Trimmed weighted Simes’ test for two one-sided hypotheses with arbitrarily correlated test statistics

Werner Brannath∗1, Frank Bretz 2, Willi Maurer 2, and Sanat Sarkar 3

1 Section of Medical Statistics, Medical University of Vienna, A-1090 Vienna, Austria
2 Statistical Methodology, Novartis Pharma AG, Basel, Switzerland
3 Department of Statistics, Temple University, Philadelphia, PA 19122, USA

Received 15 November 2004, revised 30 November 2004, accepted 2 December 2004
Published online 3 December 2004

Summary
The two-sided Simes test is known to control the type I error rate with bivariate normal test statistics. For one-sided hypotheses control of the type I error rate requires that the correlation between the bivariate normal test statistics is non-negative. In this paper we introduce a trimmed version of the one-sided weighted Simes test for two hypotheses which rejects if (i) the one-sided weighted Simes test rejects and (ii) both p-values are below one minus the respective weighted Bonferroni adjusted level. We show that the trimmed version controls the type I error rate at nominal significance level $\alpha$ if (i) the common distribution of test statistics is point symmetric and (ii) the two-sided weighted Simes test at level $2\alpha$ controls the level. These assumptions apply, for instance, to bivariate normal test statistics with arbitrary correlation. In a simulation study we compare the power of the trimmed weighted Simes test with the power of the weighted Bonferroni test and the untrimmed weighted Simes test.

An additional result of this paper ensures type I error rate control of the usual weighted Simes test under a weak version of the positive regression dependence condition for the case of two hypotheses. This condition is shown to apply to the two-sided p-values of one- or two-sample t-tests for bivariate normal endpoints with arbitrary correlation and to the corresponding one-sided p-values if the correlation is non-negative. The Simes test for such types of bivariate t-tests has not been considered before. According to our main result, the trimmed version of the weighted Simes test then also applies to the one-sided bivariate t-test with arbitrary correlation.

Key words: Hochberg procedure, multiple endpoints, weighted Bonferroni-Holm procedure, positive regression dependence condition, step up test, step down test, strong control of familywise error rate, t-test.

1 Introduction
It is well known that the Simes test (Simes, 1986) is uniformly more powerful than the Bonferroni test with the power advantage being most pronounced when the test statistics are positively correlated and/or the alternative hypotheses hold simultaneously. For example, for two null hypotheses the Simes test additionally rejects the intersection of the two hypotheses if both p-values are below the unadjusted nominal significance level $\alpha$. There are many applications where it is advantageous to use the Simes test. In Section 2 we describe a large confirmatory clinical trial comparing a new treatment with a control for two primary endpoints and where it is sufficient to show efficacy for at least one endpoint.

In general, Simes’ test does not control the type I error rate (Samuel-Cahn, 1996), and with two hypotheses the type I error rate can be inflated up to $1.5\alpha$ (Hommel, 1983; see also the example in Remark (ii) of Section 4.2). Type I error control has only been verified for special cases, such as for two-sided hypotheses with arbitrarily correlated bivariate normal test statistics and for one-sided hypotheses with non-negatively correlated bivariate normal test statistics (Samuel-Cahn, 1996, 2004; Sarkar and

* Corresponding author: e-mail: werner.brannath@meduniwien.ac.at, Phone: +43 40400 7480, Fax: +43 40400 7477
Chang, 1997; Sarkar, 1998). However, in many applications, like the one described in Section 2, type I error rate control for the intersection of one-sided hypotheses is required without making assumptions on the correlation. In this case neither the two-sided nor the one-sided Simes test may be acceptable in the regulatory setting of confirmatory clinical trials.

The goal of this paper is to modify the one-sided weighted Simes test such that it controls the type I error rate also for negative correlations but has similar power advantages as the unmodified Simes test. To this end, we consider a trimmed version of the weighted Simes test for two one-sided hypotheses, which is shown to control the type I error rate at a pre-specified significance level whenever the bivariate test statistic is point symmetric and the weighted two-sided Simes test has type I error rate two times that level. This result applies, for instance, to arbitrarily correlated bivariate normal test statistics. Accordingly, the trimmed weighted Simes test rejects whenever the weighted Simes test rejects and both p-values are below one minus the respective weighted Bonferroni-adjusted level. If at least one of the test statistics points towards the negative direction, such that the p-value is larger than one minus the respective weighted Bonferroni adjusted level, then the trimmed weighted Simes test retains both hypotheses. In a clinical trial with two primary endpoints it would be difficult to claim overall efficacy if one of the treatment effects is significantly negative. In this example, the trimmed Simes test thus provides the same conclusions of practical relevance as the untrimmed Simes test. Note that the trimmed weighted Simes test uniformly improves a similarly trimmed version of the weighted Bonferroni test. Finally, applying the closed testing principle to the trimmed weighted Simes test provides a multiple test procedure which strongly controls the familywise error rate for arbitrarily correlated bivariate normal test statistics. The motivation for trimming the Simes test is similar to the consistency requirement introduced by Huque and Alosh (2009). They considered the problem of testing a primary and a secondary endpoint and proposed a trimming strategy to avoid similar interpretation problems as described above.

In a simulation study we investigate the operation characteristics of the trimmed weighted Simes test in comparison to the weighted Bonferroni and the untrimmed weighted Simes test for normally distributed test statistics. We conclude that the loss in power due to trimming is generally very small if the effects are in the same direction and that the trimmed Simes test still improves upon the Bonferroni test.

Another result of this paper ensures that the usual (untrimmed) Simes test controls the type I error rate under a weaker than the usual version of positive regression dependency and that this condition applies to the one- or two-sample t-tests for bivariate normal endpoints with arbitrary correlation and to the corresponding one-sided t-tests if the correlation is non-negative. To our knowledge, the Simes test has not been considered yet for such types of bivariate t-tests. We also show that the trimmed weighted Simes test can be applied to such one-sided t-tests for bivariate normal endpoints with arbitrary correlation.

The paper is organized as follows. In Section 2 we present the essential features of the clinical trial example which has motivated this paper. In Section 3 we introduce a weak version of the positive regression dependence condition and verify type I error rate control of the usual weighted Simes test under this condition. In Section 4 we introduce the trimmed version of the weighted Simes test and investigate its operating characteristics by a simulation study for bivariate normal test statistics. In Section 5 we discuss the applicability of the weighted two-sided Simes and the weighted one-sided trimmed Simes test to the p-values of t-tests for normally distributed endpoints. The paper closes with a discussion in Section 6. Proofs of theorems are given in the Appendix.

2 Motivating Example

This paper is motivated by a long-term controlled clinical trial in patients with impaired glucose tolerance (IGT) who have established or are at high risk for cardiovascular disease (CVD). The primary objectives of the trial is to confirm that the treatment will delay the occurrence of CVD morbidity and mortality and of progression to diabetes (PD). The CVD events are assessed by a composite endpoint including death, myocardial infarction, stroke, or hospitalization for heart failure plus coronary revascularization and
hospitalization for unstable angina. Progression to diabetes is assessed based on six monthly laboratory tests.

The elementary one-sided hypothesis for the CV endpoint will be tested by means of a log-rank test on the time to the first CV event, while the test of the PD endpoint will be based on a proportional odds model on a discrete time scale. We aim at controlling the familywise type I error rate at the one-sided significance level $\alpha = 0.025$ by using a closed testing procedure based on a version of Simes’ test that assigns different weights to the hypotheses. PD is assigned 1/5 of the total weight, whereas 4/5 of the total weight is assigned to the CV endpoint. The rationale is that progression to diabetes is expected to occur at a much higher rate than the CV endpoint. The unequal weights allow to achieve sufficient power of the CV endpoint while still retaining reasonable power for the progression to diabetes endpoint.

If could be shown to be valid, the weighted one-sided Simes test would be preferable over the weighted one-sided Bonferroni test, in particular, because the treatment is expected to be efficient in both endpoints. For the validity of the weighted one-sided Simes test one would need to assume that the resulting test statistics (which can be assumed to be bivariate normal) are positively correlated. Although a positive correlation is plausible, a negative correlation cannot be excluded. Hence the validity of the Simes test appears questionable. In this paper we therefore introduce a modification of the weighted Simes test which is valid under arbitrary correlations and has a similar power advantage over the Bonferroni test, in particular, when the correlation is positive and the treatment is efficient in both endpoints.

### 3 Weighted Simes’ test for two hypotheses

Let $p_A$ and $p_B$ denote uniformly distributed p-values for two hypotheses $H_A$ and $H_B$ satisfying $P_{H_i}(p_i \leq u) = u$ for all $0 < u < 1$ and $i \in \{A, B\}$. We consider the weighted Simes test for the intersection hypothesis $H_0 = H_A \cap H_B$. To this end we split the significance level $\alpha$ into the adjusted levels $\alpha_A, \alpha_B \geq 0$ with $\alpha_A + \alpha_B = \alpha$. According to Benjamini and Hochberg (1997) we reject $H_0$ if either

\[
(a) \quad p_A \leq \alpha_A \quad \text{or} \quad (b) \quad p_B \leq \alpha_B \quad \text{or} \quad (c) \quad \max(p_A, p_B) \leq \alpha. \tag{1}
\]

Recall that the weighted Bonferroni test rejects $H_0$ based on either (a) or (b). Hence, the weighted Simes test enlarges the rejection region of the weighted Bonferroni test by the additional rejection rule (c). Rejection rule (c) is particularly appealing for confirmatory clinical trials with two endpoints, because a significant effect in both endpoints is often considered as convincing evidence for an overall benefit of the treatment. If $\alpha_A = \alpha_B = \alpha/2$ we denote rejection rule (1) as the unweighted Simes test.

When testing one-sided hypotheses then rejection rule (1) must be applied to one-sided p-values, whereas for two-sided hypotheses rejection rule (1) is applied to two-sided p-values $p_A$ and $p_B$. Accordingly, we refer to these procedures as the one- and two-sided Simes test, respectively.

Hochberg and Lieberman (1994) suggested another extension of Simes’ test for the possibility of incorporating weights. With this test we would reject $H_0$ if either $\min(p_A/\alpha_A, p_B/\alpha_B) \leq 1$ or $\max(p_A/\alpha_A, p_B/\alpha_B) \leq 2$. Since, with this version, we may still be forced to retain $H_0$ if both p-values are below $\alpha$ (e.g. when $\alpha_A = 0.02$, $\alpha_B = 0.005$ and $p_A = p_B = 0.022$), we prefer using rejection rule (1) and focus on Benjamini and Hochberg’s version of the weighted Simes test.

Type I error rate control of the Simes test has been mainly verified for the unweighted version under specific positive dependence conditions (e.g. Samuel-Cahn, 1996; Sarkar and Chang, 1997; Sarkar, 1998; Sarkar, 2008). In case of two hypotheses, type I error rate control has been verified under the positive regression dependence condition of Lehmann (1966). Chang (1997) showed for two hypotheses that under this condition rejection rule (1) controls the type I error rate also for $\alpha_A \neq \alpha_B$. This implies, for instance, that the weighted Simes test controls the type I error rate for one-sided hypotheses if the test statistics are bivariate normal with non-negative correlation and for two-sided hypotheses with bivariate normal and arbitrarily correlated test statistics (Sarkar and Chang, 1997).
A careful investigation of the proof in Chang (1997) reveals that the weighted Simes test (1) controls the type I error rate under the following somewhat weaker condition:

\[ P(p_B > \alpha \mid p_A = u) \text{ and } P(p_A > \alpha \mid p_B = u) \text{ are non-decreasing in } u \text{ for } 0 \leq u \leq \alpha \]  

where \( P \) is the probability under the intersection hypothesis \( H_0 \). The positive regression dependence condition of Lehmann is stronger as it requires \( P(p_B > v \mid p_A = u) \) and \( P(p_A > v \mid p_B = u) \) to be non-decreasing in \( u \) for all \( 0 \leq u \leq 1 \) and all \( 0 < v < 1 \). For completeness we verify the following theorem in the Appendix.

**Theorem 3.1** If \( p_A \) and \( p_B \) satisfy (2) then the type I error rate of the weighted Simes test with rejection rule (1) is bounded by \( \alpha \).

**Remark.** If we apply the closed testing procedure (Marcus et al., 1976) to the weighted Simes test, we obtain a multiple test procedure for \( H_A \) and \( H_B \), which controls the familywise error rate in the strong sense, see also Hommel (1988) for the unweighted case. In the case of two hypotheses we thus reject \( H_i, i \in \{A, B\} \), if the weighted Simes test (1) rejects the intersection hypothesis \( H_0 \) and \( p_i \leq \alpha \). By construction, this closed testing procedure is uniformly more powerful than the weighted Bonferroni-Holm test which rejects \( H_i \) if \( p_i \leq \alpha \) and either (a) or (b) applies. Note further that the closed testing procedure for two hypotheses based on the weighted Simes test is consonant, i.e., the rejection of \( H_0 \) automatically implies the rejection of at least one individual \( H_i, i \in \{A, B\} \). With more than two hypotheses the Simes test can fail to be consonant already in the unweighted case.

## 4 Trimmed Simes’ test for two one-sided hypotheses

Samuel-Cahn (1996) has shown that without a suitable positive dependency assumption on the p-values (or test statistics) the Simes test may not control the type I error rate. Hence, if a negative correlation between the test statistics cannot be excluded, the Simes test may be unacceptable, in particular in the regulated environment of confirmatory clinical trials. This is, for example, the case when testing one-sided hypotheses for multiple endpoints.

### 4.1 A tempting but erroneous approach

Motivated by the example from Section 2 we consider a clinical trial in which two treatments are compared for two primary endpoints with one-sided hypotheses \( H_i : \theta_i \leq 0, i \in \{A, B\} \), where the parameter \( \theta_i \) is the treatment difference for endpoint \( i \in \{A, B\} \). Assume a bivariate normal test statistic \((X_A, X_B)\) with \( X_i \sim N(\theta_i, 1) \) and \( \text{Cor}(X_A, X_B) = \rho \). For simplicity, we consider the unweighted Simes test with \( \alpha_A = \alpha_B = \alpha/2 \) and assume further that \( \theta_i < 0 \) can be excluded for good reasons, so that \( H_A \cap H_B \) consists of the single parameter configuration \( \theta_A = \theta_B = 0 \).

Often, two-sided tests at level \( 2\alpha \) provide one-sided tests at level \( \alpha \) when restricting the rejection region to cases that are consistent with the one-sided alternative. Since the two-sided Simes test controls the type I error rate also for \( \rho < 0 \) it appears tempting to use the two-sided Simes test at level \( 2\alpha \) for testing \( H_0 : \theta_A = \theta_B = 0, \) and in case that \( H_0 \) is rejected - the one-sided \( \alpha \)-level tests with rejection rule \( p_i \leq \alpha \) for the individual \( H_i \), where \( p_i \) is the one-sided p-value for the one-sided hypothesis \( H_i : \theta_i \leq 0 \). This implies that we reject \( H_A \cap H_B \) if the two-sided Simes test rejects, and \( \min(p_A, p_B) \leq \alpha \). Figure 1 illustrates the resulting rejection region for the intersection hypothesis \( H_A \cap H_B \) in terms of the test statistics \((X_A, X_B)\). The region is the complement of the acceptance region \( A \cup B \), i.e., the rejection region is the union of \( C, H_1, H_2, D_1, D_2, E, F_1, F_2, I_1, I_2, G \). This region is larger than the halved rejection region \( H_2 \cup D_2 \cup E \cup F_1 \cup I_1 \) for the two-sided Simes test (with acceptance region \( A \)). If the p-values \( p_i \) are independent (i.e., \( \rho = 0 \)) and uniformly distributed then the probability over the halved rejection region is \( \alpha \). Therefore, rejecting \( H_A \cap H_B \) if the two-sided Simes test rejects at level \( 2\alpha \) and \( \min(p_A, p_B) \leq \alpha \) gives the rejection probability \( \alpha + 2P(X_A \geq z_{\alpha}, X_A + X_B < 0) > \alpha \).
Figure 1. Illustration of the different rejection regions for $\alpha_A = \alpha_B = \alpha/2$ and standard normal test statistics $(X_A, X_B)$. The critical values corresponding to the levels $\alpha$ and $\alpha_A = \alpha_B$ are the $(1 - \alpha)$ and $(1 - \alpha/2)$ quantiles of the standard normal distribution, denoted as $z_\alpha$ and $z_{\alpha/2}$, respectively. Region A is the acceptance region of the two-sided Simes test at level $2\alpha$. Rejecting $H_A \cap H_B$ if this two-sided Simes test rejects and $\max(X_A, X_B) \geq z_\alpha$ gives the acceptance region $A \cup B$, i.e., the rejection region $C \cup H_1 \cup H_2 \cup D_2 \cup E \cup F_1 \cup F_2 \cup I_1 \cup I_2 \cup G$. The rejection region of the one-sided Simes test is $D_1 \cup D_2 \cup E \cup F_1 \cup F_2$. The trimmed Simes test has rejection region E.

For instance, when $\alpha = 0.025$ and $\alpha_A = \alpha_B = 0.0125$ (and $\rho = 0$) this probability is 0.0256. This shows that the combination of the two-sided Simes test at level $2\alpha$ with the one-sided tests at level $\alpha$ can exceed the nominal level $\alpha$ and is therefore not an appropriate method. The maximum type I error rate inflation of this procedure can in fact be substantially larger than that for the one-sided Simes test with $\rho \leq 0$ (Samuel-Cahn, 1996).

4.2 Trimmed weighted Simes’ test

We next show that a minor modification of the weighted Simes test for two one-sided hypotheses $H_i : \theta_i \leq 0$ for $i \in \{A, B\}$, but consider now $\theta_i < 0$ as possible. We further assume that the distribution of the bivariate test statistic $(X_A, X_B)$ is point symmetric for $\theta_1 = \theta_2 = 0$, which means that $(X_A, X_B)$ and $(-X_A, -X_B)$ have the same bivariate distribution. We further assume that the distribution function $F_i(x; \theta_i)$ of each $X_i$ is non-increasing in $\theta_i$ at all $x$. Note that the point symmetry of $(X_A, X_B)$ for $\theta_1 = \theta_2 = 0$ implies that the marginal distribution functions $F_i(x; 0)$, $i \in \{A, B\}$, are symmetric, i.e., the associated density or probability functions $f_i(x; 0)$ of $F_i(x; 0)$ satisfy $f_i(-x; 0) = f_i(x; 0)$ for all $x$. 

Copyright line will be provided by the publisher
For $\alpha_A, \alpha_B \geq 0$ with $\alpha_A + \alpha_B = \alpha$ we consider a trimmed version of the weighted Simes test with one-sided p-values $p_i = F(-X_i), i \in \{A, B\}$, which rejects $H_A \cap H_B$ if

$$\begin{align*}
\text{(a)} & \quad p_A \leq \alpha_A \quad \text{and} \quad p_B < 1 - \alpha_B \\
\text{or (b)} & \quad p_B \leq \alpha_B \quad \text{and} \quad p_A < 1 - \alpha_A \\
\text{or (c)} & \quad \max(p_A, p_B) \leq \alpha .
\end{align*}$$

In Figure 1, where $F_i(x; 0) = \Phi(x)$ and $\alpha_A = \alpha_B = 0.0125$, the rejection region of the trimmed weighted Simes test is region E. Note that the condition $p_i < 1 - \alpha_i$ in (a) and (b) of equation (3) “trims” the rejection regions of the weighted Simes test (1). We also note that the results in the remainder of this paper remain valid if stronger trimming requirements than $p_i < 1 - \alpha_i, i \in \{A, B\}$, are applied. For example, in clinical trials it might be necessary to demonstrate a positive trend for one variable (i.e. a p-value less than 0.5) if significance for the other variable has been achieved.

The additional condition $p_i < 1 - \alpha_i$ is natural in the example from Section 2 of a clinical study involving the comparison of two treatments for two endpoints $A$ and $B$. Using the trimmed Simes test we would not be able to claim for an overall positive effect, if we observe a significant positive effect for one endpoint and a significant negative effect for the other one. Note that a similar trimming of the Bonferroni test would lead to a rejection of $H_A \cap H_B$ with only (a) and (b) from equation (3). Hence, the trimmed Simes test uniformly improves the trimmed Bonferroni test in the same way as the untrimmed Simes test improves the untrimmed Bonferroni test. Note further that the trimming has no influence on the probability to reject both hypotheses because this event (namely $\{\max(p_A, p_B) \leq \alpha\}$) remains part of the rejection region.

We see from Figure 1 that in the unweighted case $\alpha_A = \alpha_B = \alpha/2$ the rejection region E for the trimmed Simes test is a subset of a half of the original rejection region for the two-sided Simes test. Hence, the type I error rate of the trimmed Simes test is smaller than $\alpha$. According to the following theorem, which is verified in the Appendix, this is valid also for an arbitrary $\alpha$-splitting.

**Theorem 4.1** Assume a bivariate test statistics $(X_A, X_B)$ where $(X_A, X_B)$ and $(-X_A, -X_B)$ have the same bivariate distribution under $\theta_1 = \theta_2 = 0$, and which has marginal distribution functions $F_i(x; \theta_i), i \in \{A, B\}$, that are non-increasing in $\theta_i$. Let further $\alpha \leq 0.5$. If the type I error rate of the weighted Simes test at level $2\alpha$ with two-sided p-values $q_i = 2F_i(-|X_i|; 0)$ is bounded by $2\alpha$ then the type I error rate of the trimmed weighted Simes test (3) with one-sided p-values $p_i = F_i(-X_i; 0)$ is bounded by $\alpha$ for all $\theta_A, \theta_B \leq 0$.

**Remarks.** (i) We obtain a multiple testing procedure for the two one-sided hypotheses $H_A$ and $H_B$ with strong familywise error rate control, when applying a closed testing procedure with the trimmed weighted Simes test. With this procedure we reject $H_i, i \in \{A, B\}$, if the trimmed weighted Simes test (3) rejects the intersection $H_A \cap H_B$ and $p_i \leq \alpha$. By construction, this closed testing procedure is uniformly more powerful than the trimmed weighted Bonferroni-Holm test which rejects $H_i$ if $p_i \leq \alpha$ and either (a) or (b) in (3) applies. Both closed testing procedures are consonant, i.e., the rejection of $H_A \cap H_B$ always implies rejection of at least one $H_i, i \in \{A, B\}$.

(ii) In the following we give an example which shows that the conditions of Theorem 4.1 are generally insufficient for type I error rate control of the untrimmed one-sided Simes test. From this example we conclude that the untrimmed one-sided Simes test can inflate the maximal type I error up to $1.5\alpha$ under the distributional conditions of Theorem 4.1 while the trimmed Simes test controls the nominal level $\alpha$. The example is with a discrete bivariate test statistic $(X_A, X_B)$ which for $\theta_1 = \theta_2 = 0$ takes values $(-z_{\alpha/2}, z_{\alpha/2}), (z_{\alpha/2}, -z_{\alpha/2}), (z_{\alpha/2}, z_{\alpha/2})$ and $(-z_{\alpha/2}, -z_{\alpha/2})$ each with probability $\alpha/2$, and the value $(0, 0)$ with probability $1 - 2\alpha$. For $\theta_i \neq 0$ we shift the discrete values of $X_i$ by $\theta_i$.

One can verify that the trimmed and untrimmed one-sided Simes test would have the same rejection regions as in the normal case (see Figure 1), and that the untrimmed one-sided Simes test (which has rejection region $D_1 \cup D_2 \cup E \cup F_1 \cup F_2$ in Figure 1) has type I error rate $1.5\alpha$ whereas the type I error rate of the trimmed Simes test is only $\alpha/2$. For the verification one has to take into account that, e.g.,
Figure 2. Type I error rate error rate of the Bonferroni, weighted Simes and trimmed weighted Simes test for \( \alpha = 0.025 \). Panel (a) is for the unweighted case \( \alpha_A = \alpha_B = 0.0125 \), panel (b) for \( \alpha_A = 0.02 \) and \( \alpha_B = 0.005 \).

\((-z_{\alpha/2}, z_{\alpha/2})\) belongs to the rejection region of the untrimmed one-sided Simes test but not to the region of the trimmed Simes test because of the strict and non-strict inequalities in (3).

One can also verify that the conditions of Theorem 4.1 are satisfied in this example. Other exam-

ples demonstrate that whenever any of the two distributional conditions in Theorem 4.1 are violated,

the type I error rate of the trimmed Simes test can also be up to \( 1.5\alpha \).

4.3 Simulation study

We conducted a simulation study to investigate the operating characteristics of the trimmed and untrimmed weighted Simes test in comparison to the untrimmed weighted Bonferroni test for one-sided hypotheses at nominal level \( \alpha = 0.025 \). We assume bivariate normal test statistics with varying correlation \( \rho \). We used \( 10^7 \) simulation runs for the type I error rate and \( 10^6 \) runs for the power.

Figure 2 gives the simulated type I error rate for the intersection hypothesis \( H_0 \) when using the trimmed and untrimmed weighted Simes test as well as the untrimmed weighted Bonferroni test. Panel (a) displays the results for \( \alpha_A = \alpha_B = 0.0125 \) and panel (b) for \( \alpha_A = 0.02 \) and \( \alpha_B = 0.005 \). Note that the results inherently display also the familywise error rate for the closed testing procedure based on either the trimmed weighted Simes test or the weighted Bonferroni test. As seen from Figure 2, the error rates for the trimmed Simes test are almost indistinguishable from those of the untrimmed Simes test for positive correlations (the maximum difference is smaller than \( 0.00033 \)), and they are substantially smaller for negative correlations. Note also that the type I error rate inflation of the untrimmed Simes test is almost invisible and undetectable with our simulation standard error of about \( 5 \cdot 10^{-5} \) per scenario. From Table 1 in Samuel-Cahn (1996), we expect the type I error rate inflation to be maximal for correlations between \(-0.25 \) and \( 0 \). As expected, the mean type I error rate from all simulation runs with correlation \(-0.25 \leq \rho < 0 \) (for 25 distinct values of \( \rho \) in steps of 0.01, each with \( 10^7 \) runs) indicate a small type I error inflation of the untrimmed Simes test. With equal weights (Figure 2a) these means were 0.02498 for the Bonferroni, 0.02503 for the Simes, and 0.02404 for the trimmed Simes test. For unequal weights (Figure 2b), the means were 0.02497, 0.02502, and 0.02454 for the Bonferroni, Simes and trimmed Simes test, respectively. With a type I error rate of about 0.025 in all 25 scenarios of \( \rho \) with \(-0.25 \leq \rho < 0 \) combined, the simulation standard error is about \( 10^{-5} \).
Figure 3. The probability to reject $H_A \cap H_B$ with the weighted Bonferroni and trimmed weighted Simes test at level $\alpha = 0.025$ for $\theta_B$ between 0 and $\theta_A$ and for different correlations $\rho$. Panel (a) is for the unweighted case $\alpha_A = \alpha_B = 0.0125$, panel (b) for $\alpha_A = 0.02$ and $\alpha_B = 0.005$. $\theta_A$ is fixed such that the probability to reject $H_A \cap H_B$ with the weighted Bonferroni Test is 0.8 for $\theta_A = \theta_B$. If the test statistics have variance 1, this power is achieved at $\theta_A = \theta_B = 2.37$ in the unweighted case (a) and at $\theta_A = \theta_B = 2.42$ in the weighted case (b).

Figure 3 displays the rejection probability for $H_0$ when using the trimmed weighted Simes test as well as the untrimmed weighted Bonferroni test in dependence of the non-centrally parameter $\theta_B \in [0, \theta_A]$. The value of $\theta_A$ is fixed in such a way that the rejection probability equals 80% for $\theta_A = \theta_B$ when $\rho = 0$. **Note that the displayed rejection probability of $H_0$ is also the probability** to reject at least one of the individual hypotheses $H_i$ with the closed testing procedure based on either the trimmed weighted Simes test or the weighted Bonferroni test. We consider two different correlations, $\rho = 0, 0.5$. Again, panel (a) displays the results for $\alpha_A = \alpha_B = 0.0125$, and panel (b) for $\alpha_A = 0.02$ and $\alpha_B = 0.005$. As seen from Figure 3, the trimmed weighted Simes test improves the untrimmed weighted Bonferroni test in most cases, being inferior only for $\rho = 0$ and small values of $\theta_B$. **The trimmed weighted Simes test is nevertheless uniformly more powerful than the trimmed Bonferroni test, which de facto would have to be applied also in the situation with two primary variables, because statistically significant effects into opposite directions could hardly be considered a success.** Similar simulations comparing, e.g., the trimmed and untrimmed version of the one sided Simes test show that for values of $\theta_A$ and $\theta_B$ as in Figure 3 the power difference was always smaller than 0.01 with the maximum power difference occurring at $\theta_B = 0$ and equal level split $\alpha_A = \alpha_B = \alpha/2$. For cases of interest, namely where there is moderate to large correlation between the endpoints and similar true positive effect sizes, the power gain of the trimmed weighted Simes test over the untrimmed weighted Bonferroni test is about the same as for untrimmed weighted Simes. For example for $\alpha_A = \alpha_B = 0.0125$ and $\theta_B = \theta_A = 2.37$ this gain in power is for both, the trimmed and untrimmed Simes test, 1.1% when $\rho = 0, 1.4\%$ when $\rho = 0.5$, and 2.0% when $\rho = 0.8$. 

Copyright line will be provided by the publisher
5 Trimmed and untrimmed weighted Simes’ test with t-test statistics

The above results show, for instance, that with bivariate normally distributed test statistics a minor modification to the weighted Simes test, namely, the trimmed version (3), provides type I error rate control for the intersection of two one-sided hypotheses, $H_i : \theta_i \leq 0$ for $i \in \{A, B\}$, and a valid multiple test via the closed testing principle. These assumptions apply, for instance, to the example in Section 2 and to the z-test for a comparison of two groups with regard to (approximately) normally distributed endpoints $A, B$, if the variance is known or the sample size is sufficient large. With a small or moderate sample size (and an unknown variance) the z-test can become inappropriate and the application of the t-test would be preferred over the z-test. According to Theorem 4.1, the trimmed weighted Simes test could be applied to the one-sided p-values of the t-tests if the usual weighted Simes test applied to the two-sided p-values of the t-tests would provide type I error rate control.

Unfortunately, for multiple endpoints no results on the control of the usual weighted or unweighted Simes’ test with p-values from t-tests are available yet. Accordingly, we discuss in this section the applicability of Theorem 3.1 to important examples of t-tests for multiple endpoints. We will first derive a result for a specific class of t-tests which, for instance, can then be applied to one- and two-sample t-tests for the means of two normal distributed endpoints.

To discuss the applicability of Theorem 3.1 to t-tests, let $(X_A, X_B)$ be a vector of bivariate normal distributed test statistics with $X_i \sim N(\theta_i, \sigma_i)$ and unknown standard errors $\sigma_i$. Let $(S_A, S_B)$ be a vector of estimates for $(\sigma_A, \sigma_B)$. We assume that the vector $(S_A, S_B)$ is independent from the vector $(X_A, X_B)$ and $\nu_i S_i^2 / \sigma_i \sim \chi_i^2$, where $\chi_i^2$ is the $\chi^2$-distribution with $\nu_i$ degrees of freedom. Let $F_i$ denote the cumulative t-distribution function of $X_i / S_i$.

We now consider the weighted Simes test applied to the p-values $p_i = F_i(-X_i / S_i), i \in \{A, B\}$, for the one-sided hypotheses $H_i : \theta_i \leq 0$ if $\rho = \text{Cor}(X_A, X_B) \geq 0$, and to the p-values $p_i = 2F_i(-X_i / S_i)$ of the two-sided null hypotheses $H_i : \theta_i = 0$ for arbitrary correlation $\rho$. Sarkar (2008) verified control of the type I error rate in the unweighted case for $k \geq 2$ hypotheses assuming a common standard error $\sigma_A^2 = \sigma_B^2 = \sigma^2$ and a pooled standard error estimate $S$ instead of separate estimates $S_A, S_B$. Restricting attention to the case of two hypotheses, we extend this result to the weighted Simes test and to cases with $\sigma_A \neq \sigma_B$ and non-identical $S_i$. The following result is verified in the Appendix.

**Theorem 5.1** Let $(X_A, X_B)$ be bivariate normal with unknown variances $\sigma_A, \sigma_B$ and correlation $\rho$. Let further $(S_A, S_B)$ be independent from $(X_A, X_B)$ such that $\nu_i S_i^2 / \sigma_i \sim \chi_i^2$, for $i \in \{A, B\}$, and

$$P(S_B > c \mid S_A = s) \text{ is non-decreasing in } s \text{ for all } c, s > 0.$$  

(4)

Then $P(|X_B|/S_B > c \mid |X_A|/S_A = t)$ and $P(|X_A|/S_A > c \mid |X_B|/S_B = t)$ are non-decreasing in $t > 0$ for all $c$ and arbitrary $\rho$. Furthermore, $P(X_B/S_B > c \mid X_A/S_A = t)$ and $P(X_A/S_A > c \mid X_B/S_B = t)$ are non-decreasing in $t > 0$ for all $c$ and $\rho \geq 0$.

Since $p_i = 2F_i(-X_i / S_i)$ is a decreasing function in $|X_i| / S_i$, the monotonicity of $P(|X_i|/S_i > c \mid |X_j|/S_j = t)$ in $t > 0$ for $i, j \in \{A, B\}$ and $i \neq j$ implies property (2) for the two-sided p-values $p_A, p_B$ and $\alpha \leq 1$. Similarly, the monotonicity of $P(X_i/S_i > c \mid X_j/S_j = t)$ in $t > 0$ for $i \neq j$, implies property (2) for the one-sided p-values for $\alpha \leq 0.5$. The restriction to $\alpha \leq 0.5$ is needed, because the stochastic monotonicity of $X_i/S_i$ in $t = X_j/S_j$ is verified only for positive $t$. The following corollary follows directly from Theorems 3.1 and 5.1.

**Corollary 5.2** Let $(X_A, X_B)$ be bivariate normal with correlation $\rho$ and $X_i \sim N(\theta_i, \sigma_i), i \in \{A, B\}$. Let further $(S_A, S_B)$ be independent from $(X_A, X_B)$, satisfy (4) and $\nu_i S_i^2 / \sigma_i \sim \chi_i^2$. Then the type I error rate of the weighted Simes test (1) with two-sided p-values $p_i = 2F_i(-|X_i| / S_i)$ is bounded by $\alpha$. Furthermore, if in addition $\alpha \leq 0.5$ and $\rho \geq 0$, then the type I error rate of the weighted Simes test (1) with one-sided p-values $p_i = F_i(-|X_i| / S_i)$ is bounded by $\alpha$ for $\theta_A, \theta_B \leq 0$. 

Copyright line will be provided by the publisher
Remark. (i) Type I error rate control of weighted Simes’ test with one-sided p-values \( p_i = F_i(-X_i/S_i) \), \( i \in \{A, B\} \), follows for \( \rho \geq 0 \) and for all \( \theta_A, \theta_B \leq 0 \) with \( \min(\theta_A, \theta_B) < 0 \) because Theorems 3.1 and 5.1 imply that the weighted Simes test with one-sided p-values \( p_{\theta,i} = F_i(-(X_i - \theta_i)/S_i) \) has type I error rate \( \leq \alpha \), and rejection with \( p_i, i \in \{A, B\} \), implies rejection with \( p_{\theta,i}, i \in \{A, B\} \), because \( p_{\theta,i} \leq p_i \) almost surely. (ii) According to Theorem 4.1 (and by the same arguments as in the previous remark) we can apply for arbitrary correlation \( \rho \) the trimmed weighted Simes test (3) to the one-sided p-values \( p_i = F_i(-X_i/S_i) \) and hypotheses \( H_i : \theta_i \leq 0 \).

Examples. (i) Assumption (4) is obviously satisfied with common standard errors \( \sigma_A = \sigma_B = \sigma \) and pooled standard error estimate \( \hat{S} \) for \( S_A \) and \( S_B \). (ii) Assumption (4) holds also for the standard error estimates \( S_i = \{\sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2/(n^2 - n)\}^{1/2} \) of the means \( \bar{X}_{A,j}, \bar{X}_{B,j} \) of i.i.d. bivariate normal observations, \( i \in \{A, B\}, j = 1, \ldots, n \). This can be verified by showing the TP2 property of \( (S_A^2, S_B^2) \).

To this end we assume w.l.o.g. that \( \text{var}(X_{A,j}) = \text{var}(X_{B,j}) = 1 \). Then \( X_{B,j}/(1-\rho^2)^{1/2} = Z_j + X_{A,j} \rho/ (1-\rho^2)^{1/2} \) where \( Z_1, \ldots, Z_n \) are i.i.d. standard normal and independent from \( X_{A,1}, \ldots, X_{A,n} \). This implies \( (X_{B,j} - \bar{X}_j)/(1-\rho^2)^{1/2} = (Z_j - \bar{Z}) + (X_{A,j} - \bar{X}_A) \rho/ (1-\rho^2)^{1/2} \) for all \( j \) whereby \( \bar{Z} \) denotes the mean of \( Z_1, \ldots, Z_n \). Furthermore, there exists an \( n \times (n-1) \)-matrix \( J \) such that the random vector \( \{Z_1, \ldots, Z_{n-1}\} = \{Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z}\}J \) has independent components, and the data vector \( \{X_{i,1}, \ldots, X_{i,n-1}\} = \{X_{i,1} - \bar{X}_i, \ldots, X_{i,n} - \bar{X}_i\}J \) satisfies \( S_A^2(n^2 - n) = \sum_{i=1}^{n-1} X_{i,j}^2 \) for each \( i \in \{A, B\} \). Note further that \( X_{B,j}/(1-\rho^2)^{1/2} = Z_j + X_{A,j} \rho/ (1-\rho^2)^{1/2} \) for all \( j = 1, \ldots, n-1 \). Hence, the conditional distribution of \( S_B^2(n^2 - n)/(1-\rho^2) \) given \( X_{A,1}, \ldots, X_{A,n-1} \) is the non-central \( \chi^2 \)-distribution with \( n-1 \) degrees of freedom and non-centrality parameter \( m = S_A^2(n^2 - n)/(1-\rho^2) \). This implies the same non-central \( \chi^2 \)-distribution for the conditional distribution of \( S_B^2(n^2 - n)/(1-\rho^2) \) given \( S_A^2 \). Since the non-central \( \chi^2 \)-density \( \chi^{(n-1)} \) is TP2 in \( (x, m) \) (see e.g. Lehmann and Romano, 2005), the density of \( (S_A^2, S_B^2) \) is TP2 as well. Note further that the vector \( (S_A, S_B) \) is independent from the mean vector \( (\bar{X}_A, \bar{X}_B) \). Hence, the assumptions of Theorem 3.1 and 5.1 and thereby Corollary 5.2 are valid for \( \bar{X}_i \) and \( S_i, i \in \{A, B\} \).

(iii) We consider now the two-sample version of the previous example, namely, a clinical trial where two independent treatment groups \( t \) (experimental treatment) and \( c \) (control treatment) are compared with respect to two primary endpoints \( X_{A,g,j} \) and \( X_{B,g,j} \). Here \( j \) denotes the \( j \)-th patient in treatment group \( g, j = 1, \ldots, n_g, g \in \{t, c\} \). Assume that we are interested in testing either \( H_i : \theta_i = 0 \) or \( H_i : \theta_i \leq 0 \), \( i \in \{A, B\} \), where \( \theta_i \) is now the mean difference between the treatment groups \( t \) and \( c \) for endpoint \( i \). Assume further that \( X_{A,g,j} \) and \( X_{B,g,j} \) are bivariate normally distributed with homogeneous variances \( \text{var}(X_{i,t,j}) = \text{var}(X_{i,c,j}), i \in \{A, B\} \), and equal correlation \( \text{Cor}(X_{A,g,j}, X_{B,g,j}) = \rho \) in both groups \( g \). The variance can, however, differ between the endpoints, i.e. \( \text{var}(X_{A,g,j}) \neq \text{var}(X_{B,g,j}) \). In this case we use the t-test \( (\bar{X}_{i,t} - \bar{X}_{i,c})/S_i \) with \( S_i^2 = (n_c^{-1} + n_t^{-1}) \{\sum_{j=1}^{n} (X_{i,t,j} - \bar{X}_{i,t})^2 + \sum_{j=1}^{n} (X_{i,c,j} - \bar{X}_{i,c})^2\}/(n_t + n_c - 2) \). By similar arguments as in the previous example one can show that the conditional distribution of \( S_B^2(n_t + n_c - 2)/(n_c^{-1} + n_t^{-1})(1-\rho^2) \) given \( S_A^2 \) is the non-central \( \chi^2 \)-distribution with \( n_t + n_c - 2 \) degrees of freedom and non-centrality parameter \( S_A^2(n_t + n_c - 2) \rho^2/((n_c^{-1} + n_t^{-1})(1-\rho^2)) \). Hence, the distribution of \( (S_A^2, S_B^2) \) is TP2 which implies (4). Furthermore, the vector \( (S_A, S_B) \) is independent from the mean difference vector \( (\bar{X}_{A,t} - \bar{X}_{A,c}, \bar{X}_{B,t} - \bar{X}_{B,c}) \). Hence, the assumptions of Corollary 5.2 are also fulfilled for \( X_i = \bar{X}_{i,t} - \bar{X}_{i,c} \) and \( S_i, i \in \{A, B\} \).

6 Discussion

We have introduced a minor modification of Simes’ test for two one-sided hypotheses that provides type I error rate control whenever the two-sided Simes test controls the level and the bivariate test statistic has a point symmetric distribution. These properties were shown to apply to normally distributed test statistics as well as to test statistics from one or two-sample t-tests for bivariate normally distributed endpoints. Our modification, which we called “trimmed Simes’ test”, rejects, irrespective
of the correlation between the endpoints, when the usual Simes test rejects but retains the intersection hypothesis if at least one of the p-values is above one minus the Bonferroni adjusted level, i.e., the test statistic points in the wrong direction in a highly significant manner. Such modification is particularly natural in our motivating example of a clinical trial with two primary endpoints where rejection for at least one primary endpoint is sufficient to verify overall efficacy. In such a trial, an outcome which strongly indicates inferiority of the treatment with regard to at least one endpoint would provide difficulties for treatment approval.

We did a small simulation study to investigate the operation characteristics of the trimmed weighted Simes test in comparison to the untrimmed Simes and Bonferroni test for normally distributed test statistics. As expected from the investigations in Samuel-Cahn (1996), the theoretically proven type I error rate inflation of the untrimmed Simes’ test is nearly invisible even with our large number of simulation runs \((10^7)\). Hence, one could conclude like Samuel-Cahn (1996) that, for practical purposes, the type I error rate is under control already with the untrimmed Simes test. However, according to our experience, doubts, e.g. by regulatory agencies, often remain, presumably, because the numerical investigations of Samuel and Cahn (1996) are restricted to normally distributed test statistics. Because of the pronounced conservatism of the trimmed weighted Simes test for negative (and for almost all strictly positive) correlations (see Figure 1) we expect that the trimmed Simes test is rather robust with regard to type I error rate control. For an illustration we have considered the one- and two-sample t-test and have verified type error rate control of the corresponding trimmed (and untrimmed two-sided) Simes test with and without weights.

Our power simulation study also shows that the trimmed and untrimmed Simes test have similar power for the relevant alternatives and that the power gain of the trimmed and untrimmed Simes test over the Bonferroni test is visible but small. The major advantage of the trimmed and untrimmed Simes is that they provide a uniform improvement (the trimmed Simes over the trimmed Bonferroni test), and that the corresponding closed test rejects both null hypotheses whenever both p-values are below the nominal level \(\alpha\). As mentioned before, we find the latter property particularly appealing for trials like the motivating example with two primary endpoints, where the treatment is expected to be efficient in both endpoints and the corresponding finding would be of high value. In this situation the trimming has also only an negligible influence on the power of the test.

The investigations of this paper has been restricted to two endpoints for several reasons. First of all, the closed test procedure with Simes’ test is consonant only with two hypotheses, and consonance is a reasonable property. Secondly, general results on the type I error rate control of the two-sided (and untrimmed one-sided) Simes test with unequal weights is currently limited to the case of two hypotheses, and our main result on the type I error rate of the trimmed Simes test relies on type I error rate control of the two-sided Simes test. Finally, the trimming of the Simes test with more than two hypotheses seems not as straightforward as for the case of two-hypotheses and is expected to lead to more complex and less natural procedures. We nevertheless believe that a higher dimensional version of the trimmed Simes test is an interesting topic for future research.

Acknowledgements  We thank Björn Holzhauer for providing us the case study example and for many stimulating discussions. We are also grateful to Peter Bauer, the referees and the editor for their helpful comments.

Appendix

Proof of Theorem 3.1

We assume w.l.o.g. that \(\alpha_A \leq \alpha_B \leq \alpha\) in which case rejection rule (1) of the weighted Simes test can be written as \(R = \{\max(p_A, p_B) \leq \alpha\} \cup \{\min(p_A, p_B) \leq \alpha_A\} \cup \{p_A > \alpha, \alpha_A < p_B \leq \alpha_B\}\). Note that the region \(R^*\) obtained when interchanging \(p_A\) and \(p_B\) in \(R\) satisfies \(P(R^*) \leq P(R)\) if and only if

\[
P(p_B > \alpha, \alpha_A < p_A \leq \alpha_B) \leq P(p_A > \alpha, \alpha_A < p_B \leq \alpha_B)
\]  

(5)
Hence, it suffices to show type I error rate control under assumption (5) because, invalidity of (5) implies \( P(R) < P(R') \) and would allow us to prove \( P(R') \leq \alpha \) by just the same arguments as below.

So, we assume (5). Note that the probability of the rejection region \( R \) can be written as

\[
P(p_B \leq \alpha) - P(p_A > \alpha, \alpha_B < p_B \leq \alpha) + P(p_B > \alpha, p_A \leq \alpha_A).
\]

We will show that

\[
P(p_B > \alpha, p_A \leq \alpha_A) \leq P(p_A > \alpha, \alpha_B < p_B \leq \alpha)
\]

which implies that the rejection probability (6) of the weighted Simes test is bounded by \( P(p_B \leq \alpha) = \alpha \).

Inequality (7) can be deduced as follows. Note that \( f(a) \leq \int_a^b f(u) du /(b - a) \leq f(b) \) for all \( a < b \) and non-decreasing integrable functions \( f(u) \). Since \( f(u) = P(p_A > \alpha | p_B = u) \) is non-decreasing by (2), we get \( P(p_A > \alpha | \alpha_B < p_B \leq \alpha) \leq P(p_A > \alpha | p_B = \alpha_B) \leq P(p_A > \alpha | \alpha_B < p_B \leq \alpha) \).

We see by similar arguments that \( P(p_B > \alpha | p_A \leq \alpha_A) \leq P(p_B > \alpha | \alpha_A < p_B \leq \alpha_B) \). The latter two inequalities together with (5) imply \( P(p_B > \alpha | p_A \leq \alpha_A) \leq P(p_B > \alpha | \alpha_B < p_B \leq \alpha) \). Since \( P(p_A \leq \alpha_A) = \alpha_A = \alpha - \alpha_B = P(\alpha_B < p_B \leq \alpha) \) we obtain (7).

Proof of Theorem 4.1

Since for the trimmed weighted Simes test the type I error rate is maximal for \( \theta_A = \theta_B = 0 \), it is sufficient to consider only this case. Let \( c_i = F_i^{-1}(1 - \alpha_i; 0) \) and \( d_i = F_i^{-1}(1 - \alpha_i; 0) \), \( i \in \{A, B\} \), where \( F_i^{-1}(u; 0) \) is a symmetric version of the quantile function of \( F_i(x; 0) \), i.e., \( F_i^{-1}(1 - u; 0) = -F_i^{-1}(u; 0) \) for all \( u \leq 0.5 \). Note that \( c_i > 0 \) and \( d_i \geq 0 \) because \( \alpha_i \leq \alpha \leq 0.5 \). Note further that \( d_A, d_B > 0 \) for \( \alpha < 0.5 \) and \( d_A = d_B = 0 \) for \( \alpha = 0.5 \). Let \( w = d_A/d_B \) if \( \alpha < 0.5 \) and \( w = 1 \) if \( d_A = d_B = 0 \). Then the rejection region of the trimmed weighted Simes test with one-sided p-values \( p_i \) can be written as

\[
R^+ = \{X_A \geq c_A, X_B > -c_B\} \cup \{X_B \geq c_B, X_A > -c_A\} \cup \{\min(X_A, wX_B) \geq d_A\}
\]

We also consider the “negative” counterpart \( R^- \) of \( R^+ \) where \( X_A \) and \( X_B \) are replaced by \( -X_A \) and \( -X_B \) in (8). Note that \( R^+ \) and \( R^- \) are disjoint and have equal probabilities \( P(R^+) = P(R^-) \) due to the assumption that \( (X_A, X_B) \) and \( (-X_A, -X_B) \) have the same null distribution. We furthermore have that

\[
R^+ \cup R^- \subseteq \{\min(|X_A|, w|X_B|) \geq d_A\} \cup \{|X_A| \geq c_A\} \cup \{|X_B| \geq c_B\}
\]

since

\[
\{\min(X_A, wX_B) \geq d_A\} \cup \{\min(-X_A, -wX_B) \geq d_A\} \subseteq \{\min(|X_A|, w|X_B|) \geq d_A\}
\]

and

\[
\{X_A \geq c_A, X_B > -c_B\} \cup \{-X_A \geq c_A, -X_B > -c_B\} \subseteq \{|X_A| \geq c_A\},
\]

and a similar inclusion holds when interchanging \( X_A \) and \( X_B \). Note that the event on the right side of (9) is the rejection region of the two-sided Simes test at level \( 2\alpha \) which is assumed to be bounded by \( 2\alpha \).

Hence, we get \( P(R^+) = P(R^+ \cup R^-)/2 \leq \alpha \).

Proof of Theorem 5.1

According to Lemma 4.8 in Barlow and Proschan (1975) there exists a function \( h(u, s) \) which is non-decreasing in \( s \) such that \( (S_A, S_B) \) has the same distribution as \( (S_A, h(U, S_A)) \) where \( U \) is a random variable which is independent from \( S_A, X_A \) and \( X_B \). Let either \( (Y_A, Y_B) = (|X_A|, |X_B|) \) or \( (Y_A, Y_B) = \)
We can assume, without loss of generality, that $\rho \geq 0$. Then $P(Y_B > c | Y_A = y)$ is non-decreasing in $y$ for all $c$, $y$. Hence, there exists a function $g(v, y)$ which is non-decreasing in $y$ such that $(Y_A, Y_B)$ has the same distribution as $(Y_A, g[V, Y_A])$ for a random variable $V$ which is independent from $(Y_A, S_A, U)$. Because $(Y_A/S_A, Y_B/S_B)$ has the same common distribution as $(Y_A/S_A, g[V, Y_A]/h(U, S_A))$, we can consider the latter pair of random variables. Note that $T = Y_A/S_A$, $R^2 = Y_A^2 + S_A^2$, $U$ and $V$ are jointly independent. Furthermore, since $Y_A = R \cdot T / (1 + T^2)^{1/2}$ and $S_A = R / (1 + T^2)^{1/2}$, we get that

$$P \left( \frac{Y_B}{S_B} > c \mid \frac{Y_A}{S_A} = t \right) = P_1 \left[ \frac{g[V, R \cdot t / (1 + t^2)^{1/2}]}{h[U, R / (1 + t^2)^{1/2}]} > c \right]$$

(10)

where $P_1$ is the probability with respect to $U$, $V$ and $R$. Finally, the monotonicity of $h(u, s)$ and $g(v, y)$ in $s$ and $y$, respectively, the fact that $t / (1 + t^2)^{1/2}$ is non-decreasing and $1 / (1 + t^2)^{1/2}$ is non-increasing in $t$ for all $t \geq 0$, implies that the right side of (10) is non-decreasing in $t$ for all $c, t > 0$.

Since (4) implies that also $P(S_A > c \mid S_B = s)$ is non-decreasing in $s$ for all $c, s > 0$, we can show by the same arguments as above that also $P \left( \frac{Y_A}{S_A} > c \mid \frac{Y_B}{S_B} = t \right)$ is non-decreasing in $t$.

References


Copyright line will be provided by the publisher