Powerful short-cut procedures for gatekeeping strategies

Frank Bretz
frank.bretz@novartis.com


MCP 2007, Vienna, July 08-11, 2007
Example 1

- Two dose groups versus a placebo control for primary and co-primary endpoint

- Gatekeeping procedure by Dmitrienko et al. (2003) can be used:

  “One family of hypotheses (comprising the primary objectives) is treated as a ‘gatekeeper’, and the other family or families comprising secondary and tertiary objectives) are tested only if one or more gatekeeper hypotheses have been rejected.”
Example 1

- Example: 2 primary hypotheses $H_1$, $H_2$, and 2 secondary $H_3$, $H_4$

<table>
<thead>
<tr>
<th>Intersection hypothesis</th>
<th>$H_1$</th>
<th>$H_2$</th>
<th>$H_3$</th>
<th>$H_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1 \cap H_2 \cap H_3 \cap H_4$</td>
<td>0.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_1 \cap H_2 \cap H_3$</td>
<td>0.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_1 \cap H_2 \cap H_4$</td>
<td>0.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_1 \cap H_2$</td>
<td>0.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_1 \cap H_3 \cap H_4$</td>
<td>0.5</td>
<td>0.0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$H_1 \cap H_3$</td>
<td>0.5</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_1 \cap H_4$</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>$H_1$</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_2 \cap H_3 \cap H_4$</td>
<td>0.0</td>
<td>0.5</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$H_2 \cap H_3$</td>
<td>0.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_2 \cap H_4$</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>$H_2$</td>
<td>0.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_3 \cap H_4$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>$H_3$</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$H_4$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

- Complicated and needs careful implementation ➔ short cuts?
Example 2

- Two treatments compared against each other
- Several (seven) secondary endpoints relevant for submission: Some HAs such as the FDA requires multiplicity adjustment for all variables one wants to make a claim from

- QoL = Quality of Life: summary score computed from 4 domain scores $D_1, ..., D_4$
- Two additional clinical endpoints $E_1$ and $E_2$
- $D_1, ..., D_4$ only of interest, if QoL is significant; no hierarchy among $D_i$
- QoL, $E_1$ and $E_2$ are of equal importance

- How to adjust for multiplicity?
A general concept: consonant test strategies

- Given: \( n \) null hypotheses \( H_1, \ldots, H_n \).
- Intersection hypothesis \( H_I = \bigcap \{ H_i : i \in I \} \)
- Local test of \( H_I \) defined by p-value \( p_I \):
  
  Reject \( H_I \), if and only if \( p_I \leq \alpha \)

- Consonance condition: If \( H_i \) is rejected locally, there exists at least one \( j \in I \), such that \( H_j \) can be rejected locally, for all \( J \subseteq I \) and \( j \in J \)
  (in particular \( H_j = H_{\{j\}} \), too!)
- Consequence: Performing the closure test becomes very simple
  (maximally \( n \) steps necessary).
- Question: Possible with weighted Bonferroni tests?
Weighted Bonferroni tests

- Choose for each index set \( I \subseteq \{1, \ldots, n\} \) weights \( w_i \) for all \( i \in I \) with \( \sum (w_i : i \in I) \leq 1 \).

- Local test of \( H_I = \cap \{H_i : i \in I\} \):
  
  Reject \( H_I \) if \( p_i \leq w_i \cdot \alpha \) for at least one \( i \)

- Apply the closure test for the system of all \( H_I \) \( \Rightarrow \) FWER is controlled.

- Assume now that for every pair \( I, J \) of index sets with \( J \subseteq I \),
  
  \[ w_i(I) \leq w_i(J) \quad \text{for all} \quad i \in J. \]

  Then the consonance condition is satisfied, and a short-cut can be performed, based on weighted p-values \( q_i = p_i/w_i(I) \).
Special cases

- **Weighted Bonferroni-Holm procedure:**
  
  Fixed weights $v_1, \ldots, v_n$
  
  $i \in I \Rightarrow w_i(l) = v_i / (\sum v_j : j \in I)$

- **Fixed sequence tests** (Bauer et al., 1998; Hommel & Kropf, 2005)

- The „fallback procedure“ (Wiens, 2003; Wiens & Dmitrienko, 2005)

- All „reasonable“ gatekeeping procedures
  - Maurer (1987), ROeS, Locarno
  - Maurer et al. (1995), Bauer et al. (1998)
  - Chen et al. (2005): Simes test
  - Dmitrienko and colleagues ... 2006 ... 2007 ...
Example 2 revisited

- Short cut of gatekeeping procedure:
  1. Test $E_1$ and $E_2$ separately at level $\alpha/3$
  2. If both are significant, then test QoL at $\alpha$
      - If one is significant, then test QoL at $2\alpha/3$
      - If none is significant, then test QoL at $\alpha/3$
  3. If QoL is significant, then test domain scores with either $\alpha$, $2\alpha/3$
     or $\alpha/3$ by adjusting for multiplicity, depending on step 2. The test
     for domain scores can be fixed sequence, closed or any other, as
     long as the above $\alpha$’s are used.

- Why does it control level $\alpha$?
Example 2 revisited

- The following rules define a closed test procedure:
  - When $E_1$ or $E_2$ are in an intersection hypothesis, assign weights $1/3$ to each
  - When QoL is in an intersection hypothesis, assign it the remaining weight not allocated to $E_1$ or $E_2$
  - When QoL is not in an intersection hypothesis, assign equally to $D_1, \ldots, D_4$ the remaining weights not allocated to $E_1$ or $E_2$
Example 2 revisited

- **Example:**
  - $p_{QoL} = 0.015$, $p_{E1} = 0.005$, $p_{E2} = 0.097$
  - $p_{D1} = 0.006$, $p_{D2} = 0.004$, $p_{D3} = 0.008$, $p_{D4} = 0.04$

<table>
<thead>
<tr>
<th>index set $I$</th>
<th>local p-value</th>
<th>identified endpoint</th>
<th>adjusted p-value $p_I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${QoL, E_1, E_2, D_1, \ldots, D_4}$</td>
<td>$\min{3p_{QoL}, 3p_{E1}, 3p_{E2}} = 0.015$</td>
<td>$E_1$</td>
<td>0.015</td>
</tr>
<tr>
<td>${QoL, E_2, D_1, \ldots, D_4}$</td>
<td>$\min{\frac{3}{2}p_{QoL}, 3p_{E2}} = 0.0225$</td>
<td>$QoL$</td>
<td>0.0225</td>
</tr>
<tr>
<td>${E_2, D_1, \ldots, D_4}$</td>
<td>$\min{3p_{E2}, 6p_{D1}, \ldots, 6p_{D4}} = 0.024$</td>
<td>$D_2$</td>
<td>0.024</td>
</tr>
<tr>
<td>${E_2, D_1, D_3, D_4}$</td>
<td>$\min{3p_{E2}, \frac{9}{2}p_{D1}, \frac{9}{2}p_{D3}, \frac{9}{2}p_{D4}} = 0.027$</td>
<td>$D_1$</td>
<td>0.027</td>
</tr>
<tr>
<td>${E_2, D_3, D_4}$</td>
<td>$\min{3p_{E2}, 3p_{D3}, 3p_{D4}} = 0.024$</td>
<td>$D_3$</td>
<td>0.027</td>
</tr>
<tr>
<td>${E_2, D_4}$</td>
<td>$\min{3p_{E2}, \frac{3}{2}p_{D_4}} = 0.06$</td>
<td>$D_4$</td>
<td>0.06</td>
</tr>
<tr>
<td>${E_2}$</td>
<td>$p_{E2} = 0.097$</td>
<td>$E_2$</td>
<td>0.097</td>
</tr>
</tbody>
</table>
Discussion

• Desirable:
  – Simple procedures (but as powerful as possible)
  – Study protocol should be understood not only by statistical experts („only“ necessary: understanding of sequentially rejective procedures, e.g., Holm)
  – Interpretation of results are straightforward

• Weighted Bonferroni tests: very general construction possible
  Choice of weights reasonable $\Rightarrow$ monotonicity condition satisfied

• Short cut procedures allow a very simple implementation (no need for complex programs)
Discussion

- **Improvement of Bonferroni adjustment?**
  - Šidák possible when applicable
  - Resampling methods: Romano & Wolf, 2005; Dmitrienko et al., 2007
  - Simes tests: very complicated + time-consuming; not consonant

- **Two demands on gatekeeping procedures by Dmitrienko et al. (2003)**
  - Rejection of secondary hypotheses only when primary hypothesis rejected
  - Decisions on primary hypotheses do not depend on secondary decisions
    - sum of weights often < 1
    - sometimes unnecessary loss in power ?