approach for testing the proportional hazards assumption with streams of survival data," *Biometrics*, 76, 171–182.

Yazarloo, F., Shirkoohi, R., Mobasheri, M. B., Emami, A., and Modarressi, M. H. (2013), "Expression analysis of four testis-specific genes AURKC, OIP5, PIWIL2 and TAF7L in acute myeloid leukemia: a gender-dependent expression pattern," *Medical Oncology*, 30, 368.

Received: March 2, 2021

Accepted: April 7, 2021