Multiple Testing Procedures for Hierarchically Related Hypotheses

Przemyslaw Biecek

Institute of Mathematics and Computer Science
Wroclaw University of Technology
Poland

MCP Vienna 10 July 2007

Outline

1. Brief introduction to multiple hypothesis testing,
2. Testing under a hierarchical relation,
3. Testing procedures that control:
   - FWER,
   - PFER,
   - FDR.
4. Simulation study,
5. Application to Gene Ontology.

A set of null hypotheses $\mathcal{H}$

- Consider a set of m random experiments $\{\mathcal{E}^{(i)} : i \in \{1, ..., m\}\}$. For each experiment $\mathcal{E}^{(i)}$ we introduce a family of distributions $\mathcal{P}^{(i)} = \{P^{(i)}_{\theta^{(i)}} : \theta^{(i)} \in \Theta^{(i)}\}$.
- We assume that observations generated by experiment $\mathcal{E}^{(i)}$ follow distribution $P^{(i)}_{\theta^{(i)}_0}$, for some $\theta^{(i)}_0 \in \Theta^{(i)}$.
- For each random experiment $\mathcal{E}^{(i)}$ we specify null and alternative hypotheses. The set of all null hypotheses is denoted by $\mathcal{H} = \{H^{(i)}_0 : i = 1, ..., m\}$.

Null hypothesis and alternative hypothesis

Null hypothesis $H^{(i)}_0$ is the supposition that $\theta^{(i)}_0 \in \Theta^{(i)}_0$.
Alternative hypothesis $H^{(i)}_A$ is the supposition that $\theta^{(i)}_0 \notin \Theta^{(i)}_0$.

Standard notation

The outcomes of testing the set $\mathcal{H}$ may be described in the following form.

	#accepted nulls	#rejected nulls	
#true nulls	U	V	m_0
#false nulls	T	S	m_1
sum	m - R	R	m.

Variables V and T denote number of wrong decisions.

The error rates

Per-family error rate (PFER)

$$PFER = E(V).$$

Family-wise error rate (FWER)

$$FWER = Pr(V > 0).$$

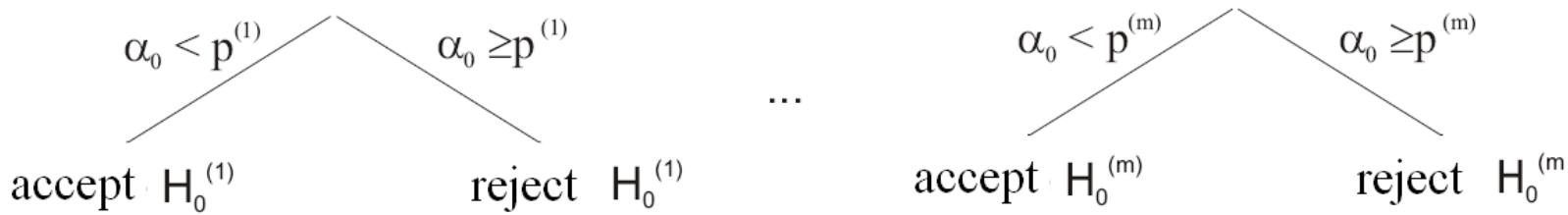
False discovery rate (FDR)

$$FDR = E(Q),$$

where

$$Q = \begin{cases} V/R & R > 0, \\ 0 & R = 0. \end{cases}$$

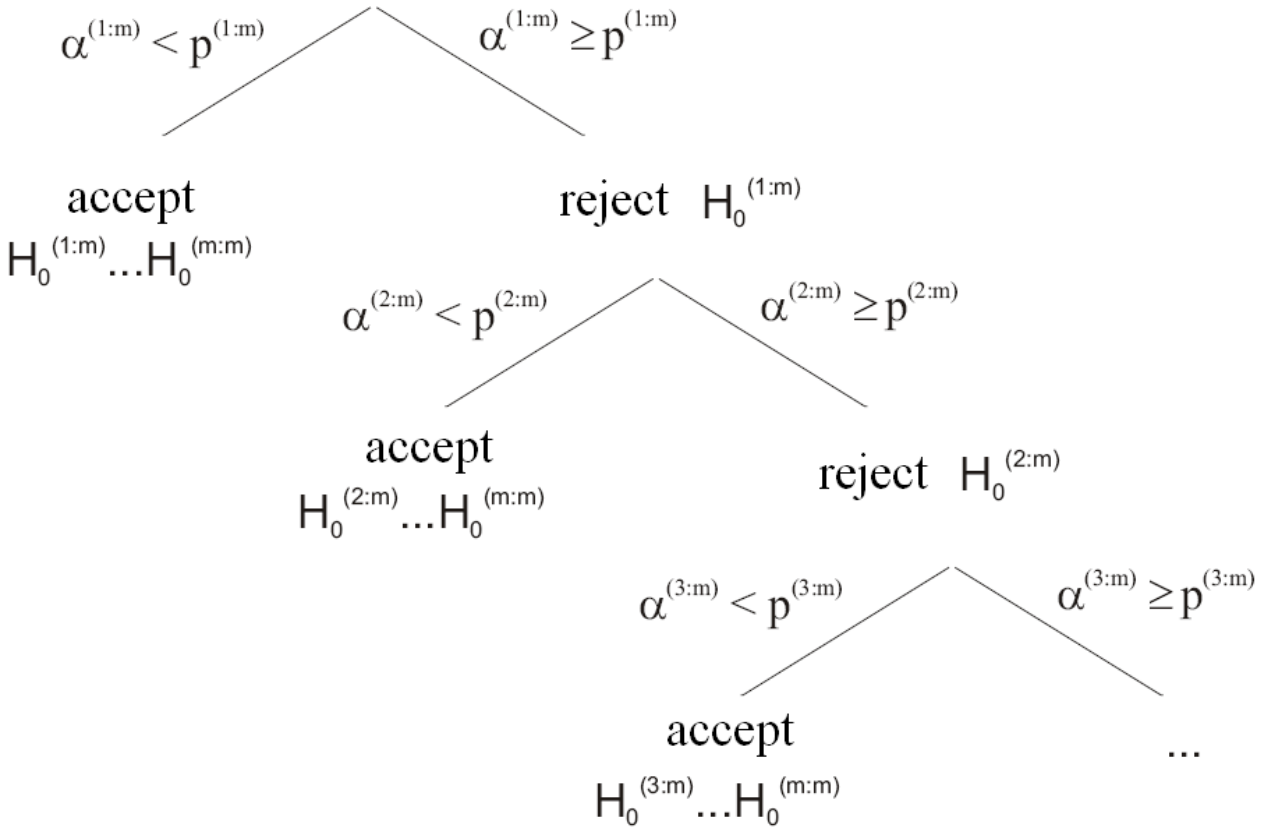
Single step procedure



Bonferroni procedure (1936)

The single step testing procedure with significance level $\alpha_0 = \alpha/m$, controls the FWER at the level α.

Step down


$\alpha^{(1:m)} < p^{(1:m)}$
$\alpha^{(1:m)} \geq p^{(1:m)}$
accept
$H_0^{(1:m)} \ldots H_0^{(m:m)}$
reject $H_0^{(1:m)}$
$\alpha^{(2:m)} < p^{(2:m)}$
$\alpha^{(2:m)} \geq p^{(2:m)}$
accept
$H_0^{(2:m)} \ldots H_0^{(m:m)}$
reject $H_0^{(2:m)}$
$\alpha^{(3:m)} < p^{(3:m)}$
$\alpha^{(3:m)} \geq p^{(3:m)}$
accept
$H_0^{(3:m)} \ldots H_0^{(m:m)}$
...

Holm procedure (1979)

The step-down testing procedure with significance levels

$$\alpha^{(i:m)} = \alpha/(m - i + 1), \tag{1}$$

controls FWER at the level α.

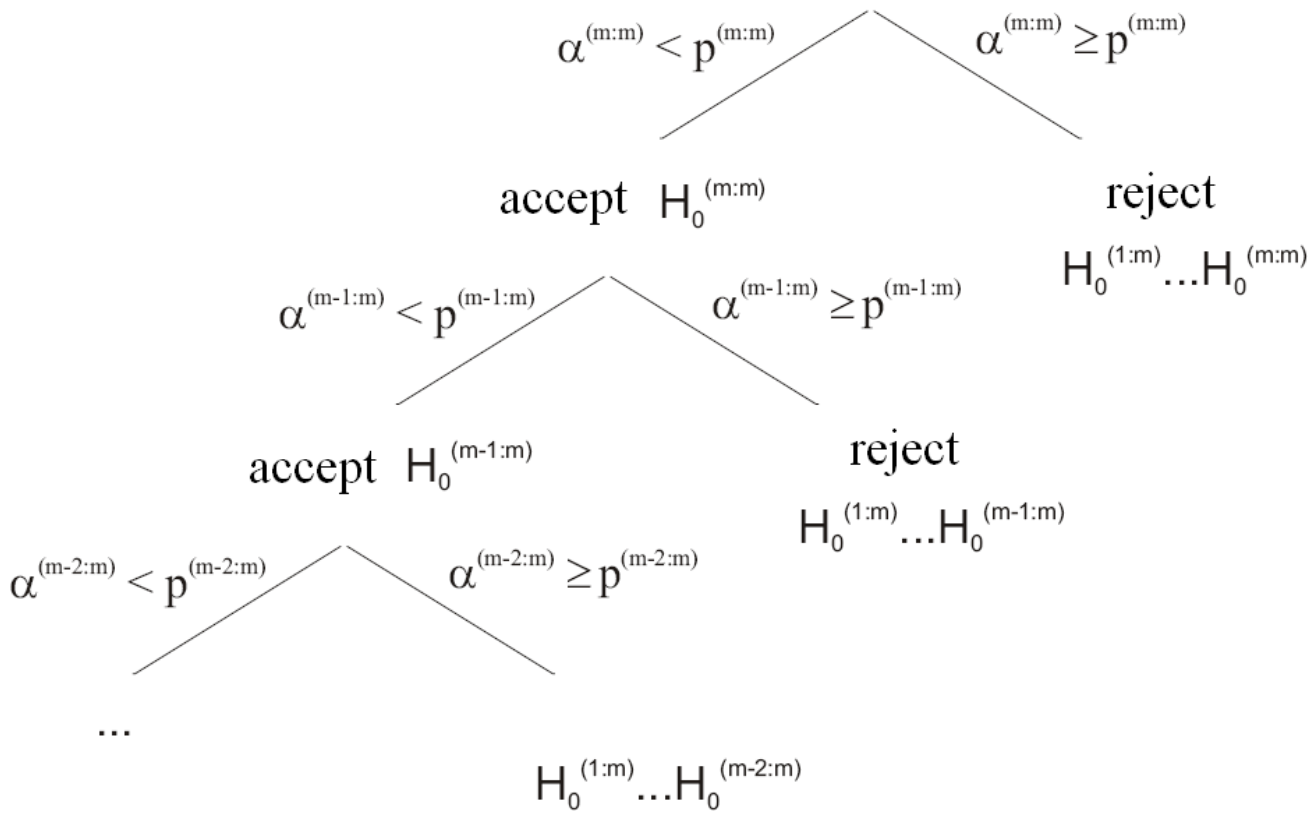
Hochberg procedure (1988)

The step-up testing procedure with significance levels

$$\alpha^{(i:m)} = \alpha/(m - i + 1), \tag{2}$$

controls FWER at the level α.

Step up


$\alpha^{(m:m)} < p^{(m:m)}$
$\alpha^{(m:m)} \geq p^{(m:m)}$
accept $H_0^{(m:m)}$
reject
$H_0^{(1:m)} \ldots H_0^{(m:m)}$
$\alpha^{(m-1:m)} < p^{(m-1:m)}$
$\alpha^{(m-1:m)} \geq p^{(m-1:m)}$
accept $H_0^{(m-1:m)}$
reject
$H_0^{(1:m)} \ldots H_0^{(m-1:m)}$
$\alpha^{(m-2:m)} < p^{(m-2:m)}$
$\alpha^{(m-2:m)} \geq p^{(m-2:m)}$
...
$H_0^{(1:m)} \ldots H_0^{(m-2:m)}$

Benjamini and Hochberg procedure (1995)

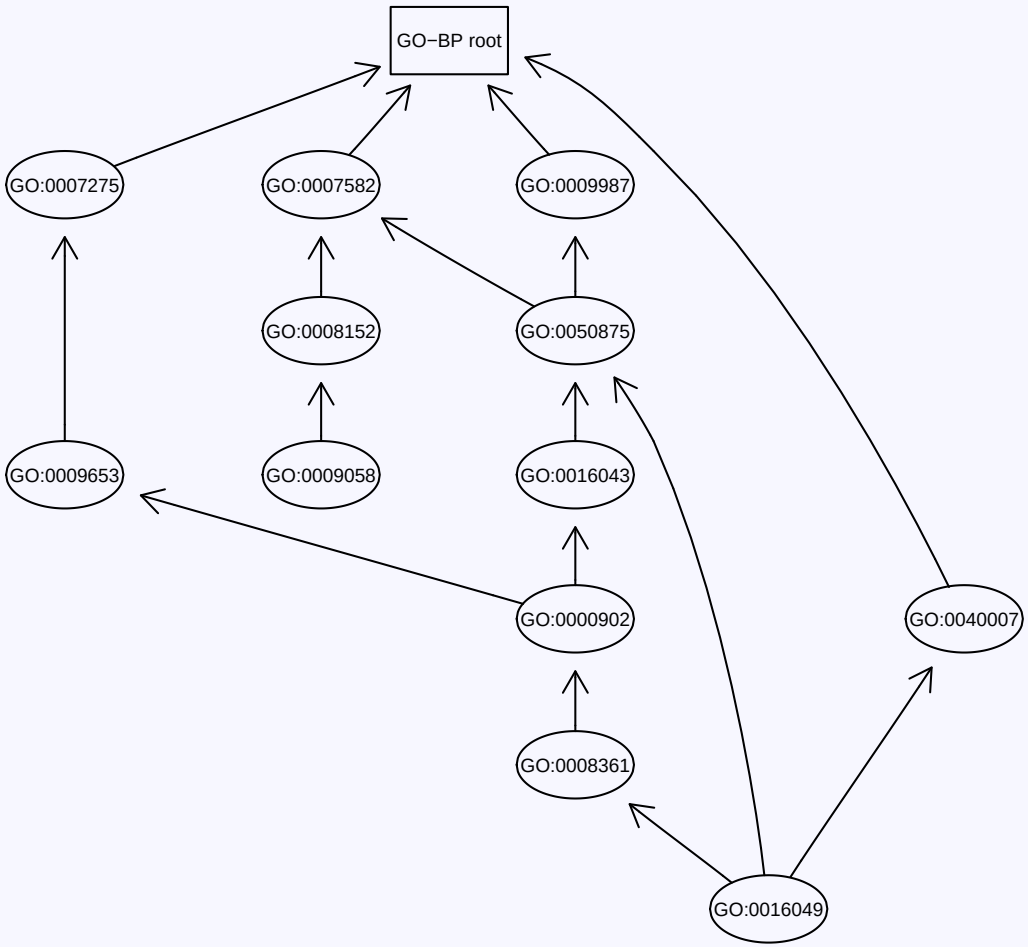
The step-up procedure with significance levels

$$\alpha^{(i:m)} = \frac{i}{m}\alpha, \tag{3}$$

controls FDR at the level α.

The control of PFER and FWER is more conservative than control of FDR (i.e. leads to a smaller number of rejections). In applications the procedures that control FDR are more popular.

Gene Ontology



- There are some papers concerned with the control of FWER under such hierarchical relations (closure testing). The first results were published by Gabriel (1969), R. Marcus (1976), U. Naik (1977), more recently by H. Finner (2002).
- In these papers one common parameter space Θ is introduced for all null hypotheses. The relation between hypotheses is derived from the relation between the sets $\Theta_0^{(i)}$, e.g. if $\Theta_0^{(i)} \subset \Theta_0^{(j)}$, then the rejection of $H_0^{(j)}$ requires the rejection of $H_0^{(i)}$.
  This approach is very interesting, but hard to apply to the analysis with the Gene Ontology, since it is not obvious how to describe the sets $\Theta_0^{(i)}$ in terms of a common parameter space.
- We propose a different approach to incorporating hierarchical relations into the testing scheme.

Set of null hypotheses

As before, consider the set of m random experiments $\{\mathcal{E}^{(i)} : i \in \{1, ..., m\}\}$, the corresponding set $\{\Theta^{(i)} : i \in \{1, ..., m\}\}$ and the set of null hypotheses $\mathcal{H} = \{H_0^{(i)} : i \in \{1, ..., m\}\}$ of the form $H_0^{(i)} : \theta_0^{(i)} \in \Theta_0^{(i)}$.

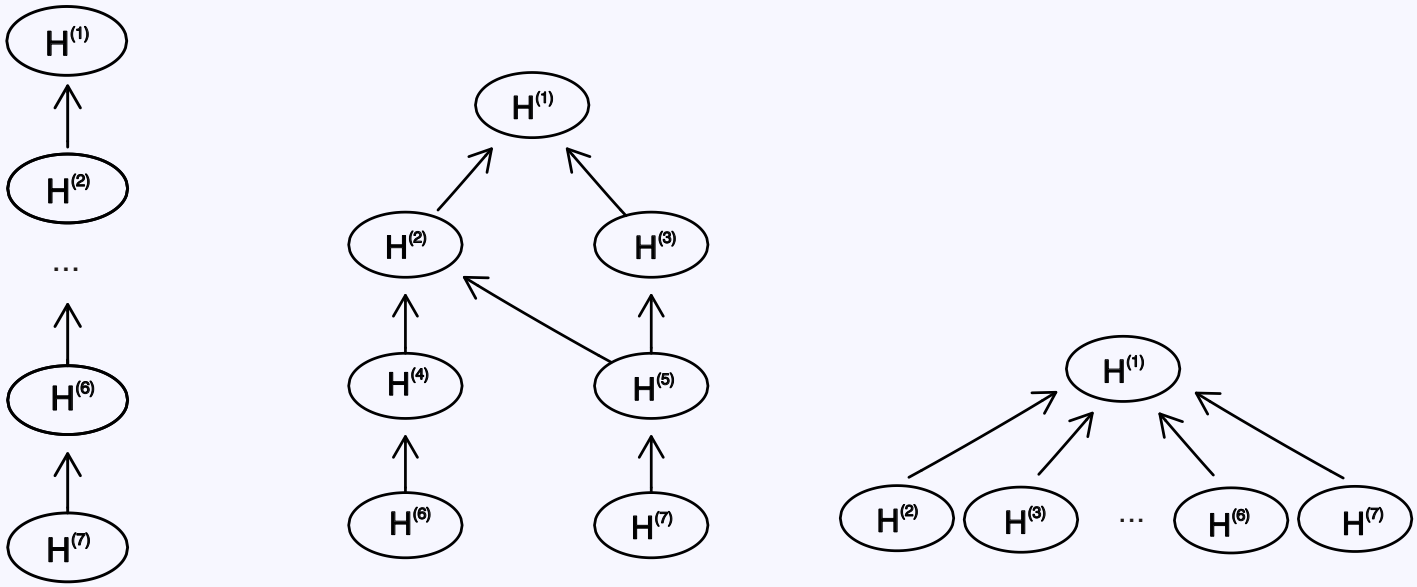
Hierarchical relation $\mathcal{R}$

Let $\mathcal{R}$ denote an irreflexive (aliorelative), asymmetric and transitive binary relation on the set $\mathcal{H}$. Relation $\mathcal{R}$ fulfills following conditions

1. $R(i, i) = 0$,
2. $(R(i, j) = R(j, k) = 1) \Rightarrow (R(i, k) = 1)$,
3. $R(i, j) * R(j, i) = 0$.

Relation $\mathcal{R}$ describes the possible states of null hypotheses.

$$(R(i,j) = 1) \Rightarrow \left(H_0^{(j)} \text{ is true } \Rightarrow H_0^{(i)} \text{ is true } \right).$$



Testing outcomes

Let $\psi = (\psi_1, ..., \psi_m)$, where $\psi_i \in \{0, 1\}$, denote the outcomes of the tests as standard.

Coherency

The outcomes of the tests ψ are coherent with relation $\mathcal{R}$ if and only if

$$(R(i,j) = 1) \Rightarrow (\psi_i \geq \psi_j)\,.$$

We want to obtain coherent results!!!

Control of FWER

Closure of the outcomes of tests

Let $\hat{\psi}$ stand for the closure of outcomes of tests ψ, where

$$\hat{\psi}_i = \max(\psi_i, \{\psi_j : R(i,j) = 1\}).$$

In other words, this means that the rejection of a test leads to rejection of all related tests.

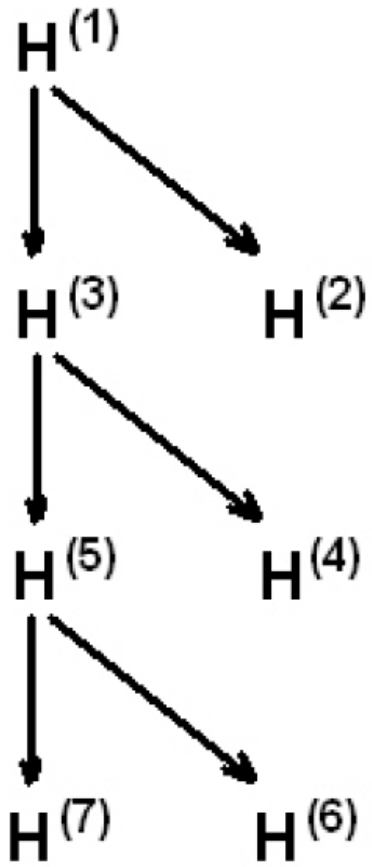
Theorem 1

Closure of the outcomes of tests does not affect FWER.

Note

Closure of the outcomes of tests results in an increase of both PFER and FDR.

Follow up testing procedure



Step 1 We test $H_0^{(7)}$

$p^{(7)} > \alpha^{(7)}$, thus we accept $H_0^{(7)}$

Step 2: We test $H_0^{(6)}$

$p^{(6)} < \alpha^{(6)}$, thus we reject $H_0^{(1)}$, $H_0^{(3)}$, $H_0^{(5)}$, $H_0^{(6)}$

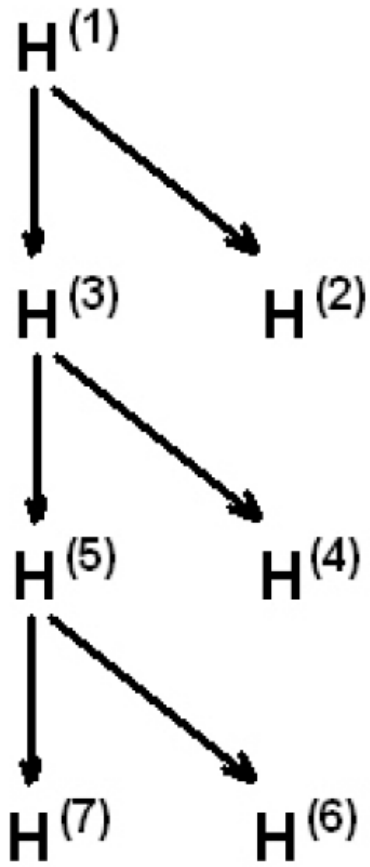
Step 3 We test $H_0^{(4)}$

$p^{(4)} > \alpha^{(4)}$, thus we accept $H_0^{(4)}$

Step 4 We test $H_0^{(2)}$

$p^{(2)} > \alpha^{(2)}$, thus we accept $H_0^{(2)}$

Follow down testing procedure



Step 1: We test $H_0^{(1)}$

$p^{(1)} < \alpha^{(1)}$, thus we reject $H_0^{(1)}$

Step 2: We test $H_0^{(3)}$

$p^{(3)} > \alpha^{(3)}$, thus we accept

$H_0^{(3)}, H_0^{(4)}, H_0^{(5)}, H_0^{(6)}, H_0^{(7)}$

Step 3: We test $H_0^{(2)}$

$p^{(2)} < \alpha^{(2)}$, thus we reject $H_0^{(2)}$

The control of PFER

Theorem 2

Let $\phi(i)$ stand for the maximum cardinality of the set of unrelated hypotheses which contains $H_0^{(i)}$

$$\phi(i) = \max_{A \in \mathcal{C}_i} \#A,$$

where

$$\mathcal{C}_i = \{C : i \in C \wedge \forall_{j,k \in C} R(j,k) = 0\}.$$

The follow down strategy with significance levels

$$\alpha^{(i)} = \alpha / \left[\phi(i)(1+\alpha)\right] \tag{4}$$

controls PFER at the level α.

The control of PFER

Theorem 3

Let $\rho(i)$ stand for the number of hypotheses related to $H_0^{(i)}$

$$\rho(i) = 1 + \#\{j : R(j, i) = 1\}.$$

The follow up procedure with significance levels

$$\alpha^{(i)} = \alpha/[m * \rho(i)] \tag{5}$$

controls PFER at the level α.

The control of FDR

Theorem 4

Let $\mathcal{C}_i$ stand for the family of sets

$$\mathcal{C}_i = \{C : i \in C \wedge \forall_{j,k \in C} R(j,k) = 0\}.$$

Let $\rho(B)$ stand for the number of hypotheses related to any hypothesis from the set B

$$\rho(B) = 1 + \#\{j : i \in B \wedge R(j,i) = 1\}.$$

The follow down strategy with significance levels

$$\alpha^{(i)} = \min_{B \in \mathcal{C}_i} \{\alpha_{lin}^{(\rho(B),m)} / \#B\}, \tag{6}$$

controls FDR at the level α.

Control of the FDR

Theorem 4 cont.

Significance levels $\alpha_{lin}^{(i,m)}$ are derived in the following way

$$
\begin{aligned}
\alpha_{lin}^{(1,m)} &= \alpha, \\
\alpha_{lin}^{(i,m)} &= \min\left(0.5, \alpha\left[\sum_{k=i}^{m-1}\frac{k-i+1}{k}(1-\alpha_{lin}^{(k+1,m)})\prod_{l=i+1}^{k}\alpha_{lin}^{(l,m)} + \right.\right. \\
&\qquad \left.\left.\frac{m-i+1}{m}\prod_{l=i+1}^{m}\alpha_{lin}^{(l,m)}\right]^{-1}\right), \text{ for } 1 < i < m, \\
\alpha_{lin}^{(m,m)} &= \min(0.5, m\alpha).
\end{aligned}
\tag{7}
$$

Example: Relation $\mathcal{R}_1$

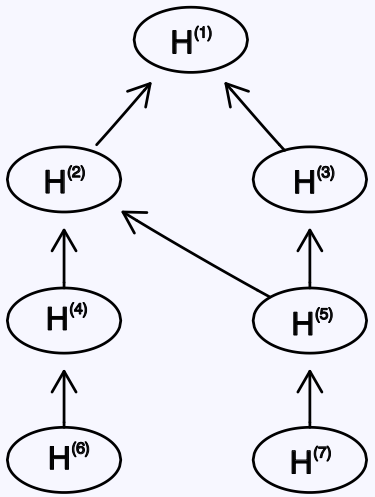
$\mathcal{R}_1(i,j)$	j=1	j=2	j=3	j=4	j=5	j=6	j=7
i=1	0	1	1	1	1	1	1
i=2	0	0	1	1	1	1	1
i=3	0	0	0	1	1	1	1
i=4	0	0	0	0	1	1	1
i=5	0	0	0	0	0	1	1
i=6	0	0	0	0	0	0	1
i=7	0	0	0	0	0	0	0

Significance levels for $\mathcal{R}_1$

	Relation $\mathcal{R}_1$				
	$\rho(i)$	$\phi(i)$	$\alpha_{FU}^{PFER}(i)$	$\alpha_{FD}^{PFER}(i)$	$\alpha_{FD}^{FDR}(i)$
i=1	1	1	0.00714	0.04761	0.05000
i=2	2	1	0.00357	0.04761	0.09523
i=3	3	1	0.00238	0.04761	0.13636
i=4	4	1	0.00178	0.04761	0.17391
i=5	5	1	0.00142	0.04761	0.20833
i=6	6	1	0.00119	0.04761	0.24000
i=7	7	1	0.00102	0.04761	0.35000

Example: Relation $\mathcal{R}_2$

$\mathcal{R}_2(i,j)$	j=1	j=2	j=3	j=4	j=5	j=6	j=7
i=1	0	1	1	1	1	1	1
i=2	0	0	0	1	0	1	0
i=3	0	0	0	0	1	0	1
i=4	0	0	0	0	0	1	0
i=5	0	0	0	0	0	0	1
i=6	0	0	0	0	0	0	0
i=7	0	0	0	0	0	0	0



Significance levels for $\mathcal{R}_2$

	Relation $\mathcal{R}_2$				
	$\rho(i)$	$\phi(i)$	$\alpha_{FU}^{PFER}(i)$	$\alpha_{FD}^{PFER}(i)$	$\alpha_{FD}^{FDR}(i)$
i=1	1	1	0.00714	0.04761	0.05000
i=2	2	2	0.00357	0.02380	0.04762
i=3	2	2	0.00357	0.02380	0.04762
i=4	3	2	0.00238	0.02380	0.06818
i=5	4	2	0.00178	0.02380	0.08696
i=6	4	2	0.00178	0.02380	0.10417
i=7	5	2	0.00142	0.02380	0.10417

Simulation design

Consider a set of 7 experiments. Observations are drawn from $\mathcal{N}(\mu, I_{7\times 7})$, where $\mu = (\mu_1, ..., \mu_7)$.
The corresponding null hypotheses are of the form $H_0^{(i)} : \mu_i = 0$.
We consider 6 different scenarios for μ

$$\begin{aligned}
\mu^{(1)} &= (0, 0, 0, 0, 0, 0, 0),\\
\mu^{(2)} &= (2, 2, 2, 2, 0, 0, 0),\\
\mu^{(3)} &= (3, 2.75, 2.5, 2.25, 0, 0, 0),\\
\mu^{(4)} &= (2, 2, 2, 2, 2, 2, 2),\\
\mu^{(5)} &= (3, 2.75, 2.5, 2.25, 2, 1.75, 1.5),\\
\mu^{(6)} &= (3, 0, 0, 0, 0, 0, 0).
\end{aligned}$$

Closure of testing procedures

		closure of single-step Bonferroni procedure	closure of step-up Benjamini-Hochberg proce
Relation	Vector μ	PFER	FDR
$\mathcal{R}_1$	$\mu^{(1)}$	0.201	0.050
$\mathcal{R}_1$	$\mu^{(2)}$	0.073	0.046
$\mathcal{R}_1$	$\mu^{(6)}$	0.151	0.061
$\mathcal{R}_2$	$\mu^{(1)}$	0.151	0.050
$\mathcal{R}_2$	$\mu^{(2)}$	0.045	0.035
$\mathcal{R}_2$	$\mu^{(6)}$	0.102	0.056
$\mathcal{R}_3$	$\mu^{(1)}$	0.093	0.050
$\mathcal{R}_3$	$\mu^{(2)}$	0.294	0.028
$\mathcal{R}_3$	$\mu^{(6)}$	0.044	0.042

Results based on 1 000 000 repetitions!

Results for proposed testing procedures

Relation	Vector μ	follow-up PFER	follow-down PFER	follow-down FDR	follow-up E(S)	follow-down E(S)	follow-down E(S)
$\mathcal{R}_1$	$\mu^{(1)}$	0.050	0.050	0.050	—	—	—
$\mathcal{R}_1$	$\mu^{(2)}$	0.007	0.008	0.017	1.712	1.433	1.852
$\mathcal{R}_1$	$\mu^{(3)}$	0.007	0.022	0.035	2.503	2.761	3.236
$\mathcal{R}_1$	$\mu^{(4)}$	—	—	—	3.384	1.634	2.670
$\mathcal{R}_1$	$\mu^{(5)}$	—	—	—	3.425	3.258	4.842
$\mathcal{R}_1$	$\mu^{(6)}$	0.031	0.045	0.045	0.712	0.909	0.912
$\mathcal{R}_2$	$\mu^{(1)}$	0.050	0.050	0.050	—	—	—
$\mathcal{R}_2$	$\mu^{(2)}$	0.006	0.008	0.010	1.552	1.432	1.722
$\mathcal{R}_2$	$\mu^{(3)}$	0.006	0.022	0.023	2.314	2.679	3.034
$\mathcal{R}_2$	$\mu^{(4)}$	—	—	—	2.975	1.638	2.265
$\mathcal{R}_2$	$\mu^{(5)}$	—	—	—	3.092	3.184	4.186
$\mathcal{R}_2$	$\mu^{(6)}$	0.028	0.044	0.043	0.712	0.909	0.912
$\mathcal{R}_3$	$\mu^{(1)}$	0.050	0.050	0.050	—	—	—
$\mathcal{R}_3$	$\mu^{(2)}$	0.011	0.015	0.006	1.448	1.272	1.303
$\mathcal{R}_3$	$\mu^{(3)}$	0.011	0.022	0.007	2.225	2.365	2.394
$\mathcal{R}_3$	$\mu^{(4)}$	—	—	—	2.345	1.915	1.967
$\mathcal{R}_3$	$\mu^{(5)}$	—	—	—	2.784	3.069	3.116
$\mathcal{R}_3$	$\mu^{(6)}$	0.021	0.043	0.022	0.714	0.909	0.912

The presented procedures were applied to functional enrichment analysis. We used a data set provided by Adam Zagdanski, which contains protein-protein interactions and GO BP annotations.
To predict functional attribute for a protein, we use Fisher's exact test which identifies overrepresented functional attributes

$$p^{(i,j)} = 1 - \sum_{l=0}^{k_{i,j}-1} \frac{\binom{k_j}{l}\binom{n-k_j}{n_i-l}}{\binom{n}{n_i}},$$

where

- n, number of proteins,
- k_j, number of proteins with functional attribute j,
- n_i, number of proteins interacting with protein i,
- $k_{i,j}$, number of proteins interacting with protein i with functional attribute j.

- We selected a subset of 1169 proteins and 258 functional attributes, which yields 301 602 null hypotheses.
- For this data set we applied the closure of the Hochberg procedure, follow up and follow down testing procedures, to control PFER, FWER and FDR, respectively, at the level $\alpha = 0.05$.
- To compute empirical error rates, the one-leave out CV scheme was used.

Results

Testing procedure	#rejected nulls	empirical error rate	#true positives
Follow up PFER controlling procedure	1706	PFER = 0.0209	1683
Follow down PFER controlling procedure	1191	PFER = 0.0164	1173
Closure of the Hochberg step up procedure	2300	FWER = 0.0382	2255
Follow down FDR controlling procedure	2892	FDR = 0.0173	2823

Acknowledgments

- Many thanks for your attention.
- Many thanks to Adam Zagdanski for arousing my interest in Gene Ontology.
- This research was financially supported by the Polish State Committee for Scientific Research, grant 1 P03A 017 29, so many thanks to the Polish Government ;-).

Literature


Marcus R., Peritz E., Gabriel K. R. 1976. On closed testing procedures with special reference to ordered analysis of variance. Biometrika 1976 63(3):655-660;


Finner H., Strassburger K. 2002. The partitioning principle: A powerful tool in multiple decision theory. The Annals of Statistics 2002, Vol. 30, No. 4, 11941213.


Benjamini Y., Hochberg Y. "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing". J. R. Statist. Soc. B (1995) 57.


Benjamini, Y., and Yekutieli, D. (2001). "The Control of the False Discovery Rate in Multiple Testing under Dependency,". The Annals of Statistics 29 (4).


Dudoit S., van der Laan M. J., Pollard K. S. (2004), Multiple Testing. Part I. Single-Step Procedures for Control of General Type I Error Rates, Statistical Applications in Genetics and Molecular Biology, 3(1).


Storey, J. D. (2002). ,,A direct approach to false discovery rates." Journal of the Royal Statistical Society, Series B 64, 479498


Hochberg, Y. and Tamhane, A. C. Multiple Comparison Procedures, (1987) , New York John Wiley and Sons Inc.