Multiple testing when there are correlated outcomes in medical research

Changchun Xie, PhD

Assistant Prof. , Division of Epidemiology and Biostatistics,

Department of Environmental Health, University of Cincinnati

The BERD Monthly Seminar, July 9, 2013

Outline

- Introduction and Motivation
- Methods
- Simulation
- R package WMTCc with examples
- Future Work

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

Alpha-exhaustive fallback (AEF)

- Weins and Dmitrienko developed BFS further by using more available alpha to provide a tesing procedure (AEF) with more power than original BFS.

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.

Research Article

Statistics in Medicine

Received 11 January 2011, Accepted 12 September 2011 Published online 23 November 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4434

Weighted multiple testing correction for correlated tests

Changchun Xie*†

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

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

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

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

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

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

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

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

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.

- The WMTCc method does not require testing sequence
- The WMTCc method can control family-wise type I error rate very well.
- The WMTCc and FS can keep the family-wise type I error rate at 5% level when the correlation increase, but the family-wise type I error rate in PAAS, AEF and the weighted Holm decrease, demonstrating decreased power when correlation increase.

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

- The FS method always has very low power for testing other hypotheses when the power for testing the first hypothesis is very low.

- WMTCc method is for multiple continuous correlated endpoints. Does it still keep its advantages when correlated binary endpoints are used?

Communications in Statistics—Simulation and Computation®, 00: 1–10, 2013

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2012.674599

Taylor & Francis
Taylor & Francis Group

Weighted Multiple Testing Correction for Correlated Binary Endpoints

CHANGCHUN XIE,[1] XUEWEN LU,[2] JANICE POGUE,[3] AND DING-GENG (DIN) CHEN[4]

Survival Data

- For continuous data or binary data, the correlation matrix can be directly estimated from the corresponding correlated endpoints

- It is challenging to directly estimate the correlation matrix from the multiple endpoints in survival data since censoring is involved

WLW method

- The WLW method proposed by Wei, Lin, and Weissfeld (1989) is based on the marginal Cox models.
- Suppose there are n subjects and each subject can have up to m potential failure times (events). Let X_{ki} be the covariate process associated with the kth event for the ith subject. The marginal Cox models are given by

$$h_k(t) = h_{k0}(t)e^{\beta'_k X_{ki}(t)}, \; k = 1, \ldots, m; \; i = 1, \ldots, n.$$

- WLW method estimates $\beta_1, \ldots, \beta_m$ by the maximum partial likelihood estimates $\hat{\beta}_1, \ldots, \hat{\beta}_m$ respectively, and uses a robust sandwich covariance matrix estimate, Σ for $(\hat{\beta}'_1, \ldots, \hat{\beta}'_m)'$ to account for the dependence of the multiple endpoints.

- After we have the estimated robust sandwich covariance matrix, the WMTCc method will be applied.

Simulation

- To check whether the proposed method (using estimated correlation matrices from WLW method) controls family-wise type I error rate when the endpoints have different correlations.
- To compare the power of the proposed method with those nonparametric methods

- N=1000 (500 per treatment group)
- 3 endpoints with w=(5,4,1)
- Based on 100,000 runs

- To generate the event times, we use the equivalent transformed regression model

$$\log(H_0(t)) = -\beta' X + \log(e),$$

where $e \sim \exp(1)$ and $H_0(t)$ is the baseline cumulative hazard function.

- Generate sample from multi-exponential distribution with a given correlation coefficient.

- Note: the correlation coefficient between $\min(ce_1, e_3)$ and $\min(ce_2, e_3)$ is $\frac{c}{c+2}$

α allocations or weight	Effect size	ρ	Proposed method	AEF	FS	Weighted Holm
α allocations (0.025, 0.02, 0.005) or weight (5, 4,1)	0.0, 0.0, 0.0	0.0	2.6, 2.1, 0.5 (5.0)	2.5, 2.1, 0.6 (5.0)	5.0, 0.2, 0.02 (5.0)	2.6, 2.1, 0.5 (5.0)
		0.3	2.7, 2.2, 0.7 (5.1)	2.6, 2.1, 0.7 (4.9)	5.1, 0.5, 0.1 (5.1)	2.6, 2.1, 0.6 (4.9)
		0.5	2.8, 2.4, 0.8 (4.9)	2.5, 2.2, 0.8 (4.4)	4.9, 0.8, 0.3 (4.9)	2.6, 2.2, 0.7 (4.4)
		0.7	3.5, 2.9, 1.3 (5.1)	2.7, 2.4, 1.2 (4.1)	5.1, 1.8, 0.9 (5.1)	2.8, 2.4, 1.1 (4.1)
		0.9	4.2, 3.7, 2.4 (5.0)	2.7, 2.5, 1.9 (3.3)	5.0, 3.0, 2.3 (5.0)	2.8, 2.5, 1.8 (3.3)

α allocations or weight	Effect size	ρ	Proposed method	AEF	FS	Weighted Holm
α allocations	0.05,	0.0	7.2, 6.3, 55.4	7.1, 6.2, 55.5	11.2, 1.3, 1.1	7.1, 6.2, 55.3
(0.025, 0.02,	0.05,	0.3	7.7, 6.9, 55.3	7.4, 6.7, 54.7	11.2, 2.5, 2.4	7.4, 6.6, 54.6
0.005) or	0.2	0.5	8.5, 7.5, 58.1	8.0, 7.0, 56.6	11.6, 3.8, 3.8	8.0, 7.0, 56.6
weight (5, 4,1)		0.7	9.0, 8.2, 57.2	8.1, 7.5, 54.2	11.4, 5.5, 5.4	8.1, 7.5, 54.2
		0.9	10.0, 9.4, 59.7	8.1, 7.7, 53.9	11.3, 8.0, 7.8	8.1, 7.7, 53.9

α allocations or weight	Effect size	ρ	Proposed method	AEF	FS	Weighted Holm
α allocations (0.025, 0.02, 0.005) or weight (5, 4,1)	0.2, 0.05, 0.05	0.0	75.5, 8.8, 3.6	75.0, 9.3, 2.7	82.9, 9.4, 1.0	75.3, 8.7, 3.6
		0.3	75.7, 9.4, 4.6	74.9, 9.8, 3.7	82.9, 10.4, 2.5	75.0, 9.1, 4.5
		0.5	77.9, 10.1, 5.5	76.6, 10.3, 4.7	84.2, 11.1, 3.9	76.6, 9.6, 5.3
		0.7	77.5, 10.4, 6.6	74.7, 10.3, 5.8	82.8, 11.1, 5.5	74.7, 9.6, 6.1
		0.9	80.1, 10.8, 7.8	74.8, 10.1, 7.3	83.0, 10.7, 7.5	74.8, 9.3, 7.2

α allocations or weight	Effect size	ρ	Proposed method	AEF	FS	Weighted Holm
α allocations	0.2, 0.2,	0.0	80.4, 79.7, 74.9	79.4, 79.9, 75.4	82.9, 68.7, 56.9	80.2, 79.7, 74.8
(0.025, 0.02,	0.2	0.3	80.0, 79.3, 74.0	78.6, 79.1, 74.1	82.9, 71.1, 62.2	79.6, 78.8, 73.6
0.005) or		0.5	81.8, 81.0, 75.9	80.2, 80.5, 75.7	84.5, 75.1, 68.5	81.0, 80.2, 75.2
weight (5, 4,1)		0.7	80.2, 79.3, 74.4	77.7, 77.8, 73.3	82.9, 75.0, 70.1	78.4, 77.5, 72.8
		0.9	81.7, 80.7, 76.8	77.0, 77.2, 74.2	83.1, 78.7, 76.1	77.6, 76.8, 74.1

R package WMTCc with examples

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

We are developing R package “WMTCc”.

Future Work #1

- Parametric multiple testing methods are uniformly more powerful than their corresponding nonparametric methods *if* the correlations are known or correctly estimated

- If the correlations are misspecified, the FWER in the parametric multiple testing methods may not be controlled

- Developing a new method, which is robust on misspecified correlation and is more powerful than nonparametric methods

Future Work #2

- As clinical trial objectives become more complex, the multiple endpoints can be hierarchically ordered and logically related
- Develop a weighted multiple testing correction for multiple families of correlated tests

Collaborators

Prof. Christopher John Lindsell

Prof. Susan M. Pinney

Prof. Rakesh Shukla

Graduate Student: John Aidoo, Wei Zhou

The work is supported by an Institutional Clinical and Translational Science Award, NIH/NCRR Grant Number UL1TR000077

Thanks