Electronic Journal of Statistics
Vol. 7 (2013) 2683–2701
ISSN: 1935-7524
DOI: 10.1214/13-EJS855

On improving some adaptive BH procedures controlling the FDR under dependence

Li He[*]

Merck Research Laboratories
West Point, PA 19486
USA
e-mail: li.he@merck.com

and

Sanat K. Sarkar[†]

Department of Statistics
Temple University
Philadelphia, PA 19122
USA
e-mail: sanat@temple.edu

Abstract: Blanchard and Roquain (2009) presented for the first time methods of adapting the Benjamini-Hochberg (BH) method to data through an estimate of the proportion of true null hypotheses that continue to control the false discovery rate (FDR) under positive dependence in a non-asymptotic setting. However, they are often too conservative to provide a real improvement of the BH method. To obtain adaptive BH methods with proven FDR control improving the original BH method in more situations than what are seen in Blanchard and Roquain (2009), we propose alternative versions of the Blanchard-Roquain methods under some additional assumptions allowing explicit use of pairwise correlations whenever they are available. We offer numerical evidence of improved performances of the proposed alternatives in two scenarios involving test statistics satisfying the positive dependence conditions assumed for the main results.

AMS 2000 subject classifications: Primary 62J15; secondary 62G10.
Keywords and phrases: False discovery rate, multiple testing, adaptive BH procedures, positive dependence, pairwise correlations.

Received December 2012.

1. Introduction

Multiple testing methods controlling an overall measure of false rejections (or discoveries) are among the standard statistical tools being used nowadays when analyzing data from modern scientific investigations. Since these investigations

[*]The research of Li He is supported by Merck Research Fellowship.
[†]The research of Sanat Sarkar is supported by NSF Grant DMS-1006344.

typically require a large number of hypotheses to be tested, the traditional notion of familywise error rate (FWER), the probability of at least one false discovery, is often too conservative. The false discovery rate (FDR), the expected proportion of false discoveries among all discoveries, introduced by Benjamini and Hochberg (1995), on the other hand, is relatively much less conservative, making it the most popular overall measure of false discoveries in modern multiple testing.

Benjamini and Hochberg (1995) introduced an FDR controlling method, known as the BH method. Its FDR at level α is equal to $\pi_0\alpha$ when the underlying test statistics are independent, and less than or equal to $\pi_0\alpha$ when these statistics are PRDS (positive regression dependent on subset of null statistics), where π_0 is the proportion of true null hypotheses (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Sarkar, 2002). Although preferred most often in practice, its performance clearly depends on the strength of positive dependence among the test statistics or the value of π_0. With strong positive dependence or π_0 much smaller than one, the BH method can be quite conservative and can potentially be improved.

A considerable amount of research has taken place to improve the BH method by suitably adapting it to the data. Such adaptation involves estimating the unknown π_0 and using it to adjust the BH procedure. Often the formation of the estimate of π_0 is based on the number of significant hypotheses observed by an initial application of a suitable multiple testing procedure to the data. A variety of these so-called adaptive BH methods have been proposed in the literature (Benjamini and Hochberg, 2000; Benjamini et al., 2006; Blanchard and Roquain, 2009; Gavrilov et al., 2009; Hochberg and Benjamini, 1990; Liang and Nettleton, 2012; Liu and Sarkar, 2011; Sarkar, 2008b; Storey et al., 2004).

However, unlike the original BH method, these methods are theoretically shown to control the FDR only under independence in a non-asymptotic setting (i.e., when the number of hypotheses is not infinitely large) and therefore developing an adaptive BH method without losing the ultimate control over the FDR in such a setting under the PRDS condition remains to be a challenging open problem. Recently, Blanchard and Roquain (2009) gave two adaptive BH methods that control the FDR under PRDS. However, as they have noted, these methods turned out to be too conservative, not even improving the original BH procedure, in many instances. This motivates us to revisit the work of Blanchard and Roquain (2009) and attempt to improve it.

Blanchard and Roquain (2009) adaptive BH methods involve adjusting the level of the original BH method as well as multiplying each p-value in it by an estimate $\hat{\pi}_0$ of π_0. The reason that such adaptive BH methods control the FDR under PRDS is that the part of the FDR that corresponds to $\hat{\pi}_0$ exceeding π_0 can be shown to be controlled at any desired level using arguments similar to proving the same result for the original BH method, while the remaining part that corresponds to $\hat{\pi}_0$ not exceeding π_0 can be bounded above by a certain specified value provided $\hat{\pi}_0$ is appropriately determined from a procedure controlling the FWER or FDR. The choice of this FWER or FDR controlling procedure seems critical as it ultimately affects how conservative the corresponding adaptive BH method is. In fact, the procedure from which $\hat{\pi}_0$ is determined does not have to be one that controls the FWER or FDR; it can be chosen to control any other type of error rate, as long as $\hat{\pi}_0$ can be determined from it in such a way that the aforementioned second part of the FDR of the resulting adaptive BH method can be bounded above by a pre-specified value. This motivates us to consider using a procedure controlling a generalized or alternative form of FWER or FDR, like the k-FWER (Lehmann and Romano, 2005), for some arbitrary $k \geq 2$, or Pairwise FDR (Sarkar, 2008b), which allows us to capture dependence into the formulation of $\hat{\pi}_0$ explicitly under certain positive dependence situations, and thereby improving the BH method in more cases than the adaptive procedures in Blanchard and Roquain (2009).

2. Preliminaries

In this section, we present some background information that are relevant to the present paper.

2.1. Notations and definitions

Suppose that there are m null hypotheses H_i, $i = 1, \ldots, m$, that are to be simultaneously tested based on their respective p-values P_i, $i = 1, \ldots, m$. A multiple testing is often carried out using a stepdown or stepup procedure. One can distinguish between these procedures depending on how the rejection region is found. Given the ordered p-values $P_{(1)} \leq \cdots \leq P_{(m)}$, with $H_{(1)}, \ldots, H_{(m)}$ being the corresponding null hypotheses, and a non-decreasing set of critical values $0 < \alpha_1 \leq \cdots \leq \alpha_n < 1$, a stepdown procedure rejects the set of null hypotheses $\{H_{(i)}, i \leq i^*_{SD}\}$, where $i^*_{SD} = \max\{1 \leq i \leq m : P_{(j)} \leq \alpha_j, \ \forall j \leq i\}$ if the maximum exists; otherwise, it accepts all the null hypotheses. A stepup procedure, on the other hand, rejects the set of null hypotheses $\{H_{(i)}, i \leq i^*_{SU}\}$, where $i^*_{SU} = \max\{1 \leq i \leq m : P_{(i)} \leq \alpha_i\}$ if the maximum exists; otherwise, it accepts all the null hypotheses. If $\alpha_1 = \cdots = \alpha_m$, the stepup or stepdown procedure reduces to what is usually referred to as a single-step procedure.

Let R and V denote the total numbers of null hypotheses that are rejected and falsely rejected, respectively, in a multiple testing procedure. Then, the FWER of this procedure is defined by $P(V \geq 1)$, while the FDR is defined by $E(V/\{R \vee 1\})$, where $R \vee 1 = \max\{R, 1\}$. The following are generalized or alternative versions of these error rates that will be of relevance in this paper: (i) Generalized FWER: k-FWER $= P(V \geq k)$, for some fixed $k \geq 2$, and (ii) Pairwise FDR $= E[V(V-1)/\{[R(R-1)] \vee 1\}]$.

2.2. Assumptions

We will assume throughout the paper that marginally the p-value corresponding to each null hypothesis is stochastically larger than $U(0, 1)$, that is, $\Pr(P_i \leq t) \leq$

$\min\{t, 1\}$, for each $i \in I_0$, where I_0 is the set of indices for the null p-values. Regarding dependence among all the p-values, we make one of the following two assumptions in our main results to be discussed in the next section:

Assumption 1. The conditional expectation $E\{\phi(P_1, \ldots, P_m) \mid P_i \leq u\}$ is non-decreasing in u for each $i \in I_0$ and any non-decreasing (coordinatewise) function ϕ.

Assumption 2. The conditional expectation $E\{\phi(P_1, \ldots, P_m) \mid P_i \leq u, P_j \leq v\}$ is non-decreasing in (u, v) for each $\{i, j\} \subseteq I_0$ and any non-decreasing (coordinatewise) function ϕ.

Assumption 1 is a slightly weaker version of the PRDS condition (defined in Benjamini and Yekutieli (2001)), satisfied by p-values generated from distributions arising in many multiple testing problems, among which the multivariate normal with non-negative correlations is the most common one (Benjamini and Yekutieli, 2001; Sarkar, 2002). Assumption 2 is satisfied by multivariate totally positive of order two distributions (MTP_2, defined in Karlin and Rinott (1980)), among which the multivariate normal with a common nonnegative common correlation is an important one. In fact, both Assumptions 1 and 2 are satisfied by the MTP_2 condition. Other distributions satisfying the MTP_2 condition include certain types of multivariate t, F and gamma distributions (see Karlin and Rinott (1980)).

2.3. Adaptive BH methods

The level α BH method in its original form is a stepup method with the critical values $\alpha_i = i\alpha/m$, $i = 1, \ldots, m$. Its actual FDR, as said in the introduction, is less than or equal to $\pi_0\alpha$ under Assumption 1, with the equality holding under independence. If π_0 were known, one would have used $\pi_0 P_i$ instead of P_i in the BH method to get a much tighter control of the FDR at α. This brings about the idea of adapting the BH method to the data, when π_0 is unknown, through an estimate $\hat{\pi}_0$ of π_0 obtained from the data. Quite often, this adaptation involves multiplying each p-value in the original level α BH method by $\hat{\pi}_0$ after having determined this $\hat{\pi}_0$ from the number of significant hypotheses observed via a multiple testing procedure applied to the data. A large number of such adaptive BH methods have been introduced in the literature (cited in Introduction). However, they have been theoretically shown to maintain the FDR control at α only when the p-values are independent.

Recently, Blanchard and Roquain (2009) proposed two such adaptive BH procedures that continue to maintain the FDR control under Assumption 1. These two newer adaptive BH methods correspond to two different types of multiple testing procedure, one controlling the FWER and the other controlling the FDR, from which π_0 is estimated before each p-value in the BH method is multiplied by this estimate. More specifically, they considered the following two estimates of π_0: (i) $\hat{\pi}_0^{\mathrm{BR1}} = (m - R_0)/m$ and (ii) $\hat{\pi}_0^{\mathrm{BR2}} = H_\eta(R_0/m)$, where

$$H_\eta(x) = \begin{cases} 1 & x \leq 1/\eta \\ \dfrac{2/\eta}{1 - \sqrt{1 - 4(1-x)/\eta}} & x > 1/\eta, \end{cases}$$

for $0 \leq x \leq 1$ and some $\eta > 1$, and R_0 is the number of significant p-values observed at stage 1 by applying the Holm (1979) FWER controlling method in $\hat{\pi}_0^{\mathrm{BR1}}$ and the BH FDR controlling method in $\hat{\pi}_0^{\mathrm{BR2}}$. We call the resulting adaptive BH methods BR1 and BR2, respectively. The levels at the two stages have been chosen appropriately to provide the ultimate control of the FDR at the desired level. Let γ_0 and γ_1, respectively, be the levels for the first stage FWER or FDR controlling method and the original BH method before being adjusted by the corresponding estimate $\hat{\pi}_0$ at the second stage. Then, as Blanchard and Roquain (2009) proved, the FDRs of BR1 and BR2 are controlled at $\gamma_0 + \gamma_1$ and $\eta\gamma_0 + \gamma_1$ (with $\gamma_0 \leq \gamma_1$), respectively, under Assumption 1.

2.4. Procedures controlling k-FWER and pairwise FDR

A number of k-FWER procedures are available in the literature under different types of distributional assumption; see Finos and Farcomeni (2011) and Guo and Rao (2010) for references on k-FWER. Among them, the generalized Hochberg's procedure of Sarkar (2008a) developed using k-dimensional joint distributions of the null p-values under certain distributional assumptions is of relevance here. We will re-construct it (in Section 3.1) under Assumption 2 with a view to capture only the pairwise joint distributions of the null p-values before developing one of our main results. The notion of Pairwise FDR along with a procedure controlling it under Assumption 2 were introduced in Sarkar (2008b).

In this article, we consider estimating π_0 from a procedure controlling the k-FWER or Pairwise FDR, instead of the FWER or FDR, in the derivation of the Blanchard-Roquain type adaptive BH methods assuming that $n_0 \geq 2$. The goal has been to improve the FDR control of the original Blanchard-Roquain type adaptive BH methods in wider situations where the dependence structure can be captured more explicitly through pairwise null distributions whenever they are available.

3. Main results

Two main results will be developed in this article. One is an alternative to BR1 where a k-FWER (Lehmann and Romano, 2005), for some fixed $k \geq 2$, rather than an FWER (i.e., 1-FWER), procedure is used to estimate π_0, while the other one is an alternative to BR2 where a Pairwise FDR (Sarkar, 2008b), instead of an FDR, procedure is used to estimate π_0. Let us paraphrase in the following the basic result from Blanchard and Roquain (2009) that guides us to the development of these alternatives that continue to control the FDR with or without some additional assumptions. A proof of it based on slightly different line of arguments is given in Appendix.

Lemma 3.1. *Given any estimate $\hat{\pi}_0$ of π_0, which is non-decreasing in each P_i, consider the adaptive level-γ_1 BH method, that is, the stepup method in terms of the adaptive p-values $Q_i = \hat{\pi}_0 P_i$, $i = 1, \ldots, m$, and the critical values $i\gamma_1/m$, $i = 1, \ldots, m$. The FDR of this adaptive BH method satisfies the following inequality under Assumption 1:*

$$\mathrm{FDR} \le \gamma_1 + E\left[\frac{V}{R} I(R > 0, \hat{\pi}_0 < \pi_0)\right], \tag{1}$$

where R and V denote the numbers of rejected and falsely rejected null hypotheses, respectively, in the adaptive method.

To control the FDR in (1) at any desired level under Assumption 1, one should be able to control the expectation in the right-hand side of (1). Of course, how this expectation can be controlled depends on the kind of multiple testing procedure being used and applied to the data to determine $\hat{\pi}_0$. While Blanchard and Roquain (2009) chose a procedure controlling the FWER or FDR, without adding any more distributional assumption, we consider using certain generalized versions of these error rates with some additional assumptions as described in the following subsections.

3.1. Alternative to BR1

Let us consider estimating π_0 using an estimate of the form $\hat{\pi}_0 = (m - R_0 + k - 1)/m$, for some fixed $k \ge 2$, from the number of significant p-values R_0 observed by applying a multiple testing procedure at the first stage. Let V_0 be the number of falsely detected significant p-values in that multiple testing procedure. Then, since $m - R_0 \ge m_0 - V_0$, where $m_0 = m\pi_0$, we have

$$\begin{aligned} & E\left[\frac{V}{R} I(R > 0, \hat{\pi}_0 < \pi_0)\right] \\ \le \quad & \Pr(\hat{\pi}_0 < \pi_0) = \Pr(m - R_0 + k - 1 < m_0) \le \Pr(V_0 \ge k), \end{aligned} \tag{2}$$

which is the k-FWER of this first stage procedure.

As said before, a number of k-FWER procedures have been proposed in the literature. However, we will consider only the one given in the following lemma, with its proof given in Appendix, that we propose in this article for the first time. It is developed with the idea of incorporating pairwise correlations into the formulation of a k-FWER procedure whenever these correlations are available and known, and can potentially be used to produce a more informative estimator of π_0.

Lemma 3.2 (k-FWER procedure at level α)**.** *Assume that the p-values have identical and known pairwise joint null distributions. Consider the stepup procedure with the critical values α_i, $i = 1, \ldots, m$, satisfying $\alpha_i = G^{-1}(k(k -$*

$1)\alpha/(m+k-i\vee k)(m+k-i\vee k-1))$, for some fixed $k \geq 2$, where G^{-1} is the inverse of the common distribution function of the maximum of two null p-values. It controls the k-FWER at level α under Assumption 2.

Thus, we have our first main result providing an alternative to BR1.

Theorem 3.1. *The FDR of the adaptive BH method given by the stepup procedure based on the adaptive p-values $Q_i = \hat{\pi}_0 P_i$, $i = 1, \ldots, m$, and the critical values $i\gamma_1/m$, $i = 1, \ldots, m$, can be controlled at $\gamma_1 + \gamma_0$ under Assumption 1 and the assumptions in Lemma 3.2 by choosing $\hat{\pi}_0 = (m - R_0 + k - 1)/m$, where R_0 is the number of rejections obtained by applying the k-FWER procedure in that lemma at level γ_0.*

3.2. Alternative to BR2

Here, we consider estimating π_0 from a procedure controlling the Pairwise FDR. To explain how to control the expectation in the right-hand side of (1) corresponding to such an estimate, we first need to state the Pairwise FDR procedure that we are going to use. This procedure is given in Sarkar (2008b).

Lemma 3.3 (Pairwise FDR procedure at level α)**.** *Assume that the p-values have identical and known pairwise joint null distributions. Consider the stepup procedure with the critical values α_i, $i = 1, \ldots, m$, satisfying $\alpha_1 = \min\{\sqrt{\alpha}/m, G^{-1}(2\alpha/m(m-1)\}$, and $\alpha_i = G^{-1}(i(i-1)\alpha/m(m-1))$, $i = 2, \ldots, m$, where G^{-1} is the inverse of the common distribution function of the maximum of two null p-values. The Pairwise FDR of this procedure is less than or equal to $m_0(m_0-1)\alpha/m(m-1)$ under Assumption 2.*

Remark 3.1. It is to be noted that the Pairwise FDR equals 0 if $V = 0$ or 1, and hence the choice of the first critical value in a stepup procedure designed to control the Pairwise FDR does not matter. Our particular choice of α_1 in the above procedure is different from the one originally considered in Sarkar (2008b), and is made only to make sure, as can be seen in the following, that certain desirable inequality holds.

Let us consider $\hat{\pi}_0 = \min\{(m - R_0 + 1)/m(1-\lambda), 1\}$, for some $\lambda \in (0, 1)$, where R_0 is the number of significant p-values observed by applying the Pairwise FDR procedure at level α stated in the Lemma 3.3. Note that under Assumption 1, $G(u) \geq u^2$ for $u \in (0, 1)$, which implies that $\alpha_i \leq \sqrt{i(i-1)\alpha/m(m-1)} \leq i\sqrt{\alpha}/m$, for all $i = 2, \ldots, m$. In other words, if $\alpha \leq \gamma_1^2$, the critical values of the level γ_1 BH method in terms of the adaptive p-values will be larger than the corresponding critical values of the level α Pairwise FDR procedure. Moreover, when $\hat{\pi}_0 \leq 1$, the adaptive p-values are stochastically smaller than the corresponding non-adaptive or original p-values. Therefore, if $\alpha \leq \gamma_1^2$ and $\hat{\pi}_0 < \pi_0$, R, the number of rejections in the level γ_1 adaptive BH method, is stochastically larger than R_0, the number of rejections in the level α Pairwise FDR procedure. With V_0 denoting the number of false rejections in this Pairwise

FDR procedure, we thus have, for the expectation in the right-hand side of (1),

$$\begin{aligned}
E\left[\frac{V}{R} I(R>0, \hat{\pi}_0<\pi_0)\right] &\leq E\left[\frac{V}{R_0} I(m-R_0+1<m_0(1-\lambda))\right] \\
&\leq m_0 E\left[\frac{1}{R_0} I(V_0>1+m_0\lambda)\right] \\
&\leq \frac{m_0(m-1)}{(1+m_0\lambda)m_0\lambda} E\left[\frac{V_0(V_0-1)}{R_0(R_0-1)} I(V_0\geq 2)\right] \\
&\leq \frac{m_0^2(m_0-1)(m-1)\alpha}{m(m-1)(1+m_0\lambda)m_0\lambda} \leq \frac{m\pi_0\alpha}{(1+m\lambda)\lambda} \\
&= \eta\pi_0\gamma_0, \qquad (3)
\end{aligned}$$

with $\eta = \lambda^{-1}$ and $\alpha = (\frac{1}{\eta} + \frac{1}{m})\gamma_0$. This leads us to our next main result.

Theorem 3.2. *The FDR of the adaptive BH method given by the stepup procedure based on the adaptive p-values* $Q_i = \hat{\pi}_0 P_i$, $i = 1, \ldots, m$, *and the critical values* $i\gamma_1/m$, $i = 1, \ldots, m$, *can be controlled at* $\gamma_1 + \eta\gamma_0$ *under Assumption 1 and the assumptions in Lemma 3.3 by choosing* $\hat{\pi}_0 = \min\{\eta(m-R_0+1)/m(\eta-1), 1\}$, *for some* $\eta > 1$, *where* R_0 *is the number of significant p-values observed by applying the Pairwise FDR procedure in that lemma at level* $\alpha = (\frac{1}{\eta} + \frac{1}{m})\gamma_0 \leq \gamma_1^2$.

4. Simulation studies

4.1. Power comparisons

Our proposed adaptive BH method in Theorem 3.1 (or Theorem 3.2) is considered to be an improvement over BR1 (or BR2) if it offers better power performance compared not only with BR1 (or BR2) but also with the BH procedure they all intend to improve. We conducted simulation studies to numerically investigate if such improvements really occur for our methods, and if so, to what extent, under the type of distributional and dependence assumptions we make in the paper. We considered two different notions of power when assessing power performances, the false non-discovery rate (FNR) and the average power. The FNR is defined as the expected proportion of acceptances (non-discoveries) that are false (see, Genovese and Wasserman (2004); Sarkar (2004)), whereas, the average power is defined as the expected proportion of false null hypotheses that are correctly rejected (see, for instance, Dudoit et al. (2003)). The results of these simulation studies are presented in this section.

We considered two different simulation settings, Settings 1 and 2, each involving test statistics satisfying the MTP_2 condition. In Setting 1, we generated m test statistics $X_i \sim N(\mu_i, 1)$, $i = 1, \ldots, m$, with a common non-negative correlation ρ, having randomly set π_0 proportion of the means at 0 and the rest at 3, for testing $\mu_i = 0$ against $\mu_i > 0$ simultaneously for $i = 1, \ldots, m$. In Setting 2, these test statistics were $X_i \sim \sigma_i^2 \mathrm{Gamma}(p, 2)$, $i = 1, \ldots, m$, with a common non-negative correlation ρ, generated with π_0 proportion of the variances randomly set at 1 and the rest at 2.56 for testing $\sigma_i^2 = 1$ against $\sigma_i^2 > 1$ simultaneously for $i = 1, \ldots, m$. These multivariate gamma statistics were generated by first generating $m+1$ independent gamma random variables $Y_0 \sim \sigma_i^2 \text{Gamma}(\rho p, 2)$, $Y_i \sim \sigma_i^2 \text{Gamma}((1-\rho)p, 2)$, $i = 1 \ldots, m$, and then setting $X_i = Y_i + Y_0$.

It should be noted that $\text{Gamma}(0, 2) = 0$ with probability one when generating the above multivariate gamma, and thus $\rho = 0$ corresponds to the independence case, as in the case of multivariate normal in Setting 1. Also, for the multivariate gamma in Setting 2 to have the desired MTP_2 property, $(1-\rho)p$ has to be greater than or equal to 1, that is, $\rho \le 1 - \frac{1}{p}$; see Karlin and Rinott (1980).

In each of the above two settings, we considered $m = 1000$ and applied the following procedures to the generated data at level $\alpha = 0.1$ for the corresponding multiple testing problem: the BR1, the proposed alternative to BR1, the BR2, the proposed alternative to BR2, the original BH method, and the oracle version of the BH method.

We chose $\gamma_0 = \gamma_1 = 0.05$ in both BR1 and our proposed alternative to BR1, and the k was conveniently chosen to be 80 in the alternative to BR1 according to Table 1. In both BR2 and and our alternative to BR2, we chose $\eta = 2$. While $\gamma_0 = 0.025$ and $\gamma_1 = 0.05$ in BR2, the γ_0 and γ_1 in the alternative to BR2 were chosen in such a way that they satisfy the following equations:

$$\eta\gamma_0 + \gamma_1 = 0.1 \quad \text{and} \quad \left(\frac{1}{\eta} + \frac{1}{m}\right)\gamma_0 = \gamma_1^2. \tag{4}$$

We chose ten different values of ρ from $\{0, 0.1, \ldots, 0.9\}$ in Setting 1, and also in Setting 2 having chosen $p = 10$.

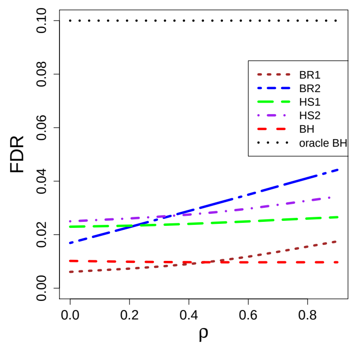
We simulated the values of the FDR, FNR and average power for each of the aforementioned procedures using 1000 simulation runs.

Figures 1 and 2 display the results in terms of the simulated FDR, FNR, and the average power obtained in Setting 1 for $\pi_0 = 0.1$ and $\pi_0 = 0.2$, respectively, while Figures 3 and 4 do the same for Setting 2.

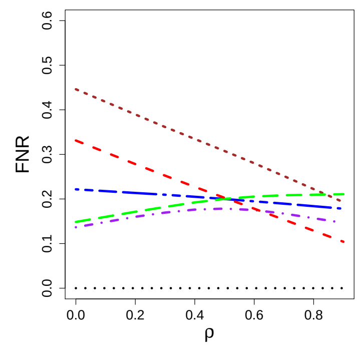
These figures clearly show that, as an adaptive BH method controlling the FDR, our proposed alternative form of the BR1, labeled HS1, has much improved power performance compared to the original BR1 for almost all of the considered values of ρ. Also our proposed alternative to the BR2, labeled HS2, is seen to have better power performance compared to the original BR2. More importantly, our proposed alternatives are seen to provide more improvements (smaller FNR or higher power) over the original BH method in more cases.

Overall, the BR2 and both of our proposed procedures are preferred to the BR1 procedure for almost all of the ρ values, with our alternative to the BR2 being the most powerful one. When correlation is not large, our proposed alternative to the BR1 can outperform the BR2 procedure. Looking at these figures, we see that with π_0 getting larger, it gets harder for all these adaptive methods to improve the original BH, which is of course expected.

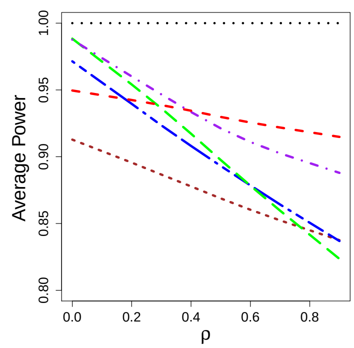
To further evaluate the extent of power improvements our proposed alternatives to the BR1 and BR2 can offer over the original BH method in comparison with similar improvements offered by the BR1 and BR2 themselves, we did some



(a)



(b)



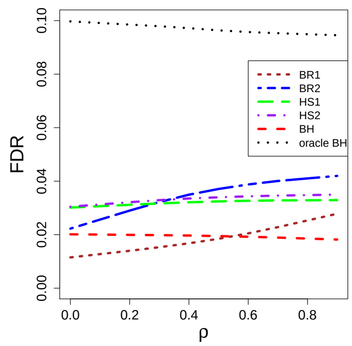
(c)

FIG 1. *Simulated FDR, FNR, and average power of the proposed alternative to BR1 (labeled HS1), the original BR1, the proposed alternative to BR2 (labeled HS2), the original BR2, the BH method, and the oracle BH method obtained in Setting 1 with* $\pi_0 = 0.1$.

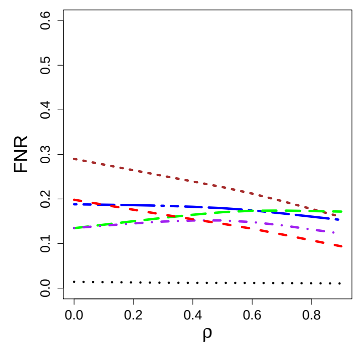
additional computations under Setting 1 with the same values of α, m, k, η, γ_0, and γ_1 as chosen before. For each of the aforementioned adaptive BH methods and for each π_0 chosen from $\{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$, we computed the crossing point $\rho(\pi_0)$ for the correlations such that, for all $\rho \leq \rho(\pi_0)$, the corresponding adaptive BH method has a smaller FNR than the BH method. The values of $\rho(\pi_0)$ were plotted in Figure 5. As seen from this figure, for most of the π_0 values considered (except when they are very small), our proposed two procedures provide improvements over the BH procedure for more values of ρ than the BR1 and BR2.

4.2. *Choice of k and η*

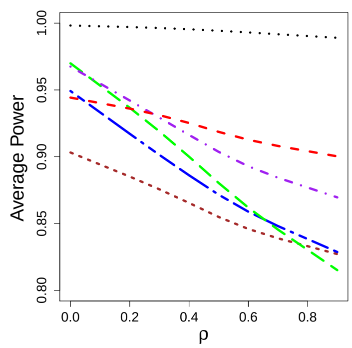
The choice of k or η for which our proposed alternative to BR1 or BR2, respectively, can have the maximum possible power is an important step in our



(a)



(b)



(c)

FIG 2. *Simulated FDR, FNR, and average power of the proposed alternative to BR1 (labeled HS1), the original BR1, the proposed alternative to BR2 (labeled HS2), the original BR2, the BH method, and the oracle BH method obtained in Setting 1 with* $\pi_0 = 0.2$.

propositions. However, if they are not pre-chosen, it would be difficult to make a general recommendation for them, since they depend in a complex way on the other parameters γ_0, γ_1, π_0 and ρ, and even the values of m and α. The best option in this case would be to get an idea about k and η through simulations. This is what we have decided to do in this paper.

Specifically, we considered the multiple testing problem as described in simulation setting 1 with $\gamma_0 = \gamma_1 = 0.05$, and determined the value of k providing the smallest possible simulated FNR based on 1000 repetitions for different combinations of π_0 and ρ values. We did these for $m = 100$ and $m = 1000$. We found that the ratio k/m to be not much different in most cases. This can be seen, for example, from Table 1 which presents the values of k/m only for the case of $m = 1000$. This table can serve as a reference for one to make a choice for k in a multiple testing situation similar to that considered in this simulation setting.

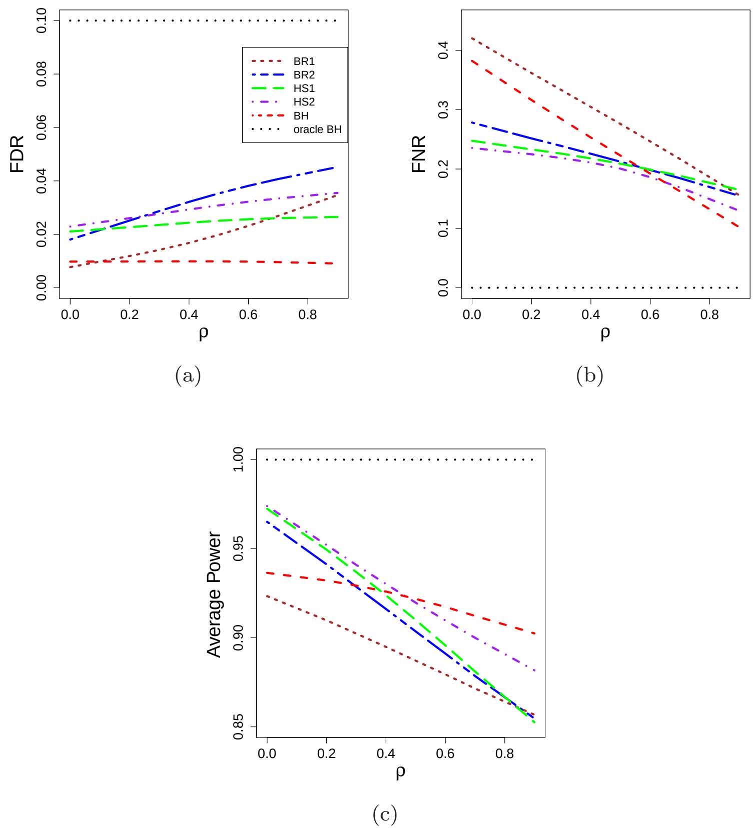


Fig 3. *Simulated FDR, FNR, and average power of the proposed alternative to BR1 (labeled HS1), the original BR1, the proposed alternative to BR2 (labeled HS2), the original BR2, the BH method, and the oracle BH method obtained in Setting 2 with* $\pi_0 = 0.1$.

We did the same for the proposed alternative to BR2, with γ_0 and γ_1 chosen according to (4), to recommend a choice for η from a targeted set $\{1.1, 1.2, \ldots, 6\}$ when $m = 1000$. It appears that $\eta = 2$ is a good choice from this set of possible values, as seen from Table 2.

5. Concluding remarks

The idea of adapting the BH method to data with a view to improving its performance has taken shape in the literature mostly in the form of drawing information about π_0 before incorporating it into the method, although there is a possibility of further improvement by also incorporating information about dependence. This paper can be viewed as one that makes an attempt for the first time to explore that possibility. We have developed newer adaptive BH methods

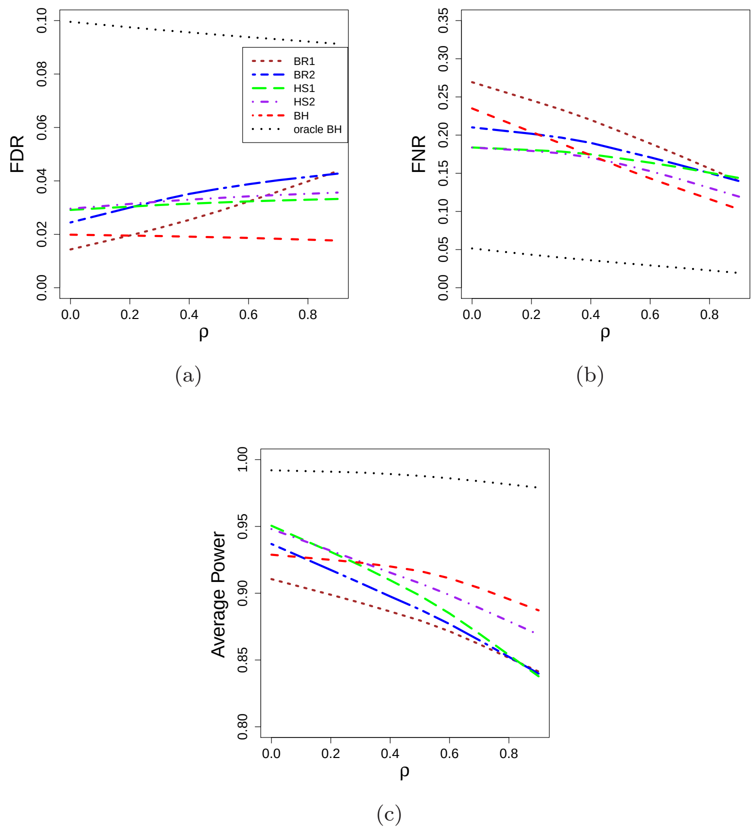


Fig 4. *Simulated FDR, FNR, and average power of the proposed alternative to BR1 (labeled HS1), the original BR1, the proposed alternative to BR2 (labeled HS2), the original BR2, the BH method, and the oracle BH method obtained in Setting 2 with* $\pi_0 = 0.2$.

by eliciting information on both π_0 and dependence without losing the ultimate control over the FDR under the dependence structure originally assumed for the data in a non-asymptotic setting (i.e., without assuming $m \to \infty$). Our simulations indicate that such newer adaptive BH methods can potentially offer significant improvements over the BH method in many instances, especially when there is weak positive dependence among the underlying test statistics and π_0 is small.

As alluded above, constructing an adaptive BH method via estimating π_0 that can capture dependence and eventually control the FDR has been our primary goal in this paper. In other words, estimating π_0 is not the main focus in our paper, rather it is designed specifically to achieve the aforementioned goal. This is similar to the idea behind estimating π_0 in all other adaptive BH methods in the literature (cited in Introduction), and even in some adaptive

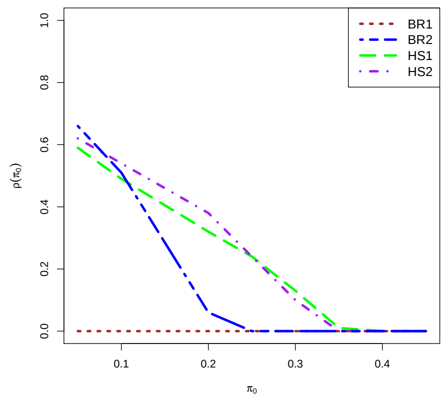


Fig 5. *The crossing point correlation value $\rho(\pi_0)$ for fixed π_0 for the proposed alternative to BR1 (labeled HS1), the original BR1, the proposed alternative to BR2 (labeled HS2) and the original BR2. The $\rho(\pi_0)$ is computed for each of the four procedures so that the corresponding procedure outperforms the original BH method in terms of FNR whenever $\rho \leq \rho(\pi_0)$.*

Table 1
Simulated values of k/m providing the least FNR for the proposed alternative to BR1 with $m = 1000$ (based on 1000 simulations runs)

ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\pi_0 = 0.1$	0.067	0.083	0.101	0.117	0.148	0.168	0.216	0.269	0.343	0.529
$\pi_0 = 0.3$	0.075	0.090	0.100	0.114	0.130	0.162	0.188	0.250	0.327	0.467
$\pi_0 = 0.5$	0.062	0.074	0.082	0.096	0.101	0.125	0.148	0.175	0.237	0.002
$\pi_0 = 0.7$	0.047	0.048	0.050	0.059	0.065	0.070	0.079	0.096	0.002	0.002
$\pi_0 = 0.9$	0.018	0.018	0.017	0.018	0.065	0.070	0.079	0.096	0.002	0.002

Bonferroni or Sidak method (Guo, 2009; Finner and Gontscharuk, 2009; Sarkar et al., 2012), although, no special efforts have been made to incorporate dependence into those estimations and the FDR (or the FWER) control of the adaptive BH (or adaptive Bonferroni or Sidak) method has been established only under independence, except in Blanchard and Roquain (2009). Several different estimates of π_0 have been considered in this process. These are mostly variants or some forms of extension of the following estimate that Schweder and Spjotvoll (1982) suggested for the first time for estimating π_0:

$$\hat{\pi}_0^{\mathrm{SS}}(\lambda) = \frac{m - R(\lambda)}{m(1-\lambda)}, \ \lambda \in (0,1), \text{ where } R(\lambda) = \sum_{i=1}^{m} I(P_i \leq \lambda),$$

with the rationale that the p-values greater than a fixed, but not too small, threshold λ should correspond to the true null hypotheses.

A number of other estimates of π_0 are available in the literature, which have been proposed without the aforementioned specific goal of constructing an adaptive BH or Bonferroni method with proven FDR or FWER control. For instance,

TABLE 2
Choice of η providing the least FNR for the proposed alternative to BR2 with $m = 1000$ (based on 1000 simulations runs)

ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\pi_0 = 0.1$	2.5	2.4	2.2	2.1	2.2	2	1.7	1.1	1.1	1.1
$\pi_0 = 0.3$	2.7	2.6	2.5	1.1	1.1	1.1	1.1	1.1	1.1	1.1
$\pi_0 = 0.5$	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
$\pi_0 = 0.7$	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
$\pi_0 = 0.9$	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1

Langaas et al. (2005) estimated π_0 using the nonparametric density estimation approach assuming that the p-values are independently distributed following a two-component mixture density function which is decreasing or both convex and decreasing. Pounds and Morris (2003) considered estimating π_0 under a parametric beta-uniform mixture model. Storey and Tibshirani (2003) provided an estimate of π_0 using smoothing techniques. Schwartzman (2008) proposed a "mode matching" method for the estimation of empirical null distribution with the theoretical null belonging to the exponential family and provided an estimate of π_0 in this framework. Jin and Cai (2007) gave an approach to estimating π_0 based on the empirical characteristic function and Fourier analysis. Most of these estimation methods rely on the assumption of independent p-values from a two-component mixture model with a uniform distribution for null p-values. However, estimators of π_0 under dependence have also been given. For instance, Friguet and Causeur (2011) modified $\hat{\pi}_0^{\mathrm{SS}}$ using factor-adjusted p-values considering a general factor analytic model framework Leek and Storey (2008) for multiple testing under dependence. Chen and Doerge (2012) developed a consistent estimator of π_0 when the test statistics follow multivariate normal distribution with a known covariance matrix representing certain types of strong dependence using the principal factor approximation developed in Han et al. (2010) and the Fourier transform method in Jin (2008).

Of course, all these other estimates of π_0 can potentially be used in adapting the BH method, or even the Bonferroni method, but whether or not the resulting BH or Bonferoni methods can provide the ultimate control over the FDR or FWER remains to be a theoretically challenging and open question.

6. Appendix

Proof of Lemma 3.1. Let I_0 be the set of indices of the true null hypotheses, and $R_{m-1}^{(-i)}$ be the number of rejections in the stepup method based on $\{Q_1, \ldots, Q_m\} \setminus \{Q_i\}$ and the critical values $i\gamma_1/m$, $i = 2, \ldots, m$. Then,

$$
\begin{aligned}
& E\left[\frac{V}{R} I(R > 0, \hat{\pi}_0 > \pi_0)\right] \\
= \quad & \sum_{i \in I_0} \sum_{r=1}^{m} \frac{1}{r} \Pr\left(\hat{\pi}_0 P_i \leq \frac{r\gamma_1}{m}, R = r, \hat{\pi}_0 > \pi_0\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{i \in I_0} \sum_{r=1}^{m} \frac{1}{r} \Pr\left(P_i \leq \frac{r\gamma_1}{m\pi_0}, R_{m-1}^{(-i)} = r-1 \right) \\
&\leq \frac{\gamma_1}{m\pi_0} \sum_{i \in I_0} \sum_{r=1}^{m} \Pr\left(R_{m-1}^{(-i)} = r-1 \mid P_i \leq \frac{r\gamma_1}{m\pi_0} \right). \qquad (5)
\end{aligned}
$$

Now,

$$
\begin{aligned}
& \sum_{r=1}^{m} \Pr\left(R_{m-1}^{(-i)} = r-1 \mid P_i \leq \frac{r\gamma_1}{m\pi_0} \right) \\
= \;& \sum_{r=0}^{m-1} \Pr\left(R_{m-1}^{(-i)} \geq r \mid P_i \leq \frac{(r+1)\gamma_1}{m\pi_0} \right) \\
& - \sum_{r=0}^{m-2} \Pr\left(R_{m-1}^{(-i)} \geq r+1 \mid P_i \leq \frac{(r+1)\gamma_1}{m\pi_0} \right) \\
\leq \;& \sum_{r=0}^{m-1} \Pr\left(R_{m-1}^{(-i)} \geq r \mid P_i \leq \frac{(r+1)\gamma_1}{m\pi_0} \right) \\
& - \sum_{r=0}^{m-2} \Pr\left(R_{m-1}^{(-i)} \geq r+1 \mid P_i \leq \frac{(r+2)\gamma_1}{m\pi_0} \right) \\
= \;& 1,
\end{aligned}
$$

with the inequality following from Assumption 1 and the fact that $\Pr(R_{m-1}^{(-i)} \geq r+1 \mid P_i \leq u) = E\{I(R_{m-1}^{(-i)} \geq r+1) \mid P_i \leq u\}$ is decreasing in u since $I(R_{m-1}^{(-i)} \geq r+1)$ is a decreasing function of $(P_1, \ldots, P_n)$. Thus, the expectation in (5) is less than or equal to γ_1, as desired. $\square$

Proof of Lemma 3.2. Since $k \leq V \leq m_0$, we have $V(m_0 - V + k) \geq m_0 k$, and similarly $(V-1)(m_0 - V + k - 1) \geq (m_0 - 1)(k-1)$. Also, $m - R \geq m_0 - V$. Thus, we have

$$
\begin{aligned}
& V(V-1)(m-R+k)(m-R+k-1) \\
\geq \;& V(V-1)(m_0-V+k)(m_0-V+k-1) \\
\geq \;& m_0(m_0-1)k(k-1),
\end{aligned}
$$

from which we get

$$
\begin{aligned}
& P(V \geq k) \\
& \leq E\left[\left\{ \frac{V(V-1)(m-R+k)(m-R+k-1)}{m_0(m_0-1)k(k-1)} \right\} I(V \geq k) \right] \\
& \leq \frac{1}{m_0(m_0-1)} \sum_{r=k}^{m} E\left[\frac{(m-r+k)(m-r+k-1)}{k(k-1)} \sum_{i \neq j \in I_0} F_2(\alpha_r, \alpha_r) \right.
\end{aligned}
$$

$$
\begin{aligned}
&P\left\{V_{m-2}^{(-i,-j)} \geq k-2,\ R_{m-2}^{(-i,-j)} = r-2 \mid P_i \vee P_j \leq \alpha_r\right\}\Big]\\
&= \frac{\alpha}{m_0(m_0-1)} \sum_{i \neq j \in I_0} \sum_{r=k}^{m} Pr\left\{V_{m-2}^{(-i,-j)} \geq k-2,\ R_{m-2}^{(-i,-j)} = r-2 \mid \right.\\
&\qquad \left. P_i \vee P_j \leq \alpha_r\right\}\\
&\leq \frac{\alpha}{m_0(m_0-1)} \sum_{i \neq j \in I_0} \sum_{r=k}^{m} \Pr\left\{R_{m-2}^{(-i,-j)} = r-2 \mid P_i \vee P_j \leq \alpha_r\right\}\\
&\leq \alpha,
\end{aligned}
$$

where $R_{m-2}^{(-i,-j)}$ is the number of rejections in the stepup test based on the $m-2$ p-values $\{P_1, \ldots, P_m\} \setminus \{P_i, P_j\}$ and the critical values α_i, $i = 3, \ldots m$, and $V_{m-2}^{(-i,-j)}$ is the number of false rejections among $R_{m-2}^{(-i,-j)}$. The last inequality follows from the result

$$
\begin{aligned}
& \sum_{r=k}^{m} \Pr\left\{R_{m-2}^{(-i,-j)} = r-2 \;\middle|\; P_i \vee P_j \leq \alpha_r\right\} \\
= \;& \sum_{r=k-2}^{m-2} \Pr\left\{R_{m-2}^{(-i,-j)} \geq r \;\middle|\; P_i \vee P_j \leq \alpha_{r+2}\right\} \\
& - \sum_{r=k-1}^{m-2} \Pr\left\{R_{m-2}^{(-i,-j)} \geq r \;\middle|\; P_i \vee P_j \leq \alpha_{r+1}\right\} \\
\leq \;& \sum_{r=k-2}^{m-2} \Pr\left\{R_{m-2}^{(-i,-j)} \geq r \;\middle|\; P_i \vee P_j \leq \alpha_{r+2}\right\} \\
& - \sum_{r=k-1}^{m-2} \Pr\left\{R_{m-2}^{(-i,-j)} \geq r \;\middle|\; P_i \vee P_j \leq \alpha_{r+2}\right\} \\
= \;& \Pr\left\{R_{m-2}^{(-i,-j)} \geq k-2 \;\middle|\; P_i \vee P_j) \leq \alpha_k\right\} \leq 1. \qquad (6)
\end{aligned}
$$

The first inequality in (6) follows from Assumption 2 and the fact that $\Pr(R_{m-2}^{(-i,-j)} \geq r \mid P_i \leq u, P_j \leq u) = E\{I(R_{m-2}^{(-i,-j)} \geq r) \mid P_i \leq u, P_j \leq u\}$ is decreasing in u since $I(R_{m-2}^{(-i,-j)} \geq r)$ is a decreasing function of $(P_1, \ldots, P_n)$. $\square$

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300. MR1325392

Benjamini, Y. and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25(1):60–83.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507. MR2261438

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, pages 1165–1188. MR1869245

Blanchard, G. and Roquain, E. (2009). Adaptive false discovery rate control under independence and dependence. *The Journal of Machine Learning Research*, 10:2837–2871. MR2579914

Chen, X. and Doerge, R. (2012). Estimating the proportion of nonzero normal means under certain strong covariance dependence.

Dudoit, S., Shaffer, J. P., and Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science*, pages 71–103. MR1997066

Finner, H. and Gontscharuk, V. (2009). Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):1031–1048. MR2750256

Finos, L. and Farcomeni, A. (2011). k-fwer control without p-value adjustment, with application to detection of genetic determinants of multiple sclerosis in italian twins. *Biometrics*, 67(1):174–181. MR2898829

Friguet, C. and Causeur, D. (2011). Estimation of the proportion of true null hypotheses in high-dimensional data under dependence. *Computational Statistics & Data Analysis*, 55(9):2665–2676. MR2802344

Gavrilov, Y., Benjamini, Y., and Sarkar, S. K. (2009). An adaptive step-down procedure with proven fdr control under independence. *The Annals of Statistics*, 37(2):619–629. MR2502645

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *The Annals of Statistics*, 32(3):1035–1061. MR2065197

Guo, W. (2009). A note on adaptive bonferroni and holm procedures under dependence. *Biometrika*, 96(4):1012–1018. MR2767287

Guo, W. and Rao, M. B. (2010). On stepwise control of the generalized familywise error rate. *Electronic Journal of Statistics*, 4:472–485. MR2657378

Han, X., Gu, W., and Fan, J. (2010). Control of the false discovery rate under arbitrary covariance dependence. arXiv preprint arXiv:1012.4397.

Hochberg, Y. and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine*, 9(7):811–818.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, pages 65–70. MR0538597

Jin, J. (2008). Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):461–493. MR2420411

Jin, J. and Cai, T. T. (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *Journal of the American Statistical Association*, 102(478):495–506. MR2325113

Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. i. multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10(4):467–498. MR0599685

Langaas, M., Lindqvist, B. H., and Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(4):555–572. MR2168204

Leek, J. T. and Storey, J. D. (2008). A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences*, 105(48):18718–18723.

Lehmann, E. and Romano, J. P. (2005). Generalizations of the familywise error rate. *The Annals of Statistics*, 33(3):1138–1154. MR2195631

Liang, K. and Nettleton, D. (2012). Adaptive and dynamic adaptive procedures for false discovery rate control and estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):163–182. MR2885844

Liu, F. and Sarkar, S. K. (2011). A new adaptive method to control the false discovery rate. *Recent Advances in Biostatistics: False Discovery Rates, Survival Analysis, and Related Topics. Series in Biostatistics*, 4:3–26.

Pounds, S. and Morris, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics*, 19(10):1236–1242.

Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *The Annals of Statistics*, 30(1):239–257. MR1892663

Sarkar, S. K. (2004). Fdr-controlling stepwise procedures and their false negatives rates. *Journal of Statistical Planning and Inference*, 125(1):119–137. MR2086892

Sarkar, S. K. (2008a). Generalizing simes' test and hochberg's stepup procedure. *The Annals of Statistics*, pages 337–363. MR2387974

Sarkar, S. K. (2008b). On methods controlling the false discovery rate. *Sankhyā: The Indian Journal of Statistics, Series A (2008–)*, pages 135–168. MR2551809

Sarkar, S. K., Guo, W., and Finner, H. (2012). On adaptive procedures controlling the familywise error rate. *Journal of Statistical Planning and Inference*, 142(1):65–78. MR2827130

Schwartzman, A. (2008). Empirical null and false discovery rate inference for exponential families. MR2655662

Schweder, T. and Spjotvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika*, pages 493–502.

Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205. MR2035766

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, 100:9440-9445. MR1994856