A tree structured gatekeeping testing procedure - a critical view.

George Kordzakhia

CDER, Office of Biostatistics
Division of Biometrics I

Introduction

Much work has been done to handle multiple testing problems while protecting the type I error rate. Some of the earlier works on this topic included the Bonferroni, Holm, Hochberg, and Hommel procedures.

More recent works on the topic introduced hierarchical structures to the testing procedure. Hypotheses were divided into families. For example, the primary family included primary hypotheses of interest, secondary and tertiary families included possible additional claims, etc.

The serial gatekeeping procedure (Bauer et al [1], Westfall et al. [10]) provided a way to test multiple families of hypotheses in a hierarchical manner where the secondary family was only examined if all hypotheses in the primary family (the gatekeeper) were rejected.

More recently, Dmitrienko et al. [4] and Chen et al. [2] weakened the requirement of serial gatekeeping procedure by allowing one to proceed to the secondary family if at least one hypothesis in the gatekeeper was rejected. They termed these procedures "parallel gatekeeping strategies."

In 2007, Dmitrienko et al. [6] introduced a tree-gatekeeping test. This procedure was designed to test multiple objectives in a more general pre-specified hierarchical order. More precisely, every null hypothesis had a serial set and a parallel set of null hypotheses assigned to it. The idea was that the null hypothesis could be tested only if all null hypotheses in the serial set were rejected and at least one was rejected in the parallel set. In 2008, an updated version of the procedure was presented in [7].

Example 1

Consider a hypothetical dose-response study with 3 doses of a treatment drug. The doses are compared with placebo in 3 endpoints. The endpoints are ordered based on their importance.

Let H_{ij} be a null hypothesis stating that jth dose of the drug is non-distinguishable from placebo in the i^{th} endpoint. Thus, null hypotheses are naturally grouped into 3 families:

$$\begin{aligned} \mathcal{F}_1 = &\quad H_{11} \quad H_{12} \quad H_{13} \\ \mathcal{F}_2 = &\quad H_{21} \quad H_{22} \quad H_{23} \\ \mathcal{F}_3 = &\quad H_{31} \quad H_{32} \quad H_{33} \end{aligned}$$

For each null hypothesis H_{ij} $(i \geq 2)$, the serial rejection set is $\{H_{1j}\}$ and the parallel rejection set $\mathcal{F}_{i-1}$.

The *serial* and *parallel* sets define the logical sequence of the tree gate-keeping. A null hypothesis H_{ij} can be tested only if

- **Condition A**: all hypotheses in the serial set are rejected;
- **Condition B**: at least one of the hypotheses in the parallel set is rejected.

Also, it is required that

- **Condition C**: rejection/acceptance of hypotheses in family $\mathcal{F}_i$ does not depend on the raw p-values of null hypotheses in families $\mathcal{F}_{i+1}, \mathcal{F}_{i+2}, \ldots$.

Application of closed testing principle:

For every intersection hypothesis H, there is an associated test procedure having the type I error at most α. The p-value of the procedure is denoted by p_H. According to the closed testing principle the adjusted p-values for null hypotheses H_{ij} are determined from

$$\tilde{p}_{ij} = \max_{\{H : H_{ij} \in H\}} p_H$$

i.e. the maximum of p_H taken over all possible H containing H_{ij}. At the end, all null hypotheses whose adjusted p-values are less than α are rejected. This approach allows to control the study-wise type I error at level α.

In order to satisfy conditions A and B the adjusted p-value for H_{ij}, $\tilde{p}_{ij}$ should be at least as large as *all* adjusted p-values in the serial set, and at least as large as *at least one* of the adjusted p-values in the parallel set.

Testing intersection hypothesis H:

Dmitrienko *et al.* use the weighted Bonferroni procedure to test H. To apply the procedure, the weights of null hypotheses H_{ij} composing H should be specified within H. The weights are denoted by $v_{ij}(H)$.

The p-value for testing H is then determined from:

$$p_H = \min_{(i,j)} p_{ij}/v_{ij}(H). \tag{1}$$

where p_{ij} are raw p-values of the null hypotheses.
Remark: if $v_{ij}(H) = 0$, it is assumed that $p_{ij}/v_{ij}(H) = 1$.

Weight assignment:

Initially, to every null hypothesis H_{ij} we assign some weight $w_{ij} > 0$ in such a way that the total weight of null hypotheses in family $\mathcal{F}_i$ $(1 \leq i \leq 3)$ is equal to 1.

Since H may contain null hypotheses from several families, the weights of null hypotheses H_{ij} composing H are adjusted from w_{ij} to $v_{ij}(H)$ with the requirement that

$$\sum v_{ij}(H) \leq 1.$$

Several algorithms were suggested to assign weights within intersection hypotheses. Below, we consider the algorithm presented in Dmitrienko *et al.* 2008 (the second version).

In order to satisfy Conditions A and B, the algorithm requires that

- $v_{ij}(H) = 0$ if H contains *at least one* elementary hypothesis from the serial set of H_{ij};
- $v_{ij}(H) = 0$ if H contains *all* elementary hypotheses from the parallel set of H_{ij}.

To ensure Condition C (independence) it is required that

- $v_{ij}(H)$ only depends on what elementary hypotheses from families $\mathcal{F}_1, \ldots \mathcal{F}_i$ are in H.

Suppose that the matrix of initial weights in our example is given by:

$$\begin{matrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{matrix}$$

Consider $H' = H_{13} \cap H_{21} \cap H_{22} \cap H_{23} \cap H_{31}$.
The weights are assigned as follows (approach by Dmitirenko *et al* 2008):

- $v_{13}(H') = 1/3$
  for the first family weights are unchanged;
- $v_{23}(H') = 0$
  since H' contains an element from the serial set for H_{23};
- $v_{21}(H') = v_{22}(H') = (1 - 1/3) \cdot 1/3 = 2/9$
  where $(1 - 1/3)$ is the weight left after the first family and $1/3$'s are the weights of H_{21} and H_{22} within the second family;
- $v_{31}(H') = 0$
  since H' contains all elements from the parallel set for H_{31}.

Consider $H'' = H_{13} \cap H_{22} \cap H_{31} \cap H_{32}$.

- $v_{13}(H'') = 1/3$
  for the first family weights are unchanged;
- $v_{22}(H'') = (1 - 1/3) \cdot 1/3 = 2/9$
  where $(1 - 1/3)$ is the weight left after the first family, and $1/3$ is the weight of H_{22} within the second family;
- $v_{31}(H'') = v_{32}(H'') = (1 - 5/9) \cdot 1/2 = 2/9$
  where $(1 - 5/9)$ is the weight left after families 1 and 2, and $1/2$'s are the relative weights of H_{31} and H_{32} within the set $\{H_{31}, H_{32}\}$.

In general,
for an intersection hypothesis H and null hypothesis $H_{ij} \in H$,
let indicator $\xi_{ij}(H)$ be 0 if H contains *at least one* elementary hypothesis from the serial set for H_{ij} **or** *all* elementary hypotheses from the parallel set for H_{ij}. Otherwise let $\xi_{ij}(H) = 1$.

Define (recursively) the adjusted weight of H_{ij}

$$v_{ij}(H) = (1 - v_1^* - \cdots - v_{i-1}^*) \cdot \xi_{ij}(H) \cdot w_{ij}$$

where v_k^* is the total weight "used" for hypotheses from family $\mathcal{F}_k$.
That is,

$$v_k^* = \sum_{H_{kj} \in H} v_{kj}(H).$$

Computation of adjusted p-values:

Suppose that the matrix of raw p-values in our example is given by:

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = \begin{pmatrix} 0.01 & 0.01 & 0.2 \\ 0.01 & 0.2 & 0.01 \\ 0.02 & 0.02 & 0.02 \end{pmatrix}$$

First, we need to compute p-values of all intersection hypotheses. For example,

$$p_{H'} =$$

$$\min\{p_{13}/v_{13}(H'), p_{21}/v_{21}(H'), p_{22}/v_{22}(H'), p_{23}/v_{23}(H'), p_{31}/v_{31}(H')\}$$

$$= 0.01 \div \frac{2}{9} = 0.045$$

(recall that $v_{13}(H') = 1/3$, $v_{21}(H') = v_{22}(H') = 2/9$, and $v_{23}(H') = v_{31}(H') = 0$)

For $H'' = H_{13} \cap H_{22} \cap H_{31} \cap H_{32}$,

$$p_{H''} =$$

$$\min\{p_{13}/v_{13}(H''), p_{22}/v_{22}(H''), p_{31}/v_{31}(H''), p_{32}/v_{32}(H'')\} =$$

$$= \min\{0.2 \div \frac{1}{3}, 0.2 \div \frac{2}{9}, 0.02 \div \frac{2}{9}, 0.02 \div \frac{2}{9}\} = 0.09.$$

After considering all intersection hypotheses, we get

$$\tilde{p}_{31} = \max_{\{H: H_{31} \in H\}} p_H = p_{H''} = 0.09.$$

Similarly, we obtain p-values for other null hypotheses in the study.

The matrix of adjusted p-values is given by:

$$\begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} & \tilde{p}_{13} \\ \tilde{p}_{21} & \tilde{p}_{22} & \tilde{p}_{23} \\ \tilde{p}_{31} & \tilde{p}_{32} & \tilde{p}_{33} \end{pmatrix} = \begin{pmatrix} 0.03 & 0.03 & 0.6 \\ 0.045 & 0.6 & 0.6 \\ 0.09 & 0.09 & 0.6 \end{pmatrix}$$

Thus, H_{11}, H_{12}, and H_{21} can be rejected at 0.05 level.

The weight assigning algorithms proposed by Dmitrienko et al. [6] and [7] may violate condition B. In some cases, a null hypothesis is rejected while failing to reject at least one hypothesis in the corresponding parallel set.

Example 2

Consider a case of four endpoints and two doses: $m = 4,\ k = 2$.

$$\begin{aligned} \mathcal{F}_1 = &\quad H_{11} \quad H_{12} \\ \mathcal{F}_2 = &\quad H_{21} \quad H_{22} \\ \mathcal{F}_3 = &\quad H_{31} \quad H_{32} \\ \mathcal{F}_4 = &\quad H_{41} \quad H_{42} \end{aligned}$$

As in Example 1, for each null hypothesis H_{ij} $(i \geq 2)$, the serial rejection set is $\{H_{1j}\}$ and the parallel rejection set $\mathcal{F}_{i-1}$.
Suppose the weight matrix is given by:

$$\begin{aligned} \mathcal{F}_1 = &\quad 3/4 \quad 1/4 \\ \mathcal{F}_2 = &\quad 1/2 \quad 1/2 \\ \mathcal{F}_3 = &\quad 1/2 \quad 1/2 \\ \mathcal{F}_4 = &\quad 1/2 \quad 1/2 \end{aligned}$$

Condition B may break if unadjusted p-values are within a certain range. Suppose that raw p-values are given by matrix below:

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \\ p_{31} & p_{32} \\ p_{41} & p_{42} \end{pmatrix} = \begin{pmatrix} 0.001 & 0.1 \\ 0.001 & 0.1 \\ 0.015 & 0.001 \\ 0.001 & 0.001 \end{pmatrix}$$

Then, both hypotheses in the third family can't be rejected at 0.05 level, but the adjusted p-value for H_{41} is less than 0.05. Indeed,

$$\tilde{p}_{31} = p_{\{H_{22},\ H_{31}\}} = 4p_{31} = 0.06 > 0.05,$$

$$\tilde{p}_{32} = p_{\{H_{12},\ H_{32}\}} = 0.4,$$

$$\tilde{p}_{41} = p_{\{H_{12},\ H_{31},\ H_{32},\ H_{41}\}} = p_{31}/(3/4 * 1/2) = 0.04 < 0.05.$$

The matrix of adjusted p-values is easily computable:

$$\begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} \\ \tilde{p}_{21} & \tilde{p}_{22} \\ \tilde{p}_{31} & \tilde{p}_{32} \\ \tilde{p}_{41} & \tilde{p}_{42} \end{pmatrix} = \begin{pmatrix} 0.0013 & 0.4 \\ 0.0026 & 0.4 \\ 0.06 & 0.4 \\ 0.04 & 0.4 \end{pmatrix}$$

Readjustment of p-values:

To satisfy condition B, and preserve conditions A and C, the testing procedures can be modified by increasing the adjusted p-values $\tilde{p}_{ij}$. The readjustment is performed by families in the increasing order from $\mathcal{F}_1$ to $\mathcal{F}_m$.

In Example 2, it is sufficient to readjust $\tilde{p}_{41}$ by making it at least as large as $\tilde{p}_{11}$ and at least as large as minimum of $\tilde{p}_{31}$ and $\tilde{p}_{32}$.

$$\hat{p}_{41} = \max\{\tilde{p}_{41}, \tilde{p}_{11}, \min\{\tilde{p}_{31}, \tilde{p}_{32}\}\} = 0.06.$$

Readjustment procedure (K. et al. 2007, unpublished)

Family $\mathcal{F}_1$: No changes are needed in the first family. For $j = 1, \ldots, k_1$,

$$\hat{p}_{ij} = \tilde{p}_{ij}$$

Family $\mathcal{F}_i$ $(2 \leq i \leq m)$: For $j = 1, \ldots, k_i$, set

$$\hat{p}_{ij} = \max\{\tilde{p}_{ij}, \max_{H_{nl} \in \mathcal{R}_{ij}^S} \hat{p}_{nl}, \min_{H_{nl} \in \mathcal{R}_{ij}^P} \hat{p}_{nl}\} \tag{2}$$

where $\mathcal{R}_{ij}^S$ and $\mathcal{R}_{ij}^P$ are respectively serial and parallel sets. Observe that $\hat{p}_{ij} \geq \tilde{p}_{ij}$, and thus the study-wise type I error is controlled. It is easy to see that conditions A and B are enforced by (2). Since the independence condition (condition C) holds for the adjusted p-values $\tilde{p}_{ij}$ $(1 \leq i \leq m,\ 1 \leq j \leq i_k)$, and since $\hat{p}_{ij}$ is determined by $\{\tilde{p}_{nl} :\ 1 \leq n \leq i,\ 1 \leq l \leq k_n\}$, the independence condition also holds for $\hat{p}_{ij}$.

Modified weight assignment (K. et al. 2007) :
Given an intersection hypothesis H and null hypothesis $H_{ij} \in H$, let $\xi_{ij}(H)$ be defined as before, set $v_0^* = 0$, and define $v_{ij}(H)$, the adjusted weight of H_{ij} within H, by

$$v_{ij}(H) = (1 - v_0^* - \cdots - v_{i-1}^*) \cdot \xi_{ij}(H) \cdot w_{ij}$$

where

$$v_k^* = (1 - v_0^* - \cdots - v_{k-1}^*) \cdot \sum_{H_{kj} \in H} w_{kj}.$$

In K. et al. 2007 it is shown that this algorithm satisfies properties A, B and C.

References

[1] Bauer, P, Rohmel, J., Maurer, W., and Hothorn, L. (1998). *Testing strategies in multi-dose experiements including active control.* Statistics in Medicine vol. 17, 2133-2146.

[2] Chen, X., Luo, X., and Capizzi, T. (2005). *The application of ehanced parallel gatekeeping strategies.* Statistics in Medicine vol. 24, 1385-1397

[3] Chi, G. (1998). *Multiple testings: multiple comparisons and multiple endpoints.* Drug Information Journal vol. 32, 1347S-1362S

[4] Dmitrienko, A., Offen W. and Westfall, P. (2003). *Gatekeeping strategies for clinical trials that do not require all primary effects to be significant.* Statistics in Medicine Vol. 22, 2387–2400

[5] Dmitrienko A., Offen W., Wang O. and Xiao D. (2006) *Gatekeeping procedures in dose-response clinical trials based on the Dunnett test.* Pharmaceutical Statistics Vol 5, 19–28

[6] Dmitrienko A., Wiens B., Tamhane A. and Wang X. (2007). *Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives.* Statistics in Medicine Vol. 26, 2465–2478

[7] Dmitrienko A., Tamhane A., Liu L. and Wiens B. (2008). *A note on tree gatekeeping proceduresd in clinical trials.* Statistics in Medicine Vol. 06, 1–6

[8] Kordzakhia G., Dinh P., Bai S., Lawrence J. and Yang P. (2007). *Bonferroni-based tree-structured gatekeeping testing procedures.* Unpublished

[9] Marcus R., Peritz E. and Gabriel K.R. (1976) *On closed testing procedures with special reference to ordered analysis of*

variance. Biometrika Vol. 63, 655–660

[10] Westfal, PH. and Krishen, A. (2001). *Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures.* Journal of Statistical Planning and Inference vol. 99, 25-40