Powerful short-cut procedures for gatekeeping strategies

Frank Bretz

frank.bretz@novartis.com

Hommel, G., Bretz, F., and Maurer, W. (2007)
Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statistics in Medicine* (in press)

MCP 2007, Vienna, July 08-11, 2007

NOVARTIS

Outline

- **Motivating examples**
- **Methods**
- **Example revisited**
- **Discussion**

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 I. Weights assigned to the intersection hypothesis tests.

Intersection hypothesis	Weights			
	H_1	H_2	H_3	H_4
$H_1 \cap H_2 \cap H_3 \cap H_4$	0.5	0.5	0.0	0.0
$H_1 \cap H_2 \cap H_3$	0.5	0.5	0.0	0.0
$H_1 \cap H_2 \cap H_4$	0.5	0.5	0.0	0.0
$H_1 \cap H_2$	0.5	0.5	0.0	0.0
$H_1 \cap H_3 \cap H_4$	0.5	0.0	0.25	0.25
$H_1 \cap H_3$	0.5	0.0	0.5	0.0
$H_1 \cap H_4$	0.5	0.0	0.0	0.5
H_1	0.5	0.0	0.0	0.0
$H_2 \cap H_3 \cap H_4$	0.0	0.5	0.25	0.25
$H_2 \cap H_3$	0.0	0.5	0.5	0.0
$H_2 \cap H_4$	0.0	0.5	0.0	0.5
H_2	0.0	0.5	0.0	0.0
$H_3 \cap H_4$	0.0	0.0	0.5	0.5
H_3	0.0	0.0	1.0	0.0
H_4	0.0	0.0	0.0	1.0

- **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 $H_1, \dots, H_n$.
- Intersection hypothesis $H_I = \cap \{ H_i : i \in I \}$
- 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 ➔ FWER is controlled.

- Assume now that for every pair I, J of index sets with $J \subseteq I$,

  $$w_i(I) \leq w_i(J) \quad \text{for all } i \in J .$$

  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 $v_1, \dots, v_n$

  $i \in I \Rightarrow w_i(I) = v_i / (\Sigma v_j : j \in I)$

- 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)
  - Westfall and Krishen (2001): „gatekeeper“
  - Dmitrienko, Offen, and Westfall (2003): „serial / parallel gatekeeping“
  - 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 α

   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 α, $2\alpha/3$ 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 α's are used.

- **Why does it control level α ?**

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, …, D_4 the remaining weights not allocated to E_1 or E_2**

Example 2 revisited

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

index set I	local p-value for H_I	identified endpoint	adjusted p-value p_I
$\{QoL, E_1, E_2, D_1, \ldots, D_4\}$	$\min\{3p_{QoL}, 3p_{E_1}, 3p_{E_2}\} = .015$	E_1	.015
$\{QoL, E_2, D_1, \ldots, D_4\}$	$\min\{\frac{3}{2}p_{QoL}, 3p_{E_2}\} = .0225$	QoL	.0225
$\{E_2, D_1, \ldots, D_4\}$	$\min\{3p_{E_2}, 6p_{D_1}, \ldots, 6p_{D_4}\} = .024$	D_2	.024
$\{E_2, D_1, D_3, D_4\}$	$\min\{3p_{E_2}, \frac{9}{2}p_{D_1}, \frac{9}{2}p_{D_3}, \frac{9}{2}p_{D_4}\} = .027$	D_1	.027
$\{E_2, D_3, D_4\}$	$\min\{3p_{E_2}, 3p_{D_3}, 3p_{D_4}\} = .024$	D_3	.027
$\{E_2, D_4\}$	$\min\{3p_{E_2}, \frac{3}{2}p_{D_4}\} = .06$	D_4	.06
$\{E_2\}$	$p_{E_2} = .097$	E_2	.097

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 straight forward

- Weighted Bonferroni tests: very general construction possible

  Choice of weights reasonable ➔ 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 ?