ELSEVIER

Journal of Neuroscience Methods 170 (2008) 158–164

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Multiple test procedures using an upper bound of the number of true hypotheses and their use for evaluating high-dimensional EEG data

Claudia Hemmelmann[a,*], Andreas Ziegler[b], Volker Guiard[c], Sabine Weiss[d], Mario Walther[a], Rüdiger Vollandt[a]

[a] *Institute of Medical Statistics, Computer Sciences and Documentation, Friedrich Schiller University of Jena, Bachstraße 18, 07743 Jena, Germany*

[b] *Institute of Medical Biometry and Statistics, University at Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany*

[c] *Research Unit Genetics and Biometry, Research Institute for the Biology of Farm Animals Dummerstorf, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany*

[d] *Institute of Cognitive Science, Neurobiopsychology, University of Osnabrück, Albrechtstraße 28, 49046 Osnabrück, Germany*

Received 5 October 2007; received in revised form 10 December 2007; accepted 22 December 2007

Abstract

Frequency analyses of EEG data yield large data sets, which are high-dimensional and have to be evaluated statistically without a large number of false positive statements. There exist several methods to deal with this problem in multiple comparisons. Knowing the number of true hypotheses increases the power of some multiple test procedures, however the number of true hypotheses is unknown, in general, and must be estimated. In this paper, we derive two new multiple test procedures by using an upper bound for the number of true hypotheses. Our first procedure controls the generalized family-wise error rate, and thus is an improvement of the step-down procedure of Hommel and Hoffmann [Hommel G., Hoffmann T. Controlled uncertainty. In: Bauer P. Hommel G. Sonnemann E., editors. Multiple Hypotheses Testing, Heidelberg: Springer 1987;ISBN 3540505598:p. 154–61]. The second new procedure controls the false discovery proportion and improves upon the approach of Lehmann and Romano [Lehmann E.L., Romano J.P. Generalizations of the familywise error rate. Ann. Stat. 2005;33:1138–54]. By Monte-Carlo simulations, we show how the gain in power depends upon the accuracy of the estimate of the number of true hypotheses. The gain in power of our procedures is demonstrated in an example using EEG data on the processing of memorized lexical items.

Keywords: Coherence; EEG data; Multiple tests; Step-down procedure; Generalized family-wise error rate; False discovery proportion; Number of true hypotheses; Average power

1. Introduction

EEG analysis procedures yield large sets of high-dimensional parameters, which have to be evaluated statistically. Thus there are m components of the observational vector, which have to be tested simultaneously. If α-level tests for each single component or endpoint of the observational vector were used then a large number of false positive statements results. However, there exist several techniques to cope with this general drawback in multiple comparisons (see, e.g., Hemmelmann et al., 2004, 2005). Traditionally, multiple test procedures are designed to control the family-wise error rate (FWE). The FWE is the probability of committing at least one type I error, i.e., FWE = P(V > 0). Here, V denotes the random number of rejected true hypotheses, i.e., of type I errors. In problems with high-dimensional data, the control of FWE appears to be too strict. The control of generalized family-wise error rate P(V > u), abbreviated as gFWE(u), i.e., the requirement $P(V>u)\leq\alpha$ for some pre-specified integer u ($0\leq u<m$), is one of the recently proposed criteria for multiple test problems when the number m of hypotheses is large (see, e.g., van der Laan et al., 2004). A further important error rate is the so-called false discovery proportion, abbreviated as FDP(γ), for some pre-specified γ ($0\leq\gamma<1$) which was introduced by Korn et al. (2004) as well as others. The FDP(γ) is given as P(Q > γ), where $Q = V/R$ if $R > 0$, and $Q = 0$, if $R = 0$, and R is the random number of rejected hypotheses. The false discovery rate FDR = E(Q), introduced by Benjamini and Hochberg (1995), is also a commonly used error rate. However its control, i.e., $E(Q)\leq\alpha$, does not prevent Q from attaining values much

[*] Corresponding author at: Institute of Medical Statistics, Computer Sciences and Documentation, University of Jena, Bachstraße 18, D-07740 Jena, Germany. Tel.: +49 3641 933610; fax: +49 3641 933200.

E-mail address: claudia.hemmelmann@mti.uni-jena.de (C. Hemmelmann).

0165-0270/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2007.12.013

greater than α in single cases. This is a disadvantage of the FDR criterion as problems in interpretation may result. Therefore, we prefer the gFWE and the FDP criterion and we consider only these two error rates in this paper. For an overview and a discussion of the different error rates in multiple comparison problems, see, e.g., Hemmelmann et al. (2005).

In this paper we propose improvements of two multiple test procedures, the gFWE procedure of Hommel and Hoffmann (1987) and the FDP procedure of Lehmann and Romano (2005), using an upper bound of the number of true hypotheses. Of course, the number m_0 of true hypotheses is unknown in practice, and it needs to be estimated. But many approaches have been proposed for estimating the proportion of true hypotheses, and good overviews have been given, e.g., by Broberg (2005), Hsueh et al. (2003) and Langaas et al. (2005). For our new procedures, we only rely on an upper bound m_0^* for the number m_0 of true hypotheses, i.e., ensuring $1 \leq m_0 \leq m_0^* \leq m$, and thus does not discuss approaches for estimating m_0 in detail. It should be noted that our procedures require an overestimate of the number of true hypotheses.

Our paper has the following aims: (a) to introduce the procedures of Hommel and Hoffmann (1987) and Lehmann and Romano (2005), (b) to improve these two methods by using an upper bound of the number of true hypotheses, (c) to demonstrate the resulting gain in power by Monte-Carlo simulations, and (d) to present the use of our new procedures in problems of multiple comparisons of coherence values obtained from EEG data recorded during the memory encoding of recalled or not-recalled abstract nouns (Weiss et al., 2000).

The procedures we discuss are not only specific to EEG data; they are equally applicable to the large data in MEG and fMRI.

2. Methods

2.1. Multiple tests

Holm's (1979) step-down procedure for control of FWE is one of the most popular approaches to adjust for multiplicity. The elegance, simplicity, and robustness of this procedure have motivated several authors to develop further improvements. Specifically, Hommel and Hoffmann (1987) and later Lehmann and Romano (2005) have derived modified constants for the stepwise comparison with the p-values which guarantee that the gFWE(u) does not exceed the significance level α for some prespecified integer u $(0 \leq u < m)$. Lehmann and Romano (2005) also modify Holm's method for controlling the FDP(γ), i.e. $P(Q > \gamma) \leq \alpha$ $(0 \leq \gamma < 1)$, under special conditions.

It is generally known that the power of FWE methods like those of the Bonferroni and Holm type can be improved if the number of true hypotheses is known (see, e.g., Hsueh et al., 2003). In a similar way, the power of the gFWE procedure of Hommel and Hoffmann (1987) and of the FDP procedure of Lehmann and Romano (2005) can be increased by utilizing information on the number m_0 $(m_0 \leq m)$ of true hypotheses. Throughout this work, we only consider step-down procedures.

2.1.1. The Hommel and Hoffmann method and its improvement

Let $H_1, \ldots, H_m$ denote the hypotheses to be tested, and let $p_1, \ldots, p_m$ be the corresponding unadjusted p-values obtained with some appropriate tests. The ordered p-values and the corresponding hypotheses are denoted by $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$ and by $H_{(1)}, \ldots, H_{(m)}$. A step-down procedure compares the smallest p-value $p_{(1)}$ with α_1, $p_{(2)}$ with α_2, etc., and continues rejecting hypotheses as long as their corresponding p-values $p_{(i)}$ are smaller than or equal to α_i. The step-down procedure of Holm (1979) compares the ordered p-values $p_{(1)}, p_{(2)}, \ldots, p_{(m)}$ with $\alpha_1 = \alpha/m$, $\alpha_2 = \alpha/(m-1)$, ..., $\alpha_m = \alpha$. Holm's procedure thus controls the FWE, i.e., $P(V > 0) \leq \alpha$.

An augmentation of Holm's method which controls the gFWE(u) for all integers u $(0 \leq u < m)$ instead of the FWE was introduced by Hommel and Hoffmann (1987). Here, one compares $p_{(i)}$ with:

$$\alpha_i^{\mathrm{HH}} = \begin{cases} \dfrac{(u+1)\alpha}{m} & \text{if } 1 \leq i \leq u+1 \\ \dfrac{(u+1)\alpha}{m+u+1-i} & \text{if } u+1 < i \leq m \end{cases} \tag{1}$$

Now, an upper bound m_0^* for the number of true hypotheses such that $1 \leq m_0 \leq m_0^* \leq m$ is given. We also assume $m_0^* > u$, otherwise one can reject any hypotheses irrespective of its p-values, because $V \leq m_0 \leq m_0^* \leq u$, thus $P(V > u) = 0$. In the case of $m_0^* \leq u$ we recommend a comparison of all $p_{(i)}$ with α. We can now improve Hommel and Hoffmann's gFWE procedure as follows:

Statement 1. *For fixed $u \geq 0$ let $m_0^* > u$ denote a known upper bound for the number m_0 of true hypotheses $(1 \leq m_0 \leq m_0^* \leq m)$. Then, the step-down procedure where $p_{(i)}$ is compared with*

$$\alpha_i^{\mathrm{HH}u} = \begin{cases} \dfrac{(u+1)\alpha}{m_0^*} & \text{if } 1 \leq i \leq m - m_0^* + u + 1 \\ \dfrac{(u+1)\alpha}{m+u+1-i} & \text{if } m - m_0^* + u + 1 < i \leq m \end{cases} \tag{2}$$

controls the gFWE(u). It thus guarantees $P(V > u) \leq \alpha$ for each integer u $(0 \leq u < m)$.

It is obvious that $\alpha_i^{\mathrm{HH}u} \geq \alpha_i^{\mathrm{HH}}$ for all $i = 1, \ldots, m$. The equality holds for $i > m - m_0^* + u + 1$ and for all i if $m_0^* = m$. The difference will be pronounced if $m_0^* \ll m$.

The proof of Statement 1 is similar to the proof of Theorem 2.2 in Lehmann and Romano (2005) and is given in Appendix A.

2.1.2. The Lehmann and Romano method and its improvement

Lehmann and Romano (2005) proposed a further modification of Holm's method which controls the FDP(γ), i.e., $P(Q > \gamma) \leq \alpha$ for any γ $(0 \leq \gamma < 1)$ under special conditions. One condition is that the p-values of the m_0 $(m_0 \geq 1)$ true hypotheses q_k $(k = 1, \ldots, m_0)$ satisfy the Simes (1986) inequality, i.e., $P(\bigcup_{k=1}^{m_0}\{q_{(k)} \leq k\alpha/m_0\}) \leq \alpha$. We point out that the Simes inequality holds true, e.g., for independent test statistics or for many positively dependent test statistics. Particularly, it is fulfilled for multivariate normal distributions and central multivariate *t* distributions with common and nonnegative correlations (Sarkar, 1998).

In the step-down procedure of Lehmann and Romano (2005) one compares $p_{(i)}$ with

$$\alpha_i^{\mathrm{LR}} = \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{m + \lfloor \gamma i \rfloor + 1 - i} (i = 1, \ldots, m), \tag{3}$$

where $\lfloor x \rfloor$ is the largest integer $\leq x$, and the other notation is as before. In the special case of $\gamma = 0$, these levels are identical to α_i^{HH} given by Eq. (1) with $u = 0$ and consequently identical to Holm's.

If an upper bound m_0^* for the number m_0 of true hypotheses is known, the procedure of Lehmann and Romano (2005) for the FDP(γ) can be improved as follows:

Statement 2. *For fixed* $\gamma \in [0, 1)$ *let* m_0^* *denote a known upper bound for the number* m_0 *of true hypotheses* $(1 \leq m_0 \leq m_0^* \leq m)$. *Furthermore, suppose that the* Simes (1986) *inequality holds for the p-values of the true hypotheses. Then the step-down procedure where* $p_{(i)}$ *is compared with*

$$\alpha_i^{\mathrm{LR}u} = \begin{cases} \dfrac{(\lfloor \gamma i \rfloor + 1)\alpha}{m_0^*} & \text{if } 1 \leq m_0^* \leq m + \lfloor \gamma i \rfloor + 1 - i \\ \dfrac{(\lfloor \gamma i \rfloor + 1)\alpha}{m + \lfloor \gamma i \rfloor + 1 - i} & \text{if } m + \lfloor \gamma i \rfloor + 1 - i < m_0^* \end{cases} \tag{4}$$

controls the FDP(γ). It thus guarantees $P(Q > \gamma) \leq \alpha$ for $0 \leq \gamma < 1$.

Thus, $\alpha_i^{\mathrm{LR}u} \geq \alpha_i^{\mathrm{LR}}$ for all $i = 1, \ldots, m$. The equality holds for $m + \lfloor \gamma i \rfloor + 1 - i < m_0^*$ and for all i if $m_0^* = m$. The difference will be pronounced if $m_0^* \ll m$.

Furthermore, if one compares α_i^{HH} (Eq. (1)) with α_i^{LR} (Eq. (3)) and $\alpha_i^{\mathrm{HH}u}$ (Eq. (2)) with $\alpha_i^{\mathrm{LR}u}$ (Eq. (4)) it is clear that these levels are stepwise identical depending on γ ($\gamma \in [0,1)$) for successive $u = 0, 1, 2, \ldots$. For example, for $\gamma = 0.1$ the α_i^{LR} and $\alpha_i^{\mathrm{LR}u}$ ($i = 1, \ldots, 9$) are identical to α_i^{HH} and $\alpha_i^{\mathrm{HH}u}$ ($i = 1, \ldots, 9$) with $u = 0$, α_i^{LR} and $\alpha_i^{\mathrm{LR}u}$ ($i = 10, \ldots, 19$) are identical to α_i^{HH} and $\alpha_i^{\mathrm{HH}u}$ ($i = 10, \ldots, 19$) with $u = 1$, etc. In particular, $\alpha_1^{\mathrm{LR}} = \alpha/m$ and $\alpha_1^{\mathrm{LR}u} = \alpha/m_0^*$ for any $\gamma \in [0,1)$ and these are identical to α_1^{HH} and $\alpha_1^{\mathrm{HH}u}$ with $u = 0$ (i.e., equal to Holm' method), respectively. This is a disadvantage compared to the gFWE methods for $u > 0$ where $\alpha_1^{\mathrm{HH}} = (u + 1)\alpha/m$ and $\alpha_1^{\mathrm{HH}u} = (u + 1)\alpha/m_0^*$.

Lehmann and Romano (2005) also suggest a FDP method without any dependence assumption, and they replace α_i^{LR} by $\alpha_i^{\mathrm{LR}*} = \alpha_i^{\mathrm{LR}}/(\sum_{k=1}^{\lfloor \gamma m \rfloor + 1} k^{-1})$. We can thus deduce the following corollary:

Corollary. *A step-down procedure with levels* $\alpha_i^{\mathrm{LR}u*} = \alpha_i^{\mathrm{LR}u}/(\sum_{k=1}^{c} k^{-1})$, *where* $\alpha_i^{\mathrm{LR}u}$ *as defined in* Eq. (4) *and* $c = \min\{\lfloor \gamma m \rfloor + 1, m_0^*\}$, *controls the* FDP($\gamma$) for any γ $(0 \leq \gamma < 1)$ *and for any dependence of the p-values.*

Similarly to the proposed gFWE method, one can reject all m hypotheses irrespective of their p-values if $m_0^* \leq \gamma m$, because $V/R = m_0/m \leq m_0^*/m \leq \gamma$, thus $\mathrm{P}(Q > \gamma) = 0$. However, we want to accept null hypotheses with large p-values and thus our proposed procedures do not utilize this. The proofs of Statement 2 and its corollary are given in Appendix A.

2.2. Simulated data and real data (EEG data)

2.2.1. Simulated data

Data were simulated using the following model: Let m_1 be the number of false hypotheses and thus $m = m_0 + m_1$. We generated m-dimensional normal distributed random vectors $\boldsymbol{x}_j \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($j = 1, \ldots, n$) with means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$', where $\mu_i = 1.5$ for $i = 1, \ldots, m_1$ and $\mu_i = 0$ for $i = m_1 + 1, \ldots, m$ and $\boldsymbol{\Sigma} = (\rho_{ik})$ for $i, k = 1, \ldots, m$. The number m_1 of false hypotheses was varied between 1 and $m - 1$. p-Values were calculated from two-sided one-sample t-tests, i.e., test the null hypotheses: $\mu_i = 0$, for each component. We chose $n = 8$ and $m = 100$, and we considered the following cases of constant correlation coefficients: $\rho_{ik} = 0$, $\rho_{ik} = 0.2$, $\rho_{ik} = 0.5$, $\rho_{ik} = 0.8$ and $\rho_{ik} = 0.9$ for $i, k = 1, \ldots, m$ and $i \neq k$. The number of replicates was 60,000, and the significance level was set to $\alpha = 0.05$.

2.2.2. EEG data

The experimental setup and the methods of EEG analysis have been described in detail elsewhere (Weiss et al., 2000; Weiss and Rappelsberger, 2000). In brief, the EEG data are $m = 171$ coherence values from $n = 23$ female German native speakers, see, e.g., Hemmelmann et al. (2005). They auditorily perceived two wordlists each containing 25 disyllabic abstract nouns. Participants had to memorize the nouns and they were asked to recall the words previously encoded immediately after the presentation of each list. The pairs of electrodes showed differences in their means of coherence values for the subsequently recalled versus non-recalled nouns for the delta1 (1–2 Hz) frequency band.

The goal of our analysis was to identify the electrode pairs with significant coherence differences. p-Values were calculated from two-sided paired t-tests for each component. For applying our new procedures to the EEG data we estimated the upper bound m_0^* of the number of true hypotheses by using the permutation method of Meinshausen and Bühlmann (2005) with the quantile bounding function, because this approach ensures $P(m_0^* \geq m_0) \geq 1 - \beta$ for a specified confidence level $1 - \beta$ under general dependence structures. Applied to our EEG data, we obtained $m_0^* = 95$ for $1 - \beta = 0.95$ with this method.

3. Results

3.1. Power comparisons

Simulations were performed to demonstrate the increase in power gained by our methods, which utilize information on the number m_0 of true hypotheses. For the FDP method we used Statement 2. The standard criterion for evaluating test procedures is the average power (Korn et al., 2004; Kwong et al., 2002; Troendle, 2000) which is the expected proportion of rejected hypotheses among the false hypotheses.

We estimated the average power of the original procedures and of our improved procedures for $u = 5$ and $\gamma = 0.1$ with

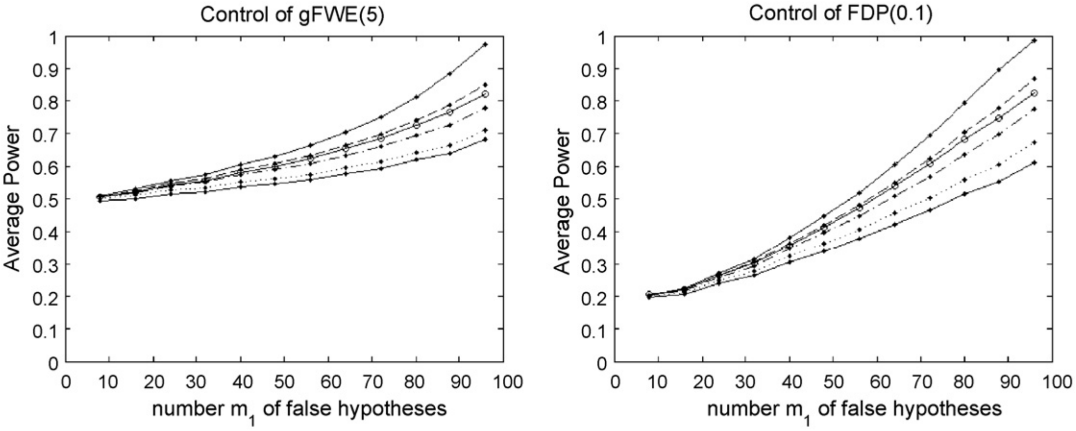

Fig. 1. Estimated average power of the original procedure (lower solid line) and our improved procedure with $m_0^* = m_0 + m_1/2$ (. . .), $m_0^* = m_0 + m_1/4$ (–·–·), $m_0^* = m_0 + m_1/8$ (– – –), $m_0^* = m_0$ (upper solid line) and m_0^* of the method of Meinshausen and Bühlmann (2005) (circle) for controlling of gFWE(5) (left side) and FDP(0.1) (right side) and $\rho_{ik} = 0.5$.

$\rho_{ik} = 0.5$ (Fig. 1). For m_0^* in formulae (2) and (4), we used the following values: $m_0^* = m_0$, $m_0^* = m_0 + m_1/8$, $m_0^* = m_0 + m_1/4$ and $m_0^* = m_0 + m_1/2$. In addition, we used the method of Meinshausen and Bühlmann (2005) to obtain a further m_0^*. As expected, our improvements resulted in a gain in power, and we obtained the maximum gain in power with $m_0^* = m_0$ assuming that $m_0 \leq m_0^* \leq m$ (see Fig. 1). This gain increased with increasing m_1 and with a decrease of the overestimation of m_0. As it can be seen in Fig. 1, the method of Meinshausen and Bühlmann (2005) also resulted in a gain in power.

3.2. *Certainty by underestimation of the number of true hypotheses*

Thus far we have considered the gain of power of our proposed procedures if the assumption $m_0 \leq m_0^* \leq m$ holds true. We also explored the case that the assumption $m_0 \leq m_0^* \leq m$ is incorrect, i.e., the number of true hypotheses is underestimated. Specifically, we investigated by how much the number of true hypotheses may be underestimated, where the significance level held, i.e., gFWE(u) $\leq \alpha$ and FDP(γ) $\leq \alpha$, respectively. Fig. 2 shows the error rates for $u = 5$ and $\gamma = 0.1$ with $n = 8$, $m = 100$, $m_0 = 50$, and m_0^* was varied between 5 and m for different constant correlation coefficients. The resulting values exceeded $\alpha = 0.05$ approximately for $m_0^* < 15$. Thus, the maximum number of underestimations of m_0, abbreviated by MNU, is about 35. The MNU can be defined as formal MNU = max{$j = 1, \ldots, m_0 - 1$: $m_0^* = m_0 - j$ whereas gFWE(u) $\leq \alpha$ and FDP(γ) $\leq \alpha$ for the proposed procedures of Statements 1 and 2, respectively}.

Table 1 shows the MNU for other values of m_0, u and γ for the "worst case" of the constant correlation coefficients, i.e., for the other correlation coefficients the MNU can be larger.

For controlling the gFWE(u), the proportion of the MNU among m_0 is approximately at least 60% and appears to be independent of u if $u > 0$. If $u = 0$ (identical to Holm's method) then no underestimation was allowed (not shown here). Furthermore, the proportion of the MNU among m_0 is not independent of γ for controlling of the FDP(γ), e.g., in analogy to Holm's method no underestimation was allowed for $\gamma = 0.01$.

To summarize, we have demonstrated that the assumption $m_0 \leq m_0^* \leq m$ may be violated in some cases, and that m_0 may be underestimated by up to 2/3 without violating the significance

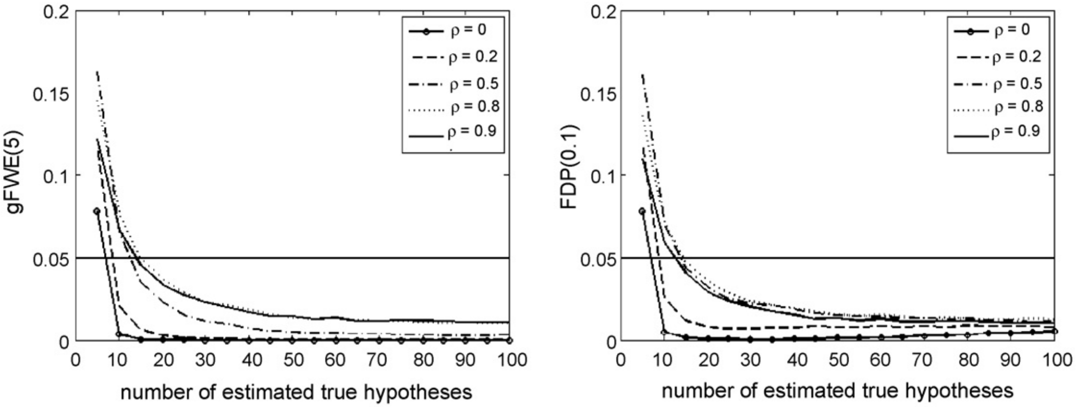

Fig. 2. Estimated gFWE(5) (left side) and FDP(0.1) (right side) for our improved procedures for $m_0 = 50$ and different constant correlation coefficients for correctly and incorrectly assumed upper bounds of the number of true hypotheses.

Table 1
List of MNU for the “worst case” of the correlation. The percentages (%) are in parenthesis

		$m_0 = 20$	$m_0 = 50$	$m_0 = 80$
gFWE(u)	$u = 1$	12 (60) for $\rho = 0.5$	31 (62) for $\rho = 0.5$	51 (63.7) for $\rho = 0.5$
	$u = 3$	13 (65) for $\rho = 0.8$	34 (68) for $\rho = 0.8$	58 (72.5) for $\rho = 0.8$
	$u = 5$	13 (65) for $\rho = 0.9$	35 (70) for $\rho = 0.8$	58 (72.5) for $\rho = 0.8$
	$u = 10$	12 (60) for $\rho = 0.9$	33 (66) for $\rho = 0.9$	56 (70.0) for $\rho = 0.8$
	$u = 15$	12 (60) for $\rho = 0.9$	32 (64) for $\rho = 0.9$	55 (68.7) for $\rho = 0.9$
FDP(γ)	$\gamma = 0.01$	0 (0) for $\rho = 0$	0 (0) for $\rho = 0$	0 (0) for $\rho = 0$
	$\gamma = 0.03$	12 (60) for $\rho = 0.5$	0 (0) for $\rho = 0.2$	0 (0) for $\rho = 0$
	$\gamma = 0.05$	13 (65) for $\rho = 0.8$	30 (60) for $\rho = 0.5$	0 (0) for $\rho = 0$
	$\gamma = 0.1$	12 (60) for $\rho = 0.9$	36 (72) for $\rho = 0.8$	25 (31.2) for $\rho = 0.2$
	$\gamma = 0.15$	12 (60) for $\rho = 0.9$	36 (72) for $\rho = 0.8$	51 (63.7) for $\rho = 0.5$

level. These results were also confirmed by further simulations including other dimensions of m, or mixed positive and negative correlation coefficients ρ_{ik}. Details are available from the corresponding author. However, we point out that nevertheless an estimation method which overestimates the number of true hypotheses with a strong certainty should be used, because multiple test procedures have to be valid for a multitude of parameter configurations. In order to ensure this, test procedures are often conservative for specific situations. This does not exclude that they hold up well in other situations. In these cases, an underestimation of the bound might be misleading.

3.3. Applications of multiple tests to EEG coherence data

The data we now evaluate are from the experiment described in Section 2.2.2. For the significance level $\alpha = 0.05$ Table 2 displays the number of significant coherence differences for the original gFWE and FDP methods and for the proposed improved procedures. For controlling the FDP(γ) we used the method of Statement 2 again. The FDP method of the corollary is very conservative and results only in a gain in power compared to the Holm method if the number of false hypotheses is very large.

Of course, the number of rejected hypotheses is small with both methods, given that there are at least $m - m_0^* = 76$ false hypotheses. Possibly, many coherence differences between recalled and non-recalled nouns are too small and the corresponding variances are too large. But our improved procedure yields at least 25% more rejections when controlling the gFWE(u) and at least 33% more when controlling the FDP(γ).

Table 2
Number of significant coherence differences when applying the original procedures and the proposed improved procedures with $m_0^* = 95$ ($\alpha = 0.05$)

	Original approach	New approach
gFWE(u) ≤ 0.05		
0	6	8
1	9	13
2	12	15
5	15	23
10	25	37
FDP(γ) ≤ 0.05		
0.01	6	8
0.05	6	8
0.10	6	8
0.15	9	15
0.20	12	18

As noted in Section 2.1.2 the first levels of the FDP methods are identical to the first levels of Holm’s method. Therefore, the FDP methods do not reject more hypotheses than the gFWE methods with $u = 0$ if they do not reject enough hypotheses to compare the p-values with the levels in analogy to $u = 1$. For example, the FDP methods with $\gamma = 0.1$ cannot reject more hypotheses than the gFWE methods with $u = 0$, if the corresponding gFWE method with $u = 0$ reject less than 9 hypotheses because the first 9 levels are identical. But the 10th levels of the FDP methods are identical to the 10th levels of the gFWE methods with $u = 1$.

The improved procedure both supports and extends significant findings of the original approach. While processing of both subsequently recalled and non-recalled nouns elicited higher coherence at anterior sites, only the recalled nouns were associated with enhanced coherence between the distant frontal and posterior (temporal, parietal and occipital) electrodes of both hemispheres. The improved statistical procedure (with $u = 1$) emphasizes these findings by showing an additional coherence increase between frontal and temporal/occipital electrode pairs. Specifically, frontal electrodes are activated, which supports the well-known role of the frontal cortex in memory processes (Tulving et al., 1994). The improved procedure also stresses the interhemispheric fronto-temporal coherence increase, which could not be found with the original statistical approach (see Fig. 3).

4. Discussion

In this article we have described the improvements of the step-down procedures of Hommel and Hoffmann (1987) and Lehmann and Romano (2005) for testing multiple hypotheses if an upper bound m_0^* for the number m_0 of true hypotheses is available. Alternative methods that also control the gFWE(u) and the FDP(γ) can be found in Korn et al. (2004), van der Laan et al. (2004) as well as others. However, the method of Korn et al. (2004) is based on permutation tests and is very complex. Possibly, these procedures can be improved in a similar way if an upper bound m_0^* for the number m_0 of true hypotheses is known.

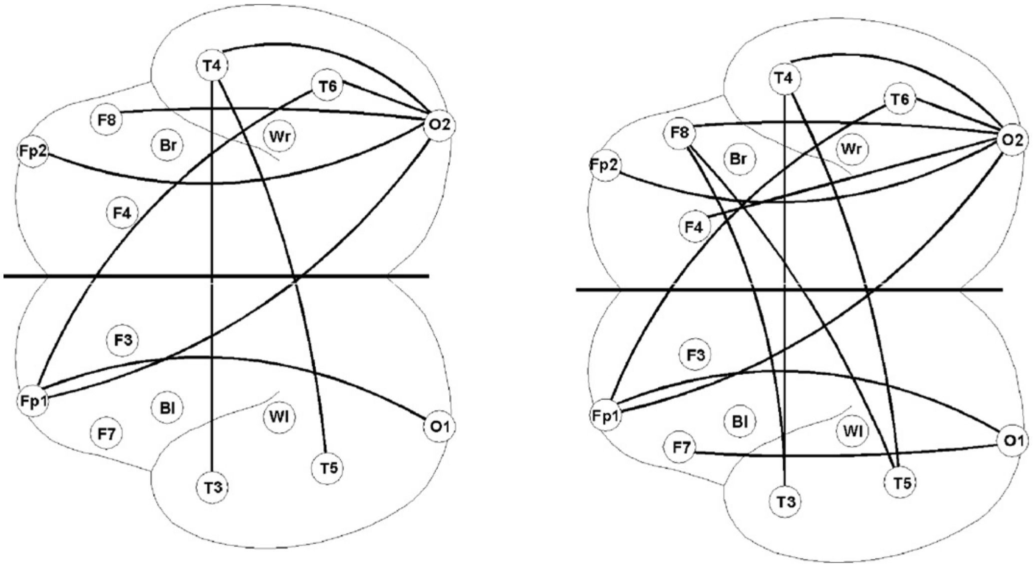

Fig. 3. View on the top of the unfolded hemispheres with electrodes used (circles) and lines denoting significant coherence differences. Left side: Hommel and Hoffmann method. Right side: our improved procedure ($u = 1$, $\alpha = 0.05$).

However, more work is needed to address the question of how these procedures can be improved.

In this work we assume that $m_0^* \geq m_0$ is available. In recent years many different estimation methods of m_0 have been introduced, e.g., by Benjamini and Hochberg (2000), Nettleton et al. (2006), Schweder and Spjotvoll (1982), Storey (2002), Storey et al. (2004) and Turkheimer et al. (2001). But these methods do not ensure that the number of true hypotheses is overestimated. In contrast, the methods of Meinshausen and Bühlmann (2005) and Meinshausen and Rice (2006) are constructed to hold $P(m_0^* \geq m_0) \geq 1 - \beta$ for a specified confidence level $1 - \beta$, i.e., they overestimated the number of true hypotheses. Therefore, we used the method of Meinshausen and Bühlmann (2005) to evaluate the EEG data. When using this method to compute m_0^* then the corresponding error rate was always controlled in our simulation analysis.

Other methods exist to control the false discovery rate which estimate m_0 in the first step, for example the method of Benjamini et al. (2006) and the method "significance analysis of microarrays" of Tusher et al. (2001). In contrast to our procedures, the estimation of m_0 is an inherent part in these methods.

We have shown theoretically and by Monte-Carlo simulations that our new procedures are more powerful than the original procedures by Hommel and Hoffmann (1987) and Lehmann and Romano (2005) if $m_0^* \ll m$. Furthermore, we have illustrated by an example using EEG data that our proposed procedures reject more hypotheses than the original procedures. It could be shown that the coherence difference between recalled and non-recalled nouns is emphasized by additional significant coherence changes. In particular, an increased relationship between signals at right frontal and left temporal electrodes is indicated by our method. Frontal and temporal regions have frequently been shown to be involved with memory formation of words (Wagner et al., 1998). Thus, the improved statistical procedure allows us to detect additional coherence changes important for the interpretation of cognitive findings.

Acknowledgements

We are grateful to Manfred Horn for helpful discussions and for many valuable suggestions that greatly improved the clarity of the paper. We would also like to thank the referees for their helpful comments. This work was supported by the Deutsche Forschungsgemeinschaft (VO 683/2-1).

Appendix A

Let $q_{(1)} \leq \ldots \leq q_{(m_0)}$ denote the ordered p-values of the m_0 true hypotheses.

Proof of Statement 1. No proof is required for $m_0 < u + 1$. Therefore, we consider $m_0 \geq u + 1$. Let j be the smallest random index satisfying $p_{(j)} = q_{(u+1)}$, so $u + 1 \leq j \leq m - m_0 + u + 1$. With $\alpha_j^{\mathrm{HH}u}$ of Eq. (2), we have gFWE(u) = P(V > u) = P(p(1) $\leq \alpha_1^{\mathrm{HH}u}$, ..., p$_{(j)} \leq \alpha_j^{\mathrm{HH}u}$) $\leq$ P($P_{(j)} \leq \alpha_j^{\mathrm{HH}u}$) for the step-down procedure. Thus, we only need to consider $p_{(j)} \leq \alpha_j^{\mathrm{HH}u}$. The following case differentiation for j is the only modification of the proof of Theorem 2.2 in Lehmann and Romano (2005).

- Case 1: let $u + 1 \leq j \leq m - m_0^* + u + 1$.
  It follows $q_{(u+1)} = p_{(j)} \leq \alpha_j^{\mathrm{HH}u} = (u+1)\alpha/m_0^* \leq (u + 1)\alpha/m_0$.
- Case 2: let $m - m_0^* + u + 1 < j \leq m - m_0 + u + 1$.
  It follows $q_{(u+1)} = p_{(j)} \leq \alpha_j^{\mathrm{HH}u} = (u+1)\alpha/(m + u + 1 - j) \leq (u+1)\alpha/m_0$.

In both cases $p_{(j)} \leq \alpha_j^{\mathrm{HH}u}$ implies $q_{(u+1)} \leq (u + 1)\alpha/m_0$, so that gFWE($u$) $\leq P(q_{(u+1)} \leq (u + 1)\alpha/m_0)$. Hommel and Hoffmann (1987) and Lehmann and Romano (2005) (proof of Theorem 2.1(i)) have shown gFWE(u) $\leq m_0\alpha/m$ when m hypotheses were tested using a single-step method with constant levels $(u + 1)\alpha/m$. Thus, gFWE(u) $\leq \alpha$ when replacing α by $\alpha m/m_0$, i.e., using constant levels $(u+1)\alpha/m_0$. We therefore finally have $\mathrm{P}(q_{(u+1)} \leq (u+1)\alpha/m_0) \leq \alpha$. □

Proof of Statement 2 and of Corollary. The proof of Statement 2 is a direct consequence of Lehmann and Romano (2005) (proof of Theorem 3.2). The range of γi $(0 \leq \gamma < 1;\ 1 \leq i \leq m)$ is divided into $\gamma i < 1$, $1 \leq \gamma i < 2$, …, $\lfloor \gamma m \rfloor \leq \gamma i < \lfloor \gamma m \rfloor + 1$. Let $j \leq m$ be the smallest random index where the proportion of false rejections Q exceeds γ for the first time. Then: $p_{(j)} \leq \alpha_j^{\mathrm{LR}u}$, $H_{(j)}$ is true and $\lfloor \gamma j \rfloor + 1 \leq m_0$. We thus have $\mathrm{FDP}(\gamma) = \mathrm{P}(Q > \gamma) \leq \mathrm{P}(\{\gamma j < 1\} \cup \{1 \leq \gamma j < 2\} \cup \ldots \cup \{b-1 \leq \gamma j < b\})$ with $b = \min\{\lfloor \gamma m \rfloor + 1, m_0\}$.

Let $k-1 \leq \gamma j < k$ for any $k \in \{1, \ldots, b\}$. Then $p_{(j)} = q_{(k)} \leq \alpha_j^{\mathrm{LR}u}$ because $(k-1)/j \leq \gamma$ and $k/j > \gamma$. This implies that $H_{(j)}$ is the kth rejected true hypothesis, and $k \leq j \leq m - m_0 + k$. We consider the case differentiation for j in analogy to the proof of Statement 1.

- Case 1: let $k \leq j \leq m - m_0^* + k$, thus $m_0^* \leq m + k - j = m + \lfloor \gamma j \rfloor + 1 - j$ and, according to Eq. (4), $q_{(k)} = p_{(j)} \leq \alpha_j^{\mathrm{LR}u} = (\lfloor \gamma j \rfloor + 1)\alpha/m_0^* \leq k\alpha/m_0$.
- Case 2: let $m - m_0^* + k < j \leq m - m_0 + k$, thus $m_0 \leq m + \lfloor \gamma j \rfloor + 1 - j < m_0^*$ and, according to Eq. (4), $q_{(k)} = p_{(j)} \leq \alpha_j^{\mathrm{LR}u} = (\lfloor \gamma j \rfloor + 1)\alpha/(m + \lfloor \gamma j \rfloor + 1 - j) \leq \frac{k\alpha}{m_0}$.

Next, $k-1 \leq \gamma j < k$ implies $q_{(k)} \leq k\alpha/m_0$ for any $k \in \{1, \ldots, b\}$, so that $\mathrm{FDP}(\gamma) = \mathrm{P}(Q > \gamma) \leq \mathrm{P}(\{\gamma j < 1\} \cup \{1 \leq \gamma j < 2\} \cup \ldots \cup \{b-1 \leq \gamma j < b\}) \leq P(\bigcup_{k=1}^{b} \{q_{(k)} \leq k\alpha/m_0\})$.

The step-down procedure with $\alpha_i^{\mathrm{LR}u}$ from Eq. (4) thus controls the $\mathrm{FDP}(\gamma)$ by α if the Simes inequality is true for the p-values $q_1, \ldots, q_{m_0}$ of the m_0 true hypotheses.

Furthermore, the inequality of Hommel (1983) $\mathrm{P}(\bigcup_{k=1}^{m_0} \{q_{(k)} \leq k\alpha/m_0\}) \leq \alpha \sum_{k=1}^{m_0} k^{-1}$ and its generalization by Lehmann and Romano (2005) $\mathrm{P}(\bigcup_{k=1}^{s} \{q_{(k)} \leq \beta_k\}) \leq m_0 \sum_{k=1}^{s} (\beta_k - \beta_{k-1}) k^{-1}$ with any $s \leq m_0$, and $0 = \beta_0 \leq \beta_1 \leq \ldots \leq \beta_s \leq 1$ hold for any dependency of q_k for $k = 1, \ldots, m_0$. From $b = \min\{\lfloor \gamma m \rfloor + 1, m_0\}$ it follows $\mathrm{P}(\bigcup_{k=1}^{b} \{q_{(k)} \leq k\alpha/m_0\}) \leq m_0 \sum_{k=1}^{b} (k\alpha/m_0 - (k-1)\alpha/m_0) k^{-1} = \alpha \sum_{k=1}^{b} k^{-1} \leq \alpha \sum_{k=1}^{c} k^{-1}$ which completes the proof of the corollary. □

References

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B Stat Methodol 1995;57:289–300.

Benjamini Y, Hochberg Y. On the adaptive control of the false discovery fate in multiple testing with independent statistics. J Educ Behav Stat 2000;25:60–83.

Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 2006;93:491–507.

Broberg P. A comparative review of estimates of the proportion unchanged genes and the false discovery rate. BMC Bioinformatics 2005:6.

Hemmelmann C, Horn M, Reiterer S, Schack B, Süße T, Weiss S. Multivariate tests for the evaluation of high-dimensional EEG data. J Neurosci Methods 2004;139:111–20.

Hemmelmann C, Horn M, Süße T, Vollandt R, Weiss S. New concepts of multiple tests and their use for evaluating high-dimensional EEG data. J Neurosci Methods 2005;142:209–17.

Holm S. A simple sequentially rejective multiple testing procedure. Scand J Stat 1979:65–70.

Hommel G. Tests of the overall hypothesis for arbitrary dependence structures. Biom J 1983;25:423–30.

Hommel G, Hoffmann T. Controlled uncertainty. In: Bauer P, Hommel G, Sonnemann E, editors. Multiple Hypotheses Testing. Heidelberg: Springer; 1987. p. 154–61.

Hsueh H-M, Chen JJ, Kodell RL. Comparison of methods for estimating the number of true null hypotheses in multiplicity testing. J Biopharm Stat 2003;13:675–89.

Korn EL, Troendle JF, McShane LM, Simon R. Controlling the number of false discoveries: application to high-dimensional genomic data. J Stat Plan Infer 2004;124:379–98.

Kwong KS, Holland B, Cheung SH. A modified Benjamini–Hochberg multiple comparisons procedure for controlling the false discovery rate. J Stat Plan Infer 2002;104:351–62.

Langaas M, Lindqvist BH, Ferkingstad E. Estimating the proportion of true null hypotheses, with application to DNA microarray data. J Roy Stat Soc Ser B Stat Methodol 2005;67:555–72.

Lehmann EL, Romano JP. Generalizations of the familywise error rate. Ann Stat 2005;33:1138–54.

Meinshausen N, Bühlmann P. Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. Biometrika 2005;92:893–907.

Meinshausen N, Rice J. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. Ann Stat 2006;34:373–93.

Nettleton D, Hwang JTG, Caldo RA, Wise RP. Estimating the number of true null hypotheses from a histogram of p-values. J Agric Biol Environ Stat 2006;11:337–56.

Sarkar SK. Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. Ann Stat 1998;26:494–504.

Schweder T, Spjotvoll E. Plots of P-values to evaluate many tests simultaneously. Biometrika 1982;69:493–502.

Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika 1986;3:751–4.

Storey JD. A direct approach to false discovery rates. J Roy Stat Soc Ser B Stat Methodol 2002;64:479–98.

Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J Roy Stat Soc Ser B Stat Methodol 2004;66:187–205.

Troendle JF. Stepwise normal theory multiple test procedures controlling the false discovery rate. J Stat Plan Infer 2000;84:139–58.

Tulving E, Kapur S, Craik FIM, Moscovitch M, Houle S. Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. Proc Natl Acad Sci USA 1994;91:2016–20.

Turkheimer FE, Smith CB, Schmidt K. Estimation of the number of “true” null hypotheses in multivariate analysis of neuroimaging data. Neuroimage 2001;13:920–30.

Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001;98:5116–21.

van der Laan MJ, Dudoit S, Pollard KS. Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. Stat Appl Genet Mol Biol 2004:3.

Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, et al. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science 1998;281:1188–91.

Weiss S, Rappelsberger P. Long-range EEG synchronization during word encoding correlates with successful memory performance. Cogn Brain Res 2000;9:299–312.

Weiss S, Müller HM, Rappelsberger P. Theta synchronization predicts efficient memory encoding of concrete and abstract nouns. NeuroReport 2000;11:2357–61.