

*Research Article*

# Sample Size Calculation for Controlling False Discovery Proportion

**Shulian Shang,<sup>1</sup> Qianhe Zhou,<sup>2</sup>  
Mengling Liu,<sup>1</sup> and Yongzhao Shao<sup>1</sup>**

<sup>1</sup> *Division of Biostatistics, New York University School of Medicine, New York, NY 10016, USA*

<sup>2</sup> *Noxartis Institutes for Biomedical Research, Cambridge, MA 02139, USA*

Correspondence should be addressed to Yongzhao Shao, [yongzhao.shao@nyumc.org](mailto:yongzhao.shao@nyumc.org)

Received 31 March 2012; Accepted 5 June 2012

Academic Editor: Xiaohua Douglas Zhang

Copyright © 2012 Shulian Shang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The false discovery proportion (FDP), the proportion of incorrect rejections among all rejections, is a direct measure of abundance of false positive findings in multiple testing. Many methods have been proposed to control FDP, but they are too conservative to be useful for power analysis. Study designs for controlling the mean of FDP, which is false discovery rate, have been commonly used. However, there has been little attempt to design study with direct FDP control to achieve certain level of efficiency. We provide a sample size calculation method using the variance formula of the FDP under weak-dependence assumptions to achieve the desired overall power. The relationship between design parameters and sample size is explored. The adequacy of the procedure is assessed by simulation. We illustrate the method using estimated correlations from a prostate cancer dataset.

## 1. Introduction

Modern biomedical research frequently involves parallel measurements of a large number of quantities of interest, such as gene expression levels, single nucleotide polymorphism (SNP) and DNA copy number variations. The scientific question can often be formulated as a multiple testing problem. In order to address the multiplicity issue, many methods have been proposed to control the family-wise error rate (FWER), false discovery rate (FDR) or false discovery proportion (FDP). Controlling FDR has been widely used in high-dimensional data analysis [1–3]. FDR is the expected value of the FDP, which is the proportion of incorrect rejections among all rejections. Controlling FDR ensures that the average of FDP from many independently repeated experiments is under control. However, the variability of FDP is ignored, and the actual FDP could be much greater than FDR with high probability, especially

when test statistics are correlated. Consequently, researchers have proposed many procedures to control FDP directly [4–11]:

$$P(\text{FDP} \leq r_1) \geq c_1 \quad (1.1)$$

for given  $r_1$  and  $c_1$ . This is a more stringent criterion than FDR because the proportion of false rejections is bounded above by  $r_1$  with high probability. The FDP controlling procedures are generally too conservative to be useful for the purpose of study design or power analysis.

When we design studies involving multiple testing, it is important to determine sample size to ensure adequate statistical power. Methods for calculating sample size have been proposed to control various criteria, for example, FWER [12–14], FDR [15–20], the number of false discoveries [19, 21] and FDP [22]. For controlling FDP, Oura et al. [22] provided a method to calculate sample size using the beta-binomial model for the sum of rejection status of true alternative hypotheses. It is assumed that only test statistics of true alternative hypotheses are dependent, with a parametric correlation structure. This assumption is restrictive because null test statistics can also be correlated and the dependence structure can be more complicated than the assumed parametric correlation structure. Furthermore, the computation is intensive because computation of the beta-binomial distribution is required. However, to our knowledge this is the only paper that directly deals with this important design problem.

In this paper, we provide a more general method of sample size calculation for controlling FDP under weak-dependence assumptions. Under some assumptions on dependence among test statistics, explicit formulas for the mean and variance of FDP have been derived for each fixed effect size [23]. The formulas elucidate the effects of various design parameters on the variance of FDP. Moreover, the formulas provide a convenient tool to calculate sample size for controlling the FDP. As in [13, 18, 19, 24], we consider the probability of detecting at least a specified proportion of true alternative hypotheses as the power criterion. An iterative computation algorithm for calculating sample size is provided. Simulation experiments indicate that studies with the resultant sample sizes satisfy the power criterion at the given rejection threshold. We illustrate the sample size calculation procedure using a prostate cancer dataset.

## 2. Methods

### 2.1. Notation

Suppose that  $m$  hypotheses are tested simultaneously. Let  $M_0$  denote the index set of  $m_0$  tests for which null hypotheses are true and  $M_1$  the index set of  $m_1 = m - m_0$  tests for which alternative hypotheses are true. Denote the proportion of true null hypotheses by  $\pi_0 = m_0/m$ . We reject a hypothesis if the  $P$  value is less than some threshold  $\alpha$ , and denote the rejection status of the  $i$ th test by  $R_i(\alpha) = I(p_i < \alpha)$ , where  $p_i$  denotes the  $P$  value of the  $i$ th test and  $I(\cdot)$  is an indicator function. The number of rejections is  $R = \sum_{i=1}^m R_i(\alpha)$ . Let the comparison-wise type II error of the  $i$ th test be  $\beta_i$  and the average type II error be

$$\bar{\beta} = \frac{1}{m_1} \sum_{i \in M_1} \beta_i. \quad (2.1)$$

Table 1 summarizes the outcomes of  $m$  tests and their expected values.

**Table 1:** Outcomes and expected outcomes of testing  $m$  hypotheses.

	Outcomes		Total
	Reject $H_0$	Accept $H_0$	
$H_0$ is true	$V$	$m_0 - V$	$m_0$
$H_1$ is true	$U$	$m_1 - U$	$m_1$
Total	$R$	$m - R$	$m$

	Expected outcomes		Total
	Reject $H_0$	Accept $H_0$	
$H_0$ is true	$m_0\alpha$	$m_0(1 - \alpha)$	$m_0$
$H_1$ is true	$m_1(1 - \bar{\beta})$	$m_1\bar{\beta}$	$m_1$
Total	$m_0\alpha + m_1(1 - \bar{\beta})$	$m_0(1 - \alpha) + m_1\bar{\beta}$	$m$

Denote the Pearson correlation coefficient of two rejection indicators by

$$\theta^{ij} = \text{corr} \{R_i(\alpha), R_j(\alpha)\}. \quad (2.2)$$

Furthermore, for  $i, j \in M_0$ , define

$$\theta_V^{ij} = \text{corr} \{R_i(\alpha), R_j(\alpha)\}. \quad (2.3)$$

Let the average correlation be denoted as

$$\bar{\theta}_V = \frac{\sum_{i,j \in M_0, i \neq j} \theta_V^{ij}}{m_0(m_0 - 1)}. \quad (2.4)$$

Similarly, for  $i, j \in M_1$ , we define

$$\theta_U^{ij} = \text{corr} \{R_i(\alpha), R_j(\alpha)\}. \quad (2.5)$$

The average correlation is

$$\bar{\theta}_U = \frac{\sum_{i,j \in M_1, i \neq j} \theta_U^{ij}}{m_1(m_1 - 1)}. \quad (2.6)$$

In addition, for  $i \in M_1, j \in M_0$ , denote

$$\theta_{UV}^{ij} = \text{corr} \{R_i(\alpha), R_j(\alpha)\}. \quad (2.7)$$

Denote the average correlation by

$$\bar{\theta}_{UV} = \frac{\sum_{i \in M_1} \sum_{j \in M_0} \theta_{UV}^{ij}}{m_0 m_1}. \quad (2.8)$$

## 2.2. The Effect of Design Parameters on the Variance of FDP

It has been shown via numerical studies that the variability of FDP increases when test statistics are dependent [25, 26]. But the relationship between design parameters and the variance of FDP has not been examined through analytical formulas. Under the assumptions of common effect size and weak dependence among test statistics, explicit formulas for the mean ( $\mu_Q$ ) and variance ( $\sigma_Q^2$ ) of the FDP have been derived [23]:

$$\mu_Q \approx \frac{\pi_0 \alpha}{\pi_0 \alpha + (1 - \pi_0)(1 - \bar{\beta})}, \quad (2.9)$$

$$\sigma_Q^2 \approx \frac{\pi_0(1 - \pi_0)^2 \alpha(1 - \alpha)(1 - \bar{\beta})}{\{\pi_0 \alpha + (1 - \pi_0)(1 - \bar{\beta})\}^4} \Sigma, \quad (2.10)$$

where

$$\begin{aligned} \Sigma = & \frac{1}{m} \left( 1 - \bar{\beta} + \frac{\pi_0}{1 - \pi_0} \omega \bar{\beta} \right) + \left( \pi_0 - \frac{1}{m} \right) (1 - \bar{\beta}) \bar{\theta}_V \\ & + \pi_0 \omega \bar{\beta} \bar{\theta}_U - 2\pi_0 \sqrt{\omega \bar{\beta} (1 - \bar{\beta})} \bar{\theta}_{UV} \end{aligned} \quad (2.11)$$

and  $\omega = \alpha / (1 - \alpha)$ .

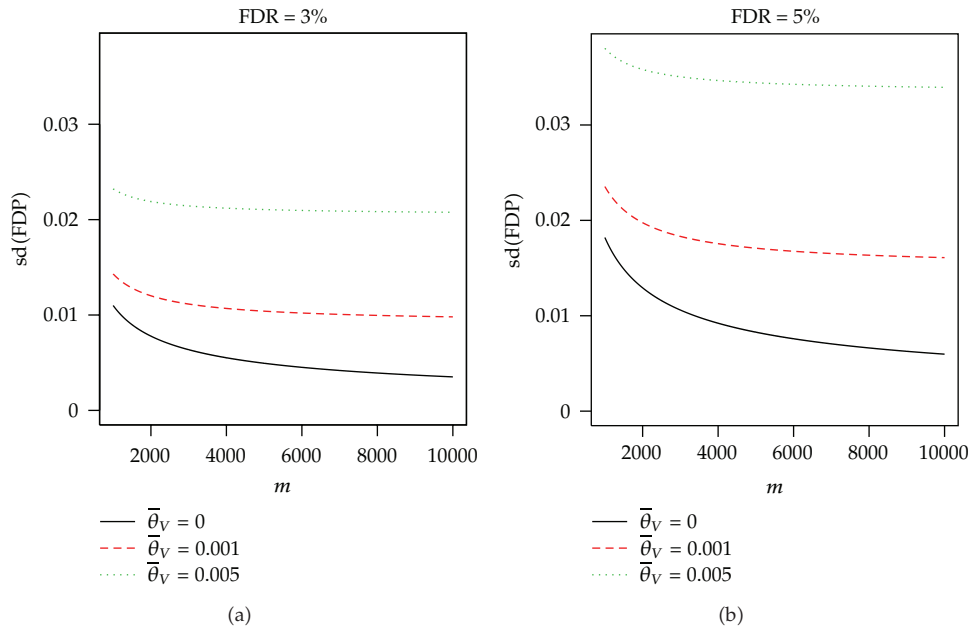
The variance formula (2.10) elucidates the effects of various design parameters on the variance of FDP. To explore the effects, in Figure 1 we calculated  $\sigma_Q$  using (2.10) and plotted it against  $m$  for different correlations  $\bar{\theta}_V$ . We set  $\pi_0 = 0.7$  and  $m$  in the range of 1000 to 10000. The average correlations  $\bar{\theta}_U$  and  $\bar{\theta}_{UV}$  are fixed to be 0.001 and 0, respectively. The levels of  $\alpha$  and  $\bar{\beta}$  are chosen such that FDR is 3% or 5%. At each value of  $\bar{\theta}_V$ ,  $\sigma_Q$  decreases as the number of tests  $m$  increases. The solid line shows the standard deviation of the FDP when  $\bar{\theta}_V$  is 0. When  $\bar{\theta}_V$  is not 0,  $\sigma_Q$  increases evidently. If test statistics are highly correlated, FDP can be much greater than its mean FDR at a given rejection threshold due to its large variability.

In Figure 2, the relationship between  $\sigma_Q$  and  $\pi_0$  was investigated. When other parameters are fixed,  $\sigma_Q$  increases as  $\pi_0$  increases.

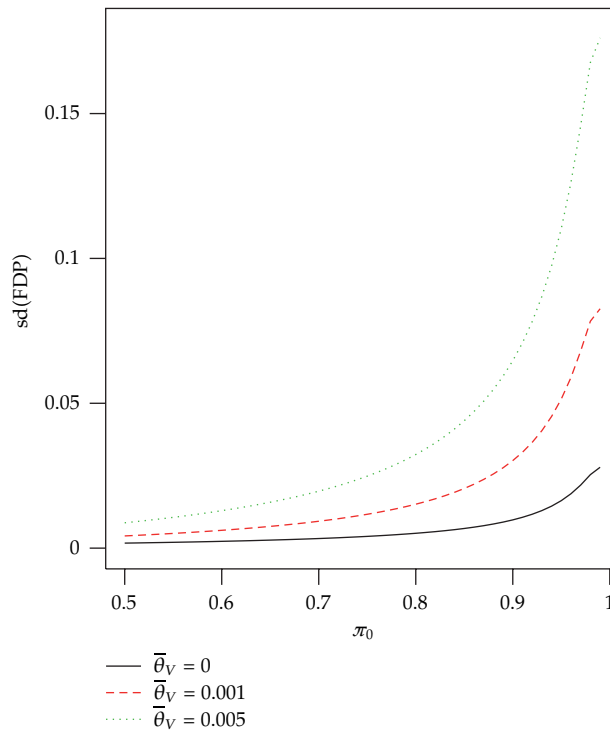
Figure 3 shows that  $\sigma_Q$  increases as  $\bar{\beta}$  increases. When other factors are fixed, the variability of FDP is smaller when the comparison-wise type II error is smaller.

## 2.3. Power and Sample Size Analysis

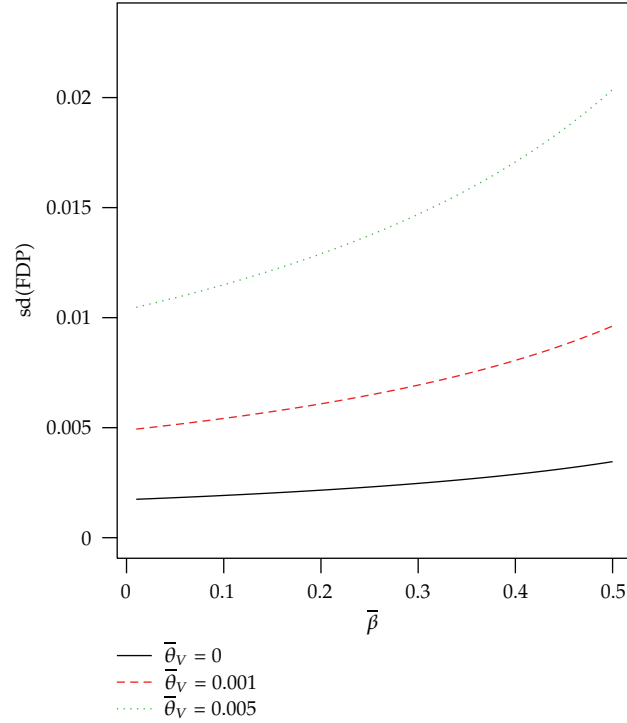
Under some general regularity conditions including weak dependence among test statistics, the FDP follows an asymptotic normal distribution  $N(\mu_Q, \sigma_Q^2)$  [23, 27]. As was pointed out by Shang et al. [23],  $Y = \log(\text{FDP})$  also has an asymptotic normal distribution by the delta



**Figure 1:** Plots of  $sd(FDP)$  at different  $m$  and  $\bar{\theta}_V$ .  $\pi_0 = 0.7$ ,  $\alpha = 0.01$ ,  $\bar{\theta}_U = 0.001$  and  $\bar{\theta}_{UV} = 0$ . (a)  $\bar{\beta} = 0.25$ , FDR = 3%; (b)  $\bar{\beta} = 0.56$ , FDR = 5%.



**Figure 2:** Plots of  $sd(FDP)$  at different  $\pi_0$  and  $\bar{\theta}_V$ .  $m = 10000$ ,  $\alpha = 0.01$ ,  $\bar{\beta} = 0.20$ ,  $\bar{\theta}_U = 0.001$  and  $\bar{\theta}_{UV} = 0$ .



**Figure 3:** Plots of  $\text{sd}(\text{FDP})$  at different  $\bar{\beta}$  and  $\bar{\theta}_V$ .  $m = 10000$ ,  $\pi_0 = 0.7$ ,  $\alpha = 0.004$ ,  $\bar{\theta}_U = 0.001$  and  $\bar{\theta}_{UV} = 0$ .

method, and under weak dependence  $\log(\text{FDP})$  is closer to normal than the FDP itself. The approximate mean and variance of  $Y = \log(\text{FDP})$  are [23]

$$\mu_Y \approx \log(\mu_Q) \approx \log \left\{ \frac{\pi_0 \alpha}{\pi_0 \alpha + (1 - \pi_0)(1 - \bar{\beta})} \right\}, \quad (2.12)$$

$$\sigma_Y^2 \approx \frac{(1 - \pi_0)^2 (1 - \alpha)(1 - \bar{\beta})}{\pi_0 \alpha \{ \pi_0 \alpha + (1 - \pi_0)(1 - \bar{\beta}) \}^2} \Sigma, \quad (2.13)$$

where  $\Sigma$  is in (2.11).

To control FDP with desired power, criterion (1.1) has to be satisfied. Asymptotic normality of  $\log(\text{FDP})$  implies that

$$\Phi \left( \frac{\log r_1 - \mu_Y}{\sigma_Y} \right) \geq c_1, \quad (2.14)$$

where  $\Phi(\cdot)$  is the cumulative distribution function (CDF) of standard normal distribution,  $\mu_Y$  is in (2.12), and  $\sigma_Y^2$  is in (2.13).

There are two commonly used power criteria in multiple testing: the average power, defined as  $E(U/m_1)$ , and the overall power, defined as  $P(U/m_1 \geq r_2)$  for given  $r_2$ . When

a study is designed using the average power criterion, the proportion of true alternative hypotheses rejected will be greater than a prespecified number on average. However, under dependence among test statistics the variability of  $U/m_1$  increases [18], and the study can be underpowered with high probability. Consequently, the overall power has been used in [13, 18, 19, 24] and we also use this power criterion,

$$P\left(\frac{U}{m_1} \geq r_2\right) \geq c_2 \quad (2.15)$$

for given  $r_2$  and  $c_2$ .

Under the weak-dependence assumptions in [18, 23],  $U/m_1$  has an asymptotic normal distribution:

$$\frac{U}{m_1} \sim N\left(1 - \bar{\beta}, \frac{\bar{\beta}(1 - \bar{\beta})}{m_1} \{1 + \bar{\theta}_U(m_1 - 1)\}\right). \quad (2.16)$$

Setting the inequality in (2.15) to equality, the following equation for  $\bar{\beta}$  can be obtained as in [18]:

$$\bar{\beta} = 1 - r_2 - \frac{1 - 2r_2 + \sqrt{4m_1^*r_2(1 - r_2) + 1}}{2m_1^* + 2}, \quad (2.17)$$

where  $m_1^* = m_1 / \{1 + \bar{\theta}_U(m_1 - 1)\} z_{1-c_2}^2$  and  $\Phi(z_{1-c_2}) = c_2$ .

For illustration, consider that a two-sample one-sided  $t$ -test is performed. Let  $\delta$  denote the effect size (mean difference divided by the common standard deviation), and  $a_1$  and  $1 - a_1$  denote the allocation proportion for two groups. We first find  $\alpha$  and  $\bar{\beta}$  which fulfill criteria (1.1) and (2.15). The required sample size  $n$  is the smallest integer satisfying the following inequality:

$$1 - \bar{\beta} \leq 1 - \Gamma_{n-2}\left(-t_{n-2,\alpha} \left| \delta \sqrt{a_1(1 - a_1)n} \right. \right), \quad (2.18)$$

where  $\Gamma_{n-2}(\cdot | b)$  is the CDF of a noncentral  $t$ -distribution with  $n - 2$  degrees of freedom and noncentrality parameter  $b$  and  $t_{n-2,\alpha}$  is the upper  $\alpha$  critical value of the central  $t$ -distribution with  $n - 2$  degrees of freedom.

Following the notation defined in Section 2.1, the correlations can be calculated as:

$$\theta_V^{ij} = \frac{\Psi_{n-2}(-t_{n-2,\alpha}, -t_{n-2,\alpha}; \rho^{ij}) - \alpha^2}{\alpha(1-\alpha)}, \quad (2.19)$$

$$\theta_U^{ij} = \frac{\Psi_{n-2}(t_{n-2,\bar{\beta}}, t_{n-2,\bar{\beta}}; \rho^{ij}) - (1-\bar{\beta})^2}{\bar{\beta}(1-\bar{\beta})}, \quad (2.20)$$

$$\theta_{UV}^{ij} = \frac{\Psi_{n-2}(-t_{n-2,\alpha}, t_{n-2,\bar{\beta}}; \rho^{ij}) - \alpha(1-\bar{\beta})}{\sqrt{\alpha(1-\alpha)\bar{\beta}(1-\bar{\beta})}}, \quad (2.21)$$

where  $\Psi_{n-2}$  is the CDF of a bivariate  $t$ -distribution with  $n - 2$  degrees of freedom and  $\rho^{ij}$  denotes the Pearson correlation between the  $i$ th and  $j$ th test statistics. As can be seen from these formulas, the correlations depend on  $\alpha$  and  $\bar{\beta}$ . No analytical solutions can be found for these two parameters. We use the following iterative computation algorithm to calculate sample size.

*Algorithm.*

- (1) Input design parameters  $r_1, c_1, r_2, c_2, m, \pi_0, \delta, a_1, \rho_U, \rho_V$  and  $\rho_{UV}$ .
- (2) Start from  $\bar{\theta}_U = 0, \bar{\theta}_V = 0$  and  $\bar{\theta}_{UV} = 0$ .
- (3) Calculate  $\bar{\beta}$  from (2.17).
- (4) Using the current values of  $\bar{\theta}_U, \bar{\theta}_V, \bar{\theta}_{UV}$  and  $\bar{\beta}$ , solve for  $\alpha$  from equation  $\Phi((\log r_1 - \mu_Y)/\sigma_Y) = c_1$ .
- (5) Using the current estimates of  $\bar{\beta}$  and  $\alpha$ , calculate  $\theta_V, \theta_U$  and  $\theta_{UV}$  from (2.19), (2.20) and (2.21), respectively. Obtain the average correlations  $\bar{\theta}_V, \bar{\theta}_U$  and  $\bar{\theta}_{UV}$ .
- (6) With updated estimates of  $\bar{\theta}_V, \bar{\theta}_U$  and  $\bar{\theta}_{UV}$ , repeat steps 3 to 5 until the estimates of  $\bar{\beta}$  and  $\alpha$  converge.
- (7) Plug the estimated  $\bar{\beta}$  and  $\alpha$  into (2.18) to solve the sample size.

The estimates of rejection threshold  $\alpha$  and comparison-wise type II error  $\bar{\beta}$  can also be obtained.

### 3. Numerical Studies

#### 3.1. Simulation

The proposed sample size calculation procedure was illustrated for one-sided  $t$ -test comparing the mean of two groups. The effect size  $\delta = 1$  and allocation proportion  $a_1 = 0.5$ . Two types of correlation structures were used: blockwise correlation and autoregressive correlation structure. In the blockwise correlation structure, a proportion of test statistics were correlated in units of blocks. The correlation coefficient within block was a constant, and test statistics were independent across blocks. True null test statistics and true alternative



test statistics were independent. In the autoregressive correlation structure, the correlation matrix  $(\sigma_{ij})$  for dependent test statistics was parameterized by  $\sigma_{ij}(\rho) = \rho^{|i-j|}$ , where  $\sigma_{ij}(\rho)$  is the Pearson correlation coefficient for the  $i$ th and  $j$ th test statistics and  $\rho$  is a correlation parameter.

Oura et al. [22] provided a sample size calculation method for controlling FDP using the beta-binomial model. Only test statistics of true alternative hypotheses are allowed to be dependent, with blockwise correlation structure. For comparison, this method and the sample size calculation procedure for controlling FDR with dependence adjustment in [18] were also assessed. Specifically, the criteria for controlling FDP are

$$P(\text{FDP} \leq 0.05) \geq 0.95, \quad P\left(\frac{U}{m_1} \geq 0.9\right) \geq 0.8. \quad (3.1)$$

The criteria for controlling FDR are

$$\text{FDR} \leq 0.05, \quad P\left(\frac{U}{m_1} \geq 0.9\right) \geq 0.8. \quad (3.2)$$

Table 2 presents the sample size estimates for the blockwise correlation structure. Several parameter configurations were used. The block size is 20 or 100, for  $m = 2000$  or 10000, respectively. We observe that the sample size increases as the correlation between test statistics gets stronger, represented by a greater correlation parameter or a larger proportion of correlated test statistics. When the correlation is fixed, as the number of tests  $m$  increases, the required sample size decreases. With the other parameters fixed, when the number of true alternative hypotheses increases ( $\pi_0$  decreases), the required sample size decreases.

The sample sizes for controlling FDP are greater than those for controlling FDR because controlling FDP is in general more stringent. In the case that  $\pi_0 = 0.9$ ,  $p_v = 0.3$ ,  $\rho_v = 0.6$  and  $m = 2000$  (see Table 2), the sample size for controlling FDP is 81, which is 23% greater than the sample size for controlling FDR. The sample sizes using the method in [22] are in parentheses and are slightly smaller than ours. In terms of computational efficiency, our algorithm converges very fast and generally within 10 steps. The computation is not heavy, and in fact, very similar and comparable to that in [18] for controlling FDR with dependence adjustment. The method of Oura et al. [22] is more computationally intensive. It becomes not feasible when the number of tests or the number of blocks of dependent test statistics is large. Simulation studies show that FDP is controlled and the power is achievable with the sample size given by our procedure at the calculated rejection threshold  $\alpha$  (results not shown).

Table 3 presents the sample sizes for the autoregressive correlation structure. Similar trends for sample size are observed as the design parameters vary. The method in [22] is not applicable to this dependence structure.

### **3.2. Sample Size Calculation Based on a Prostate Cancer Dataset**

We use a prostate cancer dataset as source of correlation structure to illustrate the proposed sample size calculation method while ensuring overall power. The study by Wang et al. [28] investigated the association between mRNA gene expression levels and the aggressive phenotype of prostate cancer. The dataset contains 13935 mRNA measured from 62 patients with aggressive prostate cancer and 63 patients with nonaggressive disease. The method in

**Table 2:** Sample size calculation for controlling FDP and FDR to achieve overall power, blockwise correlation structure.

$\pi_0$	$\rho_u$	$\rho_v$	$p_u$	$p_v$	$m = 2000$		$m = 10000$	
					FDP	FDR	FDP	FDR
0.9	0.2	0.2	0.1	0.1	75	66	68	65
	0.5	0.5	0.1	0.1	77	66	70	65
	0.8	0.8	0.1	0.1	81	67	74	66
	0	0.2	0	0.3	76	66	68	64
	0	0.5	0	0.3	78	66	71	64
	0	0.6	0	0.3	81	66	73	64
	0.2	0	1	0	77 (74)	67	70	67
	0.5	0	1	0	79 (75)	69	72	69
	0.8	0	1	0	82 (77)	72	75	72
0.7	0.2	0.2	0.1	0.1	53	49	50	48
	0.5	0.5	0.1	0.1	54	49	51	48
	0.8	0.8	0.1	0.1	55	49	53	49
	0	0.2	0	0.3	53	48	50	48
	0	0.5	0	0.3	55	48	52	48
	0	0.8	0	0.3	58	48	56	48
	0.2	0	1	0	54 (53)	49	50	49
	0.5	0	1	0	55 (54)	50	50	50
	0.8	0	1	0	56 (55)	52	50	52

$\rho_u$ : correlation between test statistics for which the alternative hypotheses are true;  $\rho_v$ : correlation between test statistics for which the null hypotheses are true;  $p_u$ : proportion of correlated test statistics for which the alternative hypotheses are true;  $p_v$ : proportion of correlated test statistics for which the null hypotheses are true.

[23] was used to estimate the correlation between gene expression levels. The estimated average correlation of expression levels of null genes, alternative genes and between null genes and alternative genes are 0.0040, 0.0043 and  $-0.0005$ , respectively.

Sample size was calculated for one-sided  $t$ -test. The total number of genes is  $m = 10000$ , and the following criteria are to be satisfied:  $P(\text{FDP} \leq 0.10) \geq 0.7$  and  $P(U/m_1 \geq 0.9) \geq 0.8$ . Table 4 presents the sample sizes for various values of  $m_1$ . We performed simulation studies to confirm that these sample sizes provided adequate power at the rejection threshold given by our algorithm. Simulation data were generated with blockwise dependence structure such that the average correlation was close to the estimated correlation from the real dataset.

#### 4. Discussion

In practice, when planning a study one typically needs to make some assumptions. For designing multiple testing studies, a common assumption is that the dependence between test statistics is weak. In this paper, we provide a computationally effective method of sample size calculation for controlling FDP under weak dependence while achieving the desired overall power. This approach uses semiparametric assumptions on dependence structure. We only need to estimate the Pearson correlation between test statistics, and thus this method is applicable to many realistic settings where weak dependence can be assumed. The variance formula of FDP provides a convenient tool to uncover the relationship between the design

**Table 3:** Sample size calculation for controlling FDP and FDR to achieve overall power, autoregressive correlation structure.

$\pi_0$	$\rho_u$	$\rho_v$	$p_u$	$p_v$	$m = 2000$		$m = 10000$	
					FDP	FDR	FDP	FDR
0.9	0.2	0.2	0.1	0.1	75	66	67	64
	0.5	0.5	0.1	0.1	75	66	67	64
	0.8	0.8	0.1	0.1	77	66	68	64
	0.2	0.2	0.4	0.4	75	66	67	64
	0.5	0.5	0.4	0.4	76	66	68	64
	0.8	0.8	0.4	0.4	81	66	69	64
0.7	0.2	0.2	0.1	0.1	53	48	49	48
	0.5	0.5	0.1	0.1	53	48	49	48
	0.8	0.8	0.1	0.1	54	48	50	48
	0.2	0.2	0.4	0.4	53	49	49	48
	0.5	0.5	0.4	0.4	54	49	50	48
	0.8	0.8	0.4	0.4	56	49	51	48

$\rho_u$ : correlation parameter for test statistics for which the alternative hypotheses are true;  $\rho_v$ : correlation parameter for test statistics for which the null hypotheses are true;  $p_u$ : proportion of correlated test statistics for which the alternative hypotheses are true;  $p_v$ : proportion of correlated test statistics for which the null hypotheses are true.

**Table 4:** Sample size calculation using the prostate cancer dataset.

$\pi_0$	$m_1$	$n$	$\hat{P}(\text{FDP} \leq 0.10)$	$\hat{P}((U/m_1) \geq 0.9)$	$\hat{\alpha}$
0.95	500	74	0.830	0.915	0.003
0.9	1000	63	0.855	0.970	0.007
0.85	1500	57	0.885	0.955	0.012
0.8	2000	52	0.875	0.945	0.018
0.75	2500	48	0.870	0.955	0.026
0.7	3000	44	0.915	0.925	0.034

$\hat{P}(\text{FDP} \leq 0.10)$  and  $\hat{P}((U/m_1) \geq 0.9)$ : empirical probability from 200 simulation runs.

parameters and the variability of FDP under the assumption of weak dependence. Simulation studies indicate that the algorithm is computationally efficient and stable.

We have used one-sided  $t$ -test to illustrate the method, and the procedure can be easily extended to two-sided  $t$ -test and other tests. Common effect size is assumed in the sample size calculation algorithm. In practice, one can try different effect sizes, see the range of sample sizes and the variability and then make a decision. Effects of variations on other parameters such as  $\pi_0$  can be examined similarly.

## Acknowledgments

We would like to thank the reviewers and the Editor for their careful reading of our paper and the constructive suggestions. This research was partially supported by a Stony Wold-Herbert Foundation grant and NIH grants 2P30 CA16087, 5P30 ES00260, UL1 TR000038 and R03 CA153083.

## References

- [1] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society B*, vol. 57, no. 1, pp. 289–300, 1995.
- [2] J. D. Storey, "A direct approach to false discovery rates," *Journal of the Royal Statistical Society B*, vol. 64, no. 3, pp. 479–498, 2002.
- [3] J. D. Storey, J. E. Taylor, and D. Siegmund, "Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach," *Journal of the Royal Statistical Society B*, vol. 66, no. 1, pp. 187–205, 2004.
- [4] A. Farcomeni, "Generalized augmentation to control the false discovery exceedance in multiple testing," *Scandinavian Journal of Statistics*, vol. 36, no. 3, pp. 501–517, 2009.
- [5] Y. Ge, S. C. Sealfon, and T. P. Speed, "Multiple testing and its applications to microarrays," *Statistical Methods in Medical Research*, vol. 18, no. 6, pp. 543–563, 2009.
- [6] C. R. Genovese and L. Wasserman, "Exceedance control of the false discovery proportion," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1408–1417, 2006.
- [7] C. Genovese and L. Wasserman, "A stochastic process approach to false discovery control," *The Annals of Statistics*, vol. 32, no. 3, pp. 1035–1061, 2004.
- [8] E. L. Korn, M.-C. Li, L. M. McShane, and R. Simon, "An investigation of two multivariate permutation methods for controlling the false discovery proportion," *Statistics in Medicine*, vol. 26, no. 24, pp. 4428–4440, 2007.
- [9] E. L. Korn, J. F. Troendle, L. M. McShane, and R. Simon, "Controlling the number of false discoveries: application to high-dimensional genomic data," *Journal of Statistical Planning and Inference*, vol. 124, no. 2, pp. 379–398, 2004.
- [10] N. Meinshausen, "False discovery control for multiple tests of association under general dependence," *Scandinavian Journal of Statistics*, vol. 33, no. 2, pp. 227–237, 2006.
- [11] M. J. van der Laan, S. Dudoit, and K. S. Pollard, "Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives," *Statistical Applications in Genetics and Molecular Biology*, vol. 3, article 15, 2004.
- [12] S. H. Jung, H. Bang, and S. Young, "Sample size calculation for multiple testing in microarray data analysis," *Biostatistics*, vol. 6, no. 1, pp. 157–169, 2005.
- [13] S. J. Wang and J. J. Chen, "Sample size for identifying differentially expressed genes in microarray experiments," *Journal of Computational Biology*, vol. 11, no. 4, pp. 714–726, 2004.
- [14] M. C. K. Yang, J. J. Yang, R. A. McIndoe, and J. X. She, "Microarray experimental design: power and sample size considerations," *Physiological Genomics*, vol. 16, pp. 24–28, 2003.
- [15] S. H. Jung, "Sample size for  $fdr$ -control in microarray data analysis," *Bioinformatics*, vol. 21, no. 14, pp. 3097–3104, 2005.
- [16] S. S. Li, J. Bigler, J. W. Lampe, J. D. Potter, and Z. Feng, "FDR-controlling testing procedures and sample size determination for microarrays," *Statistics in Medicine*, vol. 24, no. 15, pp. 2267–2280, 2005.
- [17] S. Pounds and C. Cheng, "Sample size determination for the false discovery rate," *Bioinformatics*, vol. 21, no. 23, pp. 4263–4271, 2005.
- [18] Y. Shao and C.-H. Tseng, "Sample size calculation with dependence adjustment for FDR-control in microarray studies," *Statistics in Medicine*, vol. 26, no. 23, pp. 4219–4237, 2007.
- [19] C. A. Tsai, S. J. Wang, D. T. Chen, and J. J. Chen, "Sample size for gene expression microarray experiments," *Bioinformatics*, vol. 21, no. 8, pp. 1502–1508, 2005.
- [20] X. D. Zhang, *Optimal High-Throughput Screening: Practical Experimental Design and Data Analysis for Genome-Scale Rnai Research*, Cambridge University Press, 2011.
- [21] M. L. T. Lee and G. A. Whitmore, "Power and sample size for dna microarray studies," *Statistics in Medicine*, vol. 21, no. 23, pp. 3543–3570, 2002.
- [22] T. Oura, S. Matsui, and K. Kawakami, "Sample size calculations for controlling the distribution of false discovery proportion in microarray experiments," *Biostatistics*, vol. 10, no. 4, pp. 694–705, 2009.
- [23] S. Shang, M. Liu, and Y. Shao, "A tight prediction interval for false discovery proportion under dependence," *Open Journal of Statistics*, vol. 2, no. 2, pp. 163–171, 2012.
- [24] W.-J. Lin, H.-M. Hsueh, and J. J. Chen, "Power and sample size estimation in microarray studies," *Bmc Bioinformatics*, vol. 11, article 48, 2010.
- [25] R. Heller, "Correlated  $z$ -values and the accuracy of large-scale statistical estimates," *Journal of the American Statistical Association*, vol. 105, no. 491, pp. 1057–1059, 2010.
- [26] Y. Pawitan, S. Calza, and A. Ploner, "Estimation of false discovery proportion under general dependence," *Bioinformatics*, vol. 22, no. 24, pp. 3025–3031, 2006.

- [27] A. Farcomeni, "Some results on the control of the false discovery rate under dependence," *Scandinavian Journal of Statistics*, vol. 34, no. 2, pp. 275–297, 2007.
- [28] L. Wang, H. Tang, V. Thayanithy et al., "Gene networks and microRNAs implicated in aggressive prostate cancer," *Cancer Research*, vol. 69, no. 24, pp. 9490–9497, 2009.





# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

