A test of the dependence assumption for the Simes-test-based multiple test procedures (DRAFT)

Jiangtao Gou

January 9, 2021

Abstract. The Hochberg procedure is known to be more powerful than the Bonferroni method but the test statistics need to satisfy certain dependence conditions. Due to the lack of validation of the positive dependence assumption in the Hochberg procedure, the US Food and Drug Administration guidance on multiple endpoints in clinical trials provides conservative recommendations on the Hochberg procedure, and limits its application on a small set of standard test statistics. Based on the demand of using the Hochberg procedure in a more flexible way, we develop a test to validate the dependence assumption in the Hochberg procedure and Benjamini-Hochberg procedure. A simulation study is conducted for power analysis, and a case study in metastatic breast cancer is included to illustrate how the proposed test can be applied to validating the dependence type between the progression-free survival and overall survival.

Keywords. Benjamini-Hochberg procedure, Dependence, Hochberg procedure, Multiple testing procedure, Simes test

1 Introduction

The Bonferroni inequality and Simes (1986) inequality are two prevalent tools in developing multiplicity adjustment methods. The widely-used Bonferroni-based methods include the Holm (1979) procedure and the graphical approaches (Bretz et al., 2009; Burman et al., 2009), and the commonly-applied Hochberg (1988) procedure and Benjamini and Hochberg (1995) procedure are based on the Simes test (Tamhane and Gou, 2018). Comparing with the Bonferroni test, the Simes test is more powerful. However, there is no free lunch: the Bonferroni-based testing procedures are flexible and assumption-free, but the Simes-based procedures require certain assumptions about dependence structure. To be more precise, the Simes test controls the type I error rate at the desired level under a specific type of positive dependence, which is the positive dependence through stochastic ordering (PDS). This dependence structure is considerably stricter than the general positive dependence requirement or the positive correlation condition. However, this specific positive dependence requirement is often misinterpreted as a verification of positive correlation coefficient in practice. The draft guideline on multiple endpoints in clinical trials released by the US Food and Drug Administration includes a general description about the Hochberg procedure which is based on the Simes test: "*Hochberg procedure usually will, but is not guaranteed to, control the overall type I error rate for positively-correlated endpoints*" (FDA, 2017).

The Hochberg procedure and other Simes-test-based methods require a relatively strict positive dependence condition to guarantee the type I error rate control. However, there is no statistical method available to test or verify the PDS condition in literature. In order to apply a multiple testing procedure that guarantees the type I error rate control and is

more powerful than the Bonferroni-based procedures, we construct a new goodness-of-fit test to compare the empirical multivariate distribution function with the distribution functions which satisfy the PDS condition. The dependence assumption of the Simes-based methods (e.g., the Hochberg procedure and Benjamini-Hochberg procedure) can therefore be checked by applying this test.

This article follows this strategy to provide statistical verification of the dependence assumptions in multiple testing procedures, and is organized as follows. Section 2 briefly summarizes the PDS dependence assumption and other dependence structures which are related to the Simes test. In Section 3, we develop a new goodness-of-fit test to validate if a sample of multivariate data come from a population that satisfies the PDS condition. Section 4 is an application of the proposed test of the dependence assumption to the progression-free survival (PFS) and overall survival (OS) in metastatic breast cancer patients. Section 5 contains some discussion and directions for future work.

2 Dependence assumptions

We briefly describe three positive dependence conditions in this section: the positive correlation dependence (PCD), positive orthant dependence (POD), and positive dependence through stochastic ordering (PDS). The positive correlation dependence is the least strict condition and easy to be tested. The positive dependence through stochastic ordering is the most strict condition and has no statistical test to verify its validity. The positive quadrant dependence is in the middle.

Consider a random vector $\boldsymbol{U} = (U_1, \ldots, U_n)$. The positive correlation dependence condition is satisfied if $\text{corr}(U_i, U_j) \geq 0$ for all $(i, j) \in \{1, \ldots, n\}^2$ where $i \neq j$. The Pearson's correlation coefficient $r_{ij} = \text{corr}(U_i, U_j)$ is usually used and calculated as $\text{cov}(U_i, U_j)/\sqrt{\text{var}(U_i)\text{var}(U_j)}$. One-sided tests are available to verify whether all correlation coefficients are positive based on the value of the sample correlation coefficient, including Student's t-test (Soper et al., 1917), permutation test, bootstrap, and distribution-based tests (Fisher, 1915; Hotelling, 1953).

The positive orthant dependence is a stricter condition comparing with the positive correlation dependence. The random vector $\boldsymbol{U}$ is POD if both $\Pr(U_1 > u_1, \ldots, U_n > u_n) \geq \prod_{i=1}^n \Pr(U_i > u_i)$ and $\Pr(U_1 \leq u_1, \ldots, U_n \leq u_n) \geq \prod_{i=1}^n \Pr(U_i \leq u_i)$ are satisfied for all possible $(u_1, \ldots, u_n)$. The positive orthant dependence is also called positive quadrant dependence (PQD) when $n = 2$ for bivariate cases (Lehmann, 1966). POD/PQD implies PCD. For $n \geq 3$, the pairwise PQD is stricter than PCD but less strict than POD, where the random vector $\boldsymbol{U}$ is pairwise PQD if all pairs (U_i, U_j) are PQD. Shetty and Pandit (2003), Janic-Wroblewska et al. (2004), Scaillet (2005), Gijbels and Sznajder (2013), and Ledwina and Wyłupek (2014) have proposed various methods to test the PQD condition.

The positive dependence through stochastic ordering (PDS) implies both PCD and PQD/POD, and is the most strict condition among all three dependence conditions we considered. The random vector $\boldsymbol{U}$ is PDS if $\Pr(U_1 > u_1, \ldots, U_{i-1} > u_{i-1}, U_{i+1} > u_{i+1}, \ldots, U_n > u_n \mid U_i = u_i)$ is increasing in u_i for all possible $(u_1, \ldots, u_{i-1}, u_{i+1}, u_n)$ (Shaked, 1977; Block et al., 1985). An equivalent condition to check if $\boldsymbol{U}$ is PDS is to verify if $\mathbb{E}[\psi(\boldsymbol{U}) \mid U_i = u_i]$ is increasing in u_i where ψ is any increasing function.

Consider testing n hypotheses simultaneously. Let $P_{(1)} \leq \cdots P_{(n)}$ be the ordered p-values and denote by $H_{(1)}$, ..., $H_{(n)}$ the corresponding null hypotheses. The Simes inequality $\Pr\left(\cup_{i=1}^{n}\left\{P_{(i)} \leq i\alpha/n\right\}\right) \leq \alpha$ holds if the test statistics satisfy the PDS condition (Sarkar and Chang, 1997; Sarkar, 1998; Benjamini and Yekutieli, 2001; Sarkar, 2002). The PDS condition in the Simes test can be slightly weaken by assuming that $\mathbb{E}\left[\psi\left(\boldsymbol{U}\right) \mid U_i \leq u_i\right]$ is increasing in u_i for all i and for any increasing function ψ (Finner et al., 2009). However, only assuming PQD/POD or PCD does not guarantee a valid type I error control of the Simes test, and the inflation of error rate may not be negligible. Therefore, when applying the Hochberg (1988) procedure or Benjamini and Hochberg (1995) procedure, the dependence assumption need to be checked ahead.

3 Tests for the dependence assumptions

Since correlation is not synonymous with dependence, the tests of correlation coefficient are not adequate for validating the dependence assumption in the Simes test. In literature, there are statistical tests for positive quadrant dependence (PQD), but no test for positive dependence through stochastic ordering (PDS) to the best of our knowledge. The setup of the null and alternative hypotheses follows the way of testing for distributional assumptions for normality (Anderson and Darling, 1952; Shapiro and Wilk, 1965). The distributional assumption here is the PDS condition in the Simes test, where the null distribution satisfies the PDS assumption but the alternative does not.

We start from the bivariate case and let $(X_{11}, X_{21}), \ldots, (X_{1m}, X_{2m}) \in \mathbb{R}^2$ be a total of m pairs of independent and identically distributed observations from (X_1, X_2), and let σ and τ are two permutations of the set $\{1, \ldots, m\}$ corresponding to the sequence of the order statistics $X_{1\sigma(1)} \leq \cdots \leq X_{1\sigma(m)}$ and that of the order statistics $X_{2\tau(1)} \leq \cdots \leq X_{2\tau(m)}$. Next, we divide $\mathbb{R}$ into k pieces in two ways which are $\cup_{i=1}^{k}(u_{1,i-1}, u_{1,i}]$ and $\cup_{i=1}^{k}(u_{2,i-1}, u_{2,i}]$, where $-\infty = u_{10} < u_{11} < \ldots < u_{1,k-1} < u_{1k} = +\infty$ and $-\infty = u_{20} < u_{21} < \ldots < u_{2,k-1} < u_{2k} = +\infty$. We divide m pairs of observations into k mutually exclusive and collectively exhaustive sets in two ways: $\mathcal{S}_1 = \{\{(x_{1i}, x_{2i}) : x_{1i} \in (u_{1,j-1}, u_{1,j}]\} : j = 1, \ldots, k\}$ and $\mathcal{S}_2 = \{\{(x_{1i}, x_{2i}) : x_{2i} \in (u_{2,j-1}, u_{2,j}]\} : j = 1, \ldots, k\}$. In addition, we denote intervals $(u_{1,j-1}, u_{1,j}]$ and $(u_{2,j-1}, u_{2,j}]$ by I_{1j} and I_{2j} respectively. Assuming (X_1, X_2) satisfies the PDS condition, the null hypothesis can be stated as

$$\begin{aligned} H_0 : F_1(x_1 \mid x_2 \in I_{2i}) \leq F_1(x_1 \mid x_2 \in I_{2j}) \text{ and } F_2(x_2 \mid x_1 \in I_{1i}) \leq F_2(x_1 \mid x_1 \in I_{1j}) \\ \text{for all } i \text{ and } j \text{ where } 1 \leq i < j \leq k, \end{aligned} \tag{1}$$

where $F_1(x_1 \mid x_2 \in I_{2i})$ is the conditional distribution of X_1 given X_2 falls in interval I_{2i} and $F_2(x_2 \mid x_1 \in I_{1i})$ is the conditional distribution of X_2 where $u_{1,i-1} < X_1 \leq u_{1,i}$. The total number of pairwise comparisons is $k(k-1)$ for bivariate case where the number of comparisons for $F_1(x_1 \mid x_2)$ and that for $F_2(x_2 \mid x_1)$ are both $\binom{k}{2}$. Note that the empirical distribution function converges to the true distribution according to the Glivenko–Cantelli theorem. We propose to apply the one-sided Mann-Whitney test (Wilcoxon, 1945; Mann and Whitney, 1947) for each pairwise comparison and combine the $k(k-1)$ results in a Bonferroni way for the overall null hypothesis in (1). In addition, since using unequal group sizes will reduce the statistical power of the Mann-Whitney test, we usually divide the

observed values equally into k groups in the proposed PDS condition test, where the size of set $\{(x_{1i}, x_{2i}) : x_{1i} \in I_{1j}, 1 \le i \le m\}$ and $\{(x_{1i}, x_{2i}) : x_{2i} \in I_{2j}, 1 \le i \le m\}$ are approximately equal to m/k for $j = 1, \ldots, k$.

We first evaluate the control of type I error rate when (X_1, X_2) satisfies the PDS condition. Consider a bivariate normal distribution with correlation coefficient $\rho \ge 0$. We simulate bivariate normally distributed random numbers $(x_{11}, x_{21}), \ldots, (x_{1m}, x_{2m})$ with $\rho = 0$ and $\rho = 0.3$, where the sample size $m = 100$. The number of replicas is 10^4. Under a nominal level $\alpha = 5\%$, Table 1 shows the simulated type I error rates of the proposed PDS condition test where the group number $k = 4$, 6 and 8. Besides using the Mann-Whitney test for pairwise comparisons, we also apply the parametric two-sample t-test for pairwise comparisons in the PDS condition test. The results of using Mann-Whitney and t-tests are similar in this setting.

Table 1: Simulated type I error rate of the PDS condition tests ($\alpha = 5\%$)

Group number	$\rho = 0$		$\rho = 0.3$	
	Mann-Whitney	Student's t	Mann-Whitney	Student's t
$k = 4$	0.036	0.037	0.001	0.002
$k = 6$	0.038	0.036	0.004	0.004
$k = 8$	0.031	0.033	0.003	0.004

Next we check the statistical power of the PDS condition test. We construct a bivariate distribution that is positively correlated but does not satisfy the PDS condition. Consider another bivariate distribution with distribution function $C(p_1, p_2)$ on $[0, 1]^2$. Its joint density function is

$$\frac{\partial^2 C(p_1, p_2)}{\partial p_1 \partial p_2} = \begin{cases} \sqrt{\frac{b}{p_2}}, & 0 \le p_1 \le 1/2, 4bp_1^2 \le p_2 < b \text{ and} \\ & 1/2 < p_1 \le 1, 4b(p_1 - 1/2)^2 \le p_2 < b \text{ (Region I)}, \\ \sqrt{\frac{b}{1-p_2}}, & 0 \le p_1 \le 1/2, 1 - b < p_2 \le 1 - 4b(p_1 - 1/2)^2 \text{ and} \\ & 1/2 < p_1 \le 1, 1 - b < p_2 \le 1 - 4b(p_1 - 1)^2 \text{ (Region II)}, \\ 1, & b \le p_2 \le 1 - b \text{ (Region III)}, \\ 0, & [0, 1]^2 \backslash \{\text{I} \cup \text{II} \cup \text{III}\} \text{ (Region IV)}, \end{cases} \tag{2}$$

where the parameter $b \in [0, 1/2]$, and the region for defining $C(p_1, p_2)$ is presented in Figure 1. Direct calculation shows that the marginal distribution of P_1 and P_2 are Unif(0,1), so $C(p_1, p_2)$ is a copula. The correlation coefficient is $\rho = b(1 - 3b/5)$ as an increasing function of b, where $\rho = 0$ when $b = 0$ and $\rho = 0.35$ when $b = 1/2$. The random variable (P_1, P_2) is PC but not PDS.

Simulated data are generated from the distribution in (2) and Figure 1 with $b = 0.4$ and sample size $m = 200$, which are $\{(P_{1i}, P_{2i})\}_{i=1,\ldots,200}$. We repeat this random p-value generation step 10^4 times. When $b = 0.4$, the correlation $\text{corr}(P_1, P_2) = 0.304$. A 5%-level correlation test between P_1 and P_2 concludes a positive association with a simulated chance of 99.8%. The statistical powers of detecting that the distribution of (P_1, P_2) does not satisfy

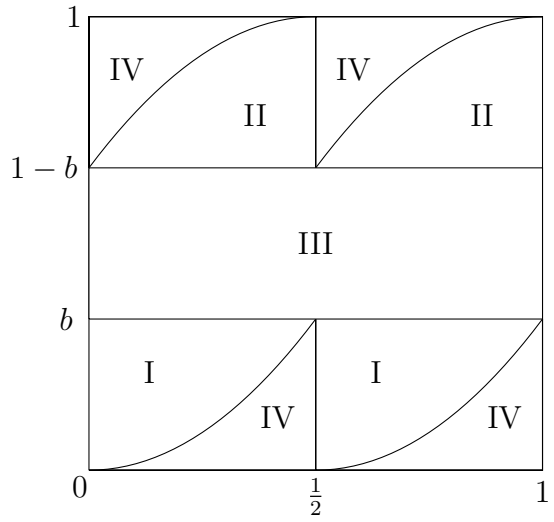

Figure 1: The region for defining $C(p_1, p_2)$

the PDS condition are listed in Table 2, where k ranges from 2 to 16. The splitting of the region of the observed p-values become finer when k increases, and the power of the PDS condition test will increase first and then decrease. When selecting the parameter k, we need to balance between the good approximation of condition distribution functions which requires a finer splitting, and enough sample points in each piece what require a coarser splitting. For bivariate distribution, we suggest using k at least 4. When the sample size m increases, you can increase the number of pieces k. However, it is usually not necessary to use k greater than 10. Table 2 also reports the summary statistics of the distribution of p-value of the PDS condition test, including mean and standard deviation (SD), and median and interquartile range (IQR).

For multivariate random variable $(X_1, \cdots, X_n)$, we use the proposed test to validate the pairwise PDS condition, which is a weaker dependence assumption comparing with the PDS condition. In other words, we verify the condition that $\mathbb{E}\left[\psi\left(X_i\right) \mid X_j = x\right]$ is increasing in x for any increasing function ψ for all $1 \leq i, j \leq n$. This strategy allows us to only compare univariate distribution functions and ensure the observations in each piece are not too few.

4 Application

In oncology clinical trials, the overall survival (OS) and progression-free survival (PFS) are two commonly-used efficacy endpoints, where OS is the time from registration to time of death and PFS is time from registration to the disease progression or death (Green et al., 2012; Gou and Xi, 2019). These two survival endpoints are often used together as observed indicators to evaluate the efficacy of treatments, and therefore the statistical multiplicity need to be considered. When a Simes-based method, for example, the Hochberg procedure, is applied to adjust the multiplicity from multiple endpoints involved in a clinical study, we can apply the proposed PDS condition test to validate its dependence assumption.

The data used here for testing the PDS condition were collected by Adunlin et al. (2015),

Table 2: Simulated power and distribution of p-value of the PDS condition test (sample size $m = 200$)

k	Distribution of p-value				Power
	mean	SD	median	IQR	
2	0.890	0.235	1	0.0204	0.007
3	0.986	0.093	1	0	0.001
4	0.042	0.161	0.0001	0.0034	0.895
5	0.313	0.389	0.0899	0.5898	0.435
6	0.039	0.148	0.0001	0.0041	0.893
7	0.053	0.170	0.0012	0.0146	0.848
8	0.023	0.103	0.0002	0.0036	0.926
9	0.024	0.103	0.0004	0.0050	0.919
10	0.021	0.096	0.0004	0.0041	0.928
11	0.019	0.083	0.0006	0.0052	0.929
12	0.021	0.088	0.0008	0.0065	0.924
13	0.024	0.094	0.0016	0.0086	0.912
14	0.024	0.089	0.0017	0.0098	0.906
15	0.029	0.099	0.0028	0.0131	0.881
16	0.030	0.094	0.0040	0.0163	0.878

and comprise hazard ratios (HR) for PFS and OS in metastatic breast cancer patients receiving anthracyclines, taxanes, or targeted therapies between experimental and standard treatment arms from randomized trials published between January 1990 and August 2015. A total of 75 pairs of hazard ratios for PFS and OS are presented in the scatter plots in Figure 2. The correlation coefficients with 95% confidence intervals are: (1) Pearson's r, 0.409 with a t CI $[0.201, 0.582]$, (2) Kendall's τ, 0.344 with a bootstrap CI $[0.183, 0.498]$, (3) Spearman's ρ, 0.461 with a jackknife CI $[0.241, 0.682]$. Therefore, the positive correlation between PFS and OS in metastatic breast cancer patients is concluded to be statistically significant based on Adunlin et al. (2015)'s systematic review.

We conduct a PDS condition test proposed in Section 3 with region-splitting parameter $k = 8$, where we partition the hazard ratio pairs equally into eight groups based on either the HR for PFS or the HR for OS to calculate the empirical conditional distribution functions. The p-value from the test of the PDS condition is equal to 1, indicating that the positively dependent through stochastic ordering (PDS) assumption in the Simes test and Hochberg procedure is acceptable. Therefore, we confirm the validity of using the Hochberg procedure in a clinical study when PFS and OS are involved as multiple endpoints. Lastly, we apply the Bonferroni method to combine the p-values of pairwise comparisons in the PDS condition test as the default setting. We can combine p-values using other methods (Fisher, 1925; Stouffer et al., 1949; Gou et al., 2014; Gou and Tamhane, 2014). For example, if we use the Simes method to combine the p-values of pairwise comparisons in this application, the p-value from the PDS condition test is 0.999.

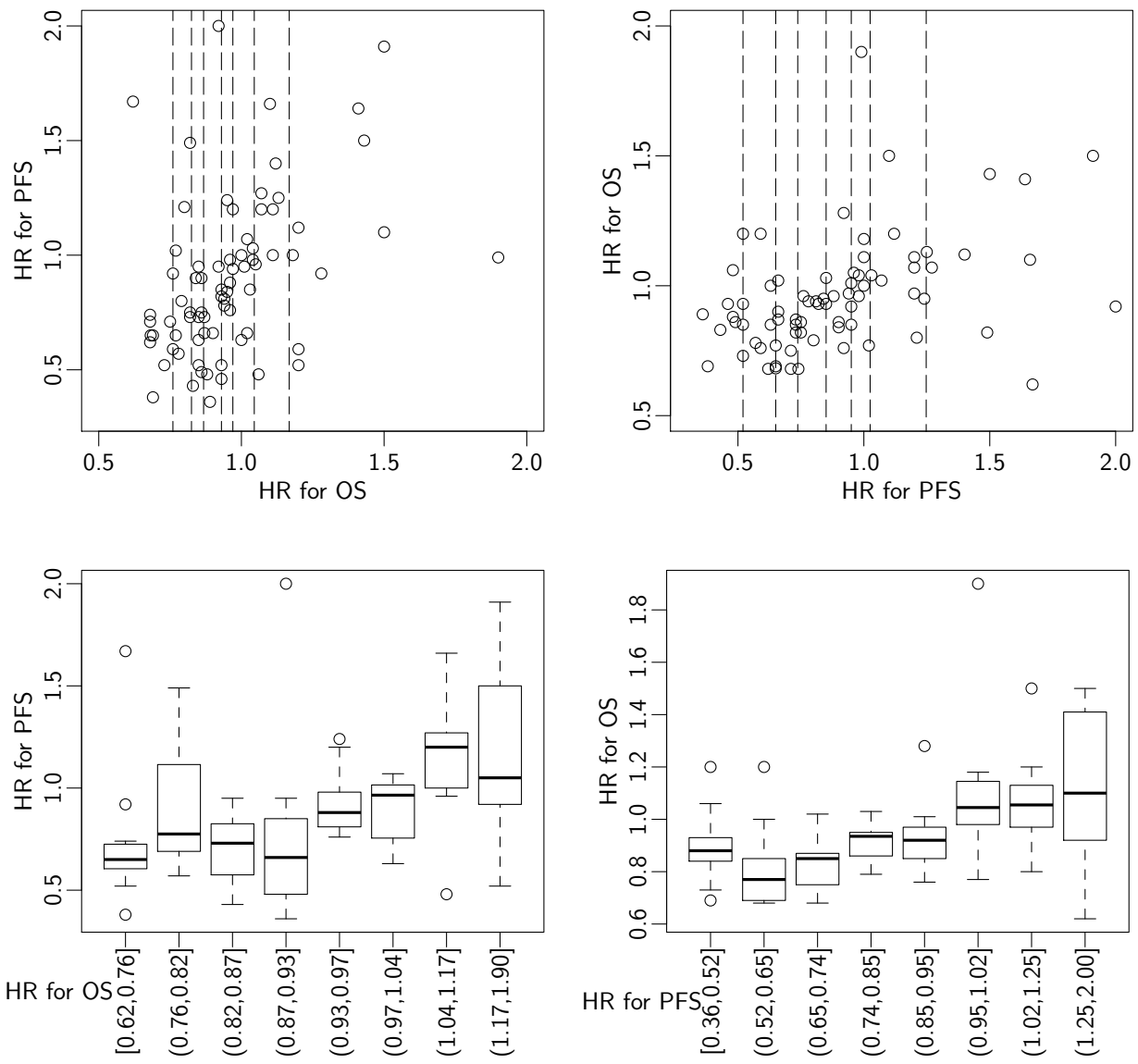

Figure 2: Hazard ratios for PFS and OS (top panels) with spitted regions (dashed lines), empirical conditional distribution function of PFS given OS (bottom left panel) and that of OS given PFS (bottom right panel)

5 Discussion

The proposed PDS condition test in this article provides practitioners a tool to test if a sample of multivariate data come from a population that satisfies positively dependent through stochastic ordering (PDS) assumption. Validating the PDS condition is essential when the Simes-test-based methods are applied for multiplicity adjustments. The prevalent Hochberg (1988) procedure that controls the familywise error rate (FWER) (Hochberg and Tamhane, 1987), and the Benjamini and Hochberg (1995) procedure that controls the false discovery rate (FDR) are both based on the Simes test. In clinical studies, this test of the PDS condition allows validating the assumption of the Hochberg procedure and extending its implementation to multiple hypothesis testing questions which can pass this assumption test, and therefore it is not necessary for practitioners to limit the application of the Hochberg procedure on "*two positively-correlated dependent tests with standard test statistics, such as the normal Z, student's t, and 1 degree of freedom chi-square*" mentioned in the US Food and Drug Administration guideline (FDA, 2017).

We include the asymptotic bounds of the type I error rate under positive and negative dependence of the Simes test and other multiple testing methods. It would be interesting to consider the dependence assumption for the type II error rate control when multiple hypotheses are involved, using appropriate models for p-values under true significance (Zhang and Gou, 2016). Another future research direction of interest is to construct additional statistics to test the dependence assumption. For example, consider a bivariate distribution of p-values (P_1, P_2) with distribution function $C(p_1, p_2)$, where Kendall's tau is calculated using $\tau = 4\int_0^1 \int_0^1 C(p_1, p_2)dC(p_1, p_2) - 1$ and Spearman's rho is given by $\rho = 12\int_0^1 \int_0^1 C(p_1, p_2)dp_1dp_2 - 3$ (Balakrishnan and Lai, 2009; Durante and Sempi, 2016). If (P_1, P_2) is PQD, then $\rho/\tau \leq 3$. If (P_1, P_2) is PDS, then ρ and τ satisfy the inequality $\rho/\tau \geq 1$ in addition to $\rho/\tau \leq 3$. Therefore, $\widehat{\rho}/\widehat{\tau}$ can be used as a test statistics for testing the dependence assumption in the Simes inequality. Another direction is to construct test statistics using the measures of the spread between permutations (Diaconis, 1988; Critchlow, 1992; Alvo and Pan, 1997; Alvo and Yu, 2014). In addition, we can consider testing the PDS assumption under a certain distribution family (Bodnar and Dickhaus, 2017; Finner et al., 2017).

References

Adunlin, G., Cyrus, J. W. W., and Dranitsaris, G. (2015). Correlation between progression-free survival and overall survival in metastatic breast cancer patients receiving anthracyclines, taxanes, or targeted therapies: a trial-level meta-analysis. *Breast Cancer Research and Treatment* **154,** 591–608.

Alvo, M. and Pan, J. (1997). A general theory of hypothesis testing based on rankings. *Journal of Statistical Planning and Inference* **61,** 219–248.

Alvo, M. and Yu, P. L. (2014). *Statistical methods for ranking data.* Frontiers in probability and the statistical sciences. Springer-Verlag, New York.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Annals of Mathematical Statistics* **23,** 193–212.

Balakrishnan, N. and Lai, C. D. (2009). *Continuous bivariate distributions*. Springer, New York; London, 2 edition.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57,** 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29,** 1165–1188.

Block, H. W., Savits, T. H., and Shaked, M. (1985). A concept of negative dependence using stochastic ordering. *Statistics and Probability Letters* **3,** 81–86.

Bodnar, T. and Dickhaus, T. (2017). On the Simes inequality in elliptical models. *Annals of the Institute of Statistical Mathematics* **69,** 215–230.

Bretz, F., Maurer, W., Brannath, W., and Posch, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine* **28,** 586–604.

Burman, C.-F., Sonesson, C., and Guilbaud, O. (2009). A recycling framework for the construction of Bonferroni-based multiple tests. *Statistics in Medicine* **28,** 739–761.

Critchlow, D. E. (1992). On rank statistics: an approach via metrics on the permutation group. *Journal of Statistical Planning and Inference* **32,** 325–346.

Diaconis, P. (1988). *Chapter 6: Metrics on Groups, and Their Statistical Uses*, volume 11 of *Lecture Notes–Monograph Series*, pages 102–130. Institute of Mathematical Statistics, Hayward, California.

Durante, F. and Sempi, C. (2016). *Principles of Copula Theory*. CRC Press, Taylor & Francis Group, Boca Raton, Florida.

FDA (2017). *Multiple endpoints in clinical trials guidance for industry (draft guidance)*. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). FDA-2016-D-4460.

Finner, H., Dickhaus, T., and Roters, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *The Annals of Statistics* **37,** 596–618.

Finner, H., Roters, M., and Strassburger, K. (2017). On the Simes test under dependence. *Statistical Papers* **58,** 775–789.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10,** 507–521.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Gijbels, I. and Sznajder, D. (2013). Positive quadrant dependence testing and constrained copula estimation. *Canadian Journal of Statistics* **41,** 36–64.

Gou, J. and Tamhane, A. C. (2014). On generalized Simes critical constants. *Biometrical Journal* **56,** 1035–1054.

Gou, J., Tamhane, A. C., Xi, D., and Rom, D. (2014). A class of improved hybrid Hochberg-Hommel type step-up multiple test procedures. *Biometrika* **101,** 899–911.

Gou, J. and Xi, D. (2019). Hierarchical testing of a primary and a secondary endpoint in a group sequential design with different information times. *Statistics in Biopharmaceutical Research* **11,** 398–406.

Green, S., Benedetti, J., Smith, A., and Crowley, J. (2012). *Clinical Trials in Oncology*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, Boca Raton, Florida, 3rd edition.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75,** 800–802.

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*. John Wiley and Sons, New York, New York.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6,** 65–70.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)* **15,** 193–232.

Janic-Wroblewska, A., Kallenberg, W., and Ledwina, T. (2004). Detecting positive quadrant dependence and positive function dependence. *Insurance: Mathematics and Economics* **34,** 467–487.

Ledwina, T. and Wyłupek, G. (2014). Validation of positive quadrant dependence. *Insurance: Mathematics and Economics* **56,** 38–47.

Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics* **37,** 1137–1153.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18,** 50–60.

Sarkar, S. K. (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *The Annals of Statistics* **26,** 494–504.

Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *The Annals of Statistics* **30,** 239–257.

Sarkar, S. K. and Chang, C. K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Asociation* **92,** 1601–1608.

Scaillet, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **33,** 415–427.

Shaked, M. (1977). A family of concepts of dependence for bivariate distributions. *Journal of the American Statistical Association* **72,** 642–650.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52,** 591–611.

Shetty, I. and Pandit, P. (2003). Distribution-free tests for independence against positive quadrant dependence: A generalization. *Statistical Methods & Applications* **12,** 5–17.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73,** 751–754.

Soper, H. E., Young, A. W., Cave, B. M., Lee, A., and Pearson, K. (1917). On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "Student" and R. A. Fisher. A cooperative study. *Biometrika* **11,** 328–413.

Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., and Williams, R. M., J. (1949). *The American soldier: Adjustment during army life. (Studies in social psychology in World War II)*, volume 1. Princeton University Press, Princeton.

Tamhane, A. C. and Gou, J. (2018). Advances in p-value based multiple test procedures. *Journal of Biopharmaceutical Statistics* **28,** 10–27.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin* **1,** 80–83.

Zhang, F. and Gou, J. (2016). A p-value model for theoretical power analysis and its applications in multiple testing procedures. *BMC Medical Research Methodology* **16,** 135.