Letter to the Editor

Changchun Xie*1,2 and Ding-Geng Chen3,4,5

1 Division of Epidemiology and Biostatistics, Department of Environmental Health, University of Cincinnati, OH 45267, USA
2 Center for Clinical and Translational Science and Training, University of Cincinnati, OH 45267, USA
3 School of Nursing, University of Rochester Medical Center, Rochester, NY 14642, USA
4 Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY 14642, USA
5 Department of Orthopaedics and Rehabilitation, School of Medicine and Dentistry, University of Rochester Medical Center, Rochester, NY 14642, USA

Regarding Paper

“Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests”

by F. Bretz, M. Posch, E. Glimm, F. Klinglmueller, W. Maurer, and K. Rohmeyer

Biometrical Journal (2011) 53(6): 894–913
Article: http://dx.doi.org/10.1002/bimj.201000239
Author’s reply: http://dx.doi.org/10.1002/bimj.201200256

Additional supporting information may be found in the online version of this article at the publisher’s web-site

In clinical trials, the investigators often collect many endpoints and these endpoints are usually correlated. Recently, many parametric multiple testing methods (Huque and Alosh, 2008; Xie, 2012; Bretz et al., 2011) have been proposed to take into account correlations among these endpoints. Theoretically, all these methods should control FWER at \( \alpha = 0.05 \) level. Our simulations (Xie, 2012) have shown that Huque and Alosh’s, and Xie’s methods do control FWER at \( \alpha = 0.05 \) level for all levels of correlations among endpoints. However, we found that the weighted parametric tests illustrated in your journal by Bretz et al. (2011) as implemented with the “gMCP” package (Rohmeyer and Klinglmueller, 2011) cannot control FWER at \( \alpha = 0.05 \) level when the correlation is greater than 0.3. The problem might be due to the implementations in the “gMCP” package. We simulate clinical trials with 4 endpoints, \( y = (y_1, y_2, y_3, y_4) \) in small \( (n = 100) \) and large \( (n = 1000) \) sample sizes. Each individual has probability 0.5 to receive both the active treatment and placebo. The 4 endpoints are generated from a multivariate normal distribution \( N(0, \Sigma) \), where

\[
\Sigma = \begin{pmatrix}
1 & \rho & \rho & \rho \\
\rho & 1 & \rho & \rho \\
\rho & \rho & 1 & \rho \\
\rho & \rho & \rho & 1
\end{pmatrix}
\]

*Corresponding author: e-mail: cxie933@gmail.com, Phone: +1-513-5580229
Table 1  Result of the FWER as function of the correlation $\rho$.

| $\rho$  | $n = 100$       | $n = 1000$       |
|--------|-----------------|------------------|
| 0.0    | 5.0184%         | 5.0205%          |
| 0.1    | 5.0499%         | 5.0567%          |
| 0.3    | 5.2048%         | 5.1365%          |
| 0.5    | 5.2278%         | 5.1840%          |
| 0.7    | 5.2579%         | 5.2247%          |
| 0.9    | 5.1956%         | 5.2191%          |

and $\rho$ is chosen as 0, 0.1, 0.3, 0.5, 0.7 and 0.9. The weighting vector for the 4 endpoints is (0.4, 0.4, 0.1, 0.1). The treatment effect size is assumed as (0, 0, 0, 0). We replicate the clinical trial 1,000,000 times independently and calculate the FWER as the number of clinical trials where at least one true hypothesis is rejected (i.e. the adjusted p-value $\leq 0.05$) by multiple two-sided tests from the 1,000,000 simulated trials. The result is shown in Table 1 and the R program is provided as supplementary material.

References

Bretz, F., Posch, M., Glimm, E., Klinglmueller, F., Maurer, W. and Rohmeyer, K. (2011). Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. Biometrical Journal 53, 894–913.

Huque, M. F. and Alosh, M. (2008). A flexible fixed-sequence testing method for hierarchically ordered correlated multiple endpoints in clinical trials. Journal of Statistical Planning and Inference 138, 321–335.

Rohmeyer, K. and Klinglmueller, F. (2011). gMCP: A graphical approach to sequentially rejective multiple test procedures. R package version 0.6–5. http://cran.r-project.org/package=gMCP.

Xie, C. (2012). Weighted multiple testing corrections for correlated tests. Statistics in Medicine 31, 341–352.