Research Article
Special Issue: 10 years of Dependence Modeling

Qingyang Zhang*

A nonparametric test for comparing survival functions based on restricted distance correlation

https://doi.org/10.1515/demo-2023-0108
received September 4, 2023; accepted November 7, 2023

Abstract: In this article, we propose an omnibus test for comparing two survival functions under non-proportional hazards. The test statistic is based on a product-limit estimate of the restricted distance correlation, which is closely related to the L_2 distance between survival curves. The strong consistency is established under mild regularity conditions. Our simulation studies show that the new test has satisfactory power under proportional hazard and various non-proportional hazards settings including delayed treatment effect, diminishing effect, and crossing survival curves; therefore, it can be a competitive alternative to the existing omnibus tests such as Kolmogorov-Smirnov test, Cramer-von Mises test, two-stage test, and the maxCombo test based on weighted log-rank statistics. Two extensions of the new test to one-sided alternatives and a Gaussian kernel are also discussed.

Keywords: non-proportional hazards, restricted distance correlation, omnibus test, strong consistency

MSC 2020: 62N03 (primary), 62H20 (secondary)

1 Introduction

To evaluate the treatment effect for survival data, we often need to compare the survival functions of the treatment and control groups. The most popular approach to comparing survival functions is the log-rank test, and it is well known that under proportional hazards, the log-rank test is optimal and equivalent to the score test in Cox regression model. When the proportional hazard assumption is moderately or severely violated, however, the log-rank test might be suboptimal. In many clinical studies, especially cancer immunotherapy trials [1,19,24], the violation of proportional hazards assumption is often encountered, and different patterns of non-proportional hazards are frequently observed, e.g., delayed treatment effect, diminishing effect, and crossing survival curves, making the traditional log-rank test underpowered. One way to address this challenge is using the weighted log-rank test, and a popular weight function is the Fleming-Harrington (FH) weight with parameters ρ and γ,

$$w_{\mathrm{FH}}(t;\, \rho, \gamma) = [S_n(t-)]^{\rho}[1 - S_n(t-)]^{\gamma},$$

where $S_n(t-)$ is the estimated survival function immediately prior to time t. The choice of ρ and γ can handle different types of treatment effect. For instance, $w_{\mathrm{FH}}(t;\, \rho > 0, \gamma = 0)$ is good for early separation, $w_{\mathrm{FH}}(t;\, \rho = 0, \gamma > 0)$ for late separation, and $w_{\mathrm{PH}}(t;\, \rho > 0, \gamma > 0)$ for middle separation. However, none of these tests is good for all situations, and a prior misspecification of the weight function may decrease the power of the test. Motivated by previous studies

* **Corresponding author: Qingyang Zhang**, Department of Mathematical Sciences, University of Arkansas, AR 72701, Fayetteville, United States, e-mail: qz008@uark.edu

[13,26], a cross-industry working group proposed a maxCombo test by taking the maximum of multiple FH-weighted log-rank statistics [14]. One such combination is

$$Z_{\max} = \max\{Z_{\mathrm{FH}}(\rho,\gamma), (\rho,\gamma) \in [(0,0),(0,1),(1,0),(1,1)]\},$$

where $Z_{\mathrm{FH}}(\rho,\gamma)$ stands for the Z-statistic of the weighted log-rank test with $w_{\mathrm{FH}}(t;\rho,\gamma)$, and it can be shown that $[Z_{\mathrm{FH}}(0,0), Z_{\mathrm{FH}}(0,1), Z_{\mathrm{FH}}(1,0), Z_{\mathrm{FH}}(1,1)]^T$ are asymptotically joint normal. Using simulated data, Lin et al. [14] showed that the maxCombo test has good statistical power under proportional hazard and different patterns of non-proportional hazards, thus it can be used as an omnibus test for a broad class of alternative hypotheses. A robust version of maxCombo based on weights $[(0,0),(0,1/2),(1/2,0),(1/2,1/2)]$ is suggested by Roychoudhury et al. [18].

In addition to the maxCombo test, there are many other omnibus tests developed for survival data. To name a few, Fleming et al. generalized the Kolmogorov-Smirnov (KS) test for arbitrarily right-censored data [8]. Koziol et al. and Schumacher modified the Crámer-von Mises (CVM) test under the assumption of randomly censoring [12,20]. All these KS- or CVM-based methods can be viewed as a weighted L_p distance ($p = 2$ for CVM and $p = \infty$ for KS) between two Kaplan-Meier (KM) curves or Nelson-Aalen curves. With the intent of addressing crossing survival curves, Qiu and Sheng proposed a two-stage procedure, where the log-rank test is used in the first stage and a particular weighted log-rank test is used in the second stage [16]. The weight function in stage two is chosen to change signs before and after a potential crossing point, boosting its power for crossing survival curves. Recently, Ditzhaus et al. proposed a permutation test based on the Nelson-Aalen-type integrals without any restrictive model assumption [3]. Fernández et al. introduced a general nonparametric independence test between right-censored survival times and covariates based on the supremum of a potentially infinite collection of weight-indexed log-rank tests, with weight functions belonging to a reproducing kernel Hilbert space (RKHS) of functions [7].

In this study, we shall develop a new omnibus test for comparing survival functions. The test statistic is based on a restricted version of distance correlation and related to the unweighted L_2 distance between two survival functions. Motivated by Edelmann et al. [5] and Zhang et al. [28], a permutation procedure based on the product-limit estimator is used for implementing our method. Simulation studies show that the new test performs well under different sample sizes and survival models.

The remainder of this study is structured as follows: Section 2 introduces the notion of restricted distance correlation, and proposes a consistent estimator. Section 3 evaluates the performance of the new and existing tests under different non-proportional hazards settings. Section 4 discusses the method with some future perspectives.

2 Restricted distance correlation test

In this section, we introduce the restricted distance correlation test for survival data and establish its statistical consistency. We begin with the notations. For subject $i \in \{1,\ldots,n\}$, let T_i denote the survival time, X_i the group index (0 for the control arm, 1 for the treatment arm), $\pi = P(X_i = 1)$, and C_i the censoring time due to administrative censoring or patient dropout. The observed event or censoring time is defined as $U_i = \min\{T_i, C_i\}$, with event indicator $\delta_i = \mathbb{I}\{T_i \le C_i\}$. Let $f_0(t), f_1(t), F_0(t), F_1(t), S_0(t), S_1(t)$ be the probability density functions (p.d.f.), cumulative distribution functions (c.d.f.), and survival functions of T in two arms, i.e., $f_0(t) = f(t|X_i = 0)$, and $f_1(t) = f(t|X_i = 1)$, $F_0(t) = F(t|X_i = 0)$, $F_1(t) = F(t|X_i = 1)$, $S_0(t) = 1 - F(t|X_i = 0)$, and $S_1(t) = 1 - F(t|X_i = 1)$. Let $G(t)$ be the c.d.f. of C_i for both arms and τ be the study duration ($\max_i C_i \le \tau$). The null and alternative hypotheses can be formulated as follows:

$$\begin{aligned} &H_0 : S_0(t) = S_1(t), \quad \text{for } 0 \le t < \infty,\\ &H_a : S_0(t) \ne S_1(t), \quad \text{for some } t. \end{aligned} \tag{1}$$

It is noteworthy that testing (1) amounts to testing the independence between the survival time T and group index X, where T is continuous and X is binary. Herein, we consider the distance correlation test by Székely et al. [23]. The distance covariance between two random vectors X and Y is defined as the square root of

$$\mathrm{dCov}^2(X,Y) = \int_{R^2} \frac{||\phi_{x,y}(t,s) - \phi_x(t)\phi_y(s)||^2}{c_{d_x}c_{d_y}||t||_{d_x}^{1+d_x}||s||_{d_y}^{1+d_y}}\mathrm{d}t\mathrm{d}s, \tag{2}$$

where $c_{d_x} = \frac{\pi^{(1+d_x)/2}}{\Gamma\{(1+d_x)/2\}}$ and $c_{d_y} = \frac{\pi^{(1+d_y)/2}}{\Gamma\{(1+d_y)/2\}}$, $||z||_{d_z}$ denotes the Euclidean norm of $z \in \mathbb{R}^{d_z}$, and $||\phi||^2 = \phi\bar{\phi}$ for a complex-valued function ϕ and its conjugate $\bar{\phi}$ [15,23]. Similar to Pearson's correlation coefficient, the distance correlation is defined as follows:

$$\mathrm{dCor}(X,Y) = \frac{\mathrm{dCov}(X,Y)}{\sqrt{\mathrm{dCov}(X,X)}\sqrt{\mathrm{dCov}(Y,Y)}}. \tag{3}$$

One remarkable property of distance correlation is that it is 0 if and only if X and Y are statistically independent, indicating that the distance correlation can detect any form of association. Székely et al. [23] also provided the following alternative definition of $\mathrm{dCov}^2(X,Y)$:

$$\mathrm{dCov}^2(X,Y) = \mathrm{Cov}(||X_1 - X_2||, ||Y_1 - Y_2||) - 2\mathrm{Cov}(||X_1 - X_2||, ||Y_1 - Y_3||),$$

where (X_1, Y_1), (X_2, Y_2), and (X_3, Y_3) stand for three independent copies of (X, Y).

As a special case of (3), in the following, we give the explicit formula of squared distance correlation between the survival time T and group index X (the detailed proof is provided in Appendix A.1):

$$\mathrm{dCor}^2(T,X) = \frac{\int_0^\infty [S_1(t) - S_0(t)]^2\mathrm{d}t}{8\int_0^\infty\int_t^\infty [\pi S_1(s) + (1-\pi)S_0(s)]^2[1 - \pi S_1(s) - (1-\pi)S_0(s)]^2\mathrm{d}s\mathrm{d}t}. \tag{4}$$

Noteworthily, the squared distance covariance between X and T (see equation (A1) in A.1) is equivalent to the energy distance between $T|X = 0$ and $T|X = 1$. In fact, up to a constant multiple, equation (A1) is also equivalent to Crámer's distance [2] between $T|X = 0$ and $T|X = 1$. Crámer's distance can be viewed as a special case of energy distance when both variables are univariate. However, as Rizzo and Székely [17] pointed out, the equivalence of energy distance with Crámer's distance cannot extend to higher dimensions, because while energy distance is rotation invariant, Crámer's distance is not.

For clinical trials with survival endpoints, an administrative censoring is often applied at the end of the study period so that no event can be observed after time τ, i.e., $\max_i C_i \le \tau$. Similar to the restricted mean survival time (RMST), we consider a restricted version of distance correlation on $[0, \tau]$ for hypothesis testing

$$\mathrm{dCor}^2(T,X;\ \tau) = \frac{\int_0^\tau [S_1(t) - S_0(t)]^2\mathrm{d}t}{8\int_0^\tau\int_t^\tau [\pi S_1(s) + (1-\pi)S_0(s)]^2[1 - \pi S_1(s) - (1-\pi)S_0(s)]^2\mathrm{d}s\mathrm{d}t}. \tag{5}$$

The null and alternative hypotheses based on the restricted distance correlation can be formulated as:

$$\begin{aligned} H_0 &: S_0(t) = S_1(t), \quad \text{for } 0 \le t \le \tau, \\ H_a &: S_0(t) \ne S_1(t), \quad \text{for some } t \in [0, \tau]. \end{aligned} \tag{6}$$

Under the restriction of total study duration, the null hypothesis in (6) can be interpreted as the independence between T and X conditioning on $0 \le T \le \tau$, i.e., $T \perp X | 0 \le T \le \tau$. With a sufficient study duration τ, e.g., $\max\{S_0(\tau), S_1(\tau)\}$ is small, (6) can be used as a proxy of (1), but results should not be over-interpreted for relatively short τ.

Assuming independent censoring, i.e., $T \perp C$, let $S_{1n}(t)$ and $S_{0n}(t)$ be some consistent estimators of $S_1(t)$ and $S_0(t)$, such as the product-limit estimator or the piecewise exponential estimator [11]. For simplicity, we consider the following product-limit estimate:

$$\mathrm{dCor}_n^2(T,X;\ \tau) = \frac{\int_0^\tau [S_{1n}(t) - S_{0n}(t)]^2\mathrm{d}t}{8\int_0^\tau\int_t^\tau [\pi S_{1n}(s) + (1-\pi)S_{0n}(s)]^2[1 - S_{1n}(s) - (1-\pi)S_{0n}(s)]^2\mathrm{d}s\mathrm{d}t}. \tag{7}$$

Theorem 1 establishes the statistical consistency of (7), under mild regularity conditions (proof is given in Appendix A.2).

Theorem 1. *Assuming independent censoring*, $0 < S(\tau) < 1$ *and* $G(\tau) < 1$, *we have*

$$\lim_{n\to\infty}|\mathrm{dCor}_n^2(T,X;\,\tau) - \mathrm{dCor}^2(T,X;\,\tau)| = 0 \quad a.s.$$

In general, the null distribution of distance correlation is impractical to derive as it depends on the underlying distributions of X and Y; therefore, we suggest a permutation procedure to evaluate significance. One may first calculate the test statistic $\mathrm{dCor}_n^2(T,X;\,\tau)$ for the observed data, then for each $b = 1,\ldots,B$, calculate the distance correlation $\mathrm{dCor}_b^2(T,X;\,\tau)$ based on the random permutation of group indices $\{X_i,\, i = 1,\,\ldots,n\}$. The permutation p-value can be computed as:

$$p = \frac{\sum_{b=1}^{B}\mathbb{I}\{\mathrm{dCor}_b^2(T,X;\,\tau) \geq \mathrm{dCor}_n^2(T,X;\,\tau)\} + 1}{B+1}. \tag{8}$$

Though the formula for $\mathrm{dCor}_n^2(T,X;\,\tau)$ seems unwieldy, for the purposes of constructing a permutation test, only the numerator is relevant. The numerator is essentially a L_2 distance between $S_{1n}(t)$ and $S_{0n}(t)$, which is closely related to the CVM criterion. The CVM statistic is

$$\mathrm{CVM}_n(\tau) = \int_0^{\tau}[S_{1n}(t) - S_{0n}(t)]^2 d[-\pi S_{1n}(t) - (1-\pi)S_{0n}(t)], \tag{9}$$

which is a L_2 distance between two estimated survival functions with weight $\pi f_{1n}(t) + (1-\pi)f_{0n}(t)$. The CVM statistic assigns more weight on time points with higher event rates; thus, for concave-up survival functions, it tends to better detect early separation than late separation. In contrast, our distance correlation statistic is an unweighted L_2 distance, targeting the difference between two survival curves for the entire study period.

3 Two extensions of the proposed test

A limitation of the restricted distance correlation test is that it is only for two-sided alternatives, therefore not suitable for superiority tests that can be formulated as:

$$\begin{aligned} H_0 &: S_0(t) = S_1(t), \text{ for } 0 \leq t \leq \tau,\\ H_a &: S_0(t) < S_1(t), \text{ for } 0 \leq t \leq \tau. \end{aligned} \tag{10}$$

To this end, we also suggest a directional test by incorporating the sign information in L_2 distance. The directional statistic for permutation test can be written as:

$$T_{n,+} = \int_0^{\tau}\mathrm{sgn}[S_{1n}(t) - S_{0n}(t)] \times [S_{1n}(t) - S_{0n}(t)]^2 \mathrm{d}t, \tag{11}$$

where $\mathrm{sgn}()$ is the sign function, i.e., $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and $\mathrm{sgn}(x) = -1$ if $x < 0$. Similar to (8), the permutation p-value can be computed as:

$$p = \frac{\sum_{b=1}^{B}\mathbb{I}\{T_{b,+} \geq T_{n,+}\} + 1}{B+1}. \tag{12}$$

Second, as suggested by the reviewers, we extend the proposed distance correlation test to the Gaussian kernel, which can be equivalently used in the distance correlation formulation [6,10,22]. We derived the distance covariance between X and T based on the Gaussian kernel with bandwidth parameter σ^2 (see Appendix A.3 for details). For illustrative purposes, we present the formula for $\pi = 1/2$ as follows:

$$\mathrm{dCov}^2(T,X) = \frac{1-\exp(-1/2\sigma^2)}{8}(D_{00} + D_{11} - 2D_{01}),$$

where

$$D_{00} = \int_0^1\int_0^1 \exp\left(-\frac{|t_1 - t_2|^2}{2\sigma^2}\right) \mathrm{d}S_0(t_1)\mathrm{d}S_0(t_2),$$
$$D_{11} = \int_0^1\int_0^1 \exp\left(-\frac{|t_1 - t_2|^2}{2\sigma^2}\right) \mathrm{d}S_1(t_1)\mathrm{d}S_1(t_2),$$
$$D_{01} = \int_0^1\int_0^1 \exp\left(-\frac{|t_1 - t_2|^2}{2\sigma^2}\right) \mathrm{d}S_0(t_1)\mathrm{d}S_1(t_2).$$

The expected distances D_{00}, D_{11}, and D_{01} can be estimated by replacing the survival functions $S_0(t)$ and $S_1(t)$ with the KM estimates $S_{0n}(t)$ and $S_{1n}(t)$. A permutation test similar to equation (12) can then be performed based on $D_{00} + D_{11} - 2D_{01}$. In general, the tuning parameter σ^2 in the Gaussian kernel affects the testing power, and the effect depends on the data. Therefore, a data-driven approach should be used for selecting σ^2. Our simulation studies have shown that the median survival of the pooled data (two arms) performs well under different settings; therefore, we suggest using it for σ^2.

4 Simulation study

In this section, we conduct simulation studies to evaluate the performance of the restricted distance correlation test under different settings. In particular, we investigate the empirical statistical power and type I error rate under both two-sided and one-sided settings.

4.1 Two-sided alternatives

We compare the distance correlation tests (based on Euclidean distance and Gaussian kernel, respectively) with five existing tests, namely, (1) the robust maxCombo test, (2) two-stage test, (3) KS test, (4) CVM test, and (5) log-rank test. The log-rank test is used as a gold standard for proportional hazard, and it was implemented using R function *survdiff* in the *survival* package. The robust maxCombo test was implemented using the *logrank.maxtest* function in the *nph* package. The two-stage test by Qiu and Sheng [16] was implemented by the *two-stage* function in the *TSHRC* package. For KS, CVM, and our restricted distance correlation tests, p-values were computed based on 2,000 random permutations. The CVM test is based on equation (9), and the KS test is based on the following statistic:

$$\mathrm{KS}_n(\tau) = \max_{0\le t\le\tau} |S_{1n}(t) - S_{0n}(t)|. \tag{13}$$

In the simulation, we set $\pi = 1/2$, and $n = 60, 100, 150$, and 200 (total sample size for two arms). For all subjects, the loss to follow-up time (in months) is assumed to be exponential with rate parameter 0.005, corresponding to a 5.8% annual dropout rate and 26% five-year dropout rate. Moreover, we assume that the accrual time follows a uniform distribution over 3 years. Four alternatives, namely, a proportional hazards setting (A) and three non-proportional hazards settings (B: delayed treatment effect, C: crossing survival curves, D: diminishing effect), were constructed using exponential mixture models (similar curves can be also constructed by other flexible models such as generalized Weibull models or piecewise exponential models). The survival function of a two-component exponential mixture model is as follows:

$$S_j(t;\, \gamma_j, \lambda_{j1}, \lambda_{j2}) = \gamma_j \exp(-\lambda_{j1}t) + (1 - \gamma_j)\exp(-\lambda_{j2}t), \quad j \in \{0, 1\}, \tag{14}$$

where γ_j and $1 - \gamma_j$ represent the proportions of two components in arm j, and λ_{j1} and λ_{j2} are the corresponding rate parameters. The parameters of each simulation setting are listed in the following, and the survival curves are sketched in Figure 1.

(A) Proportional hazards: $\gamma_0 = \gamma_1 = 1$, $\lambda_{01} = \log(2)/15$, $\lambda_{11} = \log(2)/22.5$.

(B) Delayed treatment effect: $\gamma_0 = \gamma_1 = 0.5$, $\lambda_{01} = \log(2)/20$, $\lambda_{02} = \log(2)/10$, $\lambda_{11} = \log(2)/70$, $\lambda_{12} = \log(2)/7$.

(C) Crossing survival curves: $\gamma_0 = \gamma_1 = 0.2$, $\lambda_{01} = \log(2)/20$, $\lambda_{02} = \log(2)/20$, $\lambda_{11} = \log(2)/1$, $\lambda_{12} = \log(2)/40$.
(D) Diminishing effect: $\gamma_0 = \gamma_1 = 0.1$, $\lambda_{01} = \log(2)/20$, $\lambda_{02} = \log(2)/3$, $\lambda_{11} = \log(2)/20$, $\lambda_{12} = \log(2)/5$.

Figure 2 summarizes the empirical power over 5,000 simulations at the significance level of 0.05. The two distance correlation metrics perform comparably across all settings, with the Gaussian kernel performing slightly better than Euclidean distance. For example, in the proportional hazard setting with a sample size of 100, the test based on Euclidean distance achieves a power of 39%, while the test based on the Gaussian kernel achieves a power of 41%. In the proportional hazards setting, the log-rank test has the highest power. The distance correlation tests are the best among all omnibus tests, and when $n = 200$, our tests achieve similar statistical power to the log-rank test. For Setting B, the distance correlation tests have the highest power among all methods. The maxCombo, log-rank, and two-stage tests also have good performance especially for relatively large sample sizes. It is noteworthy that the CVM test has low power in this delayed treatment effect setting, because it assigns more weight on the early stage and less weight on the late stage. For crossing survival curves (Setting C), the most powerful test is the two-stage test by Qiu and Sheng [16]. The two-stage procedure is particularly designed for crossing survival curves; thus, it is sensitive to this pattern. Our new tests have the second highest power in this setting, close to the robust maxCombo test. For the diminishing effect setting (Setting D), where the separation occurs at the early and middle stage, CVM provides the best power. The KS and distance correlation tests have slightly lower power than CVM. Overall, our distance correlation tests have satisfactory power for different settings; thus, it can be an competitive alternative to the existing ones.

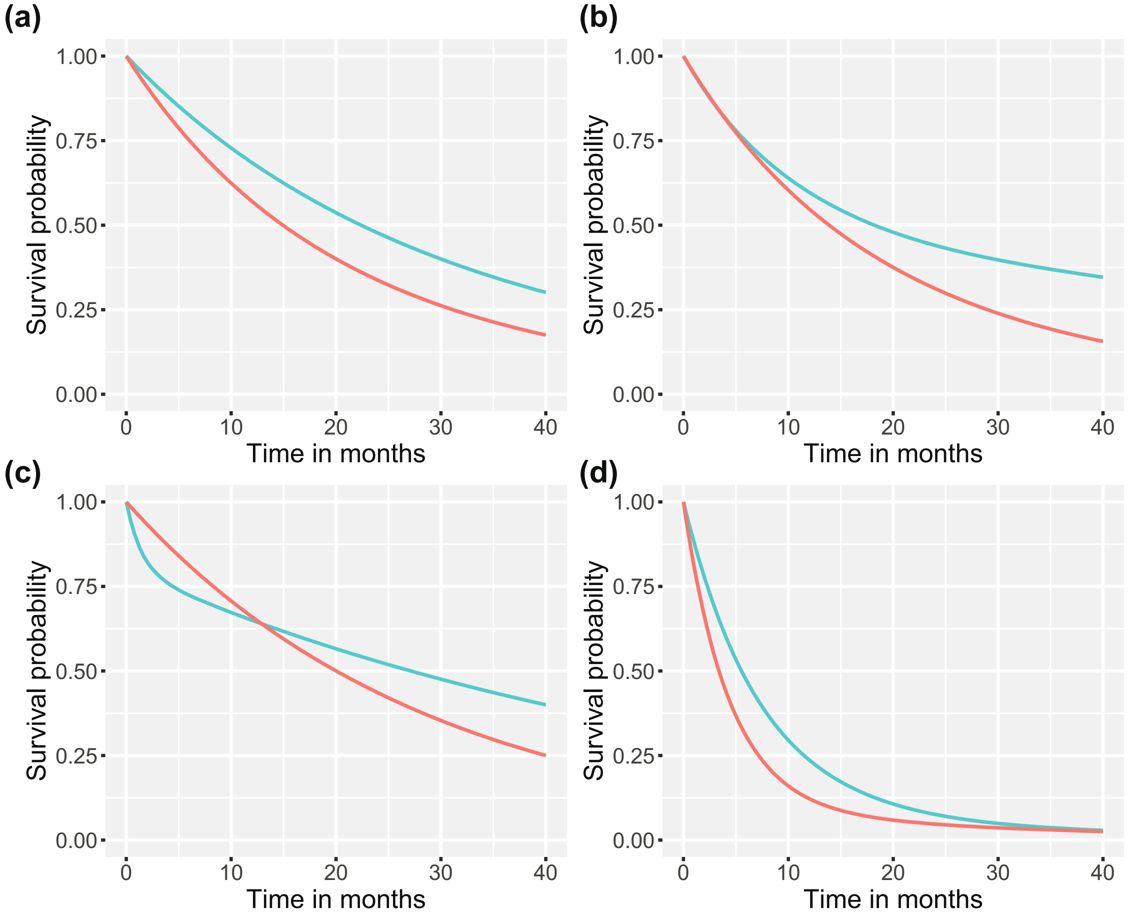

Figure 1: Survival curves in the simulation study (a: proportional hazards, b: delayed treatment effect, c: crossing survival curves, and d: diminishing effect), where red represents the control arm and blue represents the treatment arm.

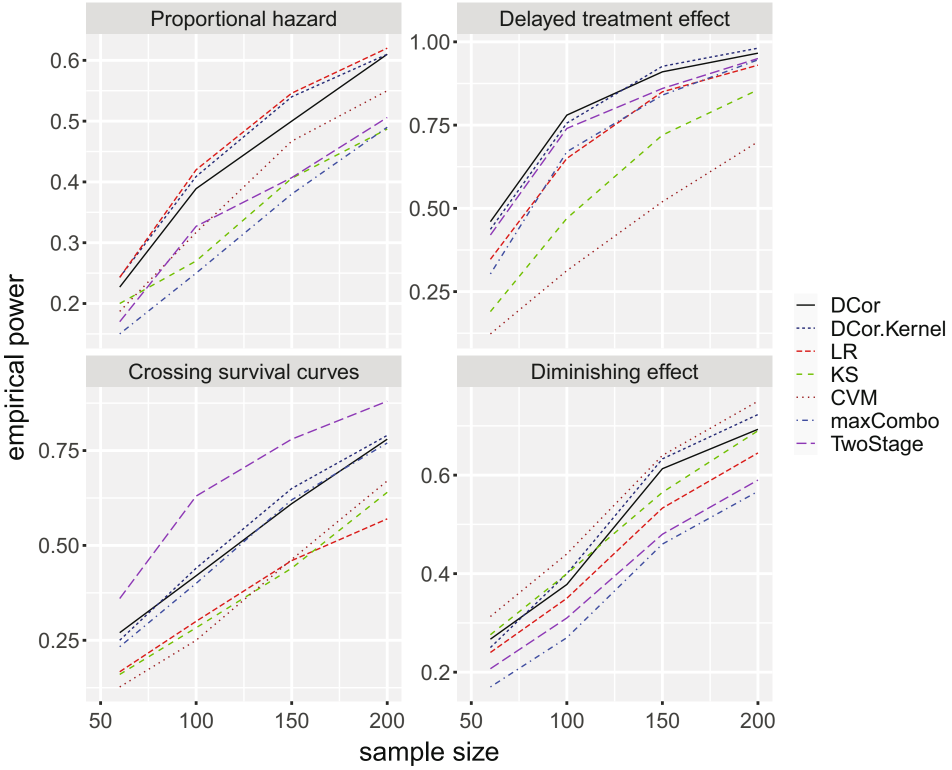

Figure 2: Empirical power over 5,000 simulations for two-sided alternatives.

We also investigated the type I error rate control under different sample sizes. Figure 3 presents the type I error rate over 10,000 simulations (under the null model $\gamma_0 = \gamma_1 = 1$, $\lambda_{01} = \lambda_{11} = \log(2)/15$). All the tests control the type I error rate. The three permutation based tests, namely KS, CVM, and restricted distance correlation tests, have type I error rates close to the nominal level of 0.05. The log-rank test and maxCombo test based on weighted log-rank statistics are slightly conservative when sample size is small, e.g., $n = 60$.

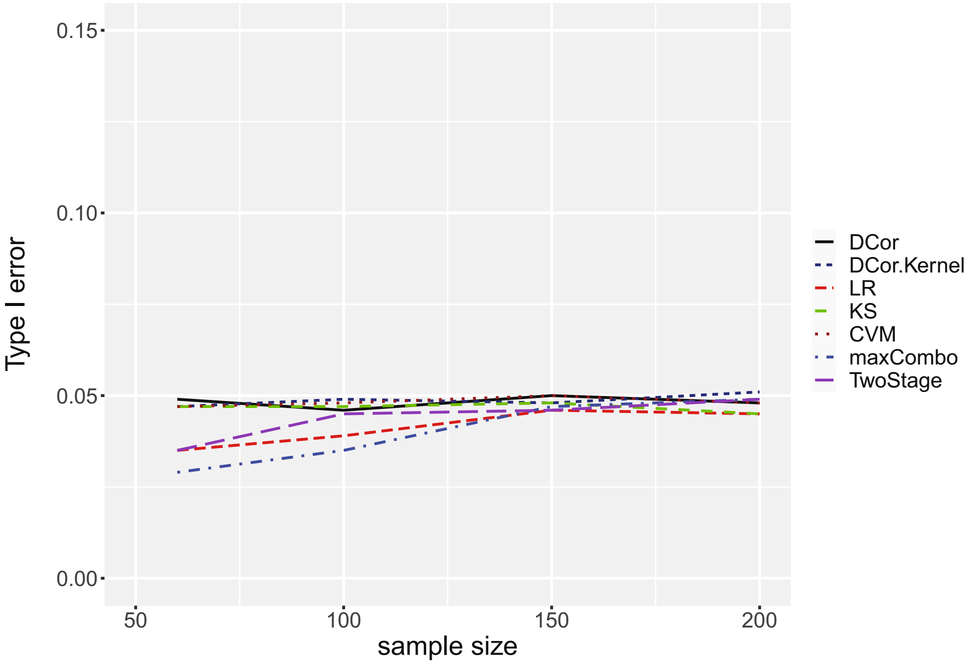

Figure 3: Type I error rate over 10,000 simulations for two-sided alternatives.

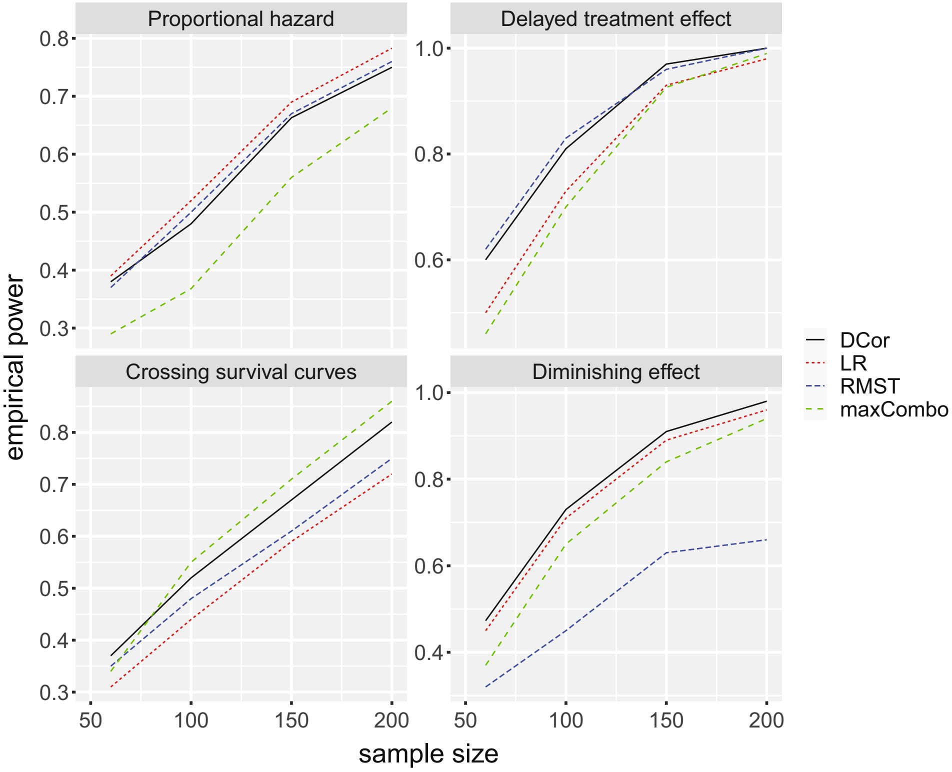

Figure 4: Empirical power over 5,000 simulations for one-sided alternatives.

4.2 One-sided alternatives

For one-sided alternatives, we compare our directional distance correlation test (equations (11) and (12)) with (1) log-rank test, (2) RMST test, and (3) the robust maxCombo, under the same simulation settings as detailed in Section 3.1. The two-stage, CVM, and KS tests are excluded in the analysis because they are not suitable for one-sided alternatives. Figure 4 displays the empirical power over 5,000 simulations at the significance level of 0.05. Same as what we observed in the two-sided case, in the proportional hazards setting, the log-rank test has the highest statistical power. Our distance correlation test has similar power to RMST, both higher than maxCombo. In the delayed treatment effect setting, the RMST and distance correlation tests substantially outperform the log-rank and maxCombo tests. Specifically, the maxCombo test is the most powerful test for detecting differences between crossing survival curves (Setting C). In the diminishing effect setting, the distance correlation test has the greatest power, slightly higher than the log-rank test. Overall, our distance correlation test have satisfactory power across different settings. Figure 5 summarizes the empirical sizes of the four tests, where it can be seen that all four tests control the type I error rate, and three RMST tests are slightly conservative.

5 Discussion and conclusions

In recent clinical studies, especially in cancer immunotherapy studies, the violation of the proportional hazards assumption is often encountered; thus, the traditional log-rank test may not be optimal. In this work, we propose a simple and versatile test to compare survival curves under non-proportional hazards. The test statistic is derived from a restricted version of the widely used distance correlation metric, which is

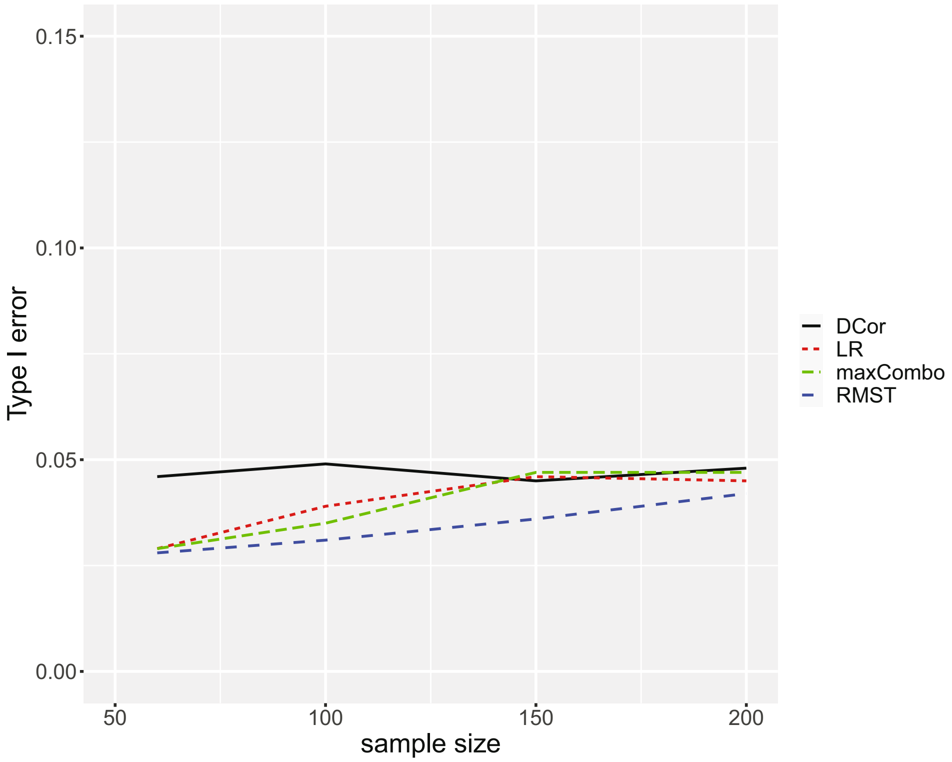

Figure 5: Type I error rate over 10,000 simulations for one-sided alternatives.

essentially the L_2 distance between the KM curves of two treatment groups. Our simulation studies show that the new test is powerful under both proportional hazards and different types of non-proportional hazards.

One major limitation of the proposed test is the lack of an analytical formula for computing p-values. Therefore, it would be of great interest to investigate the asymptotic behavior of the restricted distance correlation theoretically. While the sampling distribution of the distance correlation is generally impractical to derive, Shen et al. [21] derived a chi-square distribution that well approximates and dominates the limiting null distribution in the upper tail. They showed that under the bias-corrected estimate of the distance correlation, the chi-square test exhibits similar testing power to the standard permutation test. However, the existence of censored samples in survival data makes it difficult to obtain the bias-corrected estimate. Therefore, directly applying the chi-squared approximation based on the KM estimates may result in low power. Figure 6 presents a comparison of the power of the permutation test and Shen et al.'s chi-squared approximation. As can be seen, the chi-squared test can be very conservative, especially when the sample size is relatively small (e.g., n = 60). To circumvent this problem, we need to find a new estimate or approximating distribution function to calculate an upper bound of the p-value. We leave this as a topic for future research.

Another practical limitation is the assumption of independent censoring, meaning that the censoring time is independent of both groups and survival time. When the censoring depends on groups, the permutation test may have an inflated type I error rate; even, the survival curves are equal. To illustrate this, we performed simulations (Figure 7) and found that the type I error rate inflation is non-negligible when there is a substantial difference in the censoring rate between two arms. Therefore, it is important to check whether the two arms have similar censoring distributions before using a distance correlation test. Possible approaches for estimating censoring distributions or censoring rates include the reverse KM curve and the person-time follow-up rate [25].

There are several possible extensions of our test. Throughout this study, for illustrative purposes, we have focused on the two-sample comparison. However, our method can be readily applied to compare multiple survival functions. In the K-sample case, the restricted distance correlation based on Euclidean distance can be expressed as:

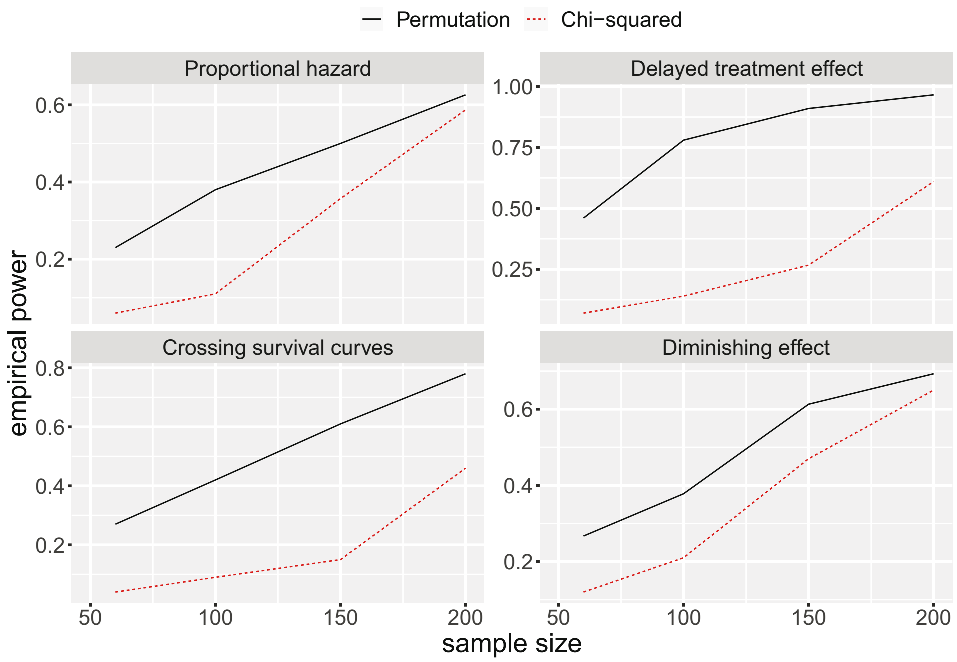

Figure 6: Power comparison for the permutation test and chi-squared test.

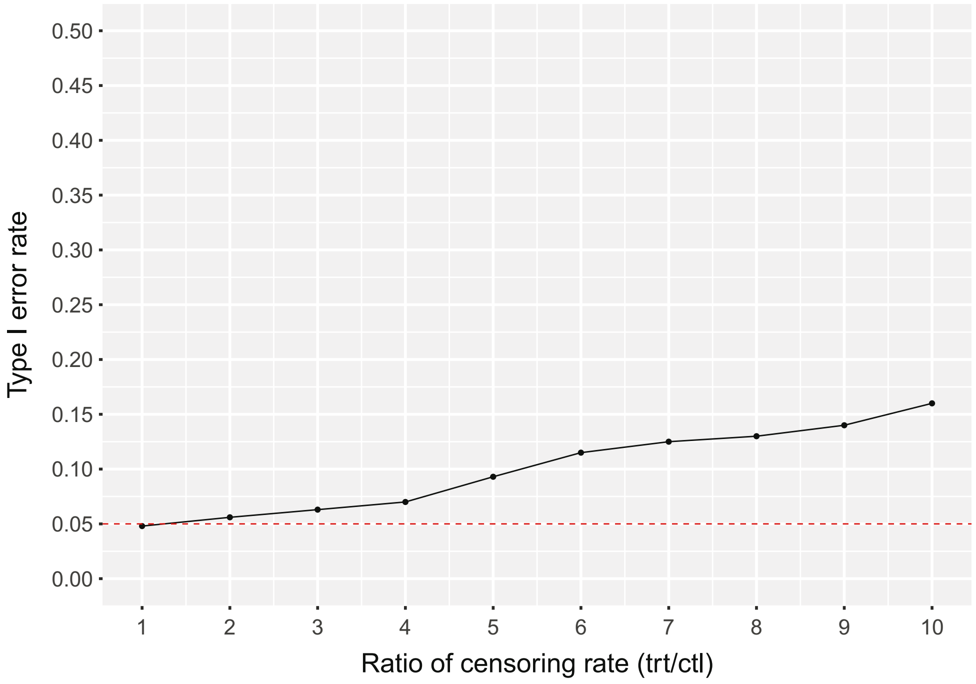

Figure 7: Type I error rate inflation under different censoring rates in two arms (the x-axis represents the ratio of censoring rates of two arms, ranging from 1 to 10).

$$\mathrm{dCor}^2(T, X;\ \tau) = \frac{\sum_{k=1}^{K} \pi_k^2 \int_0^\tau [S_k(t) - S(t)]^2 \mathrm{d}t}{4\int_0^\tau \int_t^\tau [S(t)]^2 [1 - S(t)]^2 \mathrm{d}s\mathrm{d}t}, \tag{15}$$

where $\pi_k = P(X = k)$, $S_k(t) = S(t|X = k)$, and $S(t) = \sum_{k=1}^{K} \pi_k S_k(t)$. Similar to equation (7), one can use the product-limit method to estimate the restricted distance correlation, and a permutation test based on the numerator of (13) can be used to obtain p-values. In the case of ordinal X, e.g., age groups or dosage levels, one can derive the restricted distance correlation based on the predefined distance between categories.

In addition to right-censored data, our test might also be applicable to other censoring types. For instance, when the data are left-censored, one may utilize the Left-KM (LeftKM) method to estimate the survival functions in the distance correlation. Under independent censoring, Gomez et al. [9] proved the consistency of the LeftKM estimator, and we may use this result to establish the statistical consistency of the restricted distance correlation test.

Acknowledgement: The author would like to thank the editor and two reviewers for their valuable suggestions and remarks.

Funding information: The work was supported by an NSF DBI Biology Integration Institute (BII) Grant (Award No. 2119968; PI-Ceballos).

Conflict of interest: The author states that there is no conflict of interest.

Appendix

A.1 Derivation of equation (4)

The squared distance covariance between T and X can be written as:

$$\mathrm{dCov}^2(T, X) = E(|T_1 - T_2||X_1 - X_2|) + E(|T_1 - T_2|)E(|X_1 - X_2|) - 2E(|T_1 - T_2||X_1 - X_3|),$$

where (T_1, X_1), (T_2, X_2), and (T_3, X_3) are three independent copies of (T, X). Let

$$\begin{aligned} D_{00} &= E(|T_1 - T_2||X_1 = 0, X_2 = 0), \\ D_{01} &= E(|T_1 - T_2||X_1 = 0, X_2 = 1), \\ D_{11} &= E(|T_1 - T_2||X_1 = 1, X_2 = 1), \end{aligned}$$

the following results can be shown using elementary probability:

$$\begin{aligned} E(|T_1 - T_2|) &= \pi^2 D_{11} + (1 - \pi)^2 D_{00} + 2\pi(1 - \pi)D_{01}, \\ E(|X_1 - X_2|) &= 2\pi(1 - \pi), \\ E(|T_1 - T_2||X_1 - X_2|) &= 2\pi(1 - \pi)D_{01}, \\ E(|T_1 - T_2||X_1 - X_3|) &= \pi^2(1 - \pi)D_{11} + \pi(1 - \pi)^2 D_{00} + \pi(1 - \pi)D_{01}. \end{aligned}$$

Furthermore, we can show

$$\begin{aligned} D_{00} &= 2\int_0^\infty S_0(t)[1 - S_0(t)]\mathrm{d}t, \\ D_{11} &= 2\int_0^\infty S_1(t)[1 - S_1(t)]\mathrm{d}t, \\ D_{01} &= \int_0^\infty S_0(t) + S_1(t) - 2S_0(t)S_1(t)\mathrm{d}t. \end{aligned}$$

Summarizing the aforementioned results, we have

$$\mathrm{dCov}^2(T,X) = 4\pi^2(1-\pi)^2\int_0^\infty[S_1(t)-S_0(t)]^2\mathrm{d}t. \tag{A1}$$

It is also straightforward to show

$$\mathrm{dCov}^2(X,X) = 4\pi^2(1-\pi)^2. \tag{A2}$$

Finally, by Theorem 5.1 of Edelmann et al. [4], we have

$$\mathrm{dCov}^2(T,T) = 8\int_0^\infty\int_t^\infty S(s)^2[1-S(s)]^2\mathrm{d}s\mathrm{d}t, \tag{A3}$$

where $S(t)$ stands for the overall survival function and $S(t) = \pi S_1(t) + (1-\pi)S_0(t)$. Combining (16)–(18), we have

$$\mathrm{dCor}^2(T,X) = \frac{\int_0^\infty[S_1(t)-S_0(t)]^2\mathrm{d}t}{8\int_0^\infty\int_t^\infty[\pi S_1(s)+(1-\pi)S_0(s)]^2[1-\pi S_1(s)-(1-\pi)S_0(s)]^2\mathrm{d}s\mathrm{d}t}.$$

A.2 Proof of Theorem 1

By Yu and Li [27], for any τ such that $S(\tau) > 0$ and $G(\tau) < 1$, we have

$$\lim_{n\to\infty}\sup_{t<\tau}|S_{1n}(t)-S_1(t)| = 0 \quad \text{a.s.} \tag{A4}$$

and

$$\lim_{n\to\infty}\sup_{t<\tau}|S_{0n}(t)-S_0(t)| = 0 \quad \text{a.s.} \tag{A5}$$

As $4\pi^2(1-\pi)^2 \le 1/4$ and $|S_{1n}(t)+S_1(t)-S_{0n}(t)-S_0(t)| < 4$, we have

$$\begin{aligned}
|\mathrm{dCov}_n^2(T,X;\,\tau) - \mathrm{dCov}^2(T,X;\,\tau)| &\le \frac{1}{4}\int_0^\tau|[S_{1n}(t)-S_{0n}(t)]^2-[S_1(t)-S_0(t)]^2|\mathrm{d}t\\
&\le \int_0^\tau|S_{1n}(t)-S_1(t)-S_{0n}(t)+S_0(t)|\mathrm{d}t\\
&\le \int_0^\tau|S_{1n}(t)-S_1(t)|\mathrm{d}t + \int_0^\tau|S_{0n}(t)-S_0(t)|\mathrm{d}t\\
&\le \tau\sup_{t<\tau}|S_{1n}(t)-S_1(t)| + \tau\sup_{t<\tau}|S_{0n}(t)-S_0(t)|.
\end{aligned}$$

By equations (A4) and (A5), $\tau\sup_{t<\tau}|S_{1n}(t)-S_1(t)|$ and $\tau\sup_{t<\tau}|S_{0n}(t)-S_0(t)|$ both converge to 0 almost surely; therefore,

$$\mathrm{dCov}_n^2(T,X;\,\tau) \xrightarrow{a.s.} \mathrm{dCov}^2(T,X;\,\tau).$$

Next, we show the almost sure convergence of the denominator, i.e., $\mathrm{dCov}^2(T,T;\tau)$. First, we bound

$$\Delta := \left|\int_0^\tau\int_t^\tau S_n(s)^2[1-S_n(s)]^2\mathrm{d}s\mathrm{d}t - \int_0^\tau\int_t^\tau S(s)^2[1-S(s)]^2\mathrm{d}s\mathrm{d}t\right|.$$

Similar to the proof for $\mathrm{dCov}_n^2(T, X;\, \tau)$,

$$\begin{aligned}
\varDelta &\le \int_0^{\tau}\int_t^{\tau} |S_n(s)^2[1 - S_n(s)]^2 - S(s)^2[1 - S(s)]^2| \mathrm{d}s\mathrm{d}t\\
&\le 2\int_0^{\tau}\int_t^{\tau} |S_n(s) - S(s)| \mathrm{d}s\mathrm{d}t + 2\int_0^{\tau}\int_t^{\tau} |S_n^2(s) - S^2(s)| \mathrm{d}s\mathrm{d}t\\
&\le 2\int_0^{\tau}\int_0^{\tau} |S_n(s) - S(s)| \mathrm{d}s\mathrm{d}t + 2\int_0^{\tau}\int_0^{\tau} |S_n(s) - S(s)||S_n(s) + S(s)| \mathrm{d}s\mathrm{d}t\\
&\le 6\int_0^{\tau}\int_0^{\tau} |S_n(s) - S(s)| \mathrm{d}s\mathrm{d}t\\
&\le 6\tau^2 \sup_{t<\tau} |S_n(t) - S(t)|.
\end{aligned}$$

Again by equations (A4) and (A5), $6\tau^2 \sup_{t<\tau}|S_n(t) - S(t)|$ converges almost surely to 0; therefore,

$$\mathrm{dCov}_n^2(T, T;\tau) \overset{a.s.}{\longrightarrow} \mathrm{dCov}^2(T, T;\, \tau).$$

To show the almost sure convergence of $\mathrm{dCor}_n^2(T, X;\, \tau)$, we only need to show that $\mathrm{dCov}^2(T, T;\, \tau)$ is strictly positive. Since we assume $1 > S(\tau) > 0$ and $S(t)$ is non-increasing, there exists $0 < \omega_{\min} < 1$ such that $1 - \omega_{\min} > S(t) > \omega_{\min}$ uniformly for $0 \le t \le \tau$; thus,

$$\int_0^{\tau}\int_t^{\tau} S(s)^2[1 - S(s)]^2 \mathrm{d}s\mathrm{d}t > \omega_{\min}^4 \tau^2/2.$$

This completes the proof.

A.3 Derivation for the Gaussian kernel

Let $K(x, y;\, \sigma^2) = \exp(-|x - y\,|^2/2\sigma^2)$ be the Gaussian kernel with bandwidth parameter σ^2. By elementary probability, we have

$$\begin{aligned}
E[K(X_1, X_2)] &= \pi^2 + (1 - \pi)^2 + 2\pi(1 - \pi)e^{-\frac{1}{2\sigma^2}},\\
E[K(T_1, T_2)] &= (1 - \pi)^2 D_{00} + \pi^2 D_{11}\pi^2 + 2\pi(1 - \pi)D_{01},\\
E[K(T_1, T_2)K(X_1, X_2)] &= (1 - \pi)^2 D_{00} + \pi^2 D_{11}\pi^2 + 2\pi(1 - \pi)e^{-\frac{1}{2\sigma^2}}D_{01},\\
E[K(T_1, T_2)K(X_1, X_3)] &= [\pi^2(1 - \pi) + \pi^3 e^{-\frac{1}{2\sigma^2}}]D_{00} + [\pi(1 - \pi)^2 + (1 - \pi)^3 e^{-\frac{1}{2\sigma^2}}]D_{11}\\
&\quad + 2[\pi^2(1 - \pi)e^{-\frac{1}{2\sigma^2}} + \pi(1 - \pi)^2]D_{01}.
\end{aligned}$$

The squared distance covariance based on $K(x, y;\, \sigma^2)$ is

$$\mathrm{dCov}^2(T, X) = E[K(X_1, X_2)]E[K(T_1, T_2)] + E[K(T_1, T_2)K(X_1, X_2)] - 2E[K(T_1, T_2)K(X_1, X_3)], \qquad \text{(A6)}$$

where

$$D_{00} = \int_0^1\int_0^1 \exp\left(-\frac{|t_1 - t_2|^2}{2\sigma^2}\right)\mathrm{d}S_0(t_1)\mathrm{d}S_0(t_2),$$

$$D_{11} = \int_0^1\int_0^1 \exp\left(-\frac{|t_1 - t_2|^2}{2\sigma^2}\right)\mathrm{d}S_1(t_1)\mathrm{d}S_1(t_2),$$

$$D_{01} = \int_0^1\int_0^1 \exp\left(-\frac{|t_1 - t_2|^2}{2\sigma^2}\right)\mathrm{d}S_0(t_1)\mathrm{d}S_1(t_2).$$

When $\pi = 1/2$, equation (A6) can be simplified to:

$$\mathrm{dCov}^2(T, X) = \frac{1 - \exp(-1/2\sigma^2)}{8}(D_{00} + D_{11} - 2D_{01}),$$

References

[1] *A study of idasanutlin with cytarabine versus cytarabine plus placebo in participants with relapsed or refractory acute myeloid leukemia*. https://clinicaltrials.gov/ct2/show/NCT02545283.

[2] Crámer, H. (1928). On the composition of elementary errors. *Skand Aktuar*, *11*, 141–180.

[3] Ditzhaus, M. Genuneit, J., Janssen, A. & Pauly, M. (2021). CASANOVA: Permutation inference in factorial survival designs. *Biometrics*, *79*, 203–215.

[4] Edelmann, D., Richards, D., & Vogel, D. (2020). The distance standard deviation. *Annals of Statistics*, *48*(6), 3395–3416.

[5] Edelmann, D., Welchowski, T., & Benner, A. (2022). A consistent version of distance covariance for right-censored survival data and its application in hypothesis testing. *Biometrics*, *78*, 867–879.

[6] Edelmann, D., & Goeman, J. (2022). A Regression Perspective on Generalized Distance Covariance and the Hilbert-Schmidt Independence Criterion. *Statistical Science*, *37*(4), 562–579.

[7] Fernandez, T., Gretton, A., Rindt, D., & Sejdinovic, D. (2023). A Kernel log-rank test of independence for right-censored data. *Journal of the American Statistical Association*, *118*, 542, 925–936.

[8] Fleming, T. R., O'Fallon, J., & O'Brien, P. (1980). Modified Kolmogorov-Smirnov test procedure with application to arbitrarily right-censored data. *Biometrics*, *36*(4), 607–625.

[9] Gomez Julia, O., Utzet, F., & Moeschberger, M. (1992). Survival analysis for left censored data. *Survival Analysis: State of the Art.* Springer, (pp 269–288).

[10] Gretton, A. Herbrich, R., Smola, A., Bousquet, O., & Scholkopf, B. (2005). Kernel methods for measuring independence. *Journal of Machine Learning Research*, *6*, 2075–2129.

[11] Kim, J. S. (1991). Piecewise exponential estimator of the survivor function. *IEEE Transactions on Reliability*, *40*(2), 134–2794.

[12] Koziol, J. A. (1978). A two sample Cramer-von Mises test for randomly censored data. *Biometrical Journal*, *20*(6), 603–608

[13] Lee, S. H. (2007). On the versatility of the combination of the weighted log-rank statistics. *Computational Statistics and Data Analysis*, *51*(12), 6557–6564.

[14] Lin, R. S. Lin, J., Roychoudhury, S., Anderson, K., Hu, T., & Huang, B. (2020). Alternative analysis methods for time to event endpoints under nonproportional hazards: A comparative analysis. *Statistics in Biopharmaceutical Research*, *12*(2), 187–198.

[15] Panda, S., Shen, C., Perry, R., Zorn, J., Lutz, A., & Priebe, C. (2023). *High-dimensional and universally consistent k-sample tests*. https://arxiv.org/abs/1910.08883.

[16] Qiu, P. & Sheng, J. (2008). A two-stage procedure for comparing hazard rate functions. *Journal of Royal Statistical Society - Series B*, *70*(1), 191–208.

[17] Rizzo, M. L., & Székely, G. J. (2016). Energy distance. *WIREs Computational Statistics*, *8*, 27–38.

[18] Roychoudhury, S., Anderson, K., Ye, J., & Mukhopadhyay, P., (2023). Robust Design and Analysis of Clinical Trials With Nonproportional Hazards: A Straw Man Guidance From a Cross-Pharma Working Group. *Statistics in Biopharmaceutical Research*, *15*(2), 280–294.

[19] Rufibach, K., Heinzmann, D., & Monnet, A. (2020). Integrating phase 2 into phase 3 based on an intermediate endpoint while accounting for a cure proportion-With an application to the design of a clinical trial in acute myeloid leukemia. *Pharmaceutical Statistics*, *19*, 44–58.

[20] Schumacher, M. (1984). Two-sample tests of Cramer-von Mises and Kolmogorov-Smirnov type for randomly censored data. *International Statistical Review*, *52*(3), 263–281.

[21] Shen, C., Panda, S., & Vogelstein, J. (2021). The Chi-square test of distance correlation. *Journal of Computational and Graphical Statistics*, *31*(1), 254–262.

[22] Shen, C., & Vogelstein, J. T. (2005). The exact equivalence of distance and kernel methods in hypothesis Testing. *AStA Advances in Statistical Analysis*, *105*(3), 385–403.

[23] Székely, G., Rizzo, M., & Bakirov, N., (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, *35*(6), 2769–2794.

[24] Wolchok, J. D. (2017). Overall survival with combined nivolumab and iplimumab in advanced melanoma. *New England Journal of Medicine*, *377*, 1345–1356.

[25] Xue, X., Agalliu, I., Kim, M., Wang, T., Lin, J., & Ghavamian, R. (2017). New methods for estimating follow-up rates in cohort studies. *BMC Medical Research Methodology*, *17*, 155.

[26] Yang, S. & Prentice, R. (2010). Improved logrank-type tests for survival data using adaptive weights. *Biometrics*, *66*(1), 30–38.

[27] Yu, Q. & Li, L. (1994). On the strong uniform consistency of the product limit estimator. *Sankhyaaa A*, *56*(3), 416–430.

[28] Zhang, J., Liu, Y., & Cui, H. (2021). Model-free feature screening via distance correlation for ultrahigh dimensional survival data. *Statistical Papers*, *62*, 2711–2738.