Leibniz
Universität Hannover

Reports of the Institute of Biostatistics

No 05 / 2008

Leibniz University of Hannover

Natural Sciences Faculty

Titel: *Equivalence with respect to a control*

Author: *Mario Hasler*

1 Data situation

Let there be a negative control C and some doses $D_1, \ldots, D_k$ with commonly $k = 3$. For $j = 1, \ldots, n_i$ and $i = 0, \ldots, k$, let X_{ij} denote the jth observation under the ith dose/ treatment D_i, where $i = 0$ represents the control C. If not acting on explicitly different assumptions, suppose the X_{ij} to be independently normal with means μ_i and variances σ_i^2. The related estimators are $\bar{X}_i$ and S_i^2, respectively.

2 Methods

The doses are declared to be harmless if they both do not undershoot a certain lower limit of the control and do not overshoot a certain upper limit of it, respectively. We additionally assume balancedness for the sample sizes of the dose groups, say $n_1 = \ldots = n_k$, and homogeneous group variances σ_i^2. The resulting hypotheses to be tested are

$$\begin{aligned} H_{0i}: &\quad |\mu_i - \mu_0| \geq \delta \quad \text{(harmful)} \quad \text{vs.} \\ H_{Ai}: &\quad |\mu_i - \mu_0| < \delta \quad \text{(harmless)} \quad (1 \leq i \leq k) \end{aligned}$$

with a relevant threshold $\delta > 0$. The limits of the two-sided $(1-\alpha)100\%$ simultaneous confidence intervals according to Bofinger and Bofinger [1] are given as

$$\begin{aligned} \hat{\delta}_i^{lower} &= \min\left(\bar{X}_i - \bar{X}_0 - t_{k,1-\alpha}(\nu, \mathbf{R})\, S\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}, 0\right), \\ \hat{\delta}_i^{upper} &= \max\left(\bar{X}_i - \bar{X}_0 + t_{k,1-\alpha}(\nu, \mathbf{R})\, S\sqrt{\frac{1}{n_i} + \frac{1}{n_0}}, 0\right) \quad (1 \leq i \leq k) \end{aligned} \tag{1}$$

with the pooled sample variance S^2 and the lower $(1-\alpha)$ quantile $t_{k,1-\alpha}(\nu, \mathbf{R})$ of an underlying k-variate t-distribution with $\nu = \sum_{i=0}^{k}(n_i - 1)$ degrees of freedom and correlation matrix $\mathbf{R} = (r_{im})_{i,m}$ according to Tong [5] and Bofinger and Bofinger [1], where

$$r_{im} = \begin{cases} 1, & i = m, \\ \rho, & i \neq m, \quad i, m \in \{1, 2, ..., t\} \quad \text{or} \quad i, m \in \{t+1, t+2, ..., k\}, \\ -\rho, & i \neq m, \quad i \in \{1, 2, ..., t\} \quad \text{and} \quad m \in \{t+1, t+2, ..., k\}, \\ -\rho, & i \neq m, \quad m \in \{1, 2, ..., t\} \quad \text{and} \quad i \in \{t+1, t+2, ..., k\} \end{cases} \tag{2}$$

with

$$\rho = \frac{1}{\sqrt{(1 + n_0/n_1)(1 + n_0/n_1)}} \tag{3}$$

and $t = \lfloor k/2 \rfloor$ (the integral part of $k/2$).

Note that the approach of Bofinger and Bofinger [1] is only correct for balancedness within the non-control group doses and for homogeneous group variances. If this is not fulfilled, one can not derive a k-variate t-distribution. Taken the special structure of the correlation matrix in Equation (2) into account, it becomes clear that a Bonferroni-type alternative [3] does not loose much power even if all assumptions are fulfilled, is much simpler and can be generalized for several situations. So, a

Bonferroni-adjusted TOST approach is an alternative. **T**wo **O**ne-**S**ided **T**ests against the upper threshold δ_i^{upper} and the lower threshold δ_i^{lower}, respectively, are applied to test the hypotheses

$$\begin{aligned} H_{0i}: &\quad \mu_i - \mu_0 \leq \delta_i^{lower} \quad \text{or} \quad \mu_i - \mu_0 \geq \delta_i^{upper} \quad \text{(harmful)} \quad \text{vs.} \\ H_{Ai}: &\quad \mu_i - \mu_0 > \delta_i^{lower} \quad \text{and} \quad \mu_i - \mu_0 < \delta_i^{upper} \quad \text{(harmless)} \quad (1 \leq i \leq k). \end{aligned}$$

t-tests are applied for homogeneous group variances. The related confidence intervals then have confidence level $(1-2\alpha)$ according to the TOST principle. The confidence limits stay the same but the quantiles to be used change into Bonferroni-adjusted $(1-\alpha/k)$-quantiles of univariate t-distributions, $t_{\nu_i,1-\alpha/k}$. The degrees of freedom are $\nu_i = n_i + n_0 - 2$ but $\nu = \sum_{i=0}^{k}(n_i - 1)$ can also be used, which leads to a substantial gain in power.

If the group variances are heterogeneous, Welch t-tests [6] have to be applied. The confidence limits then are given by

$$\begin{aligned} \hat{\delta}_i^{lower} &= \min\left(\bar{X}_i - \bar{X}_0 - t_{k,1-\alpha}(\tilde{\nu}_i, \mathbf{R})\sqrt{\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}}, 0\right), \\ \hat{\delta}_i^{upper} &= \max\left(\bar{X}_i - \bar{X}_0 + t_{k,1-\alpha}(\tilde{\nu}_i, \mathbf{R})\sqrt{\frac{S_i^2}{n_i} + \frac{S_0^2}{n_0}}, 0\right) \quad (1 \leq i \leq k) \end{aligned} \tag{4}$$

Degrees of freedom $\tilde{\nu}_i$ according to Welch [6] have to be used here, of course.

Moreover, if interest is in ratios to control (not differences), the testing problem can be formulated by the hypotheses

$$\begin{aligned} H_{0i}: &\quad \mu_i/\mu_0 \leq \theta_i^{lower} \quad \text{or} \quad \mu_i/\mu_0 \geq \theta_i^{upper} \quad \text{(harmful)} \quad \text{vs.} \\ H_{Ai}: &\quad \mu_i/\mu_0 > \theta_i^{lower} \quad \text{and} \quad \mu_i/\mu_0 < \theta_i^{upper} \quad \text{(harmless)} \quad (1 \leq i \leq k) \end{aligned}$$

with relevant relative thresholds $\theta_i^{lower} < 1$ and $\theta_i^{upper} > 1$. For $A_i > 0$, the limits of the approximate two-sided $(1-2\alpha)100\%$ simultaneous confidence intervals are given by

$$\begin{aligned} \hat{\theta}_i^{lower} &= \min\left(\frac{-B_i - \sqrt{B_i^2 - 4A_iC_i}}{2A_i}, 1\right), \\ \hat{\theta}_i^{upper} &= \max\left(\frac{-B_i + \sqrt{B_i^2 - 4A_iC_i}}{2A_i}, 1\right) \quad (1 \leq i \leq k) \end{aligned} \tag{5}$$

with

$$\begin{aligned} A_i &= \bar{X}_0^2 - t_{\nu,1-\alpha/k}^2 S^2/n_0, \\ B_i &= -2\bar{X}_i\bar{X}_0, \\ C_i &= \bar{X}_i^2 - t_{\nu,1-\alpha/k}^2 S^2/n_i \end{aligned} \tag{6}$$

for homogeneous group variances and

$$\begin{aligned} A_i &= \bar{X}_0^2 - t^2_{\tilde{\nu}_i,1-\alpha/k} S_0^2/n_0, \\ B_i &= -2\bar{X}_i\bar{X}_0, \\ C_i &= \bar{X}_i^2 - t^2_{\tilde{\nu}_i,1-\alpha/k} S_i^2/n_i \end{aligned} \tag{7}$$

for heterogeneous ones with degrees of freedom $\tilde{\nu}_i$ according to Welch [6], but for ratio t-tests.

These testing problems are realized in the R package `ETC` [4, 2] by the functions `etc.diff` and `etc.rat`. The option `method` for `etc.diff` allows distinction between the Bofinger method and methods for homogeneous or heterogeneous group variances. A non-parametric method based on the Wilcoxon test is available, too.

3 Simulation study

A simulation study was performed to get some impressions about the conservatism of the Bonferroni-adjusted TOST approach compared with the method of Bofinger and Bofinger [1]. Several situations were considered for the difference based methods, and they are:

- C+2, homo: 10 N$(100, 10^2)$, 10 N$(80, 10^2)$, 10 N$(120, 10^2)$,
- C+2, hetero 1: 15 N$(100, 5^2)$, 10 N$(80, 10^2)$, 5 N$(120, 15^2)$,
- C+2, hetero 2: 5 N$(100, 5^2)$, 10 N$(80, 10^2)$, 15 N$(120, 15^2)$,
- C+4, homo: 10 N$(100, 10^2)$, 10 N$(80, 10^2)$, 10 N$(80, 10^2)$, N$(120, 10^2)$, N$(120, 10^2)$,
- C+4, hetero 1: 15 N$(100, 5^2)$, 12 N$(80, 8^2)$, 10 N$(80, 10^2)$, 8 N$(120, 12^2)$, 5 N$(120, 15^2)$,
- C+4, hetero 2: 5 N$(100, 5^2)$, 8 N$(80, 8^2)$, 10 N$(80, 10^2)$, 12 N$(120, 12^2)$, 15 N$(120, 15^2)$

with $\delta_i^{upper} = -\delta_i^{lower} = 20$ for all $i = 1, \ldots, k$. One control (C) and two (+2) or 4 (+4) dose groups with homogeneous (homo) or heterogeneous (hetero) variances, respectively, were taken. Equivalent situations were considered for the ratio based methods, too:

- C+2, homo: 10 N$(100, 10^2)$, 10 N$(80, 10^2)$, 10 N$(125, 10^2)$,
- C+2, hetero 1: 15 N$(100, 5^2)$, 10 N$(80, 10^2)$, 5 N$(125, 15^2)$,
- C+2, hetero 2: 5 N$(100, 5^2)$, 10 N$(80, 10^2)$, 15 N$(125, 15^2)$,
- C+4, homo: 10 N$(100, 10^2)$, 10 N$(80, 10^2)$, 10 N$(80, 10^2)$, N$(125, 10^2)$, N$(125, 10^2)$,
- C+4, hetero 1: 15 N$(100, 5^2)$, 12 N$(80, 8^2)$, 10 N$(80, 10^2)$, 8 N$(125, 12^2)$, 5 N$(125, 15^2)$,
- C+4, hetero 2: 5 N$(100, 5^2)$, 8 N$(80, 8^2)$, 10 N$(80, 10^2)$, 12 N$(125, 12^2)$, 15 N$(125, 15^2)$

with $\theta_i^{upper} = 1/\theta_i^{lower} = 1.25$ for all $i = 1, \ldots, k$. Test decisions based on p-values were evaluated. The conclusions are applicable to the described confidence intervals because of a strict one-to-one relation to the tests. Only the intervals for ratios of means with heterogeneous group variances do not have this strict relation. But the discrepancy is negligible here. The simulation results were obtained by 10000 simulation runs with the same starting seed (seed 100000) using the package `ETC` [2] in R [4].

Tables 1 and 2 show the related results. All methods keep the α-level, of course, for their conditions. For the considered situations both the Bonferroni-adjusted TOST method (TOST) and the related Welch method (Welch-TOST) yield conservative test decisions. As is known, this conservatism increases with the number of comparisons/ groups (which may not be seen here explicitly). This conservatism is in acceptable ranges, especially against the background of flexibility gain compared to the Bofinger approach (Bof). Only the Bonferroni-adjusted TOST method based on Wilcoxon tests shows an obvious decrease of the familywise error rate (FWER) for increasing number of comparisons/ groups.

Situation	Bof	TOST	Welch-TOST	Wilc-TOST
C+2, homo	0.0487	0.0487	0.0485	0.0431
C+2, hetero 1			0.0513	
C+2, hetero 2			0.0469	
C+4, homo	0.0481	0.0439	0.0439	0.0342
C+4, hetero 1			0.0503	
C+4, hetero 2			0.0446	

Table 1: Simulated global FWER for several test procedures (for differences) and situations; $\alpha = 0.05$.

Situation	TOST	W-TOST
C+2, homo	0.0480	0.0467
C+2, hetero 1		0.0526
C+2, hetero 2		0.0467
C+4, homo	0.0479	0.0468
C+4, hetero 1		0.0503
C+4, hetero 2		0.0454

Table 2: Simulated global FWER for several test procedures (for ratios) and situations; $\alpha = 0.05$.

References

[1] E. Bofinger and M. Bofinger. Equivalence with respect to a control: Stepwise tests. *Journal of the Royal Statistical Society B*, 57(4):721–733, 1995.

[2] M. Hasler. *ETC: Tests and simultaneous confidence intervals for equivalence to control*, 2008. R package version 1.1.

[3] D. Hauschke, M. Kieser, and L. A. Hothorn. Proof of safety in toxicology based on the ratio of two means for normally distributed data. *Biometrical Journal*, 41(3):295–304, 1999.

[4] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[5] Y. L. Tong. On partitioning a set of normal populations by their locations with respect to a control. *Annals of Mathematical Statistics*, 40(4):1300–1324, 1969.

[6] B. L. Welch. The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29:350–362, 1938.