Weighted Multiple Testing Corrections for Correlated Tests

Changchun Xie, PhD

Assistant Prof. , Population Genomic Program, Dept. of CEB

Associate Member, Dept. of Mathematics & Statistics

Senior Statistical Analyst, Population Health Research Institute

McMaster University

CANNeCTIN Biostatistics Videoconference

May 13, 2011

Outline

- Introduction and Motivation
- Method
- Simulation
- Example
- Conclusion and Discussion

Motivation

- It is well known that ignoring multiple testing issue can cause false positive results.
- Many medical researchers still do not pay much attention to it. Benjamini (Biometrical Journal 2010, 52:6, 708-721) examined a sample of 60 papers from NEJM (2000-2004) and found 47/60 had no multiplicity adjustment at all, even though all needed it in some form or the other.
- Some researchers only use Bonferroni correction, which can be conservative if tests are correlated.

Problem

	not rejected	rejected	Total
True H_0	**U**	**V**	$\mathbf{m_0}$
True H_1	**T**	**S**	$\mathbf{m_1}$
Total	**m-R**	**R**	**m**

Error Rate control

- Family-wise Error Rate
- FWER=P(V≥1)

- False Discovery Rate
- FDR=E(V/R|R>0)P(R>0)

- When m_0=m, FDR is equivalent to FWER
- When m_0<m, FDR≤FWER.

Bonferroni Correction

- Adjusting individual testing significance level to be α/m

- ---- does not require the tests are independent
  ---- can be conservative if tests are correlated
  ---- equally weighted tests

Fixed Sequence (FS)

- tests each null hypothesis at the same α without any adjustment in a pre-specified testing sequence and further testing stops when the null hypothesis in the testing sequence is not rejected

---- require the pre-specified testing sequence

---- if the first null hypothesis cannot be rejected, the second null hypothesis cannot be reject even the p-value is very small.

Weighted Bonferroni

- Moyé (2000) developed the prospective alpha allocation scheme (PAAS). For example, 0.045 for the first endpoint and 0.005 for the second endpoint

---- independent tests

Bonferroni Fixed Sequence (BFS)

- Wiens (2003) proposed a Bonferroni fixed sequence (BFS) procedure. For example, 0.045 for the first endpoint and 0.005 for the second endpoint. If the first null hypothesis is rejected, the significance level for the second test will be 0.045+0.005=0.05.

---- require the pre-specified testing sequence

---- ignore correlation between the tests

---- has more power for the second or later tests

Flexible Fixed Sequence (FFS)

- Huque and Alosh (2008) suggested a flexible fixed-sequence (FFS) testing method to improve the BFS approach by taking into account correlation among endpoints.

- For example, 0.045 for the first endpoint and 0.005 for the second endpoint. If the first null hypothesis is rejected, the significance level for the second test will be 0.045+0.005=0.05. If the first null hypothesis is not rejected, the significance level for the second test will be α_2 , where α_2 is chosen to satisfy

$$0.005 = P(T_1 < C_{0.045}, T_2 \geq C_{\alpha_2} \mid H_0^{(1)}, H_0^{(2)})$$

Flexible Fixed Sequence (FFS)

---- They only provided critical values and significant levels for testing up to three endpoints, which followed a multi-normal distribution with equal correlations

---- The first endpoint in the sequence cannot get the benefit for its power either from the correlation among the endpoints or from the rejection of other null hypotheses.

Weighted Holm

- Assume that $p_1,\ldots,p_m$ are the unadjusted p-values and $w_i>0$, i=1,…,m are the corresponding weights that add to 1. Let $q_i=p_i/w_i$, i=1,…,m. Without loss of generality, suppose $q_1 \le q_2 \le \ldots \le q_m$. Then the adjusted p-value for the first hypothesis is $P_{adj\_1} = \min(1, q_1)$. Inductively, the adjusted p-value for the *j*th hypothesis is

$$P_{adj\_j} = \min(1, \max(P_{adj\_(j-1)}, (w_j + \ldots + w_m)q_j)) \quad , \quad j=2,\ldots,m.$$

The method rejects a hypothesis if the adjusted p-value is less than the family-wise error rate α.

The Proposed Method

Let $p_1,\ldots,p_m$ be the observed p-values for m tests and $w_i>0$, $i=1,\ldots,m$ be the corresponding weights. Calculate $q_i=p_i/w_i$, $i=1,\ldots,m$. Then the adjusted p-value for p_i is

$$P_{adj\_i} = P(\min_j q_j \le q_i)$$

$$= 1 - P(all \quad q_j > q_i)$$

$$= 1 - P(all \quad p_j > p_i w_j / w_i)$$

$$= 1 - P\left(\bigcap_{j=1}^{m} a_j \le X_j \le b_j\right)$$

The Proposed Method

where X_j, j=1,…,m are standardized multivariate normal with correlation matrix $\sum$ and for the two-sided case,

$$a_j = \Phi^{-1}\left(\frac{p_i w_j}{2w_i}\right),$$

$$b_j = \Phi^{-1}\left(1 - \frac{p_i w_j}{2w_i}\right)$$

The Proposed Method

If the adjusted p-values $\leq \alpha$, reject the null hypothesis. Suppose k_1 null hypotheses have been rejected, we then adjust the remaining m-k_1 observed p-values for multiple testing after removing the rejected k_1 null hypotheses, using the corresponding correlation matrix and weights.

Continue the procedures above until there is no null hypothesis left after removing the rejected null hypotheses or there is no null hypothesis which can be rejected.

Computation Issue

Computation of the adjusted P-values requires integration of the multivariate normal density function, which has no closed-form solution.

However, a recently developed package "mvtnorm" in R can be used.

Can handle hundreds of endpoints

Simulation 1

- To check whether the proposed method (using estimated correlation matrices) controls family-wise type I error rate when the endpoints have different correlation structures.

- N=1000 (500 per treatment group)
- 4 endpoints with w=(4,4,1,1)
- 5 correlation structures, which are called as I, II, III, IV, V respectively:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^3 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & \rho \\ 0 & 0 & \rho & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & \rho & -\rho & -\rho \\ \rho & 1 & -\rho & -\rho \\ -\rho & -\rho & 1 & \rho \\ -\rho & -\rho & \rho & 1 \end{pmatrix}$$

where ρ is chosen as 0.2, 0.7 and 0.9 .

Results of simulation 1

ρ	Correlation structures			
	II	III	IV	V
0.9	0.050261	0.049982	0.049788	0.050285
0.7	0.050302	0.049872	0.050119	0.049883
0.2	0.050226	0.049977	0.050525	0.050207
0.0	0.050048			

Family-wise type I error rate is defined as the number of runs, where at least one hypothesis is rejected (i.e. the adjusted p-value ≤ 0.05), divided by 1,000,000.

Simulation 2

- To compare the power, the significance level and the family-wise type I error rate of the proposed method (using estimated correlations) with those of FFS, BFS, PAAS, FS and Weighted Holm.

- N=240 (120 per treatment group)
- 2 endpoints
- Correlation structure: equal correlations, II

Results of Simulation 2

α allocations or weight	Effect size	ρ	Proposed method	FFS	BFS	PAAS	FS	Weighted Holm
α allocations	0.0, 0.0	0.0	3.1, 2.1	3.0, 2.2	3.0, 2.1	3.0, 2.0	5.0, 0.2	3.0, 2.1
(0.03, 0.02)			(5.0)	(5.0)	(4.9)	(4.9)	(5.0)	(4.9)
or weight		0.3	3.2, 2.3	3.0, 2.3	3.0, 2.2	3.0, 2.0	5.0, 0.5	3.1, 2.2
(3, 2)			(5.0)	(5.0)	(4.9)	(4.9)	(5.0)	(4.9)
		0.5	3.4, 2.4	3.0, 2.6	3.0, 2.3	3.0, 2.0	5.0, 0.9	3.1, 2.3
			(5.0)	(5.0)	(4.7)	(4.7)	(5.0)	(4.7)
		0.7	3.7, 2.8	3.0, 3.2	3.0, 2.5	3.0 2.0	5.0, 1.6	3.2, 2.5
			(5.0)	(5.0)	(4.3)	(4.3)	(5.0)	(4.3)
		0.9	4.3, 3.5	3.0, 4.2	3.0, 2.8	3.0, 2.0	5.0, 3.0	3.3, 2.8
			(5.0)	(5.0)	(3.7)	(3.7)	(5.0)	(3.7)

The first cell entry (%) is for the first endpoint and the second entry is for the second endpoint. The family-wise type I error rate (%) is given in brackets

α allocations or weight	Effect size	ρ	Proposed method	FFS	BFS	PAAS	FS	Weighted Holm
α allocations (0.03, 0.02) or weight (3, 2)	0.1, 0.4	0.0	11.21, 78	8.26, 78	8.26, 78	8.26, 77	12.03, 10	11.18, 78
		0.3	11.54, 78	8.22, 78	8.22, 78	8.22, 77	12.02, 11	11.50, 78
		0.5	11.73, 78	8.21, 79	8.21, 77	8.21, 77	12.00, 12	11.69, 77
		0.7	11.90, 79	8.25, 81	8.25, 77	8.25, 77	12.04, 12	11.85, 77
		0.9	11.98, 81	8.25, 85	8.25, 77	8.25, 77	12.05, 12	11.90, 77
	0.4, 0.4	0.0	85.8, 85.2	81.8, 85.2	81.8, 85.1	81.8, 77.3	86.9, 75.5	85.7, 85.1
		0.3	85.4, 84.6	81.8, 84.6	81.8, 84.4	81.8, 77.3	86.8, 77.0	85.2, 84.4
		0.5	85.3, 84.4	81.8, 84.6	81.8, 84.0	81.8, 77.4	86.9, 78.5	84.9, 84.0
		0.7	85.2, 84.2	81.8, 84.7	81.8, 83.4	81.8, 77.3	86.8, 80.3	84.4, 83.4
		0.9	85.8, 84.7	81.8, 85.6	81.8, 82.7	81.8, 77.2	86.9, 83.1	83.6, 82.7

α allocations or weight	Effect size	ρ	Proposed method	FFS	BFS	PAAS	FS	Weighted Holm
α allocations	0.0, 0.0	0.0	4.5, 0.7	4.5, 0.7	4.5, 0.7	4.5, 0.5	5.0, 0.3	4.5, 0.7
(0.045, 0.005)			(5.0)	(5.0)	(5.0)	(5.0)	(5.0)	(5.0)
or weight		0.3	4.6, 0.9	4.5, 0.9	4.5, 0.9	4.5, 0.5	5.0, 0.5	4.5, 0.9
(9, 1)			(5.0)	(5.0)	(5.0)	(5.0)	(5.0)	(5.0)
		0.5	4.6, 1.2	4.5, 1.4	4.5, 1.2	4.5, 0.5	5.0, 0.9	4.5, 1.2
			(5.0)	(5.0)	(4.8)	(4.8)	(5.0)	(4.8)
		0.7	4.8, 1.8	4.5, 2.0	4.5, 1.7	4.5, 0.5	5.0, 1.6	4.5, 1.7
			(5.0)	(5.0)	(4.7)	(4.7)	(5.0)	(4.7)
		0.9	5.0, 3.0	4.5, 3.3	4.5, 2.8	4.5, 0.5	5.0, 3.0	4.5, 2.8
			(5.0)	(5.0)	(4.5)	(4.5)	(5.0)	(4.5)

α allocations or weight	Effect size	ρ	Proposed method	FFS	BFS	PAAS	FS	Weighted Holm
α allocations (0.045, 0.005) or weight (9, 1)	0.1, 0.4	0.0	11.65, 63	11.09, 64	11.09, 63	11.09, 60	11.97, 10	11.63, 63
		0.3	11.81, 62	11.14, 64	11.14, 62	11.14, 60	12.03, 11	11.78, 62
		0.5	11.84, 62	11.07, 65	11.07, 61	11.07, 60	11.97, 12	11.79, 61
		0.7	11.95, 61	11.10, 69	11.10, 61	11.10, 60	11.99, 12	11.90, 61
		0.9	12.01, 62	11.12, 78	11.12, 60	11.12, 60	12.01, 12	11.97, 60
	0.4, 0.4	0.0	86.5, 83.1	85.9, 83.1	85.9, 83.1	85.9, 60.2	86.9, 75.4	86.5, 83.1
		0.3	86.4, 82.2	85.8, 82.4	85.8, 82.1	85.8, 60.3	86.8, 77.0	86.3, 82.1
		0.5	86.5, 81.8	85.9, 82.3	85.9, 81.7	85.9, 60.3	86.9, 78.5	86.3, 81.7
		0.7	86.5, 81.9	85.9, 82.8	85.9, 81.6	85.9, 60.3	86.8, 80.3	86.1, 81.6
		0.9	86.8, 83.2	85.9, 84.2	85.9, 82.6	85.9, 60.2	86.9, 83.0	85.9, 82.6

Example 1

- A two-sided trial with $\alpha=0.05$, three endpoints, and equal correlation with $\rho=0.7$ has the following p-values with corresponding α allocations (0.025, 0.02, 0.005):

  0.03, 0.026, 0.001.

- The proposed method with the corresponding weight (5,4,1) gives the adjusted p-values 0.0486, 0.0524 and 0.0087
- Removing the two rejected null hypotheses (the first and the third)
- Reject the second null hypothesis since $0.026 < 0.05$

- BFS, PAAS and weighted Holm reject the third null hypothesis only.
- FFS rejects the second and the third null hypotheses ($0.026<0.0274352$).
- The proposed method rejects all the three null hypotheses
- FS rejects all the three null hypotheses

Example 2

- The same as example 1 except $\rho=0.99$
- The proposed method gives the adjusted p-values: 0.0316, 0.0342 and 0.0054 respectively. All the null hypotheses are rejected in the first step.
- FFS, BFS, PAAS and weighted Holm still cannot reject the first null hypothesis

Conclusion and Discussion

- The proposed method does not require testing sequence
- The proposed method can control family-wise type I error rate very well.
- The significance level for the first hypothesis in FFS, BFS, PAAS and FS does not change, while the significance levels for all the hypotheses in the proposed method increase, thus allowing increased power of the tests for all endpoints, when the correlation between endpoints increases

- The proposed method has higher power for testing the first hypothesis than FFS, BFS and PAAS.

- The gain in power in testing the first hypothesis depends on the weights (or α allocations) and the extent of correlation between endpoints.

- The proposed method has higher power for each individual hypothesis than the weighted Holm procedure, especially when the correlation between endpoints is high.

- The proposed method might still have high power for testing other hypotheses when the power for testing the first hypothesis is very low.

Thanks