Adjustment method to address type I error and power issues with outcome multiplicity and correlation

Richard E. Blakesley, BS [1], Sati Mazumdar, PhD [1, 2], Patricia R. Houck, MSH [2]

[1] Department of Biostatistics, University of Pittsburgh [2] Department of Psychiatry, University of Pittsburgh School of Medicine

5th International Conference on Multiple Comparison Procedures, Vienna, Austria, July 8-11, 2007

Motivation

Multiple hypothesis testing calls into question the validity of individual hypothesis test results due to type I error inflation. Several adjustment methods exist in statistical literature to protect type I error. However, their type I error and power performance suffer with increasing outcome multiplicity and correlation. Single-step approaches (Bonferroni, Sidak) protect type I error for independent outcomes, but become conservative with increasing correlation and lack power. Stepwise approaches (Holm, Hochberg, Hommel) demonstrate improved power over single-step methods. Methods which use correlation components in the adjustment formulae (Dubey/Armitage-Parmar [D/AP] and R-Squared Adjustment [RSA]) address overcorrection of type I error to a limited extent. Resampling methods (Bootstrap MinP and Step-Down MinP) incorporate correlation structure, but there exist caveats and implementation limits. In examinations of the D/AP and RSA methods, curvilinear type I error patterns suggest the inefficiency of the linear correlation function in the exponent. Furthermore, the correlation between raw data vectors may not reflect the null hypotheses, and normality assumptions may be violated. We propose combining a stepwise approach with a new correlation component to address some limitations of existing methods.

Proposal: The Blakesley Method

For p-values $p_1 \geq p_2 \geq \ldots \geq p_k \geq \ldots \geq p_{K-1} \geq p_K$, corresponding to two-sample t-tests for hypotheses H_{01} through H_{0K} and outcomes X_1 through X_K, where $X_k = (X'_{k1} \;\; X'_{k2})'$, define the adjusted p-value p_{ak} as:

$$p_{a1} = p_1,\ p_{ak} = \min\left\{1-(1-p_k)^{g(k)}, p_{a(k-1)}\right\},\ g(k) = k^{\sqrt{1-\left[\sum_{i<j\leq k}\lambda_{ij}^2 \Big/ \binom{k}{2}\right]}},\ \lambda_{ij} = \operatorname{corr}\left\{T_i(X_i), T_j(X_j)\right\}$$

$$T_k(X_k)_{n\times 1} = \left\{c_1\left(X'_{k1} - \overline{X}_{k1}\right)_{1\times n_1}\ c_2\left(X'_{k2} - \overline{X}_{k2}\right)_{1\times n_2}\right\}',\ i,j,k \in \{1, 2, \ldots, K\},\ n = n_1 + n_2$$

where $T_k(X_k)$ is a transformation of the data vector X_k to reflect the null hypothesis H_{0k}, such that the data values are centered by their group-specific means and multiplied by the test statistic contrast. In this simplified two-sample case, define contrast $c = (c_1 \;\; c_2) = (1 \;\; -1)$.

Method

Simulation trial design

- Trial parameters:
  - Replicates: 10,000
  - Correlation structure/magnitude
    - Compound Symmetry (CS)
    - Block Symmetry (BS)
    - Decreasing Dependence (DD)
  - Hypothesis set
  - K = 4, 8, 12, or 24 multivariate normal outcomes for two groups
  - $\alpha = 0.05$
- Equal-variance, two-sided, two-sample t-tests conducted on each replicate to generate raw p-values, and p-value adjustment methods were applied.
- Performance measures:
  - Type I Error (FWER)
    - Pr[reject at least 1 TN]
  - Average Power
    - Mean Pr[reject each FN]
  - Minimal Power
    - Pr[reject at least 1 FN]
  - Maximal Power
    - Pr[reject all K FNs]

Correlation Structures

$$CS = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}_{K\times K} \quad \rho = 0.0, 0.1, \ldots, 0.9$$

$$BS = \begin{bmatrix} \Sigma_W & \Sigma_B \\ \Sigma_B & \Sigma_W \end{bmatrix}_{K\times K} \quad \Sigma_W = \begin{bmatrix} 1 & W & \cdots & W \\ W & 1 & \cdots & W \\ \vdots & \vdots & \ddots & \vdots \\ W & W & \cdots & 1 \end{bmatrix}_{\frac{K}{2}\times\frac{K}{2}} \quad \Sigma_B = \begin{bmatrix} B & B & \cdots & B \\ B & B & \cdots & B \\ \vdots & \vdots & \ddots & \vdots \\ B & B & \cdots & B \end{bmatrix}_{\frac{K}{2}\times\frac{K}{2}}$$

$B = 0.0, 0.2, 0.5$

$W = 0.2, 0.5, 0.8$

$W > B$

$$DD = \begin{bmatrix} 1 & \eta & \eta^2 & \cdots & \eta^{K-1} \\ \eta & 1 & \eta^2 & \cdots & \eta^{K-1} \\ \eta^2 & \eta^2 & 1 & \cdots & \eta^{K-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \eta^{K-1} & \eta^{K-1} & \eta^{K-1} & \cdots & 1 \end{bmatrix}_{K\times K} \quad \eta = 0.5, 0.6, \ldots, 0.9$$

Outcome Types

	Block 1		Block 2	
Hypothesis Sets	$X_1, \ldots, X_{K/4}$	$X_{1+(K/4)}, \ldots, X_{K/2}$	$X_{1+(K/2)}, \ldots, X_{3K/4}$	$X_{1+(3K/4)}, \ldots, X_K$
Uniform – True Null	TN	TN	TN	TN
Uniform – False Null	FN	FN	FN	FN
Split	TN	TN	FN	FN

Outcomes Types: TN = True Null (effect size = 0.0), FN = False Null (effect size = 0.5)

Conclusion

The Blakesley method, as demonstrated in the above plots, showed stable type I error protection ranging values of K, correlation structures and magnitudes. It also demonstrated similar or greater average and maximal power compared to nearly all methods with conservative type I error protection. The Hommel method was shown to have greater power under the uniform hypothesis set with lower correlation magnitude and higher values of K, but the Blakesley method performed better under the split hypothesis set, with increasingly improved power with higher correlation magnitude. The Blakesley method followed similar patterns compared to the resampling methods, typically with greater power. The new method holds promise to allow high power with controlled type I error when analyzing multiple correlated hypotheses.

Funding Sources

NIMH T32 MH073451, NIMH P30 MH071944

Supplementary Formulae

For ordered p-values, $p_1 \geq p_2 \geq \ldots \geq p_k \geq \ldots \geq p_{K-1} \geq p_K$

Method	Adjusted p-value p_{ak}
Bonferroni	$\min(1, Kp_k)$
Holm	$\min\left(1, \max\left\{kp_k, p_{a(k+1)}\right\}\right), p_{aK} = Kp_K$
Hochberg	$\min\left(kp_k, p_{a(k-1)}\right), p_{a1} = p_1$
Hommel	$\max\left(p_{S_{k,1}}, p_{S_{k,2}}, \ldots, p_{S_{k,2^{K-1}}}\right)^*$
Sidak	$1-(1-p_k)^K$
TCH	$1-(1-p_k)^{\sqrt{K}}$
D/AP	$1-(1-p_k)^{K^{1-(K-1)^{-1}\sum_{i\neq k}\rho_{ik}}}$
RSA	$1-(1-p_k)^{K^{1-R_k^2}}$
minP	$\Pr\left[\min_{j\in\{1,2,\ldots,K\}} P_j \leq p_k \mid H_0^C\right]$
Step-down minP	$\max\left(\Pr\left[\min_{j\in\{1,2,\ldots,k\}} P_j \leq p_k \mid H_0^C\right], p_{a(k+1)}\right), p_{aK} = \text{minP } p_{aK}$

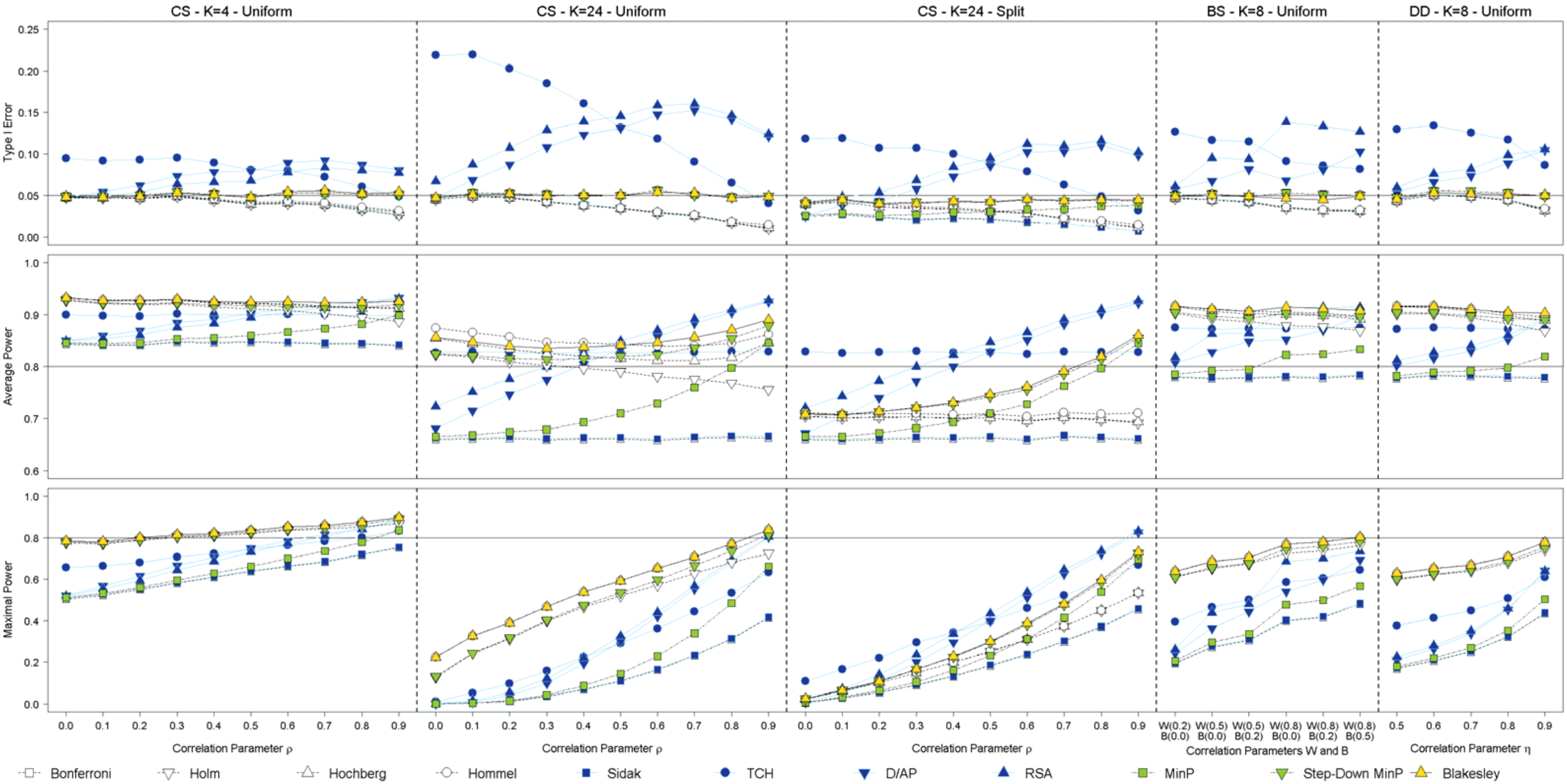
*$p_{S_{k,i}}$ is the Simes p-value for the i^{th} subset of hypotheses including the k^{th} of the K hypotheses, $i = 1, 2, \ldots, 2^{K-1}$

References

1. Dudoit, S., J.P. Shaffer, and J.C. Boldrick, *Multiple hypothesis testing in microarray experiments*. Statistical Science, 2003. **18**(1): p. 71-103.
2. Hochberg, Y., *A sharper Bonferroni procedure for multiple tests of significance*. Biometrika, 1988. **75**(4): p. 800-802.
3. Holm, S., *A simple sequentially rejective multiple test procedure*. Scandinavian Journal of Statistics, 1979. **6**: p. 65-70.
4. Hommel, G., *A stagewise rejective multiple test procedure based on a modified Bonferroni test*. Biometrika, 1988. **75**(2): p. 383-386.
5. R Development Core Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* 2006, URL http://www.R-project.org.
6. Sankoh, A.J., M.F. Huque, and S.D. Dubey, *Some comments on frequently used multiple endpoint adjustment methods in clinical trials*. Statistics in Medicine, 1997. **16**(22): p. 2529-2542.
7. Sidak, Z., *Rectangular confidence regions for the means of multivariate normal distributions*. Journal of the American Statistical Association, 1967. **62**(318): p. 626-633.
8. Simes, R.J., *An improved Bonferroni procedure for multiple tests of significance*. Biometrika, 1986. **73**(3): p. 751-754.
9. Venables, W.N. and B.D. Ripley, *Modern applied statistics with S*. Fourth ed. 2002, New York: Springer.
10. Westfall, P.H. and S.S. Young, *Resampling-based multiple testing: Examples and methods for p-value adjustment*. 1993, New York: John Wiley & Sons, Inc.
11. Wright, S.P., *Adjusted P-values for simultaneous inference*. Biometrics, 1992. **48**(4): p. 1005-1013.

Results

Heading format: Correlation structure – Number of outcomes – Hypothesis set



Blakesley Method Observations

- •Type I error: Maintains values near 0.05 across number of outcomes K and correlation structure/magnitude. Type I error slightly reduced under split hypothesis set conditions.
- •Average Power: Consistently similar or higher power compared to methods with consistent type I error ≤ 0.05. Increasing power with increasing correlation magnitude under split hypothesis set conditions.
- •Maximal Power: Designed to have identical maximal power to Hochberg/Hommel methods under uniform hypothesis set conditions. Demonstrated greater maximal power under split hypothesis set conditions.
- •Minimal Power: Results not shown. Power greater than 0.80 for all methods under all conditions. Relative power patterns similar to average power. Least informative power measure.