

Statistical Applications in Genetics and Molecular Biology

Volume 8, Issue 1

2009

Article 23

Weighted Multiple Hypothesis Testing Procedures

Guolian Kang* Keying Ye[†] Nianjun Liu[‡]
David B. Allison** Guimin Gao^{††}

*University of Alabama at Birmingham, gkang@ms.soph.uab.edu

[†]University of Texas at San Antonio, keying.ye@utsa.edu

[‡]University of Alabama at Birmingham, nliu@uab.edu

**University of Alabama at Birmingham, dallison@ms.soph.uab.edu

^{††}University of Alabama at Birmingham, ggao@ms.soph.uab.edu

Copyright ©2009 The Berkeley Electronic Press. All rights reserved.

Weighted Multiple Hypothesis Testing Procedures*

Guolian Kang, Keying Ye, Nianjun Liu, David B. Allison, and Guimin Gao

Abstract

Multiple hypothesis testing is commonly used in genome research such as genome-wide studies and gene expression data analysis (Lin, 2005). The widely used Bonferroni procedure controls the family-wise error rate (FWER) for multiple hypothesis testing, but has limited statistical power as the number of hypotheses tested increases. The power of multiple testing procedures can be increased by using weighted p-values (Genovese et al., 2006). The weights for the p-values can be estimated by using certain prior information. Wasserman and Roeder (2006) described a weighted Bonferroni procedure, which incorporates weighted p-values into the Bonferroni procedure, and Rubin et al. (2006) and Wasserman and Roeder (2006) estimated the optimal weights that maximize the power of the weighted Bonferroni procedure under the assumption that the means of the test statistics in the multiple testing are known (these weights are called optimal Bonferroni weights). This weighted Bonferroni procedure controls FWER and can have higher power than the Bonferroni procedure, especially when the optimal Bonferroni weights are used. To further improve the power of the weighted Bonferroni procedure, first we propose a weighted Šidák procedure that incorporates weighted p-values into the Šidák procedure, and then we estimate the optimal weights that maximize the average power of the weighted Šidák procedure under the assumption that the means of the test statistics in the multiple testing are known (these weights are called optimal Šidák weights). This weighted Šidák procedure can have higher power than the weighted Bonferroni procedure. Second, we develop a generalized sequential (GS) Šidák procedure that incorporates weighted p-values into the sequential Šidák procedure (Scherrer, 1984). This GS Šidák procedure is an extension of and has higher power than the GS Bonferroni procedure of Holm (1979). Finally, under the assumption that the means of the test statistics in the multiple testing are known, we incorporate the optimal Šidák weights and the optimal Bonferroni weights into the GS Šidák procedure and the GS Bonferroni procedure, respectively. Theoretical proof and/or simulation studies show that the GS Šidák procedure can have higher power than the GS Bonferroni procedure when their corresponding optimal weights are used, and that both of these GS procedures can have much higher power than the weighted Šidák and the weighted

*We thank the editor and two referees for their helpful comments and useful suggestions. This research was supported by grant GM073766, GM077490, and GM081488 from the National Institute of General Medical Sciences. Address for correspondence: Dr. Guimin Gao, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294. email: ggao@ms.soph.uab.edu. Phone: 205-975-9188.

Bonferroni procedures. All proposed procedures control the FWER well and are useful when prior information is available to estimate the weights.

KEYWORDS: weight, multiple hypothesis testing, Bonferroni procedure, Šidák procedure, family-wise error rate

1 Introduction

Multiple hypothesis testing involves testing multiple hypotheses simultaneously; each hypothesis is associated with a test statistic (Rubin *et al.*, 2006). Multiple hypothesis testing is a common problem in genome research, such as genome-wide studies and gene expression data analysis (Lin, 2005). For multiple hypothesis testing, a traditional criterion for error (type I) control is the family-wise error rate (FWER), which is the probability of rejecting one or more true null hypotheses (Hochberg and Tamhane, 1987; Lin, 2005).

The Bonferroni procedure (Bonferroni, 1937) and the Šidák procedure (Šidák, 1967) are two well-known methods for controlling FWER with computational simplicity and wide applicability (Olejnik *et al.*, 1997). However, both of these methods have limited statistical power as the number of hypotheses tested (m) increases (Nakagawa, 2004). Holm (1979) proposed a (step-down) sequential Bonferroni procedure which has slightly higher power than the Bonferroni procedure but there is little difference between these two procedures when the number of tests (m) is large (Lin, 2005). As an extension of the (step-down) sequential Bonferroni procedure, Holm (1979) proposed a generalized sequential (GS) Bonferroni procedure by using different weights for hypotheses of different importance. Although Holm did not show how to estimate the weights, the method has the potential to improve the power of multiple hypothesis testing when prior information is available to estimate the weights.

Rubin *et al.* (2006) and Wasserman and Roeder (2006) proposed a weighted Bonferroni procedure that adjusts p-values by using optimal weights. These optimal weights were calculated by maximizing the average power of the weighted Bonferroni procedure under the assumption that the means of all test statistics are known, and these weights are called optimal Bonferroni weights. Under such assumption, the average power of the weighted Bonferroni procedure is much higher than that of the Bonferroni procedure (Rubin *et al.*, 2006; Genovese *et al.*, 2006; Wasserman and Roeder, 2006). In practice, the means of the test statistics are unknown. However, if some prior information is available to estimate the means, this weighted Bonferroni procedure can be more powerful than the Bonferroni procedure (Rubin *et al.*, 2006; Wasserman and Roeder, 2006; Roeder *et al.*, 2006; Roeder *et al.*, 2007).

The purpose of this study is to develop more powerful weighted hypothesis testing procedures as extensions of the weighted Bonferroni procedure. First, we propose a weighted Šidák procedure, and then under the assumption that the means of all test statistics are known, we estimate the optimal weights maximizing the average power of the weighted Šidák procedure (these weights are called optimal Šidák weights). The weighted Šidák procedure has slightly higher power than the weighted Bonferroni procedure. Second, we develop a GS

Šidák procedure as an extension of the GS Bonferroni procedure of Holm (1979) and the sequential Šidák procedure (Scherrer, 1984). Finally, assuming that the means of all test statistics are known, we incorporate the optimal Šidák weights and the optimal Bonferroni weights into the GS Šidák procedure and the GS Bonferroni procedure, respectively. Theoretical proof and/or simulation studies show that, using their corresponding optimal weights, the GS Šidák procedure has slightly higher power than the GS Bonferroni procedure, and that both GS Šidák procedure and GS Bonferroni procedure have much higher power than the weighted Šidák procedure and the weighted Bonferroni procedure. All the proposed procedures can control the FWER well.

2 Methods

2.1 Notations

Consider testing m (null) hypotheses $\mathbf{H} = (H_1, H_2, \dots, H_m)$ with corresponding test statistics $\mathbf{Z} = (Z_1, Z_2, \dots, Z_m)$, where we assume that Z_i follows normal distribution of $N(\mu_i, 1)$, and all Z_i 's are independent. For simplicity, in this paper, we only present the results for one-sided tests. Similar results for two-sided tests can easily be obtained. Thus, for the i -th test, the (null) hypothesis is $H_i : \mu_i = 0$, and the corresponding alternative hypothesis is $\bar{H}_i : \mu_i > 0$. Suppose that there are m_1 true null hypotheses and m_2 false null hypotheses among all hypotheses in \mathbf{H} , where $m_2 = m - m_1$. Let \mathbf{H}_0 denote all the true null hypotheses in \mathbf{H} . Let $\mathbf{p} = (p_1, p_2, \dots, p_m)$ denote the p-values associated with the hypotheses (H_1, H_2, \dots, H_m) . Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_m)$ denote the means of the test statistics \mathbf{Z} .

As described earlier, FWER is the probability of falsely rejecting at least one true null hypothesis (Hochberg and Tamhane, 1987), which can be written as

$$\text{FWER} = \Pr(\text{rejecting at least one } H_i \mid H_i \in \mathbf{H}_0).$$

A multiple testing procedure is said to control the family-wise error rate at a significance level α if $\text{FWER} \leq \alpha$.

The power for a single test is called *per-hypothesis* power. For a single test with hypothesis H_i , the *per-hypothesis* power is the probability of rejecting H_i given that the alternative hypothesis \bar{H}_i is true, i.e., $\Pr(\text{rejecting } H_i \mid \mu_i > 0)$. For multiple hypotheses testing, Roeder *et al.* (2007) defined the average power of a

testing procedure as the average value of *per-hypothesis* powers of the m_2 tests associated with the false null hypotheses: $\frac{1}{m_2} \sum_{i: \mu_i > 0} \Pr(\text{rejecting } H_i \mid \mu_i > 0)$.

2.2 Weighted Bonferroni procedure and optimal Bonferroni weights

2.2.1 Weighted Bonferroni procedure

In the Bonferroni procedure, if $p_j \leq \frac{\alpha}{m}$, then reject the null hypothesis H_j ; otherwise, it is failed to reject H_j ($j = 1, \dots, m$). The power of multiple testing procedures can be increased by using weighted p-values (Genovese *et al.*, 2006). Holm appears to be the first one proposing the idea of the weighted Bonferroni procedure, which incorporates weighted p-values into the Bonferroni procedure (Holm, 1979). Wasserman and Roeder (2006) provided a clear description of the weighted Bonferroni procedure as follows. Given nonnegative weights (w_1, w_2, \dots, w_m) for the tests associated with the hypotheses (H_1, H_2, \dots, H_m) , where

$$\frac{1}{m} \sum_{j=1}^m w_j = 1. \quad (1)$$

For hypothesis H_j ($1 \leq j \leq m$), when $w_j > 0$, reject H_j if $\frac{p_j}{w_j} \leq \frac{\alpha}{m}$, and fail to reject H_j when $w_j = 0$.

This procedure controls FWER at level α . The weights (w_1, w_2, \dots, w_m) can be specified by using certain prior information available to the researcher. For example, in genome-wide association studies, the prior information can be linkage signals or results from gene expression analyses. Roeder *et al.* (2006) proposed a method to estimate weights by using linkage data to weight association p-values in association studies. However, how to estimate the optimal weights in multiple testing is still a topic to be further investigated (see also the section on discussion).

2.2.2 Optimal Bonferroni weights

Rubin *et al.* (2006) and Wasserman and Roeder (2006) independently proposed very similar approaches to estimate the optimal weights by maximizing the average power of the procedure, assuming that the means $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_m)$ are known. We call these optimal weights optimal Bonferroni weights and they are calculated (Wasserman and Roeder, 2006) by

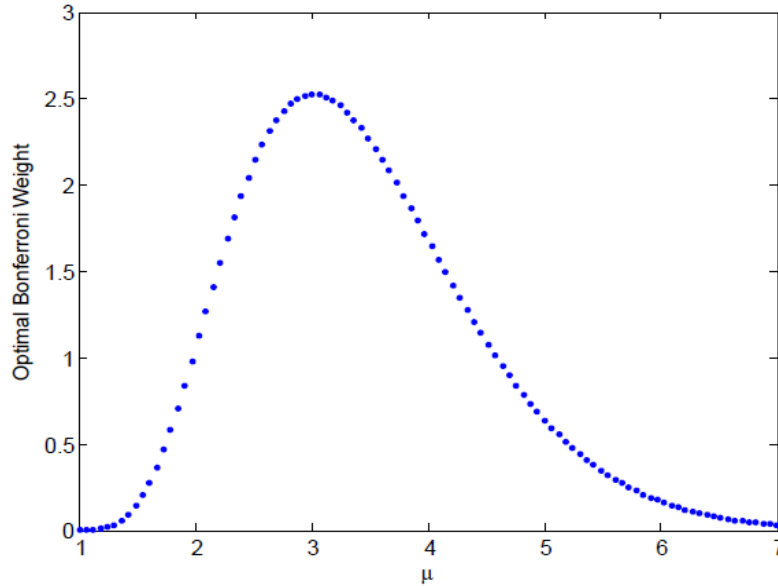
$$w_j = \frac{m}{\alpha} \bar{\Phi} \left(\frac{\mu_j}{2} + \frac{\Delta}{\mu_j} \right) I(\mu_j > 0), \quad (2)$$

where $\bar{\Phi}(x)$ is the upper tail probability of a standard normal cumulative distribution function (CDF) (i.e., $\bar{\Phi}(x) = 1 - \Phi(x)$ and $\Phi(x)$ denotes the CDF of the standard normal distribution) and Δ is the constant that satisfies equations (1) and (2) i.e.

$$\frac{1}{m} \sum_{j=1}^m \frac{m}{\alpha} \bar{\Phi} \left(\frac{\mu_j}{2} + \frac{\Delta}{\mu_j} \right) I(\mu_j > 0) = \alpha. \quad (3)$$

As an illustrative example, Figure 1 shows the optimal Bonferroni weights as a function of the means μ_j in a multiple testing with 100 tests. The means μ vary from 1 to 7 in increment of $6/99 = 0.0606$. When the means μ_j are small, the optimal weights increase with the increase of μ_j but when μ_j are large enough, the optimal weights decrease with the increase of μ_j . In other words, the weighted Bonferroni procedure offers large weights (often > 1) to the tests with midrange of means and offers small weights (often < 1) to tests with small or large means (Wasserman and Roeder, 2006). Dividing the p-value by a weight $w > 1$ increases the probability of rejecting the corresponding null hypothesis, and dividing the p-value by a weight $0 < w < 1$ decreases the probability of rejecting the corresponding null hypothesis. However, in most situations, even though the tests with large means are assigned small weights (< 1), the corresponding hypotheses can still be rejected because the related p-values are very small. The weighted Bonferroni procedure using these optimal weights can have much higher power than the Bonferroni procedure when the means ($\boldsymbol{\mu}$) of the test statistics are given or given prior information that can be used for estimating the means (Roeder *et al.*, 2006, 2007; Rubin *et al.*, 2006; Wasserman and Roeder, 2006).

Figure 1. Distribution of optimal Bonferroni weights for a multiple testing procedure with $m = 100$ tests. μ are the means of the test statistics, and vary from 1 to 7.



2.3 Weighted Šidák procedure and optimal Šidák weights

Since the Šidák procedure has higher power than the Bonferroni procedure for independent tests (Simes, 1986), we propose a weighted Šidák procedure that incorporates weighted p-values into the Šidák procedure as an extension of the weighted Bonferroni procedure. We also describe how to calculate the optimal weights for the weighted Šidák procedure assuming means of the test statistics are known.

2.3.1 Weighted Šidák procedure

In the Šidák procedure (Šidák, 1967), for any null hypothesis H_j ($1 \leq j \leq m$), if

$p_j \leq 1 - (1 - \alpha)^{\frac{1}{m}}$, then reject H_j . The Šidák procedure controls the FWER at level α . Now we propose a *weighted Šidák procedure* by using weighted p-values as follows: given a set of nonnegative weights (w_1, w_2, \dots, w_m) specified for independent tests associated with the hypotheses (H_1, H_2, \dots, H_m) such that equation (1) holds (i.e., $m^{-1} \sum_i w_i = 1$), for hypothesis H_j ($1 \leq j \leq m$), when $w_j > 0$, if

$$p_j \leq 1 - (1 - \alpha)^{\frac{w_j}{m}} \text{ or equivalently } (1 - p_j)^{\frac{1}{w_j}} \geq (1 - \alpha)^{\frac{1}{m}}, \quad (4)$$

then reject the null hypothesis H_j ; on the other hand, when $w_j = 0$, do not reject H_j . In this article we denote the weighted p-value $(1 - p_j)^{\frac{1}{w_j}}$ as S_j . Therefore, (4) can be written as $S_j \geq (1 - \alpha)^{\frac{1}{m}}$.

Theorem 1. *Suppose m tests are independent, then the weighted Šidák procedure controls the family-wise error rate at a significance level α .*

Proof. P(failing to reject any true null hypotheses in \mathbf{H}_0)

$$= \prod_{j: H_j \in \mathbf{H}_0} \mathbb{P}(p_j > 1 - (1 - \alpha)^{w_j/m}) = \prod_{j: H_j \in \mathbf{H}_0} (1 - \alpha)^{w_j/m} = (1 - \alpha)^{\sum_{j: H_j \in \mathbf{H}_0} w_j/m} = 1 - \alpha,$$

where, p_j follows standard uniform distribution when $H_j \in \mathbf{H}_0$. Since FWER = $1 - \mathbb{P}(\text{failing to reject any true null hypotheses in } \mathbf{H}_0)$, then Theorem 1 follows. ■

From the Taylor series expansions, we obtain

$$\frac{\alpha}{m} w_j \leq 1 - (1 - \alpha)^{\frac{w_j}{m}}.$$

Based on this inequality, when the same pre-determined weights (w_1, w_2, \dots, w_m) are used by the weighted Šidák procedure and the weighted Bonferroni procedure, if any hypothesis H_j is rejected by the weighted Bonferroni procedure (i.e.,

$p_j \leq \frac{\alpha}{m} w_j$), then it must be rejected by the weighted Šidák procedure (i.e.,

$p_j \leq 1 - (1 - \alpha)^{\frac{w_j}{m}}$). Thus, we have Theorem 2.

Theorem 2. *For m independent tests, if the same pre-determined weights (w_1, w_2, \dots, w_m) are used in the weighted Šidák procedure and the weighted Bonferroni procedure, then the weighted Šidák procedure has higher average power than the weighted Bonferroni procedure.*

Remark 1. If all weights $w_j = 1$, then the weighted Šidák procedure becomes the Šidák procedure. The weighted Šidák procedure can have higher power than the Šidák procedure if the weights are selected appropriately.

2.3.2 Optimal Šidák weights

As stated earlier, how to estimate optimal weights by using the prior information still needs further investigation. Here, we derive the optimal weights that maximize the average power of the weighted Šidák procedure under the assumption that the means $(\mu_1, \mu_2, \dots, \mu_m)$ are known. These optimal weights are called optimal Šidák weights, which is an extension of the optimal Bonferroni weights of Wasserman and Roeder (2006).

For any specified weights (w_1, w_2, \dots, w_m) , the per-hypothesis power for the single test with hypothesis H_j in the weighted Šidák procedure is

$$Power_j = P\left(p_j < 1 - (1 - \alpha)^{\frac{w_j}{m}} \mid \mu_j > 0\right) = \bar{\Phi}\left(\bar{\Phi}^{-1}\left(1 - (1 - \alpha)^{\frac{w_j}{m}}\right) - \mu_j\right).$$

The average power of the weighted Šidák procedure is

$$PW_{\text{average}} = \frac{1}{m_2} \sum_{j: \mu_j > 0} \bar{\Phi}\left(\bar{\Phi}^{-1}\left(1 - (1 - \alpha)^{\frac{w_j}{m}}\right) - \mu_j\right).$$

To find the optimal weights that maximize this average power subject to constraint of equation (1), Lagrange method was used to obtain conditional extremum of PW_{average} .

Theorem 3. Given FWER being α and known means $(\mu_1, \mu_2, \dots, \mu_m)$ of the m independent test statistics (Z_1, Z_2, \dots, Z_m) , the optimal non-negative weights (w_1, w_2, \dots, w_m) that maximize the average power of the weighted Šidák procedure subject to constraint of equation (1) can be obtained by solving inequalities $w_i \geq 0$, equations (1) and

$$c - \frac{w_i}{m} \ln(1 - \alpha) = \mu_i \bar{\Phi}^{-1}\left(1 - (1 - \alpha)^{w_i/m}\right) - \frac{\mu_i^2}{2}, \text{ for } i=1, \dots, m, \quad (5)$$

where c is a constant (given in Appendix A).

The proof of this theorem is given in Appendix A. The inequalities and equations can be solved by using the “**nlminb()**” function in R.

In the following simulation studies we will show that the weighted Šidák procedure using the optimal Šidák weights can have higher power than the weighted Bonferroni procedure using the optimal Bonferroni weights and that the weighted Šidák procedure using the optimal Šidák weights can have much higher power than the Šidák procedure.

2.4 GS Bonferroni procedure and GS Šidák procedure

Holm (1979) introduced a GS Bonferroni procedure that is a step-down procedure using ordered weighted p-values. If the (unknown) weights used in the procedure are estimated appropriately by using prior information, the procedure can have higher power than the weighted Bonferroni procedure (also see below). In this section, we first review this GS Bonferroni procedure, and then we propose a GS Šidák procedure as an extension of the GS Bonferroni procedure.

When assuming that the means of the statistics are known, it is difficult to derive the optimal weights by maximizing the average power of these GS procedures as done before for the weighted Bonferroni and the weighted Šidák procedures. We incorporate the optimal Bonferroni (Šidák) weights described in Section 2.2 (2.3) into the GS Bonferroni (Šidák) procedure. We will show below that when these optimal weights are used, the GS Bonferroni (Šidák) procedure has higher power than the weighted Bonferroni (Šidák) procedure.

2.4.1 GS Bonferroni procedure

Given nonnegative weights (w_1, w_2, \dots, w_m) for the m tests associated with hypotheses (H_1, H_2, \dots, H_m) , (note that it is not necessary to satisfy the condition $m^{-1} \sum_{i=1}^m w_i = 1$), if any weight $w_i = 0$, then do not reject the corresponding

hypothesis H_i . For the remaining hypotheses with weights $w_i > 0$, define $B_i = \frac{p_i}{w_i}$ ($i = 1, 2, \dots, m$), which are called B -values (i.e., weighted p-values). Let $B_{(1)} \leq B_{(2)} \leq \dots \leq B_{(m)}$ be the ordered B -values, $H_{(1)}, H_{(2)}, \dots, H_{(m)}$ be the corresponding hypotheses and $w_{(1)}, w_{(2)}, \dots, w_{(m)}$ be the corresponding weights. Then the GS Bonferroni procedure (Holm, 1979) can be described as follows:

Step 1. If $B_{(1)} > \frac{\alpha}{\sum_{i=1}^m w_{(i)}}$, stop the procedure; otherwise, reject $H_{(1)}$ and go to the next step.

...

Step j . If $B_{(j)} > \frac{\alpha}{\sum_{i=j}^m w_{(i)}}$, stop the procedure; otherwise, reject $H_{(j)}$ and go to the next step.

....

Continue these steps until the procedure is stopped or all B -values have been processed.

This procedure controls FWER at level α . If we set all weights w_j equal to 1, the inequality $B_{(j)} > \frac{\alpha}{\sum_{i=j}^m w_{(i)}}$ in step j becomes $B_{(j)} > \frac{\alpha}{m-j+1}$ and the GS

Bonferroni procedure becomes the sequential Bonferroni procedure (Holm, 1979). This GS Bonferroni procedure can have higher power than the sequential Bonferroni procedure when the weights are chosen properly (Holm, 1979).

Now we compare the power of the GS Bonferroni procedure and the weighted Bonferroni procedure when the same pre-determined weights are used in these two procedures. For pre-specified weight $(w_{(1)}, w_{(2)}, \dots, w_{(m)})$ associated with hypotheses $(H_{(1)}, H_{(2)}, \dots, H_{(m)})$ such that $m^{-1} \sum_{j=1}^m w_j = 1$ (i.e., $m^{-1} \sum_{j=1}^m w_{(j)} = 1$), if any false hypothesis $H_{(j)}$ is rejected by the weighted Bonferroni procedure, that is, $B_{(j)} \leq \frac{\alpha}{m}$ is true, then $B_{(j)} \leq \frac{\alpha}{\sum_{i=j}^m w_{(i)}}$. Since $\sum_{i=j}^m w_{(i)} \leq m$, we have

$$B_{(1)} \leq B_{(2)} \leq \dots \leq B_{(j)} \leq \frac{\alpha}{m} \leq \frac{\alpha}{\sum_{i=j}^m w_{(i)}}.$$

Thus, $H_{(j)}$ will also be rejected by the GS Bonferroni procedure. Therefore, we have Theorem 4.

Theorem 4. Given weights (w_1, w_2, \dots, w_m) for m independent tests such that $m^{-1} \sum_{i=1}^m w_i = 1$, the GS Bonferroni procedure has higher average power than the weighted Bonferroni procedure.

2.4.2 GS Bonferroni procedure using the optimal Bonferroni weights

As stated earlier, it is difficult to estimate the optimal weights that maximize the average power of the GS Bonferroni procedure under the assumption that the means of statistics are known. Here we propose to use the optimal Bonferroni weights described in Section 2.2. When these optimal Bonferroni weights are used, from Theorem 4, we know that the GS Bonferroni procedure has higher average power than the weighted Bonferroni procedure. Our simulation studies will confirm this.

2.4.3 GS Šidák procedure

The GS Bonferroni procedure is based on the Bonferroni procedure. As stated earlier, the Šidák procedure has higher power than the Bonferroni procedure. Therefore, we propose a GS Šidák procedure.

Given nonnegative weights (w_1, w_2, \dots, w_m) for the m tests associated with hypotheses (H_1, H_2, \dots, H_m) (note that it is not necessary to satisfy the condition $m^{-1} \sum_{i=1}^m w_i = 1$), if any weight $w_i = 0$, do not reject the corresponding null hypothesis H_i . For the remaining hypotheses with weights $w_i > 0$, let $S_i = (1 - p_i)^{\frac{1}{w_i}}$ ($i = 1, 2, \dots, m$) which are called S-values (i.e., weighted p-values). Let $S_{(1)} \geq S_{(2)} \geq \dots \geq S_{(m)}$ be the ordered S-values, $H_{(1)}, H_{(2)}, \dots, H_{(m)}$ be the corresponding hypotheses, and $w_{(1)}, w_{(2)}, \dots, w_{(m)}$ be the corresponding weights. The GS Šidák procedure can be described as the following steps.

Step 1. If $S_{(1)} < (1 - \alpha)^{\frac{1}{\sum_{i=1}^m w_{(i)}}}$, then stop the procedure; otherwise reject $H_{(1)}$ and go to the next step.

.....

Step j . When $H_{(1)} \dots, H_{(j-1)}$ have been tested and rejected: if

$$S_{(j)} < (1 - \alpha)^{\frac{1}{\sum_{i=j}^m w_{(i)}}}, \quad (6)$$

stop the procedure; otherwise reject the hypothesis $H_{(j)}$, and go to the next step.

...

Continue these steps until the procedure is stopped or all S-values have been processed.

Theorem 5. *Suppose m tests are independent, then the GS Šidák procedure controls family-wise error rate at a significant level α .*

Proof. Let \mathbf{I}_0 be the set of index subscripts for the true null hypotheses, $\mathbf{I}_0 = \{t: H_t \in \mathbf{H}_0\}$. Let $S_{\mathbf{I}_0}^l = \max_{t \in \mathbf{I}_0} S_t$ denote the largest S-value among all S_t with $t \in \mathbf{I}_0$. Among the ordered S-values, $S_{(1)} \geq S_{(2)} \geq \dots \geq S_{(m)}$, suppose at integer k , $S_{(k)} = S_{\mathbf{I}_0}^l$, then $S_{(k)}$ is first ordered S-value (from large to small) which is corresponding to a true null hypothesis in \mathbf{H}_0 (i.e., all the previous $k-1$ ordered S-values $S_{(1)}, \dots, S_{(k-1)}$ are corresponding to false null hypothesis), where, $1 \leq k \leq m - m_1 + 1$, and m_1 is the number of true hypotheses. According to the GS Šidák procedure, the event of failing to reject any true null hypotheses in \mathbf{H}_0 is equal to event that equation (6) holds for some $j < k$. The family-wise error rate is $\text{FWER} = 1 - \text{P}(\text{failing to reject any true null hypotheses in } \mathbf{H}_0)$, and

$\text{P}(\text{failing to reject any true null hypotheses in } \mathbf{H}_0)$

$$\begin{aligned} &= P\left(\bigcup_{j=1}^k (S_{(j)} < (1-\alpha)^{1/\sum_{i=j}^m w_{(i)}})\right) \geq P\left(S_{(k)} < (1-\alpha)^{1/\sum_{i=k}^m w_{(i)}}\right) \\ &= \prod_{t \in \mathbf{I}_0} P\left(S_t < (1-\alpha)^{1/\sum_{i=k}^m w_{(i)}}\right) = \prod_{t \in \mathbf{I}_0} P\left((1-p_t)^{1/w_t} < (1-\alpha)^{1/\sum_{i=k}^m w_{(i)}}\right) \\ &= \prod_{t \in \mathbf{I}_0} P\left((1-p_t) < (1-\alpha)^{w_t/\sum_{i=k}^m w_{(i)}}\right) = (1-\alpha)^{\sum_{t \in \mathbf{I}_0} w_t/\sum_{i=k}^m w_{(i)}} > 1-\alpha, \end{aligned}$$

where $\sum_{t \in \mathbf{I}_0} w_t / \sum_{i=k}^m w_{(i)} \leq 1$, and $1 - p_t$ follows uniform distribution when $t \in \mathbf{I}_0$. ■

Now we compare the power of the GS Šidák procedure to that of the weighted Šidák procedure when both procedures use the same weights w_j that satisfy $m^{-1} \sum_{j=1}^m w_j = 1$ (i.e., $m^{-1} \sum_{j=1}^m w_{(j)} = 1$). If $H_{(j)}$ is rejected by the weighted Šidák

procedure, that is, $S_{(j)} \geq (1 - \alpha)^{\frac{1}{m}}$ is true (see inequality (4)), then $S_{(j)} \geq (1 - \alpha)^{\frac{1}{\sum_{i=j}^m w_{(i)}}}$. Since $\sum_{i=j}^m w_{(i)} \leq m$, we have

$$S_{(1)} \geq S_{(2)} \geq \dots \geq S_{(j)} \geq (1 - \alpha)^{1/m} \geq (1 - \alpha)^{\frac{1}{\sum_{i=j}^m w_{(i)}}}.$$

Thus, $H_{(j)}$ will also be rejected by the GS Šidák procedure, and we have Theorem 6.

Theorem 6. *Given weights (w_1, w_2, \dots, w_m) for m independent tests that satisfy $m^{-1} \sum_{j=1}^m w_j = 1$, then the GS Šidák procedure has higher power than the weighted Šidák procedure.*

Furthermore, we compare the power of the GS Šidák procedure to that of the GS Bonferroni procedure when the same pre-specified weights are used in these two procedures.

Theorem 7. *For m independent tests, if the same weights (w_1, w_2, \dots, w_m) are used in the GS Šidák procedure and GS Bonferroni procedure, then the GS Šidák procedure has higher average power than the GS Bonferroni procedure.*

Proof. From the definition of B-value ($B_i = p_i / w_i$) and S-value ($S_i = (1 - p_i)^{1/w_i}$), we know B_i (S_i) is a monotonically decreasing (increasing) function of w_i . Suppose that $B_{(1)} \leq B_{(2)} \leq \dots \leq B_{(m)}$ and $S_{(1)} \geq S_{(2)} \geq \dots \geq S_{(m)}$ are associated with the same hypotheses $H_{(1)}, H_{(2)}, \dots, H_{(m)}$. Below we show that for

any hypothesis $H_{(j)}$, if $B_{(j)} \leq \alpha / \sum_{i=j}^m w_{(i)}$, then $S_{(j)} \geq (1 - \alpha)^{\frac{1}{\sum_{i=j}^m w_{(i)}}}$, from which we know that the GS Šidák procedure has higher power than the GS Bonferroni procedure.

If $B_{(j)} \leq \alpha / \sum_{i=j}^m w_{(i)}$, from the Taylor series expansions, we have

$$p_{(j)} \leq \alpha w_{(j)} / \sum_{i=j}^m w_{(i)} \leq 1 - (1 - \alpha)^{w_{(j)} / \sum_{i=j}^m w_{(i)}}.$$

Thus, $S_{(j)} = (1 - p_i)^{1/w_i} \geq (1 - \alpha)^{1/\sum_{i=j}^m w_{(i)}}.$ ■

Remark 2. If setting all the weights equal (to 1), then the GS Šidák procedure becomes the sequential Šidák procedure (Scherrer, 1984).

2.4.4 GS Šidák procedure using the optimal Šidák weights

In the GS Šidák procedure, a major issue is how to calculate the weights. As we stated before, it is difficult to derive optimal weights that maximize the average power of the GS Šidák procedure under the assumption that the means of the statistics are known. Here, under this assumption, we suggest using the optimal Šidák weights calculated by equation (5). From Theorem 7, the GS Šidák procedure has higher average power than the weighted Šidák procedure when the optimal Šidák weights are used by these two procedures.

We will show that the GS Šidák procedure using the optimal Šidák weights has higher power than the GS Bonferroni procedure using the optimal Bonferroni weights by simulation studies (see below). It appears to be difficult to prove this statement theoretically because the optimal Šidák weights are not the same as the Bonferroni weights.

3 Simulation studies and results

To further evaluate the performance of the proposed testing procedures, we compared by simulation studies the average power of six multiple testing procedures: the Šidák procedure, the Bonferroni procedure, the weighted Šidák (Bonferroni) procedure using the optimal Šidák (Bonferroni) weights, and the GS Šidák (Bonferroni) procedure using the optimal Šidák (Bonferroni) weights.

3.1 Assuming true means $\mu = (\mu_1, \mu_2, \dots, \mu_m)$ known

When we assume that the means of statistics are known, for each true null hypothesis $H_j: \mu_j = 0$, the weight is assigned to zero in each procedure using weights. Thus, all true null hypotheses will not be rejected in these procedures. In other words, the FWER is equal to zero in the simulation studies with known

means of statistics. Thus, we only compare the average power among these six procedures.

We simulated datasets in a similar way to Rubin *et al.* (2006). Each simulated dataset $\mathbf{X} = (X_{i,j})_{n \times m}$ consisted of $n = 100$ i.i.d observations, where each observation corresponded to a subject and was a vector of measurements of $m = 1,000$ independent covariates, and $X_{i,j}$ was a measurement of covariate j at the i -th observation. For each covariate j , we assumed that $X_{i,j} \sim N(\gamma_j, 1)$, $i = 1, \dots, n$, and we implemented a test with statistic $Z_j = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,j}$ to test the null hypothesis $\gamma_j = 0$ against the alternative $\gamma_j > 0$. Using the central limit theorem, Z_j follows an asymptotic distribution of $N(u_j, 1)$, where $u_j = \sqrt{n}\gamma_j$ (Rubin *et al.*, 2006). Thus, testing hypothesis $\gamma_j = 0$ is equivalent to testing hypothesis $H_j: u_j = 0$.

When generating each dataset that was associated with 1,000 covariates, we randomly chose 50 covariates and set the means $\gamma_j > 0$ for these 50 covariates (i.e., $\mu_j > 0$ for the corresponding test statistic Z_j), and set $\gamma_j = 0$ for the other 950 covariates. In our simulation studies, we considered two scenarios for the 50 non-zero γ_j . In Scenario 1, we set the 50 non-zero γ_j equal to a common value γ that varies as a simulation parameter. We considered γ between 0.1 and 0.5, in increments of 0.1 (correspondingly, $u = \sqrt{n}\gamma$ are between 1 and 5, in increments of 1). We simulated 1,000 datasets corresponding to each of these γ values. In Scenario 2, for the 50 non-zero γ_j , we set the first ten $\gamma_j = 0.1$ ($\mu_j = 1$), the second ten $\gamma_j = 0.2$ ($\mu_j = 2$), ..., and the fifth ten $\gamma_j = 0.5$ ($\mu_j = 5$). We simulated 1,000 datasets for Scenario 2.

Table 1 shows the results of the estimated average power of the six multiple testing procedures in Scenarios 1 and 2. From Table 1, we can see that the GS Šidák, weighted Šidák, and Šidák procedures have slightly higher estimated average power than the corresponding GS Bonferroni, weighted Bonferroni, and Bonferroni procedures, and that the GS Šidák procedure is most powerful among the six procedures. The GS Šidák procedure and the GS Bonferroni procedure can have much higher power than both the weighted Šidák procedure and the weighted Bonferroni procedure. For example, in Scenario 1, when $\mu = 3$, the estimated average power of the GS Šidák procedure, GS Bonferroni procedure, the weighted Šidák procedure, and the weighted Bonferroni procedure is 0.5820, 0.5792, 0.4670 and 0.4639, respectively (see Table 1).

Table 1. The estimated average power of the six multiple testing procedures over 1,000 replicated data sets when given the means of the test statistics. ($\alpha = 0.05$, $m_2 = 50$, $m = 1000$ and $n = 100$).

Scenario	μ	Bonf ^a	Šidák	W-Bonf ^b	W-Šidák ^c	GS-Bonf ^d	GS-Šidák ^e
Scenario	1	0.0019	0.0020	0.0183	0.0186	0.0453	0.0459
	2	0.0293	0.0298	0.1378	0.1394	0.2253	0.2274
	3	0.1869	0.1886	0.4639	0.4670	0.5792	0.5820
	4	0.5436	0.5460	0.8185	0.8205	0.8781	0.8796
	5	0.8664	0.8677	0.9719	0.9724	0.9838	0.9841
Scenario		0.3256	0.3267	0.4988	0.5002	0.5419	0.5432

^a Bonf = Bonferroni; ^b W-Bonf = Weighted Bonferroni; ^c W-Šidák = Weighted Šidák; ^d GS-Bonf = Generalized sequential Bonferroni; ^e GS-Šidák = Generalized sequential Šidák

3.2 True means $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_m)$ unknown

In previous sections, all weights are calculated under the assumption that the means $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_m)$ of statistics are known. However, in real data analysis, the means $\boldsymbol{\mu}$ are usually unknown. The means $\boldsymbol{\mu}$ can be estimated by using certain prior information. How to effectively estimate $\boldsymbol{\mu}$ is still a topic to be further investigated. Rubin *et al.* (2006) described a data-splitting method, which splits the full data \mathbf{X} into two parts, \mathbf{X}_1 and \mathbf{X}_2 , with proportion π and $1 - \pi$ of the original data \mathbf{X} , respectively. Data \mathbf{X}_1 is used to estimate the means of the standardized test statistics and data \mathbf{X}_2 is used to test the hypotheses. They showed that if some prior information, such as the order of means $(\mu_1, \mu_2, \dots, \mu_m)$, is available, by using the data-splitting method the weighted Bonferroni procedure has higher power than the Bonferroni procedure. In genetic association studies, Roeder *et al.* (2007) described a two-step approach to estimate the means $(\mu_1, \mu_2, \dots, \mu_m)$ by using prior information such as reported linkage peaks, results of previously genome wide association studies, or results of gene expression studies.

It is beyond of the scope of this study to determine how to effectively estimate the means $(\mu_1, \mu_2, \dots, \mu_m)$ by using prior information. To show the performance of our proposed procedures when estimated means are used, as an example, we implemented our proposed procedures by incorporating the data-splitting method and applied these methods to the simulated Scenarios 1-2 data sets described in the previous section. The only exception is that we assume here that the first 950 covariates have means equal to zero and the last 50 covariates

have the common mean value $\gamma > 0$. This follows the assumption of Rubin *et al.* (2006) that the order of means $(\mu_1, \mu_2, \dots, \mu_m)$ is known.

Table 2. The estimated FWERs and average power of the six multiple procedures over 1,000 replicated data sets when the means of the test statistics are unknown ($\alpha = 0.05$, $m = 1000$, $m_2 = 50$, $n = 100$, and $\pi = 0.1$ for data-splitting).

Scenario	μ	Bonf ^a	Šidák	W-Bonf ^b	W-Šidák ^c	GS-Bonf ^d	GS-Šidák ^e	
Scenario 1	1	FWER	0.0610	0.0650	0.0180	0.0180	0.0180	0.0180
		Power	0.0017	0.0017	0.0091	0.0093	0.0092	0.0095
	2	FWER	0.0360	0.0370	0.0020	0.0020	0.0020	0.0020
		Power	0.0278	0.0283	0.0996	0.1009	0.1071	0.1085
	3	FWER	0.0410	0.0430	0.0010	0.0010	0.0030	0.0030
		Power	0.1835	0.1854	0.3766	0.3795	0.4523	0.4564
	4	FWER	0.0480	0.0500	0.0030	0.0030	0.0260	0.0270
		Power	0.5422	0.5445	0.7427	0.7449	0.8707	0.8719
	5	FWER	0.0490	0.0490	0.0110	0.0110	0.0530	0.0540
		Power	0.8746	0.8715	0.9342	0.9350	0.9772	0.9780
Scenario 2	FWER	0.0610	0.0630	0.0010	0.0010	0.0010	0.0010	
	Power	0.3239	0.3251	0.4566	0.4581	0.5247	0.5262	

^a Bonf = Bonferroni; ^b W-Bonf = Weighted Bonferroni; ^c W-Šidák = Weighted Šidák; ^d GS-Bonf = Generalized sequential Bonferroni; ^e GS-Šidák = Generalized sequential Šidák

For each simulated dataset, the Bonferroni procedure and the Šidák procedure were implemented on the entire dataset, while the other four procedures used the data-splitting method (under the assumption that the order of means $(\mu_1, \mu_2, \dots, \mu_m)$ is known). Table 2 shows the estimated average power and family-wise error rates of the six procedures for Scenarios 1 and 2. We only show the results with the proportion π of the first part \mathbf{X}_1 equal to 0.1.

From Table 2, we can find that the weighted Šidák and the GS Šidák procedures have slightly higher estimated average power than the weighted Bonferroni and the GS Bonferroni procedures, respectively, and that the GS Šidák procedure has the highest estimated average power among these six procedures. For example, when μ is equal to 4, the estimated average power of the GS Šidák procedure is 0.8719. It is nearly 13% more than that of the weighted Bonferroni procedure (0.7427). In addition, it is interesting that the estimated average power of the six procedures is smaller than their estimated FWERs when μ is equal to 1. This occurs because the average power is the average (not cumulative value) of

per-hypothesis powers for the 50 false null hypotheses, and the FWER is a cumulative value (not average) of type I error rates for 950 tests.

From Table 2, we can also find that the six procedures can control FWERs quite well. Interestingly, the estimated FWERs are much lower in the four procedures using weights (i.e. the weighted Bonferroni, weighted Šidák, GS Bonferroni and GS Šidák) than in the two procedures without using weights (Bonferroni and Šidák). The reason is that the four weighted procedures used the prior information of the order of means of the test statistics.

4 Discussion

In this article, we propose a weighted Šidák procedure and a GS Šidák procedure for multiple hypotheses testing based on the weighted Bonferroni procedure. Under the assumption that the means of the test statistics are known, we further describe how to estimate the optimal Šidák weights which maximize the average power of the weighted Šidák procedure. We show that the weighted Šidák procedure using the optimal Šidák weights can have higher power than the weighted Bonferroni procedure using the optimal Bonferroni weights. Furthermore, we incorporate the optimal Šidák (Bonferroni) weights into the GS Šidák (Bonferroni) procedure. Using these optimal weights the GS Šidák (Bonferroni) procedures can have higher power than the corresponding weighted Šidák (Bonferroni) procedures, respectively, and the GS Šidák procedure often has the highest power among these procedures.

For the multiple procedures using weights described in this article, how to estimate the weights (w_1, w_2, \dots, w_m) by using prior information is still an open problem. Several investigations have been reported in the literature. Roeder *et al.* (2006) used linkage data to estimate weights and adjust p-values in genome-wide association studies. Ionita-Laza *et al.* (2007) used between-family information to estimate weights and weighted association p-values calculated by use of within-family information in family-based genome-wide association studies.

It appears that the optimal Šidák weights and optimal Bonferroni weights have better property than the weights described in the previous paragraph because these optimal weights are based on maximizing the average power of the procedures. However, the optimal Šidák weights and optimal Bonferroni weights are calculated assuming that the means of test statistics are known, and in practice, these means are unknown. The means of test statistics may be estimated by using prior information (Roeder *et al.*, 2007). When certain prior information is available to estimate the means of statistics, the procedures proposed in this paper are useful and can have much higher power than the widely used Bonferroni procedure. However, how to use prior information to estimate the

optimal Šidák weights and optimal Bonferroni weights is still a challenge. We will pursue studies on this topic in the future.

Most of the proposed methods focus on the normal distribution model and one-sided tests. It is trivial to modify the formulas to handle two-sided tests for normal distribution and χ^2 distribution. All the proposed methods assume independence among the multiple tests. This assumption is very conservative. In a real data analysis, multiple tests are often highly correlated. For example, in genome-wide association studies, the tests for different markers may be correlated due to linkage disequilibrium among the markers (Conneely and Boehnke, 2007; Nyholt, 2004). How to extend our proposed method to account for correlation among tests is another issue we will pursue in the future.

All the proposed methods focus on the control of the family-wise error rate for multiple testing. However, a similar idea can be applied to control false discovery rate by using weighed p-value (see also Genovese *et al.*, 2006).

Appendix A. Proof of Theorem 3.

Proof. For the m independent test statistics (Z_1, Z_2, \dots, Z_m) , we estimate the optimal weights w_j that maximize the average power with the constraint $\sum_{j=1}^m w_j = m$. We set $w_j = 0$ if $\mu_j = 0$. For the remaining test statistics with $\mu_j > 0$, the corresponding Lagrange function is

$$G(\lambda, \mathbf{w}) = \frac{1}{m_2} \sum_{j: \mu_j > 0} \bar{\Phi}(\bar{\Phi}^{-1}(1 - (1 - \alpha)^{w_j/m}) - \mu_j) - \lambda \left(m - \sum_{j: \mu_j > 0} w_j \right).$$

By setting the derivatives, with respect to w_i , for $i = 1, 2, \dots, m$, to zero,

$$\frac{\partial G(\lambda, \mathbf{w})}{\partial w_i} = -(1 - \alpha)^{w_i/m} \frac{\phi(\bar{\Phi}^{-1}(1 - (1 - \alpha)^{w_i/m}) - \mu_i)}{\phi(\bar{\Phi}^{-1}(1 - (1 - \alpha)^{w_i/m}))} \frac{1}{mm_2} \ln(1 - \alpha) + \lambda = 0,$$

that is

$$\frac{\lambda mm_2}{\ln(1 - \alpha)} = \frac{\phi(\bar{\Phi}^{-1}(1 - (1 - \alpha)^{w_i/m}) - \mu_i)}{\phi(\bar{\Phi}^{-1}(1 - (1 - \alpha)^{w_i/m}))} (1 - \alpha)^{w_i/m}, \quad (\text{A.1})$$

where $\phi(x)$ is the probability density function of the standard normal distribution. From (A.1), we have

$$\frac{\lambda mm_2}{\ln(1-\alpha)} = \exp\left(\mu_i \bar{\Phi}^{-1}\left(1 - (1-\alpha)^{w_i/m}\right) - \frac{\mu_i^2}{2}\right) (1-\alpha)^{w_i/m}.$$

Taking logarithm on both sides, we obtain

$$\ln\left(\frac{\lambda mm_2}{\ln(1-\alpha)}\right) - \frac{w_i}{m} \ln(1-\alpha) = \mu_i \bar{\Phi}^{-1}\left(1 - (1-\alpha)^{w_i/m}\right) - \frac{\mu_i^2}{2}, \quad (\text{A.2})$$

or

$$c - \frac{w_i}{m} \ln(1-\alpha) = \mu_i \bar{\Phi}^{-1}\left(1 - (1-\alpha)^{w_i/m}\right) - \frac{\mu_i^2}{2},$$

where, $c = \ln\left(\frac{\lambda mm_2}{\ln(1-\alpha)}\right)$. Therefore, \mathbf{w} satisfies the equations (5).

To make sure that equations (5) provide optimal values, we need to investigate the second derivatives of the Lagrange function for w_i .

$$\begin{aligned} \frac{\partial^2 G(\lambda, \mathbf{w})}{\partial w_i^2} &= \frac{\partial}{\partial w_i} \left[- (1-\alpha)^{w_i/m} \frac{\phi(\delta - \mu_i)}{\phi(\delta)} \frac{1}{mm_2} \ln(1-\alpha) + \lambda \right] \\ &= \frac{1}{m_2} \left(\frac{1}{m} \ln(1-\alpha) \right)^2 (1-\alpha)^{w_i/m} \frac{\phi(\delta - \mu_i)}{\phi(\delta)} \\ &\quad + \frac{1}{mm_2} (1-\alpha)^{w_i/m} \ln(1-\alpha) \left\{ \frac{m^{-1} (1-\alpha)^{w_i/m} \ln(1-\alpha) \phi(\delta - \mu_i) (\delta - \mu_i)}{[\phi(\delta)]^2} \right. \\ &\quad \left. + \frac{m^{-1} (1-\alpha)^{w_i/m} \ln(1-\alpha) \phi(\delta - \mu_i) (-\delta)}{[\phi(\delta)]^2} \right\} \\ &= \frac{m^{-1} \left(m^{-1} \ln(1-\alpha) \right)^2 (1-\alpha)^{w_i/m} \phi(\delta - \mu_i) \left[-\phi(\delta) - \mu_i (1-\alpha)^{w_i/m} \right]}{[\phi(\delta)]^2} < 0, \end{aligned}$$

where, $\delta = \bar{\Phi}^{-1}(1 - (1 - \alpha)^{w_i/m})$. Note that the off-diagonal elements of the Hessian matrix are all zeroes. We conclude that the Hessian matrix is negative definite. Consequently, the solutions of the weights are optimal.

References

- Bonferroni, C. E. (1937). Teoria statistica delle classi e calcolo delle probabilita. In "Volume in Onore di Ricarrdo dalla Volta," Universita di Firenze, 1-62.
- Conneely, K. N. and Boehnke, M. (2007). So many correlated tests, so little time! Rapid adjustment of p values for multiple correlated tests. *Am J Hum Genet* 81:1158-1168.
- Genovese, C. R., Roeder, K. and Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika* 93:509-524.
- Hochberg, Y. and Tamhane, A. C. (1987). Multiple comparison procedures. New York: Wiley.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65-70.
- Ionita-Laza, I., McQueen, M. B., Laird, N. M. and Lange, C. (2007). Genomewide weighted hypothesis testing in family-based association studies, with an application to a 100K scan. *Am J Hum Genet* 81: 607-614.
- Lin, D. Y. (2005). An efficient Monte Carlo approach to assessing statistical significance in genomic studies. *Bioinformatics* 21:781-787.
- Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology* 14:1044-1045.
- Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide-polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 74: 765-769.
- Olejnik, S., Li, J. M., Huberty, C. J., Supattathum S. (1997). Multiple testing and statistical power with modified Bonferroni procedures. *J Educat Behavioral Statist* 22: 389-406.

- Roeder, K., Bacanu, S., Wasserman, L., and Devlin, B. (2006). Using linkage genome scans to improve power of association scans. *Am J Hum Genet* 78: 243-252.
- Roeder, K., Devlin, B. and Wasserman, L. (2007). Improving power in genome-wide association studies: weights tip the scale. *Genet Epidemiol* 31: 741-747.
- Rubin, D., Dudoit, S. and van der Laan, M. J. (2006). A method to increase the power of multiple testing procedures through sample splitting. *U.C. Statistical Applications in Genetics and Molecular Biology* 5: article 19.
- Scherrer, B. (1984). *Biostatistique*. G. Morin, Quebec, 850 pp.
- Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc* 62:626-633.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73:751-754.
- Wasserman, L. and Roeder, K. (2006). Weighted hypothesis testing. (<http://arxiv.org/abs/math.ST/0604172>) (accessed July 5, 2007)