A note on tree gatekeeping procedures in clinical trials

Alex Dmitrienko (Eli Lilly)
Ajit C. Tamhane (Northwestern University)
Lingyun Liu (Northwestern University)
Brian L. Wiens (Gilead Colorado)

Abstract

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.

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). The hypotheses included in \( F_i \), \( i = 1, \ldots, m \), are denoted by \( H_{i1}, \ldots, H_{im} \) 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 \( \alpha \).

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 \( R^S_{ij} \)) is accepted or all hypotheses in its parallel rejection set (denoted by \( R^P_{ij} \)) are accepted. Here \( R^S_{ij} \) and \( R^P_{ij} \) 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 $\tilde{p}_{ij}$) is defined as $\tilde{p}_{ij} = \max_H p(H)$, where the maximum is taken over all intersection hypotheses $H \ni H_{ij}$. The hypothesis $H_{ij}$ is rejected if $\tilde{p}_{ij} \leq \alpha$.

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 $\alpha$-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 $i$ ($1 \leq i \leq m - 1$) 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.

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 \( v_{21}^*(H) = 2/3 \). 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


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 \( \alpha \), 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 \( H_{ijr}, t = 1, \ldots, u \), denote the distinct hypotheses contained in \( H_{rs}^{**} \), i.e., \( H_{rs}^{**} = \bigcap_{t=1}^{u} H_{ijr} \).

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, \( v_{ijr}(H_{rs}^*) = v_{ijr}(H_{rs}^{**}) \) since \( H_{ijr} \) is contained in both \( H_{rs}^{**} \) and \( H_{rs}^* \). Hence \( p(H_{rs}^{**}) \geq p(H_{rs}^*) > \alpha \). Now consider the intersection hypothesis

\[
H^* = H_{ij} \cap H_{rs}^{**} = H_{ij} \cap \left( \bigcap_{t=1}^{u} H_{ijr} \right).
\]

Note that this intersection hypothesis contains at least one hypothesis in \( R_s^* \) (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

\[
p(H^*) = \min_{t=1, \ldots, u} \left\{ \frac{p_{ijr}}{v_{ijr}(H^*)} \right\}.
\]

By Condition 3, \( v_{ijr}(H^*) \leq v_{ijr}(H_{rs}^{**}), t = 1, \ldots, u \), since \( H_{ijr} \) is contained in both \( H^* \) and \( H_{rs}^{**} \) and \( H_{ijr} \) 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 \( R_{ij}^p \) be the parallel rejection set of the hypothesis \( H_{ij}, i = 2, \ldots, m, j = 1, \ldots, n_i \). Let \( H_{v,rj}, r = 1, \ldots, s, \) denote the hypotheses in \( R_{ij}^p \) and suppose none of them is rejected. This implies that there exist intersection hypotheses, denoted by \( H_{v,rj}^*, r = 1, \ldots, s \), such that \( H_{v,rj}^* \) contains \( H_{v,rj} \) and \( p(H_{v,rj}^*) > \alpha \). If \( H_{ij} \) is contained in at least one intersection \( H_{v,rj}^*, r = 1, \ldots, s, \) then \( H_{ij} \) is not rejected. If \( H_{ij} \) is not contained in any intersection \( H_{v,rj}^*, r = 1, \ldots, s, \) then let \( H_{v,rj}^{**}, r = 1, \ldots, s, \) be the intersection hypotheses obtained by eliminating hypotheses included in the last family from \( H_{v,rj}^* \). Since the weights are sequentially assigned by Condition 2, the weight of any hypothesis in \( H_{v,rj}^{**}, r = 1, \ldots, s, \) is equal to its weight in \( H_{v,rj}^* \). Hence \( p(H_{v,rj}^{**}) \geq p(H_{v,rj}^*) > \alpha, r = 1, \ldots, s. \) Let

\[
H^* = H_{ij} \cap \left( \bigcap_{r=1}^{s} H_{v,rj}^{**} \right).
\]

Let \( H_{klt}, t = 1, \ldots, u, \) be the distinct hypotheses in \( \bigcap_{r=1}^{s} H_{v,rj}^{**}, \) i.e., \( \bigcap_{r=1}^{s} H_{v,rj}^{**} = \bigcap_{t=1}^{u} H_{klt} \). To compute the \( p \)-value for \( H^* \), note first that \( H^* \) includes all hypotheses from \( R_{ij}^p \). By Condition 1, this implies that \( v_{ij}(H^*) = 0 \) and the \( p \)-value for \( H^* \) is given by

\[
p(H^*) = \min_{t=1, \ldots, u} \left\{ \frac{p_{klt}}{v_{klt}(H^*)} \right\}.
\]
Further, for any hypothesis \( H_{kt} \), \( t = 1, \ldots, u \), identify the intersection \( H_{irj}^{**} \) that contains \( H_{kt} \). Recall that the Bonferroni \( p \)-value for any \( H_{irj}^{**}, \) \( r = 1, \ldots, s \), is greater than \( \alpha \), which implies that \( p_{kt}/v_{kt}(H_{irj}^{**}) > \alpha \). By Condition 3, \( v_{kt}(H^*) \leq v_{kt}(H_{irj}^{**}) \) since \( H_{kt} \) is contained in both \( H^* \) and \( H_{irj}^{**} \) and \( H_{kt} \) is not from the last family. Thus, \( p_{kt}/v_{kt}(H^*) > \alpha \) 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).

<table>
<thead>
<tr>
<th>Family</th>
<th>Null hypothesis</th>
<th>Raw $p$-value</th>
<th>Serial rejection set</th>
<th>Parallel rejection set</th>
<th>Adjusted $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>$H_{11}$</td>
<td>0.003</td>
<td></td>
<td></td>
<td>0.009*</td>
</tr>
<tr>
<td></td>
<td>$H_{12}$</td>
<td>0.011</td>
<td></td>
<td></td>
<td>0.033*</td>
</tr>
<tr>
<td></td>
<td>$H_{13}$</td>
<td>0.038</td>
<td></td>
<td></td>
<td>0.114</td>
</tr>
<tr>
<td>$F_2$</td>
<td>$H_{21}$</td>
<td>0.019</td>
<td>${H_{11}}$</td>
<td></td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>$H_{22}$</td>
<td>0.006</td>
<td>${H_{12}, H_{13}}$</td>
<td></td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>$H_{23}$</td>
<td>0.012</td>
<td>${H_{13}}$</td>
<td></td>
<td>0.114</td>
</tr>
<tr>
<td>$F_3$</td>
<td>$H_{31}$</td>
<td>0.007</td>
<td>${H_{21}, H_{22}}$</td>
<td></td>
<td>0.036*</td>
</tr>
<tr>
<td></td>
<td>$H_{32}$</td>
<td>0.013</td>
<td>${H_{21}, H_{23}}$</td>
<td></td>
<td>0.039*</td>
</tr>
<tr>
<td></td>
<td>$H_{33}$</td>
<td>0.023</td>
<td>${H_{22}, H_{23}}$</td>
<td></td>
<td>0.114</td>
</tr>
</tbody>
</table>

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