The Annals of Statistics
2002, Vol. 30, No. 1, 220–238

MULTIPLE HYPOTHESES TESTING AND EXPECTED NUMBER OF TYPE I ERRORS

BY H. FINNER AND M. ROTERS

Deutsches Diabetes-Forschungsinstitut Düsseldorf and Universität Potsdam

The performance of multiple test procedures with respect to error control is an old issue. Assuming that all hypotheses are true we investigate the behavior of the expected number of type I errors (ENE) as a characteristic of certain multiple tests controlling the familywise error rate (FWER) or the false discovery rate (FDR) at a prespecified level. We derive explicit formulas for the distribution of the number of false rejections as well as for the ENE for single-step, step-down and step-up procedures based on independent p-values. Moreover, we determine the corresponding asymptotic distributions of the number of false rejections as well as explicit formulae for the ENE if the number of hypotheses tends to infinity. In case of FWER-control we mostly obtain Poisson distributions and in one case a geometric distribution as limiting distributions; in case of FDR control we obtain limiting distributions which are apparently not named in the literature. Surprisingly, the ENE is bounded by a small number regardless of the number of hypotheses under consideration. Finally, it turns out that in case of dependent test statistics the ENE behaves completely differently compared to the case of independent test statistics.

1. Introduction. In this paper interest is focused on the expected number of type I errors of certain multiple test procedures $\varphi = (\varphi_i : i \in I)$ for a given family of hypotheses $\mathcal{H} = \{H_i : i \in I\}$, where $I \neq \varnothing$ denotes an arbitrary index set and φ_i denotes a nonrandomized test for H_i, $i \in I$. We always assume $\varnothing \neq H_i \subset \Theta$ for all $i \in I$, where Θ denotes a parameter set referring to distributions P_ϑ, $\vartheta \in \Theta$. Many criteria of error control for multiple tests have been discussed in the literature. Discussions of basic concepts can be found, for example, in Miller (1981) and Hochberg and Tamhane (1987). Nowadays most authors prefer the (strong) familywise error rate (FWER) criterion; that is, the probability of any type I error is bounded by a prespecified level $\alpha \in (0, 1)$ irrespective of which hypotheses are true or false. We call a test with this property a multiple level α test. Formally, a multiple test $\varphi = (\varphi_i : i \in I)$ for $\mathcal{H}$ is said to control the multiple

Received December 2000; revised June 2001.

AMS 2000 subject classifications. Primary 62J15, 62F05; secondary 62F03, 60F99.

Key words and phrases. Asymptotic critical value behavior, Ballot theorem, Bolshev's recursion, Bonferroni test procedure, Dempster's formula, DKW inequality, empirical distribution function, familywise error rate, false discovery rate, independent p-values, Lagrange–Bürmann theorem, multiple comparisons, multiple level, multiple test procedure, order statistics, Schur–Jabotinski theorem, step-down test, step-up test.

level α, if

$$\forall\, \vartheta \in \Theta: \quad P_\vartheta\left(\bigcup_{i \in I(\vartheta)} \{\varphi_i = 1\}\right) \leq \alpha,$$

where $I(\vartheta) = \{i \in I : \vartheta \in H_i\}$ denotes the index set of true hypotheses given $\vartheta \in \Theta$.

Although Spjøtvoll (1972) developed some theory about the expected number of type I errors this approach keeps a more or less shadowy existence in theory and practice. In the introduction of his paper Spjøtvoll (1972) argued that the expected number of false rejections is technically easier to work with and more instructive than thinking in terms of the probability of at least one false rejection. However, this paper will show that it can be a hard job to calculate the expected number of false rejections of multiple test procedures.

There are various possibilities for defining a criterion referring to expected type I errors. For a finite index set I with $|I| = n \in \mathbb{N}$ (say) and fixed $\vartheta \in \Theta$ we define $V_n = |\{1 \leq i \leq n : \varphi_i = 1 \text{ and } \vartheta \in H_i\}|$. Then the *expected number of type I errors* (ENE for short) is defined by $\mathrm{E}_\vartheta V_n$ while the *expected* (*type I*) *error rate* (EER for short) is defined by $\mathrm{E}_\vartheta V_n/n$. In the literature the ENE is called per family error rate (PFE) and the EER is called per comparison error rate (PCE), respectively, [cf., e.g., Hochberg and Tamhane (1987)]. It is well known that if the ENE of a multiple test φ is bounded by some $\alpha \in (0, 1)$ for all $\vartheta \in \Theta$, then φ is a multiple level α test, too.

Recently, Benjamini and Hochberg (1995) rediscovered a nearly forgotten idea of Eklund [cf. Eklund and Seeger (1965) or Seeger (1966)] to combine V_n and the number of all rejected hypotheses to an error rate control criterion, that is, control of the false discovery rate (FDR) at some level $\alpha \in (0, 1)$. Letting R_n denote the number of false hypotheses being rejected, the FDR is defined to be the expectation of a random quantity Q_n, where $Q_n = 0$ in case of $V_n + R_n = 0$ and $Q_n = V_n/(V_n + R_n)$ otherwise. Since FDR-control at level α is more liberal than controlling a multiple level α it has led to a considerable number of papers dealing with this topic.

In this paper we are not interested in the ENE (or EER) as error control criteria but as characteristics describing the performance of multiple level-α tests and FDR-controlling procedures, especially when all hypotheses are assumed to be true provided that $\bigcap_{i \in I} H_i \neq \varnothing$. For example, a large ENE of a multiple level-α test may be a hint for good power performance. The distribution of V_n may be of interest, too. We know that a multiple level-α test φ satisfies $P_\vartheta(V_n > 0) \leq \alpha$, but what about $P_\vartheta(V_n > i)$ for $i = 1, \ldots, n$? Moreover, we can ask for the limiting distribution of V_n as well as the limit of the ENE (or EER) for certain types of tests if the number n of hypotheses tends to infinity. The aim of this paper is to answer some of these questions in a theoretical manner. A comparison of the main results of this paper for multiple tests based on order statistics of independent

p-values with the few results available for multiple tests based on order statistics of dependent test statistics yields some (sometimes surprising) general insight into the behavior of multiple comparison procedures at all. Another possibility to study the performance of multiple tests is to compare the critical values of various procedures [cf., e.g., Finner and Roters (1998, 1999, 2000)].

In Section 2 we first introduce some notation and summarize some results on the joint distribution of order statistics which will be used subsequently. Among others, we reformulate and discuss a result on the probability that the empirical distribution function falls below a straight line. Section 3 is concerned with the (limiting) distribution of V_n and its expectation for single-step and stepwise multiple test procedures based on independent p-values including a stepwise test based on Simes' test proposed by Hommel (1988). Depending on the test procedure we obtain a Poisson distribution as the limiting distribution of V_n or in one case a geometric distribution. Astonishingly, it turns out that $\mathrm{E}_\vartheta V_n$ tends to a finite limit approximately equal to α for small values of α if n tends to infinity. We also discuss the question whether there exists a universal upper bound for $\limsup_{n\to\infty} \mathrm{E}_\vartheta V_n$, which may be of interest in itself. Note that all considerations are carried out under the global null hypothesis $\bigcap_{i\in I} H_i$. The case where some of the hypotheses are assumed to be false is discussed at the end of Section 3. In Section 4 we study the ENE of the FDR-controlling procedure, proposed by Benjamini and Hochberg (1995) and earlier by Eklund, and its step-down counterpart. Here the distribution of V_n can be identified with a certain boundary crossing distribution appearing in connection with the one-sided Kolmogorov–Smirnov test. The resulting limiting distributions are apparently not named in the literature. Finally, in Section 5 we briefly discuss the behavior of the ENE (EER) in case of dependent test statistics and relate the results for the independence and dependence cases. Several technical proofs are deferred to the Appendix.

2. Notation and some useful tools. Let $U_{1:n} \le \cdots \le U_{n:n}$ denote the order statistics of n independent and uniformly distributed random variables $U_1, \dots, U_n$ on $[0, 1]$. We study various multiple tests based on independent p-values under the assumption that all hypotheses are true. Additionally, we assume that in this situation the p-values are distributed as $U_1, \dots, U_n$. Therefore, we define all tests under consideration in terms of $U_1, \dots, U_n$. Special attention is focused on so-called single-step (SS) tests, step-down (SD) tests and step-up (SU) tests. Let $\gamma_{i,n}$, $i = 1, \dots, n$, denote a sequence of critical values with $1 \ge \gamma_{1,n} \ge \cdots \ge \gamma_{n,n} \ge 0$, $n \in \mathbb{N}$. A SS-procedure based on a critical value $\gamma_{n,n}$ will be denoted by SS $(\gamma_{n,n})$ and rejects a hypothesis H_i if $U_i \le \gamma_{n,n}$. SD- and SU-procedures based on critical values $\gamma_{1,n} \ge \cdots \ge \gamma_{n,n}$ will be denoted by SD $(\gamma_{i,n},\ i = 1, \dots, n)$ and SU $(\gamma_{i,n},\ i = 1, \dots, n)$. If $H_{1:n}, \dots, H_{n:n}$ denote the ordered hypotheses with respect to the order of the p-values, a SD $(\gamma_{i,n},\ i = 1, \dots, n)$-procedure rejects $H_{i:n}$ if and only if $U_{j:n} \le \gamma_{n-j+1,n}$ for all $j = 1, \dots, i$, while a SU $(\gamma_{i,n},\ i = 1, \dots, n)$-procedure rejects $H_{i:n}$ if and only if $U_{j:n} \le \gamma_{n-j+1,n}$ for some $j \in$

$\{i, \ldots, n\}$. An appropriate choice of the critical values leads to a multiple level-α test procedure. Note that a SU-procedure based on the same set of critical values as its SD-counterpart rejects at least all hypotheses rejected by SD, possibly more. In other words, the random variable V_n for the SU-procedure is pointwise larger than the corresponding V_n for SD. The exact critical values for the step-down procedure will be denoted by $\bar{\alpha}_k$ and are given by $\bar{\alpha}_k = 1 - (1-\alpha)^{1/k}$, $k = 1, \ldots, n$. The exact critical values for the step-up procedure, denoted by $\alpha_1, \ldots, \alpha_n$, can be determined in terms of the joint distribution function F_k (say) of the order statistics $U_{1:k} \leq \cdots \leq U_{k:k}$ of $U_1, \ldots, U_k$ by successively solving the system of equations $F_k(1-\alpha_1, \ldots, 1-\alpha_k) = 1-\alpha,\ k = 1, \ldots, n$. Dalal and Mallows (1992) showed that the resulting α_k's are decreasing in k, which is important for the step-up algorithm. An explicit formula for the α_k's is given by [cf. Rom (1990)]

$$\alpha_k = \frac{1}{k}\left(\sum_{i=1}^{k-1} \alpha^i - \sum_{i=1}^{k-2} \binom{k}{i} \alpha_{i+1}^{k-i}\right), \qquad k = 2, \ldots, n, \tag{2.1}$$

with $\alpha_1 = \alpha$. Another valid choice for SD- as well as SU-procedures is $\gamma_{i,n} = \alpha/i$, $i = 1, \ldots, n$. In the SD-case this procedure is known as the Bonferroni–Holm test, in the SU-case as Hochberg's SU-test [Hochberg (1988)]. A further interesting multiple level-α test procedure based on an idea of Simes (1986) proposed by Hommel (1988) is the closed Simes test denoted by Simes $((m-i+1)\alpha/m$, $1 \leq i \leq m \leq n)$, which is more complicated than SD and SU. The Simes $((m-i+1)\alpha/m,\ 1 \leq i \leq m \leq n)$-procedure rejects $H_{1:n}, \ldots, H_{r:n}$, where r is determined as follows. Let $J = \{1 \leq i \leq n : U_{n-i+k:n} > k\alpha/i \text{ for all } k = 1, \ldots, i\}$. If $J \neq \varnothing$, set $j' = \sup J$, otherwise $j' = 1$. Then $r = \sup\{1 \leq i \leq n : U_{i:n} \leq \alpha/j'\}$, where $\sup \varnothing = 0$ (say). We note that the closed Simes test rejects at least all hypotheses rejected by Hochberg's SU-test, possibly more.

Under the assumption that $1 \geq \gamma_{1,n} \geq \cdots \geq \gamma_{n,n} \geq 0$, $n \in \mathbb{N}$, a general recursive formula for the joint cdf F_n^k of $U_{1:n}, \ldots, U_{n-k:n}$, $0 \leq k \leq n-1$, is given by

$$\begin{aligned} &F_n^k(1-\gamma_{1,n}, \ldots, 1-\gamma_{n-k,n}) \\ &\qquad = 1 - \sum_{j=0}^{n-k-1} \binom{n}{j} F_j(1-\gamma_{1,n}, \ldots, 1-\gamma_{j,n}) \gamma_{j+1,n}^{n-j} \end{aligned} \tag{2.2}$$

with $F_n^0 = F_n$ and $F_0^0 \equiv F_n^n \equiv 1$. This is essentially Bolshev's recursion, which is proved in different ways in Shorack and Wellner [(1986), pages 366–367] and in Finner, Hayter and Roters (1993). A useful formula for the evaluation of $F_n(1-\gamma_{1,n}, \ldots, 1-\gamma_{n,n})$, which is in connection to the classical Ballot theorem [cf. Karlin and Taylor (1981), pages 107–137] and which is used later to obtain explicit formulas for the ENE of some specific multiple test procedures, is given in the special case when $\gamma_{1,n}, \ldots, \gamma_{n,n}$ decrease linearly. More specifically, we have:

LEMMA 2.1. *Let* $n \in \mathbb{N}$, $a, \gamma \in \mathbb{R}$ *such that* $0 \le a + \gamma \le a + n\gamma \le 1$. *Then*

$$F_n(a+\gamma, \dots, a+n\gamma) = (a+\gamma)\big(a+(n+1)\gamma\big)^{n-1}.$$

PROOF. We only sketch the proof of this lemma since the assertion of it is well known [cf. Rényi (1973), page 293, and Shorack and Wellner (1986), page 344, formula (4)], although it has apparently never been stated in this utmost generality. Rényi (1973) considers the case $a + n\gamma = 1$ and Shorack and Wellner (1986) need the restriction $a + (n+1)\gamma < 1$ in their proof which is essentially due to Dempster (1959).

Similarly, but a little more generally as in Rényi (1973), we derive a recursive formula for $P_n(a, \gamma) = F_n(a+\gamma, \dots, a+n\gamma)$, $n \in \mathbb{N}$, by conditioning on the values of the smallest order statistic $U_{1:n}$. An easy calculation then yields

$$P_n(a,\gamma) = \int_0^{a+\gamma} P_{n-1}\left(\frac{a+\gamma-y}{1-y}, \frac{\gamma}{1-y}\right) n(1-y)^{n-1}\,dy, \qquad n \ge 2,$$

which immediately proves the lemma by induction on $n \in \mathbb{N}$. □

3. Results for independent *p*-values. We start this section with results for SS-, SD- and SU-procedures all keeping a multiple level $\alpha \in (0, 1)$. All probability and expectation computations are understood to be carried out under the global null hypothesis of the underlying multiple testing problem.

For a SS $(\gamma_{n,n})$-procedure it is obvious that

$$P(V_n = i) = \binom{n}{i}\gamma_{n,n}^i(1-\gamma_{n,n})^{n-i} \qquad \text{for } i = 0, \dots, n, \tag{3.1}$$

$$\mathrm{E}V_n = n\gamma_{n,n}, \tag{3.2}$$

that is, V_n follows a binomial distribution with parameters n and $\gamma_{n,n}$. For stepwise procedures the situation is somewhat more complicated. Therefore, we present two lemmas with formulas for the distribution and the expectation of V_n for SD- and SU-procedures.

LEMMA 3.1. *For a SD* $(\gamma_{i,n}, i = 1, \dots, n)$*-procedure it holds that*

$$P(V_n \ge i) = F_n^{n-i}(\gamma_{n,n}, \dots, \gamma_{n-i+1,n}), \tag{3.3}$$

$$P(V_n = i) = \binom{n}{i} F_i(\gamma_{n,n}, \dots, \gamma_{n-i+1,n})(1-\gamma_{n-i,n})^{n-i} \qquad \textit{for } i = 0, \dots, n, \tag{3.4}$$

$$\mathrm{E}V_n = \sum_{i=1}^{n} F_n^{n-i}(\gamma_{n,n}, \dots, \gamma_{n-i+1,n}). \tag{3.5}$$

In case of $\gamma_{i,n} = \bar{\alpha}_i$ *the expression* $(1-\gamma_{n-i+1,n})^{n-i+1}$ *on the right-hand side of* (3.4) *simplifies to* $1-\alpha$ *for* $i = 1, \dots, n$.

PROOF. Equation (3.3) is obvious. Equation (3.4) follows from (2.2) with (3.3), while (3.5) follows by applying (3.3). □

Similarly, we obtain the following result for the SU-procedure.

LEMMA 3.2. *For a SU* $(\gamma_{i,n}, i = 1, \dots, n)$*-procedure it holds that*

$$P(V_n \le i) = F_n^i(1-\gamma_{1,n}, \dots, 1-\gamma_{n-i,n}), \tag{3.6}$$

$$P(V_n = i) = \binom{n}{i} F_{n-i}(1-\gamma_{1,n}, \dots, 1-\gamma_{n-i,n})\gamma_{n-i+1,n}^i \qquad \textit{for } i = 0, \dots, n, \tag{3.7}$$

$$\mathrm{E}V_n = \sum_{i=0}^{n-1}\bigl(1 - F_n^i(1-\gamma_{1,n}, \dots, 1-\gamma_{n-i,n})\bigr). \tag{3.8}$$

In case of $\gamma_{i,n} = \alpha_i$ the expression $F_{n-i+1}(1-\gamma_{1,n}, \dots, 1-\gamma_{n-i+1,n})$ on the right-hand side of (3.7) simplifies to $1-\alpha$ for $i = 1, \dots, n$.

The following theorem provides us with the limiting distribution of V_n as well as with the limit of $\mathrm{E}V_n$ for n tending to infinity.

THEOREM 3.3. *Let* $c = -\log(1-\alpha)$. *For the* SS $(\bar{\alpha}_n)$-, *the* SD $(\bar{\alpha}_i, i = 1, \dots, n)$- *and the* SU $(\alpha_i, i = 1, \dots, n)$*-procedures it holds that*

$$\lim_{n\to\infty} P(V_n = i) = \exp(-c)c^i/i! \qquad \textit{for } i \in \mathbb{N}_0, \tag{3.9}$$

$$\lim_{n\to\infty} \mathrm{E}V_n = c. \tag{3.10}$$

For the SS (α/n)-, *the* SD $(\alpha/i, i = 1, \dots, n)$- *and the* SU $(\alpha/i,\ i = 1, \dots, n)$*-procedures we obtain*

$$\lim_{n\to\infty} P(V_n = i) = \exp(-\alpha)\alpha^i/i! \qquad \textit{for } i \in \mathbb{N}_0, \tag{3.11}$$

$$\lim_{n\to\infty} \mathrm{E}V_n = \alpha. \tag{3.12}$$

PROOF. For all $\gamma_{i,n} \in \{\bar{\alpha}_n, \bar{\alpha}_i, \alpha_i\}$, $i \in \{1, \dots, n\}$, $n \in \mathbb{N}$, we have $\lim_{n\to\infty} n \times \gamma_{n-i,n} = c$ for all $i \in \mathbb{N}_0$, and for all $\gamma_{i,n} \in \{\alpha/n, \alpha/i\}$, $i \in \{1, \dots, n\}$, $n \in \mathbb{N}$, we have $\lim_{n\to\infty} n\gamma_{n-i,n} = \alpha$ for all $i \in \mathbb{N}_0$. From this fact and (3.1) as well as (3.2) we immediately get the result for the SS $(\bar{\alpha}_n)$-procedure. A similar argument applies for the SS (α/n)-procedure.

For the SD-procedures under consideration, observe that in (3.4) the term $F_i(\gamma_{n,n}, \dots, \gamma_{n-i+1,n})$ can be bounded as follows: for all $i \in \{1, \dots, n\}$, $n \in \mathbb{N}$, we have $\gamma_{n,n}^i \le F_i(\gamma_{n,n}, \dots, \gamma_{n-i+1,n}) \le \gamma_{n-i+1,n}^i$; hence, for example, Fatou's lemma implies that

$$\liminf_{n\to\infty} \mathrm{E}V_n \ge \sum_{i=1}^{\infty} i \lim_{n\to\infty} \binom{n}{i} \gamma_{n,n}^i \lim_{n\to\infty} (1-\gamma_{n-i,n})^{n-i} = c \quad \text{or} \quad = \alpha,$$

according to whether the SD $(\bar{\alpha}_i,\ i = 1, \dots, n)$- or the SD $(\alpha/i,\ i = 1, \dots, n)$-procedure is considered. For this, the expectation of the Poisson distribution has to be kept in mind. To obtain an upper bound for $\limsup_{n\to\infty} \mathrm{E}V_n$, observe that for all $n - i \in \mathbb{N}$ the term $(1-\gamma_{n-i,n})^{n-i} = \exp(-c)$ or $\le \exp(-\alpha)$, according to whether the SD $(\bar{\alpha}_i, i = 1, \dots, n)$- or the SD $(\alpha/i, i = 1, \dots, n)$-procedure is considered. Moreover, in both cases we have $\lim_{n\to\infty} nP(V_n = n) = 0$, so that for computing $\limsup_{n\to\infty} \mathrm{E}V_n$ we may neglect this last term in the sum of the formula for $\mathrm{E}V_n$. But finally, the remaining term to be coped with is $\lim_{n\to\infty} \sum_{i=1}^{n-1} \binom{n}{i} i \gamma_{n-i+1,n}^i$, which can easily be evaluated by using Lemma A.2(i) in Finner and Roters (1999) with $k = 0$, $r = 1$ and $m = 1$. This yields the desired results (3.10) and (3.12).

Consider now the SU-procedures. From Finner and Roters (1994) we obtain that $\lim_{n\to\infty} F_{n-i}(1-\gamma_{1,n}, \dots, 1-\gamma_{n-i,n}) = \exp(-c)$ or $= \exp(-\alpha)$ for all $i \in \mathbb{N}_0$, according to whether the SU $(\alpha_i,\ i = 1, \dots, n)$- or the SU $(\alpha/i,\ i = 1, \dots, n)$-procedure is considered. So, as before, Fatou's lemma yields $\liminf_{n\to\infty} \mathrm{E}V_n \ge c$ or $\ge \alpha$, accordingly. Since in (3.7) for all $n - i \in \mathbb{N}$ the term $F_{n-i}(1-\gamma_{1,n}, \dots, 1-\gamma_{n-i,n}) = \exp(-c)$ for the SU $(\alpha_i,\ i = 1, \dots, n)$-procedure and $F_{n-i}(1-\gamma_{1,n}, \dots, 1-\gamma_{n-i,n}) \le (1-\gamma_{n-i,n})^{n-i} \le \exp(-\alpha)$ for the SU $(\alpha/i,\ i = 1, \dots, n)$-procedure, we may repeat the argument used before to finally obtain the desired results (3.10) and (3.12) also in the SU-cases. □

In Theorem 3.3 it is obvious from (3.9) and (3.11) that the sequence of random variables $(V_n)_{n\in\mathbb{N}}$ converges in distribution to a Poisson distributed random variable V (say) with parameter $c = -\log(1-\alpha)$ (in case of exact critical values) or with parameter α (in case of Bonferroni-adjusted critical values). It may be somewhat surprising that in the limit there is no difference between an SS-procedure and its SD- and SU-counterparts. Moreover, in the limit there is no difference between SD and SU, although an SU-procedure rejects at least all hypotheses rejected by an SD-procedure based on the same set of critical values. However, we have to keep in mind that our results refer to the situation where all hypotheses are assumed to be true.

We conclude this section with limiting results for the closed Simes test. In view of the complexity of this test procedure no explicit formulas are given for $\mathrm{E}V_n$ and $P(V_n = i)$.

THEOREM 3.4. *For the* Simes $(\frac{m-i+1}{m}\alpha,\ 1 \le i \le m \le n)$*-procedure it holds that*

$$\lim_{n\to\infty} P(V_n = i) = \exp(-\alpha)\alpha^i/i! \qquad \text{for } i \in \mathbb{N}_0, \tag{3.13}$$

$$\lim_{n\to\infty} \mathrm{E}V_n = \alpha, \tag{3.14}$$

which coincides with the results for the SS (α/n)-, *the* SD $(\alpha/i, i = 1, \dots, n)$- *and the* SU $(\alpha/i, i = 1, \dots, n)$*-procedures*, *respectively*.

PROOF. Let $U_{1:n} \le \cdots \le U_{n:n}$ denote the ordered p-values for testing $H_1, \dots, H_n$ and denote by $H_{1:n}, \dots, H_{n:n}$ the ordered hypotheses according to the ordered p-values. The components of the closed Simes test for $H_{i:n}$ are denoted by $\varphi_{i:n}$ and the components of Hochberg's SU-test procedure are referred to as $\varphi_{i:n}^{\mathrm{HC}}$, $i = 1, \dots, n$. Moreover, let V_n and V_n^{HC} denote the ENEs of the closed Simes test and of Hochberg's SU-test procedure, respectively. Since $\varphi_{i:n} \ge \varphi_{i:n}^{\mathrm{HC}}$ we obtain $P(V_n \ge i) = P(\varphi_{i:n} = 1) \ge P(\varphi_{i:n}^{\mathrm{HC}} = 1) = P(V_n^{\mathrm{HC}} \ge i)$ for all $i = 1, \dots, n$, hence, for all $i \in \mathbb{N}_0$ and $n \ge i$,

$$\begin{aligned} \mathrm{E}V_n - \mathrm{E}V_n^{\mathrm{HC}} &= \sum_{j=1}^{n} \bigl(P(V_n \ge j) - P(V_n^{\mathrm{HC}} \ge j)\bigr) \\ &\ge P(V_n \ge i) - P(V_n^{\mathrm{HC}} \ge i) \ge 0. \end{aligned} \tag{3.15}$$

But now $\lim_{n\to\infty} \mathrm{E}V_n^{\mathrm{HC}} = \alpha$ [cf. (3.12) and (3.14)], which is proved below, entails that

$$\lim_{n\to\infty} \bigl(P(V_n \ge i) - P(V_n^{\mathrm{HC}} \ge i)\bigr) = 0 \qquad \text{for } i \in \mathbb{N}_0,$$

so that with (3.11) the assertion (3.13) follows.

To prove (3.14) we first note that $\liminf_{n\to\infty} \mathrm{E}V_n \ge \alpha$ due to (3.15) and (3.12). It remains to show that $\limsup_{n\to\infty} \mathrm{E}V_n \le \alpha$.

For arbitrary $\theta \in (0, 1-\alpha)$ let $(j_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} j_n/n = \theta$. The closed Simes test rejects a hypothesis $H_{i:n}$ at most, if the intersection hypothesis

$$H_{i:n} \cap H_{j_n+1:n} \cap \cdots \cap H_{n:n}$$

is rejected by the corresponding Simes test with critical values

$$\frac{1}{n-j_n+1}\alpha, \dots, \frac{n-j_n+1}{n-j_n+1}\alpha \qquad \text{for } 1 \le i \le j_n$$

and with critical values

$$\frac{1}{n-j_n}\alpha, \dots, \frac{n-j_n}{n-j_n}\alpha \qquad \text{for } j_n < i \le n,\ n \in \mathbb{N}.$$

This means that with

$$A_i^n = \left\{U_{i:n} \le \frac{\alpha}{n - j_n + 1}\right\} \quad \text{and} \quad B_r^n = \bigcup_{j=j_n+1}^{n} \left\{U_{j:n} \le \frac{j - j_n + r}{n - j_n + r}\alpha\right\}, \qquad r = 0, 1,$$

we get $\{\varphi_{i:n} = 1\} \subseteq A_i^n \cup B_1^n$ for all $1 \le i \le j_n$ and $\{\varphi_{i:n} = 1\} \subseteq B_0^n \subseteq B_1^n$ for all $j_n < i \le n$, which entails for all $n \in \mathbb{N}$,

$$\begin{aligned} \mathrm{E}V_n &= \sum_{i=1}^{n} P(\varphi_{i:n} = 1) \le \sum_{i=1}^{j_n} P(A_i^n \cup B_1^n) + \sum_{i=j_n+1}^{n} P(B_0^n) \\ &\le \sum_{i=1}^{n} P(A_i^n) + nP(B_1^n). \end{aligned}$$

In view of $\sum_{i=1}^n P(U_{i:n} \le \beta) = n\beta$, $\beta \in [0, 1]$, we obtain

$$\begin{aligned} \limsup_{n\to\infty} \mathrm{E}V_n &\le \limsup_{n\to\infty} \left(n\frac{\alpha}{n - j_n + 1} + nP(B_1^n)\right) \\ &\le \frac{\alpha}{1-\theta} + \limsup_{n\to\infty} nP(B_1^n). \end{aligned}$$

The next step is to show $\lim_{n\to\infty} nP(B_1^n) = 0$. Denoting by $G_n(t) = \sum_{i=1}^n I_{[0,t]}(U_i)/n$, $t \in [0, 1]$, $n \in \mathbb{N}$, the empirical distribution function with respect to $U_1, \ldots, U_n$ and noting that

$$\frac{j - j_n + 1}{n - j_n + 1}\alpha - \frac{j}{n} \le -\frac{\theta'}{1-\theta'}\alpha = -\varepsilon(\text{say}) \tag{3.16}$$

for all $0 < \theta' < \theta$ ($\in (0, 1-\alpha)$) and all $1 \le j \le n$, where n exceeds a sufficiently large number $n_0 \in \mathbb{N}$, we get for these n,

$$\begin{aligned} P(B_1^n) &\le P\left(\bigcup_{j=1}^{n} \left\{U_{j:n} - \frac{j}{n} \le \frac{j - j_n + 1}{n - j_n + 1}\alpha - \frac{j}{n}\right\}\right) \\ &\le P\left(\bigcup_{j=1}^{n} \left\{U_{j:n} - \frac{j}{n} \le -\varepsilon\right\}\right) \\ &\le P\left(\bigcup_{j=1}^{n} \left\{\left(|G_n(U_{j:n}) - U_{j:n}|\right) \ge \varepsilon\right\}\right) \\ &\le P\left(\sup_{t\in[0,1]} \left|\sqrt{n}\,(G_n(t) - t)\right| \ge \sqrt{n}\varepsilon\right) \\ &\le 2\exp(-2n\varepsilon^2) \end{aligned}$$

by virtue of an improved version of the DKW inequality for the empirical processes $(\sqrt{n}\,(G_n(t) - t))_{t\in[0,1]}$, $n \in \mathbb{N}$ [cf., e.g., Csörgő and Horváth (1993), page 119]. From this, it is now evident that $\lim_{n\to\infty} nP(B_1^n) = 0$, which implies $\limsup_{n\to\infty} \mathrm{E}V_n \le \alpha/(1-\theta)$. Letting $\theta \searrow 0$ completes the proof. $\square$

REMARK 3.5. All multiple level-α test procedures considered up to now satisfy $\limsup_{n\to\infty} \mathrm{E}V_n \le c = -\log(1-\alpha)$, which is approximately equal to α for small values of α. It may be surprising that, for almost all n, $\mathrm{E}V_n$ is bounded by a number slightly larger than α for common values of α. The question arises whether there exist multiple level-α test procedures with $\limsup_{n\to\infty} \mathrm{E}V_n > c$. If we allow randomized multiple level-α test procedures, the following procedure has $\mathrm{E}V_n = n\alpha$. Choose $\varphi_i \equiv \alpha$ for all $i = 1, \dots, n$ and reject H_i, $i = 1, \dots, n$, if the realization of a standard uniform variate is less than or equal to α. Then $P(V_n = n) = \alpha$ and $P(V_n = 0) = 1 - \alpha$, hence $\mathrm{E}V_n = n\alpha$. But what happens if we only allow nonrandomized tests? To this end we consider the following well-known procedure which we call the *particular order* procedure. Before the experiment starts, fix a particular order of the hypotheses with the purpose to test the hypotheses in this order all at level α until the first acceptance occurs. For this procedure we obviously get $P(V_n \ge i) = \alpha^i$ for $i = 0, \dots, n$, hence the limiting distribution of V_n is a geometric distribution with parameter α and

$$\mathrm{E}V_n = \sum_{i=1}^{n} P(V_n \ge i) = \alpha \frac{1-\alpha^n}{1-\alpha}, \qquad n \in \mathbb{N},$$

$$\lim_{n\to\infty} \mathrm{E}V_n = \frac{\alpha}{1-\alpha} > -\log(1-\alpha) \qquad \text{for all } \alpha \in (0,1).$$

A nonrandomized multiple level-α test with $\limsup_{n\to\infty} \mathrm{E}V_n > \alpha/(1-\alpha)$ is not known to the authors. So one may formulate the conjecture that all members of the class of nonrandomized multiple level-α tests based on independent p-values satisfy $\limsup_{n\to\infty} \mathrm{E}V_n \le \alpha/(1-\alpha)$ and that all tests in the class of all nonrandomized multiple level α tests being permutation invariant satisfy $\limsup_{n\to\infty} \mathrm{E}V_n \le -\log(1-\alpha)$. Here a multiple test $\varphi = (\varphi_i : i = 1, \dots, n)$ based on p-values $(p_1, \dots, p_n)$ is said to be permutation invariant if for all permutations π it holds that

$$\forall\, i = 1, \dots, n: \quad \varphi_i(p_1, \dots, p_n) = \varphi_{\pi(i)}(p_{\pi(1)}, \dots, p_{\pi(n)}).$$

Finally, the question arises whether a sequence of test procedures satisfying $\limsup_{n\to\infty} \mathrm{E}V_n \le c$ (such as the closed Simes test) can be improved upon.

REMARK 3.6. One may ask how the ENE of multiple level α tests behaves if some of the hypotheses are false. Clearly, for a single-step procedure the ENE will decrease if the number of false hypotheses increases. If, for instance, a proportion θ of hypotheses is false and the corresponding p-values are 0 (with probability 1), the ENE of the SS $(\bar{\alpha}_n)$-procedure tends to $(1-\theta)c = -(1-\theta)\log(1-\alpha)$ for n tending to infinity. The ENE-behavior of stepwise procedures is different. For example, in the case of the SD $(\bar{\alpha}_i, i = 1, \dots, n)$-procedure the ENE still tends to $c = -\log(1-\alpha)$ if a proportion θ of hypotheses is false and the corresponding p-values are 0 (with probability 1). However, for stepwise procedures it is generally not clear whether the ENE in case of some false hypotheses is smaller or larger than the ENE in the case that all hypotheses are true because $\mathrm{E}V_n$ is not necessarily increasing in the number of hypotheses n. However, it is evident that for the FWER-controlling procedures considered in this paper, the ENE for a family of n hypotheses, where k hypotheses are false and the remaining hypotheses are true, is less than or equal to the ENE of these procedures when the family of hypotheses is restricted to the $n-k$ true hypotheses. In many cases, $c = -\log(1-\alpha)$ seems to be an upper bound for $\limsup_{n\to\infty} \mathrm{E}_\vartheta V_n$ for all nonrandomized permutation invariant multiple level-α tests whether or not some of the hypotheses are false.

4. Results for two FDR-controlling procedures. In this section we consider the SD- and SU-procedures based on Simes' critical values, that is, the SD $((n-i+1)\alpha/n,\ i = 1, \dots, n)$- and the SU $((n-i+1)\alpha/n,\ i = 1, \dots, n)$-procedure, respectively. Although these procedures are no longer multiple level-α tests (except for $n \leq 2$) they both control the so-called false discovery rate (FDR) at level α [cf. Benjamini and Hochberg (1995) for the SU-case and Sarkar (2002) for the SD-case]. The SD-FDR-controlling procedure can be viewed as a conservative counterpart of the SU-FDR-controlling procedure. As mentioned in the introduction, letting R_n denote the number of false hypotheses being rejected, the FDR is defined to be the expectation of a random quantity Q_n, where $Q_n = 0$ in case of $V_n + R_n = 0$ and $Q_n = V_n/(V_n + R_n)$ otherwise. Since there is some controversy about the FDR-concept, it may be of interest to study the behavior of V_n in this case, too.

We note that Lemma 3.1 and Lemma 3.2 are still applicable for the SU-FDR- and SD-FDR-controlling procedures. But it turns out that we can derive some more attractive formulas. We start with two lemmas dealing with the distribution and the expectation of V_n for a larger class of step-down and step-up procedures including the FDR-controlling procedures mentioned before. Proofs of these results can be found in the Appendix.

It should be mentioned that the probability parts of these lemmas are essentially known as Dempster's (1959) formula for the probability that the graph of the empirical distribution function G_n of uniformly distributed random variables on $[0, 1]$ intersects a certain straight line at a certain height for the first time [cf. Shorack and Wellner (1986), pages 344–346]. Such *barrier crossing distributions* appear for instance as helpful tools for the derivation of the distribution of the one-sided Kolmogorov–Smirnov statistics.

LEMMA 4.1. *Let* $\beta, \tau \in [0,1]$, $n \in \mathbb{N}$, *such that* $\beta \geq (n-1)\tau$ *and let* $\gamma_{i,n} = \beta - (i-1)\tau$, $i = 1, \dots, n$. *Then we have for* $V_n^{\mathrm{SD}}(\beta,\tau) = \sup\{1 \leq i \leq n : U_{j:n} \leq \gamma_{n-j+1,n}$ *for all* $j = 1, \dots, i\}$ *and* $i = 1, \dots, n$,

$$P\big(V_n^{\mathrm{SD}}(\beta,\tau) = i\big) = \binom{n}{i}\big(\beta - (n-1)\tau\big)\big(\beta - (n-i-1)\tau\big)^{i-1} \times \big(1 - \beta + (n-i-1)\tau\big)^{n-i}, \tag{4.1}$$

$$P\big(V_n^{\mathrm{SD}}(\beta,\tau) = 0\big) = \big(1 - \beta + (n-1)\tau\big)^n, \tag{4.2}$$

$$\mathrm{E}V_n^{\mathrm{SD}}(\beta,\tau) = n\big(\beta - (n-1)\tau\big)\sum_{i=0}^{n-1}\binom{n-1}{i} i!\tau^i. \tag{4.3}$$

LEMMA 4.2. *Let* β, $\tau \in [0,1]$, $n \in \mathbb{N}$, *such that* $\beta \geq (n-1)\tau$ *and let* $\gamma_{i,n} = \beta - (i-1)\tau$, $i = 1, \dots, n$. *Then we have for* $V_n^{\mathrm{SU}}(\beta,\tau) = \sup\{1 \leq i \leq n : U_{i:n} \leq \gamma_{n-i+1,n}\}$ *and* $i = 0, \dots, n-1$,

$$P\big(V_n^{\mathrm{SU}}(\beta,\tau) = i\big) = \binom{n}{i}(1-\beta)\big(1 - \beta + (n-i)\tau\big)^{n-i-1} \times \big(\beta - (n-i)\tau\big)^{i}, \tag{4.4}$$

$$P\big(V_n^{\mathrm{SU}}(\beta,\tau) = n\big) = \beta^n, \tag{4.5}$$

$$\mathrm{E}V_n^{\mathrm{SU}}(\beta,\tau) = n\left(1 - (1-\beta)\sum_{i=0}^{n-1}\binom{n-1}{i} i!\tau^i\right) \tag{4.6}$$

$$= n\sum_{i=0}^{n-1}\big(\beta - (n-i-1)\tau\big)\binom{n-1}{i} i!\tau^i. \tag{4.7}$$

REMARK 4.3. The expectations in Lemmas 4.1 and 4.2 can also be calculated by

$$\mathrm{E}V_n^{\mathrm{SD}}(\beta,\tau) = \exp(1/\tau)\tau^{n-1} n\big(\beta - (n-1)\tau\big)\int_{1/\tau}^{\infty}\exp(-t)t^{n-1}\,dt,$$

$$\mathrm{E}V_n^{\mathrm{SU}}(\beta,\tau) = n\left(1 - (1-\beta)\exp(1/\tau)\tau^{n-1}\int_{1/\tau}^{\infty}\exp(-t)t^{n-1}\,dt\right).$$

Moreover, for technical reasons we derived recursive formulas for these expectations in the proofs of Lemmas 4.1 and 4.2 [cf. (A.1) and (A.2) in the Appendix].

We now return to the SD $((n-i+1)\alpha/n, i=1,\dots,n)$- and the SU $((n-i+1)\alpha/n, i=1,\dots,n)$-procedures, respectively. In the next two theorems we also have convergence in distribution of $(V_n)_{n\in\mathbb{N}}$, but with the difference that the limiting values $\lim_{n\to\infty} P(V_n=i)$, $i\in\mathbb{N}_0$, do not evidently sum up to 1 at first sight. However, from complex analysis one can utilize the so-called Lagrange–Bürmann or Schur–Jabotinski theorems [cf. Henrici (1974), pages 55–59] to evaluate the corresponding infinite series. As a matter of fact, the SD-case is explicitly treated as an example on pages 57 and 58 in Henrici's (1974) book.

To give a probabilistic argument for the convergence in distribution of $(V_n)_{n\in\mathbb{N}}$ in both cases we use the additional assumption that besides the possibly improper convergence of the distributions of $(V_n)_{n\in\mathbb{N}}$ [cf. Feller (1971), page 248, for the definition of improper convergence], which is given in the two theorems below by (4.10) and (4.14), the sequence $(\mathrm{E}V_n^\rho)_{n\in\mathbb{N}}$ is bounded for some $\rho>0$ [cf. Feller (1971), pages 251 and 252], which is fulfilled below for $\rho=1$ by (4.11) and (4.15).

THEOREM 4.4. *For the* SD $(\frac{n-i+1}{n}\alpha,\ i=1,\dots,n)$*-procedure it holds that*

$$P(V_n=i)=\binom{n}{i}(i+1)^{i-1}\left(\frac{\alpha}{n}\right)^i\left(1-\frac{i+1}{n}\alpha\right)^{n-i}\qquad \textit{for } i=0,\dots,n, \tag{4.8}$$

$$\mathrm{E}V_n=\alpha\sum_{i=0}^{n-1}\binom{n-1}{i}i!\left(\frac{\alpha}{n}\right)^i\qquad \textit{for } n\in\mathbb{N} \tag{4.9}$$

and

$$\lim_{n\to\infty}P(V_n=i)=\frac{(i+1)^{i-1}}{i!}\alpha^i\exp\bigl(-(i+1)\alpha\bigr)\qquad \textit{for } i\in\mathbb{N}_0, \tag{4.10}$$

$$\lim_{n\to\infty}\mathrm{E}V_n=\alpha\sum_{i=0}^{\infty}\alpha^i=\frac{\alpha}{1-\alpha}. \tag{4.11}$$

PROOF. Setting $\beta=\alpha$ and $\tau=\alpha/n$ in Lemma 4.1, we immediately get (4.8) and (4.9) from (4.1), (4.2) and (4.3), respectively. From (4.8), also (4.10) is evident. A look at (4.9) shows that all the terms of the sum appearing on the right-hand side of the formula for $\mathrm{E}V_n$ are nondecreasing in n, hence one easily obtains (4.11) by using monotone convergence. $\square$

REMARK 4.5. (i) As mentioned before Theorem 4.4, the right-hand side limiting values of (4.10) constitute a genuine probability distribution over $\mathbb{N}_0$. It should be mentioned that the complex analysis argument for this statement already occurs in Rényi's (1973) paper, page 293, in connection with the convergence in distribution of the random variables $\max_{1\le k\le n} n\,U_{k:n}/k$, $n\in\mathbb{N}$, but without elaboration.

We now state some ingredients of this argument in order to show that the expectation of the aforementioned limiting distribution equals the value on the right-hand side of (4.11).

It is easy to see that the radius of convergence of the power series $f(x) = \sum_{i=1}^{\infty}(i^{i-1}/i!)x^i$ is $r = 1/e$, where e denotes Euler's constant. Moreover, it can be shown by using the aforementioned Schur–Jabotinski theorem that $f(g(\alpha)) = \alpha$ for all $\alpha \in (0, 1)$, where $g(x) = x/\exp(x)$, $x \in \mathbb{R}$.

Let V^{SD} denote a random variable having the limiting distribution just mentioned and differentiate the equation $f(g(\alpha)) = \alpha$ with respect to $\alpha \in (0, 1)$. This yields $1 = f'(g(\alpha))g'(\alpha) = \exp(\alpha)\mathrm{E}(V^{\mathrm{SD}} + 1)g'(\alpha) = \mathrm{E}(V^{\mathrm{SD}} + 1)(1 - \alpha)$, hence $\mathrm{E}V^{\mathrm{SD}} = \alpha/(1 - \alpha)$.

THEOREM 4.6. *For the* SU $(\frac{n-i+1}{n}\alpha, i = 1, \dots, n)$*-procedure it holds that*

$$P(V_n = i) = \binom{n}{i}(1-\alpha)\left(\frac{i}{n}\alpha\right)^i\left(1 - \frac{i}{n}\alpha\right)^{n-i-1} \qquad \text{for } i = 0, \dots, n, \tag{4.12}$$

$$\mathrm{E}V_n = \alpha \sum_{i=0}^{n-1}\binom{n-1}{i}(i+1)!\left(\frac{\alpha}{n}\right)^i \qquad \text{for } n \in \mathbb{N}. \tag{4.13}$$

Finally,

$$\lim_{n\to\infty} P(V_n = i) = \frac{i^i}{i!}(1-\alpha)\alpha^i \exp(-i\alpha) \qquad \text{for } i \in \mathbb{N}_0, \tag{4.14}$$

$$\lim_{n\to\infty} \mathrm{E}V_n = \alpha\sum_{i=0}^{\infty}(i+1)\alpha^i = \alpha\left(\sum_{i=0}^{\infty}\alpha^i\right)^2 = \frac{\alpha}{(1-\alpha)^2}. \tag{4.15}$$

PROOF. Setting $\beta = \alpha$ and $\tau = \alpha/n$ in Lemma 4.2, we immediately get (4.12) and (4.13) from (4.4), (4.5) and (4.7), respectively. From (4.12), also (4.14) is evident. A look at (4.13) shows that, as in the SD-case, all the terms of the sum appearing on the right-hand side of the formula for $\mathrm{E}V_n$ are nondecreasing in n, so, similarly as before, we obtain (4.15) by using monotone convergence. □

REMARK 4.7. Also in the SU-case the sequence $(V_n)_{n\in\mathbb{N}}$ converges in distribution to a random variable V^{SU} with expectation $\alpha/(1-\alpha)^2$. As mentioned before Theorem 4.4, we have already proved the convergence result by appealing to (4.14) and (4.15). However, we want to give another proof without using (4.15). To this end we add up the values given in (4.14). Defining $h(x) = \sum_{i=1}^{\infty}(i^i/i!)x^i$, $x \in (-1/e, 1/e)$, we have to prove $h(g(\alpha)) = \alpha/(1-\alpha)$. But this is true since by Remark 4.5 we have

$$\alpha/(1-\alpha) = \alpha\mathrm{E}(V^{\mathrm{SD}} + 1) = h\big(g(\alpha)\big).$$

To prove the second statement observe that

$$\begin{aligned} \mathrm{E}V^{\mathrm{SU}} &= h'\big(g(\alpha)\big)g(\alpha)(1-\alpha) = h'\big(g(\alpha)\big)g'(\alpha)\alpha \\ &= (h \circ g)'(\alpha)\alpha = \alpha/(1-\alpha)^2. \end{aligned}$$

At first sight the FDR-criterion seems to behave nearly as conservatively with respect to the ENE as the FWER-criterion. Unfortunately, this is only true if all hypotheses are true. If a fixed proportion of hypotheses is assumed to be false, the ENE of the FDR-controlling procedures considered before may tend to infinity if the number of hypotheses tends to infinity. Detailed results on this issue can be found in Finner and Roters (2001b).

5. Dependent test statistics. As far as the authors know, not many (if any) theoretical results are known on the behavior of expected type I errors of multiple level-α test procedures and FDR-controlling procedures when the underlying test statistics are dependent. A brief discussion of this issue for single-step procedures based on exchangeable test statistics and range statistics can be found in a recent paper by Finner and Roters (2001a). A comparison of step-down and step-up procedures based on exchangeable test statistics is carried out in Finner and Roters (1998). Some general results on FDR-control for dependent test statistics can be found in Benjamini and Yekutieli (2001) and Sarkar (2002). Theoretical results concerning the behavior of the ENE for stepwise procedures based on dependent test statistics do not seem to exist in the literature. In view of the results of Section 3 one might presume that the ENE of stepwise multiple level-α test procedures based on dependent test statistics behaves rather similarly to their single-step counterparts. In any case, as reported in Finner and Roters (2001b), even the ENE of single-step procedures based on dependent test statistics seems to have a completely different behavior from the case of independent test statistics, that is, in one case of dependent test statistics the ENE often tends to infinity when the number of hypotheses tends to infinity. This is the case, for example, for Tukey's range test for all pairwise comparisons of population means as well as for Dunnett's test for multiple comparisons with a control when the sample means follow, for instance, a normal distribution. Moreover, depending on the underlying distributions, the rate of convergence may differ considerably. It is intuitively clear that a multiple level-α test procedure for n hypotheses may have an ENE near $n\alpha$ if the test statistics are almost totally dependent. Consider, for example, an exchangeable sequence of standard normally distributed random variables $(Y_n)_{n\in\mathbb{N}}$ with common correlation coefficient $\rho \in [0, 1)$ under the probability distribution P_ρ. Moreover, for all $n \in \mathbb{N}$ let c_n be defined by $P_\rho(Y_n \le c_n) = 1-\alpha$ for some $\alpha \in (0, 1)$ and let d_n be defined by $P_0(Y_1 \le d_n) = 1 - (1-\alpha)^{1/n}$. Then c_n (d_n) is the critical value of a single-step procedure for testing $H_i : \mu_i \le 0$ versus $K_i : \mu_i > 0$, $i = 1, \dots, n$, in a corresponding normal model with means $\mu_1 = \vartheta_1 - \vartheta_0, \dots, \mu_n = \vartheta_n - \vartheta_0$, which typically appears in multiple comparisons with a control. To calculate $\lim_{n\to\infty} \mathrm{E}_\rho V_n$

for $\rho \in (0,1)$, Finner and Roters (2001a) determined a sequence $(m(n))_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} m(n)P_\rho(Y_1 > c_n) = \lim_{n\to\infty} nP_0(Y_1 > d_n) = -\log(1-\alpha) = c$, which is given by $m(n) = n^{1-\rho+\gamma_n}$ with $\lim_{n\to\infty}\gamma_n = 0$ (of some complicated rate of convergence). This results in $\lim_{n\to\infty} \mathrm{E}_\rho V_n/n^{\rho-\gamma_n} = \lim_{n\to\infty} nP_\rho(Y_1 > c_n)/n^{\rho-\gamma_n} = \lim_{n\to\infty} m(n)P_0(Y_1 > c_n) = c$, which determines how fast $\mathrm{E}_\rho V_n$ tends to infinity for $n\to\infty$ in case of positive correlation. We refer the interested reader to the aforementioned paper by the authors for more detailed results concerning single-step procedures in models with underlying exchangeable test statistics or range statistics. The limiting behavior of $\mathrm{E}V_n$ for stepwise procedures in the case of dependent test statistics is not yet clear, but it may differ from the limiting behavior of single-step procedures. However, some of the formulas given in Section 3 for a fixed number of n hypotheses offer the possibility of deriving corresponding formulas in the case of exchangeable test statistics by conditioning on a suitable variable.

Often the FWER-criterion is rejected because of its low power performance. On the other hand, testing all hypotheses at level α increases the expected number of type I errors. If n hypotheses are all tested at level α and all hypotheses are true, the ENE can be $n\alpha$. This may be the reason why the FDR-criterion is currently in vogue because it seems to offer a compromise between the FWER-criterion and testing everything at level α. However, the FDR has many undesirable properties and difficulties, too [cf. Finner and Roters (2001b)]. However, as also illuminated in this paper, the control of a multiple level α may have different bearings on the expected number of false rejections in different models. Sometimes it is very conservative (in case of independent test statistics), sometimes it may be quite liberal (in case of dependent test statistics) with respect to the ENE, but a more balanced method does not seem to be in sight yet.

APPENDIX

PROOF OF LEMMA 4.1. From (3.4) in Lemma 3.1 and by choosing $a = \beta - n\tau$, $\gamma = \tau$ in Lemma 2.1 we get for $i = 1,\dots,n$,

$$\begin{aligned} P\big(V_n^{\mathrm{SD}}(\beta,\tau) = i\big) &= \binom{n}{i} F_i\big(\beta-(n-1)\tau,\dots,\beta-(n-i)\tau\big) \\ &\quad\times \big(1-\beta+(n-i-1)\tau\big)^{n-i} \\ &= \binom{n}{i}\big(\beta-(n-1)\tau\big)\big(\beta-(n-i-1)\tau\big)^{i-1} \\ &\quad\times \big(1-\beta+(n-i-1)\tau\big)^{n-i}, \end{aligned}$$

hence (4.1). Formula (4.2) follows in the same manner by keeping in mind that $F_0 \equiv 1$. In order to prove (4.3), we first show the validity of the recursive formula,

$$\mathrm{E}V_n^{\mathrm{SD}}(\beta,\tau) = n\big((\beta-(n-1)\tau) + \tau \mathrm{E}V_{n-1}^{\mathrm{SD}}(\beta-\tau,\tau)\big) \quad \text{for } n \geq 2,\ \beta \geq (n-1)\tau. \tag{A.1}$$

Since $\mathrm{E}V_n^{\mathrm{SD}}(\beta,\tau)=0$ for all $n \in \mathbb{N}$ whenever $\beta=(n-1)\tau$ [cf. (4.1) and (4.2)], it suffices to restrict attention to the case $\beta > (n-1)\tau$. In this case we have

$$\begin{aligned}
\mathrm{E}V_n^{\mathrm{SD}}(\beta,\tau) &= \sum_{i=0}^{n} i\binom{n}{i}\big(\beta-(n-1)\tau\big)\big(\beta-(n-i-1)\tau\big)^{i-1} \\
&\qquad \times \big(1-\beta+(n-i-1)\tau\big)^{n-i} \\
&= n\sum_{i=0}^{n}\left(\binom{n}{i}-\binom{n-1}{i}\right)\big(\beta-(n-1)\tau\big)\big(\beta-(n-i-1)\tau\big)^{i-1} \\
&\qquad \times \big(1-\beta+(n-i-1)\tau\big)^{n-i} \\
&= n\sum_{i=0}^{n}\big(P\big(V_n^{\mathrm{SD}}(\beta,\tau)=i\big) - \big(1-\beta+(n-i-1)\tau\big) \\
&\qquad \times P\big(V_{n-1}^{\mathrm{SD}}(\beta-\tau,\tau)=i\big)\big) \\
&= n\big((\beta-(n-1)\tau) + \tau \mathrm{E}V_{n-1}^{\mathrm{SD}}(\beta-\tau,\tau)\big).
\end{aligned}$$

Setting $a_{i,n} = \mathrm{E}V_i^{\mathrm{SD}}(\beta-(n-i)\tau,\tau)$, $b=\beta-(n-1)\tau$, $c=\tau$ we get the recursive formula $a_{i,n}=i(b+ca_{i-1,n})$, $i=1,\dots,n$, where $a_{0,n}=0$, $n\in\mathbb{N}$, by applying (A.1) and (4.1), (4.2) for $n=1$. Since the (unique) solution of this recursion is $a_{n,n}=nb\sum_{i=0}^{n-1}\binom{n-1}{i}i!c^i$ for all $n\in\mathbb{N}$, (4.3) is proved. □

PROOF OF LEMMA 4.2. From (3.7) in Lemma 3.2 and by choosing $a = 1-\beta-\tau$, $\gamma=\tau$ in Lemma 2.1 we immediately get (4.4). Formula (4.5) follows in the same manner by keeping in mind that $F_0\equiv 1$. In order to prove (4.6), we first note that from the formulas already derived we can conclude that for all $i=0,\dots,n$, $n\in\mathbb{N}$, and $\beta\geq(n-1)\tau$ it holds that

$$P\big(V_n^{\mathrm{SU}}(\beta,\tau)=i\big) = P\big(V_n^{\mathrm{SD}}\big(1-\beta+(n-1)\tau,\tau\big)=n-i\big).$$

Hence, for all $n\in\mathbb{N}$ we obtain

$$\begin{aligned}
\mathrm{E}V_n^{SU}(\beta,\tau) &= n - \mathrm{E}V_n^{\mathrm{SD}}\big(1-\beta+(n-1)\tau,\tau\big) \\
&= n\left(1-(1-\beta)\sum_{i=0}^{n-1}\binom{n-1}{i}i!\tau^i\right),
\end{aligned}$$

which is (4.6). Simple algebraic manipulations now yield the remaining assertion (4.7). □

REMARK. Similarly to the SD case, one can prove the recursive formula

$$(1-\beta+\tau)\mathrm{E}V_n^{\mathrm{SU}}(\beta,\tau)=n\big(\tau+(1-\beta)(\beta-n\tau) + (1-\beta)\tau\mathrm{E}V_{n-1}^{\mathrm{SU}}(\beta-\tau,\tau)\big) \tag{A.2}$$

for $n \geq 2$, $\beta \geq (n-1)\tau$.

Acknowledgment. We are grateful for some valuable comments by an Associate Editor and a referee.

REFERENCES

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300.

BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188.

CSÖRGŐ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics.* Wiley, Chichester.

DALAL, S. R. and MALLOWS, C. L. (1992). Buying with exact confidence. *Ann. Appl. Probab.* **2** 752–765.

DEMPSTER, A. P. (1959). Generalized D_n^+ statistics. *Ann. Math. Statist.* **30** 593–597.

EKLUND G. and SEEGER, P. (1965). Massignifikansanalys. *Statistisk Tidskrift Stockholm, 3rd Ser.* **4** 355–365.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

FINNER, H., HAYTER, A. J. and ROTERS, M. (1993). On the joint distribution function of order statistics with reference to step-up multiple test procedures. Technical report 93-19, 1-11, Univ. Trier.

FINNER, H. and ROTERS, M. (1994). On the limit behaviour of the joint distribution function of order statistics. *Ann. Inst. Statist. Math.* **46** 343–349.

FINNER, H. and ROTERS, M. (1998). Asymptotic comparison of step-down and step-up multiple test procedures based on exchangeable test statistics. *Ann. Statist.* **26** 505–524.

FINNER, H. and ROTERS, M. (1999). Asymptotic comparison of the critical values of step-down and step-up multiple comparison procedures. *J. Statist. Plann. Inference* **79** 11–30.

FINNER, H. and ROTERS, M. (2000). On the critical value behaviour of multiple decision procedures. *Scand. J. Statist.* **27** 563–573.

FINNER, H. and ROTERS, M. (2001a). Asymptotic sharpness of product-type inequalities for maxima of random variables with applications in multiple comparisons. *J. Statist. Plann. Inference* **98** 127–144.

FINNER, H. and ROTERS, M. (2001b). On the false discovery rate and expected type I errors. *Biom. J.* **43** 985–1005.

HENRICI, P. (1974). *Applied and Computational Complex Analysis* **1**. Wiley, New York.

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

HOCHBERG, Y. and TAMHANE, A. C. (1987). *Multiple Comparisons Procedures*. Wiley, New York.

HOMMEL, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75** 383–386.

KARLIN, S. and TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*. Academic Press, Boston.

MILLER, R. G., JR. (1981). *Simultaneous Statistical Inference*, 2nd ed. Springer, New York.

RÉNYI, A. (1973). On a group of problems in the theory of ordered samples. *Selected Translations in Math. Statist. Probab.* **13** 289–298.

ROM, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77** 663–665.

SARKAR, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.* **30** 240–258.

SEEGER, P. (1966). *Variance Analysis of Complete Designs. Some Practical Aspects*. Almqvist and Wiksell, Uppsala.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

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

SPJØTVOLL, E. (1972). On the optimality of some multiple comparison procedures. *Ann. Math. Statist.* **43** 398–411.

DEUTSCHES DIABETES-FORSCHUNGSINSTITUT UND
DER HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF
ABTEILUNG BIOMETRIE UND EPIDEMIOLOGIE
AUF'M HENNEKAMP 65
D-40225 DÜSSELDORF
GERMANY
E-MAIL: finner@ddfi.uni-duesseldorf.de

INSTITUT FÜR MATHEMATIK
UNIVERSITÄT POTSDAM
POSTFACH 601 553
D-14415 POTSDAM
GERMANY
E-MAIL: roters@math.uni-potsdam.de