STATISTICS IN MEDICINE
Statist. Med. 2008; **06**:1–6
Prepared using simauth.cls *[Version: 2002/09/18 v1.11]*

A note on tree gatekeeping procedures in clinical trials

Alex Dmitrienko*[1], Ajit C. Tamhane[2], Lingyun Liu[2], Brian L. Wiens[3]

[1] *Eli Lilly and Company, Indianapolis, IN 46285, U.S.A.*
[2] *Northwestern University, Evanston, IL 60208, U.S.A.*
[3] *Gilead Colorado, Westminster, CO 80021, U.S.A.*

SUMMARY

Dmitrienko et al. [1] proposed a tree gatekeeping procedure for testing logically related hypotheses in hierarchically ordered families which uses weighted Bonferroni tests for all intersection hypotheses in a closure method [3]. An algorithm was given to assign weights to the hypotheses for every intersection. The purpose of this note is to show that *any* weight assignment algorithm that satisfies a set of sufficient conditions can be used in this procedure to guarantee gatekeeping and independence properties. The algorithm used in [1] may fail to meet one of the conditions, namely monotonicity of weights, which may cause it to violate the gatekeeping property. An example is given to illustrate this phenomenon. A modification of the algorithm is shown to rectify this problem. Copyright © 2008 John Wiley & Sons, Ltd.

Key words and phrases: Multiple tests, Multiple endpoints, Clinical trials.

1. Introduction

Dmitrienko et al. [1] proposed a general formulation of multiple testing problems arising in clinical trials with hierarchically ordered/logically related multiple objectives and proposed the so-called *tree gatekeeping procedures* to address multiplicity issues in these problems. They gave a procedure based on the closure method that uses a weighted Bonferroni test for testing each intersection hypothesis. In this note we give a set of sufficient conditions on the weights assigned to the hypotheses in each intersection hypothesis in order to satisfy the gatekeeping and independence properties. We show that the weight assignment algorithm used in [1] (labelled Algorithm 1) may fail the monotonicity condition and, as a result, Algorithm 1 may fail to satisfy the gatekeeping property (the monotonicity condition was introduced by Hommel, Bretz and Maurer [2] to obtain shortcuts to Bonferroni-based closed procedures). A modification of the algorithm is shown to rectify this problem.

Consider n null hypotheses corresponding to multiple objectives in a clinical trial and suppose they are grouped into m families $F_1, \ldots, F_m$ to reflect the hierarchical structure of the testing problem (e.g., F_1 may contain hypotheses associated with a set of primary analyses and the other families may include hypotheses for sequentially ordered secondary analyses).

*Correspondence to: Alex Dmitrienko, Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Drop Code 2233, Indianapolis, IN 46285. Phone: (317) 277-1979, E-mail: dmitrienko_alex@lilly.com.

Received 1 January 2008
Revised 1 January 2008

The hypotheses included in F_i, $i = 1, \ldots, m$, are denoted by $H_{i1}, \ldots, H_{in_i}$ with $\sum_{i=1}^m n_i = n$. These hypotheses are to be tested by a procedure that controls the Type I *familywise error rate (FWER)* at a designated level α.

We consider Bonferroni-type procedures based on the raw p-values, p_{ij}, associated with the hypotheses H_{ij}. We allow for differential weighting of the hypotheses, with weight $w_{ij} > 0$ assigned to the hypothesis H_{ij} such that $\sum_{j=1}^{n_i} w_{ij} = 1$ for $i = 1, \ldots, m$. The procedures are required to satisfy the following two properties which follow from the logical relations between the hypotheses.

Gatekeeping property. A hypothesis H_{ij} in F_i, $i = 2, \ldots, m$, cannot be rejected (i.e., is automatically accepted) if at least one hypothesis in its serial rejection set (denoted by R_{ij}^S) is accepted or all hypotheses in its parallel rejection set (denoted by R_{ij}^P) are accepted. Here R_{ij}^S and R_{ij}^P consist of relevant hypotheses (determined by logical relations) from families F_k for $k < i$.

Independence property. A decision to reject a hypothesis in F_i, $i = 1, \ldots, m-1$, is independent of decisions made for hypotheses in $F_{i+1}, \ldots, F_m$ (i.e., the adjusted p-values for hypotheses in F_i, $i = 1, \ldots, m-1$, do not depend on the raw p-values for the hypotheses in $F_{i+1}, \ldots, F_m$).

2. A General Bonferroni Tree Gatekeeping Procedure

The following Bonferroni tree gatekeeping procedure was proposed in [1] for performing multiplicity adjustments in this problem using the closure method. Consider the closed testing family associated with the hypotheses in $F_1, \ldots, F_m$ and let H be any non-empty intersection of the hypotheses H_{ij}. If $v_{ij}(H)$ is the weight assigned to the hypothesis $H_{ij} \in H$ then the Bonferroni p-value for testing H is given by $p(H) = \min_{i,j}\{p_{ij}/v_{ij}(H)\}$. The multiplicity-adjusted p-value for the null hypothesis H_{ij} (denoted by $\widetilde{p}_{ij}$) is defined as $\widetilde{p}_{ij} = \max_H p(H)$, where the maximum is taken over all intersection hypotheses $H \ni H_{ij}$. The hypothesis H_{ij} is rejected if $\widetilde{p}_{ij} \leq \alpha$.

We now state the conditions on the weight vector $v_{ij}(H)$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$. First we define two indicator variables. Let $\delta_{ij}(H) = 0$ if $H_{ij} \notin H$ and 1 otherwise, and $\xi_{ij}(H) = 0$ if H contains any hypothesis from R_{ij}^S or all hypotheses from R_{ij}^P and 1 otherwise. The weight vector is chosen to satisfy the following conditions.

Condition 1. For any intersection hypothesis H, $v_{ij}(H) \geq 0$, $\sum_{j=1}^{n_i} v_{ij}(H) \leq 1$ and $v_{ij}(H) = 0$ if $\delta_{ij}(H) = 0$ or $\xi_{ij}(H) = 0$.

Condition 2. For any intersection hypothesis H, the weights are defined in a sequential manner, i.e., the subvector $v_i(H) = (v_{i1}(H), \ldots, v_{in_i}(H))$ is a function of the subvectors $v_1(H), \ldots, v_{i-1}(H)$ $(i = 2, \ldots, m)$ and does not depend on the subvectors $v_{i+1}(H), \ldots, v_m(H)$ $(i = 1, \ldots, m-1)$.

Condition 3. The weights for the hypotheses from the families, $F_1, \ldots, F_{m-1}$, meet the monotonicity condition, i.e., $v_{ij}(H) \leq v_{ij}(H^*)$, $i = 1, \ldots, m-1$, if $H_{ij} \in H$, $H_{ij} \in H^*$ and $H^* \subseteq H$ (i.e., if H implies H^*). For example, if $H^* = H_{11}$ and $H = H_{11} \cap H_{12}$ then $H_{11} \subseteq H_{11} \cap H_{12}$, and we require $v_{11}(H_{11} \cap H_{12}) \leq v_{11}(H_{11})$.

Note that Condition 3 is not required to be met for the hypotheses from F_m.

Proposition 1. *Conditions 1–3 are sufficient to guarantee that the Bonferroni tree gatekeeping procedure meets the gatekeeping and independence properties.*

Proof. Given in the Appendix.

A weight assignment algorithm that meets Conditions 1-3 is given below (it will be labelled Algorithm 2), but any other scheme for assigning weights satisfying these conditions also may be used. In this sense, the Bonferroni tree gatekeeping procedure proposed here is more general than that proposed in [1].

Algorithm 2 differs from Algorithm 1 in that it does not employ normalization in the first $m-1$ steps. Normalization in the final step makes the procedure α-exhaustive and hence more powerful. Although, this last normalization can violate Condition 3 by the weights assigned to the hypotheses in F_m, the gatekeeping properties are still maintained since these hypotheses can be eliminated from consideration when evaluating the Bonferroni p-values of intersection hypotheses, as the proof of Proposition 1 shows.

Algorithm 2 uses the following weight assignment scheme. It is assumed in the algorithm that $0/0 = 0$.

Step 1. Family F_1. Let $v_{1j}(H) = v_1^*(H)w_{1j}\delta_{1j}(H)$, $j = 1, \ldots, n_1$, where $v_1^*(H) = 1$, and $v_2^*(H) = v_1^*(H) - \sum_{j=1}^{n_1} v_{1j}(H)$.

Step $i = 2, \ldots, m-1$**.** Family F_i. Let $v_{ij}(H) = v_i^*(H)w_{ij}\delta_{ij}(H)\xi_{ij}(H)$, $j = 1, \ldots, n_i$, and $v_{i+1}^*(H) = v_i^*(H) - \sum_{j=1}^{n_i} v_{ij}(H)$.

Step m**.** Family F_m. Let

$$v_{mj}(H) = v_m^*(H)w_{mj}\delta_{mj}(H)\xi_{mj}(H)/\sum_{k=1}^{n_m} w_{mk}\delta_{mk}(H)\xi_{mk}(H),\ j = 1, \ldots, n_m.$$

3. Example of Violation of Gatekeeping Property

The weight assignment scheme in Algorithm 1 may not meet Condition 3 of monotonicity of weights. This is because the weight $v_{ij}(H)$ at Step i $(1 \le i \le m-1)$ includes normalization

$$v_{ij}(H) = v_i^*(H)w_{ij}\delta_{ij}(H)\xi_{ij}(H)/\sum_{k=1}^{n_i} w_{ik}\xi_{ik}(H).$$

Hence it is possible to get $v_{ij}(H) > v_{ij}(H^*)$ for $H^* \subseteq H$ if

$$\sum_{k=1}^{n_i} w_{ik}\xi_{ik}(H) < \sum_{k=1}^{n_i} w_{ik}\xi_{ik}(H^*).$$

Violation of the monotonicity condition does not always imply violation of the gatekeeping property since it is not a necessary condition, but for some configurations of the p_{ij}-values it does so as the following example shows.

Consider a clinical trial with nine hypotheses that are grouped into three families, $F_i = \{H_{i1}, H_{i2}, H_{i3}\}$, $i = 1, 2, 3$. The hypotheses are equally weighted within each family ($w_{ij} = 1/3$, $i, j = 1, 2, 3$) and the raw p-values associated with the hypotheses are displayed in Table 1. The logical restrictions in this multiple testing problem are defined in Table 1 using serial and parallel rejection sets.

[Insert Table 1 here]

To see that Condition 3 is not met when Algorithm 1 is used, consider two intersection hypotheses, $H = H_{13} \cap H_{21} \cap H_{22} \cap H_{23}$ and $H^* = H_{21}$, so that $H^* \subseteq H$. We will show that $v_{21}(H) > v_{21}(H^*)$. First note that $v_{21}(H^*) = w_{21} = 1/3$. Next, note that $v_{11}(H) = v_{12}(H) = 0, v_{13}(H) = 1/3$ and so $v_2^*(H) = 2/3$. Furthermore, $\xi_{21}(H) = 1, \xi_{22}(H) = 0, \xi_{23}(H) = 0$ since $H_{13} \in H$ belongs to R_{22}^S and R_{23}^S. Therefore $v_{21}(H) = (2/3)w_{21}/w_{21} = 2/3$.

The adjusted p-values produced by the Bonferroni tree gatekeeping procedure based on Algorithm 1 are displayed in Table 1. We see that two adjusted p-values in F_3 are significant at the 0.05 level despite the fact that no hypotheses can be rejected at this level in F_2. This implies that the procedure does not satisfy the gatekeeping property in this example. On the other hand, as shown in Table 1, the Bonferroni tree gatekeeping procedure based on Algorithm 2 does not violate the gatekeeping property (there are no significant adjusted p-values in F_3 since all adjusted p-values are non-significant in F_2).

REFERENCES

1. Dmitrienko A, Wiens BL, Tamhane AC, Wang X. Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Statistics in Medicine*. 2007; **26**, 2465–2478.
2. Hommel G, Bretz F, Maurer W. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statistics in Medicine*. 2007; **26**, 4063–4073.
3. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976; **63**, 655–660.

Appendix

Proof of Proposition 1. We will begin with the serial gatekeeping property and consider the hypothesis H_{ij}, $i = 2, \ldots, m$, $j = 1, \ldots, n_i$. Let R_{ij}^S denote its serial rejection set and suppose that at least one hypothesis, say, H_{rs}, $r < i$, is not rejected in R_{ij}^S. This means that there exists an intersection hypothesis, say H_{rs}^*, that contains H_{rs} and whose Bonferroni p-value is greater than α, i.e., $p(H_{rs}^*) > \alpha$. If H_{rs}^* also includes H_{ij}, this immediately implies that H_{ij} is not rejected. If H_{rs}^* does not include H_{ij}, let H_{rs}^{**} denote the intersection hypothesis obtained by eliminating hypotheses included in the last family F_m from H_{rs}^*. Let $H_{i_t j_t}$, $t = 1, \ldots, u$, denote the distinct hypotheses contained in H_{rs}^{**}, i.e., $H_{rs}^{**} = \cap_{t=1}^u H_{i_t j_t}$.

According to Condition 2, the weights are defined sequentially and thus the weight of any hypothesis in H_{rs}^{**} is equal to its weight in H_{rs}^*. More specifically, $v_{i_t j_t}(H_{rs}^{**}) = v_{i_t j_t}(H_{rs}^*)$ since $H_{i_t j_t}$ is contained in both H_{rs}^{**} and H_{rs}^*. Hence $p(H_{rs}^{**}) \geq p(H_{rs}^*) > \alpha$. Now consider the intersection hypothesis

$$H^* = H_{ij} \cap H_{rs}^{**} = H_{ij} \cap \left(\bigcap_{t=1}^{u} H_{i_t j_t} \right).$$

Note that this intersection hypothesis contains at least one hypothesis in R^S_{ij} (e.g., it contains H_{rs}). Thus, by Condition 1, $v_{ij}(H^*) = 0$, which implies that the p-value for H^* is given by

$$p(H^*) = \min_{t=1,\ldots,u} \left\{ \frac{p_{i_t j_t}}{v_{i_t j_t}(H^*)} \right\}.$$

By Condition 3, $v_{i_t j_t}(H^*) \leq v_{i_t j_t}(H^{**}_{rs})$, $t = 1, \ldots, u$, since $H_{i_t j_t}$ is contained in both H^* and H^{**}_{rs} and $H_{i_t j_t}$ is not from the last family. Therefore, $p(H^*) \geq p(H^{**}_{rs}) > \alpha$. Since H^* contains H_{ij}, we conclude that H_{ij} is not rejected.

Now consider the parallel gatekeeping property. Let R^P_{ij} be the parallel rejection set of the hypothesis H_{ij}, $i = 2, \ldots, m$, $j = 1, \ldots, n_i$. Let $H_{i_r j_r}$, $r = 1, \ldots, s$, denote the hypotheses in R^P_{ij} and suppose none of them is rejected. This implies that there exist intersection hypotheses, denoted by $H^*_{i_r j_r}$, $r = 1, \ldots, s$, such that $H^*_{i_r j_r}$ contains $H_{i_r j_r}$ and $p(H^*_{i_r j_r}) > \alpha$. If H_{ij} is contained in at least one intersection $H^*_{i_r j_r}$, $r = 1, \ldots, s$, then H_{ij} is not rejected. If H_{ij} is not contained in any intersection $H^*_{i_r j_r}$, $r = 1, \ldots, s$, then let $H^{**}_{i_r j_r}$, $r = 1, \ldots, s$, be the intersection hypotheses obtained by eliminating hypotheses included in the last family from $H^*_{i_r j_r}$. Since the weights are sequentially assigned by Condition 2, the weight of any hypothesis in $H^{**}_{i_r j_r}$, $r = 1, \ldots, s$, is equal to its weight in $H^*_{i_r j_r}$. Hence $p(H^{**}_{i_r j_r}) \geq p(H^*_{i_r j_r}) > \alpha$, $r = 1, \ldots, s$. Let

$$H^* = H_{ij} \cap \left(\bigcap_{r=1}^{s} H^{**}_{i_r j_r} \right).$$

Let $H_{k_t l_t}$, $t = 1, \ldots, u$, be the distinct hypotheses in $\cap_{r=1}^s H^{**}_{i_r j_r}$, i.e., $\cap_{r=1}^s H^{**}_{i_r j_r} = \cap_{t=1}^u H_{k_t l_t}$. To compute the p-value for H^*, note first that H^* includes all hypotheses from R^P_{ij}. By Condition 1, this implies that $v_{ij}(H^*) = 0$ and the p-value for H^* is given by

$$p(H^*) = \min_{t=1,\ldots,u} \left\{ \frac{p_{k_t l_t}}{v_{k_t l_t}(H^*)} \right\}.$$

Further, for any hypothesis $H_{k_t l_t}$, $t = 1, \ldots, u$, identify the intersection $H^{**}_{i_r j_r}$ that contains $H_{k_t l_t}$. Recall that the Bonferroni p-value for any $H^{**}_{i_r j_r}$, $r = 1, \ldots, s$, is greater than α, which implies that $p_{k_t l_t}/v_{k_t l_t}(H^{**}_{i_r j_r}) > \alpha$. By Condition 3, $v_{k_t l_t}(H^*) \leq v_{k_t l_t}(H^{**}_{i_r j_r})$ since $H_{k_t l_t}$ is contained in both H^* and $H^{**}_{i_r j_r}$ and $H_{k_t l_t}$ is not from the last family. Thus, $p_{k_t l_t}/v_{k_t l_t}(H^*) > \alpha$ for all $t = 1, \ldots, u$ and $p(H^*) > \alpha$. Since H^* contains H_{ij}, this immediately implies that H_{ij} is not rejected.

To prove that the independence property is satisfied, one can utilize arguments used in [1] (this proof relies on the fact that, according to Condition 2, the weights, $v_{ij}(H)$, are determined solely by the higher ranked hypotheses contained in the intersection hypothesis H). The proof of Proposition 1 is complete.

Acknowledgements

Prof. Ajit Tamhane's research was supported by grants from the National Institute of National Heart, Lung and Blood Institute and National Security Agency.

Table 1. Adjusted p-values produced by the Bonferroni tree gatekeeping procedure based on Algorithm 1 (proposed in [1]) and Algorithm 2 (given in Section 2).

Family	Null hypothesis	Raw p-value	Serial rejection set	Parallel rejection set	Adjusted p-value	
					Algorithm 1	Algorithm 2
F_1	H_{11}	0.003			0.009*	0.009*
	H_{12}	0.011			0.033*	0.033*
	H_{13}	0.038			0.114	0.114
F_2	H_{21}	0.019	$\{H_{11}\}$		0.057	0.086
	H_{22}	0.006	$\{H_{12}, H_{13}\}$		0.114	0.114
	H_{23}	0.012	$\{H_{13}\}$		0.114	0.114
F_3	H_{31}	0.007		$\{H_{21}, H_{22}\}$	0.036*	0.086
	H_{32}	0.013		$\{H_{21}, H_{23}\}$	0.039*	0.086
	H_{33}	0.023		$\{H_{22}, H_{23}\}$	0.114	0.114

The asterisk identifies the adjusted p-values that are significant at the 0.05 level.