

# Multiple and repeated testing of primary, co-primary and secondary hypotheses

Ekkehard Glimm, Willi Maurer, Novartis  
with contributions by Frank Bretz

**BBS Spring Seminar, 7 March 2011, Basel**



# Outline

---

- Analysis of primary and secondary endpoints in multiple treatment arms
  - Importance of primary and secondary variables in regulatory submissions
  - „Consistency“ and separate control of type I error for primary and secondary hypotheses
  - Properties of two consistent strategies

Maurer W, Glimm E, Bretz F. Multiple and repeated testing of primary, co-primary and secondary hypotheses.

*Statistics in Biopharmaceutical Reserach 2010 (published online)*

# Analysis of primary and secondary endpoints

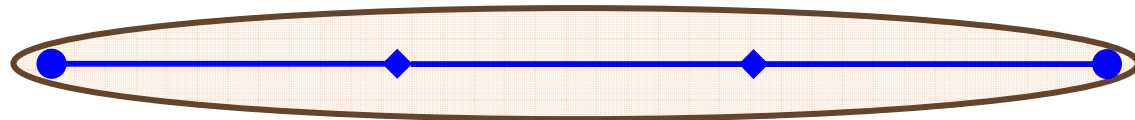
---

- Case: parallel group trial with
  - 2 treatment arms (e.g. high/low dose)
  - a primary and a secondary endpoint (e.g. HBA1C reduc. / weight loss)
- Serial gatekeeping
  - Test primary hypotheses, secondary only if both primary are rejected
- Hung and Wang (2009, J. of Biopharm. Statistics, 19:1, 1 - 11):
  - does not make „common sense“ to condition the rejection of a secondary endpoint on the rejection of the primary endpoint in another dose group
  - Parallel gatekeeping procedures can help
  - *“Fundamental question is whether the studywise type I error rate needs to be tied in with testing secondary endpoints.”*

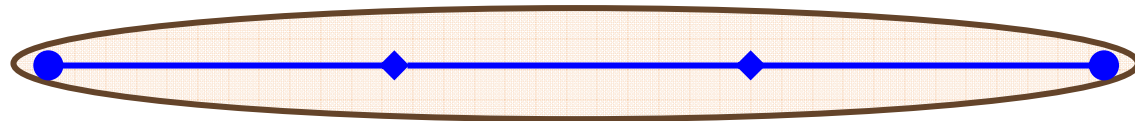
# Primary and Secondary hypotheses

- General gatekeeping procedures assume „unrelated blocks“ to be tested hierarchically

*Primary hypotheses*

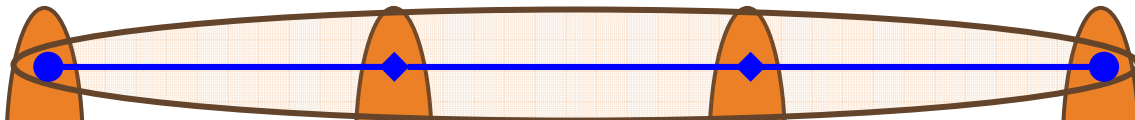


*Secondary hypotheses*

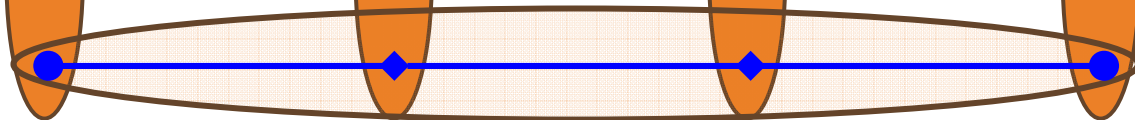


- As just seen, concrete examples often impose more structure („parent-descendant“ relation of primary and a corresponding secondary)

*Primary hypotheses*



*Secondary hypotheses*



# Structured families of primary and secondary hypotheses

---

- Consider a family of hypotheses that can be partitioned into
  - a primary family and a secondary family
  - for each primary hypothesis there is a set of “descendant” secondary hypotheses.
  - each secondary hypothesis has at least one “parent” primary hypothesis.
- A secondary hypothesis is only of interest (in a confirmatory sense) if one of the respective “parent” primary hypothesis is rejected.

# Importance of primary and secondary variables in regulatory submissions

---

- Purpose of successful claims on
  - **primary variables**: usually the basis for regulatory approval
  - **secondary variables**: often only for the qualification of an established primary effect
    - only of interest if a “related primary” null-hypothesis could be rejected.
- Definition of consistent and successive procedures:
  - **Successive**: A secondary hypothesis can only be rejected if at least one of its parent primary hypotheses is rejected,
  - **Consistent**: successive + retention of a secondary hypothesis cannot preclude the rejection of a primary hypothesis.

# Analysis of primary and secondary endpoints

## Notation

---

- primary endpoints  $P$ , secondary endpoint  $S$ 
  - **Primary hypotheses**  $F_P = \{H_{1,P}, H_{2,P}\}$ ; rejection of either  $H_{1,P}$  or  $H_{2,P}$  is a prerequisite for a positive study
  - Primary type I error to be controlled at level  $\alpha_P = \alpha$  (e.g.,  $\alpha = 0.025$ )
  - **Descendant secondary hypotheses**  $F_S = \{H_{1,S}, H_{2,S}\}$ ; rejection of either  $H_{1,S}$  or  $H_{2,S}$  is not sufficient for a positive study, but may allow label claims
  - Secondary type I error to be controlled at level  $\alpha_S$  (can be larger than  $\alpha$ )
  - Family of prim. and sec. hypotheses  $F_{PS} = F_S \cup F_P$
  - $R(H)$  = rejection of a null hypothesis  $H$
- $R(F)$  = erroneous rejection of at least one true null hypothesis from a family  $F$  of hypotheses

# Analysis of primary and secondary endpoints

## Consistent strategies

---

- Test  $F_P$  with a closed test at multiple level  $\alpha_P = \alpha$ 
  - Let  $cF_S$  denote the (consistent) family of secondary hypotheses with rejected parent primary hypothesis
- Strategies:
  - $S_1$  : Test  $cF_S$  at multiple level  $\alpha_S = \alpha$
  - $S_2$  : Test the members of  $cF_S$  individually at level  $\alpha_S = \alpha$
- Strategies  $S_1$  and  $S_2$  are consistent and
  - 1)  $P(R(F_P)) \leq \alpha$
  - 2)  $P(R(F_S)) \leq 2\alpha$
  - 3)  $P(R(F_P) \text{ or } R(F_S)) = P(R(F_{PS})) \leq 2\alpha$
  - 4) Boundaries cannot be tightened in general

# Typical scenarios

---

„Generators“ of such multiple endpoint scenarios:

1. Two endpoints
    - e.g. Ophthalmology: visual acuity, eye inflammation
  2. Two treatments (e.g. doses) with same drug vs control
  3. Two subpopulations
    - E.g. Cardiovascular: HBA1C change from baseline in Chinese subpopulation/  
all patients
  4. Non-inferiority and superiority in an endpoint
- Situation arises whenever **2 of these 4** occur in combination.
  - One is strictly „hierarchical“ (primary/secondary), the other may have any preference structure.

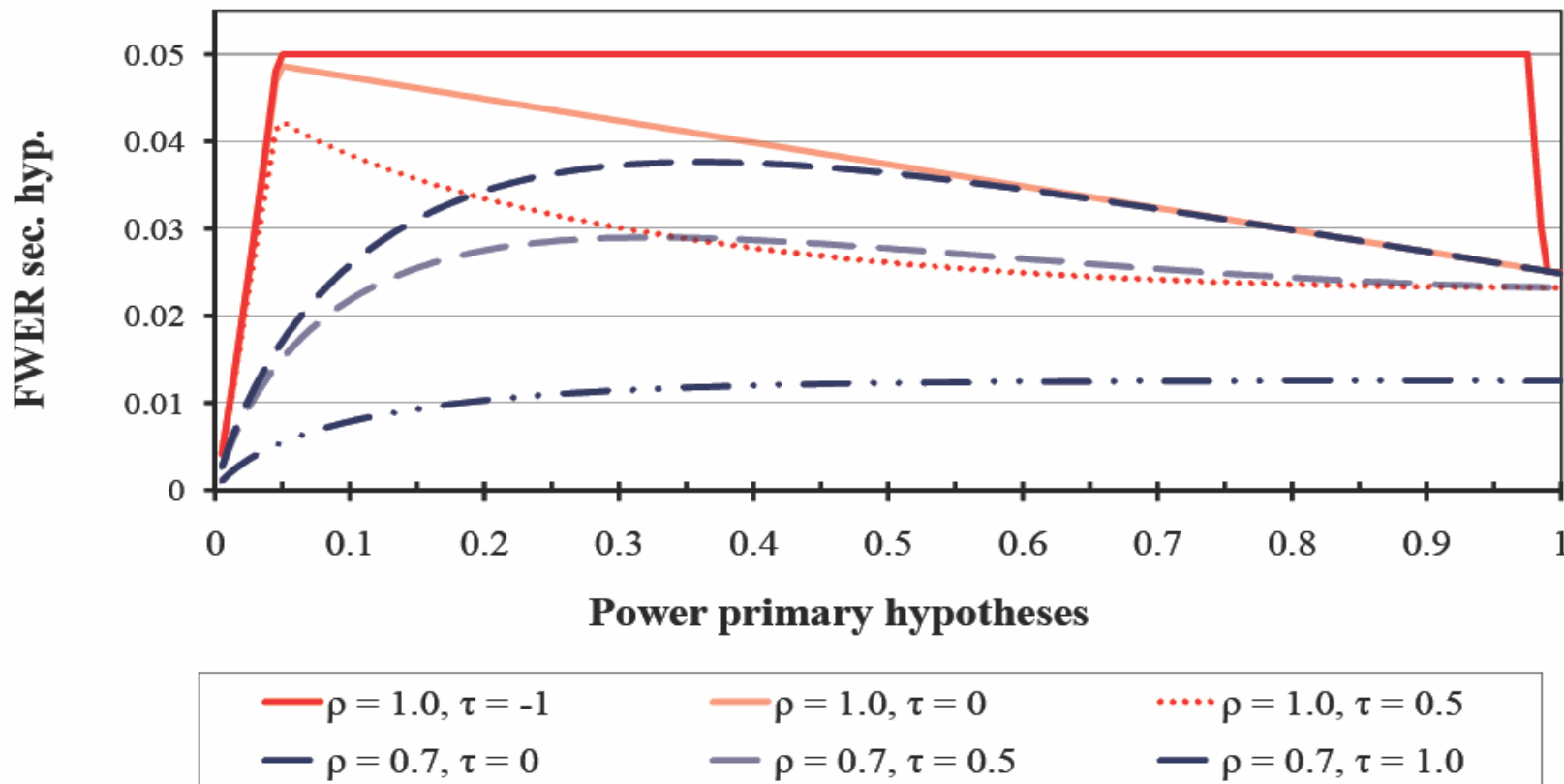
## Typical scenarios

Assume the corresponding test statistics are normally distributed with correlation  $\rho$  and  $\tau$ , respectively.

Strictly Hierarchical (prim/sec)	Any preference structure	Corr between prim and sec	Corr within prim and within sec
Endpoints	Treatment	$\rho \in (-1,1)$	$n_{trt}/(n_{trt}+n_{ctrl})$
Endpoints	Subpop	$\rho \in (-1,1)$	$n_{comm}/(n_1n_2)^{1/2}$
Endpoints	Endpoints	$\rho_1, \rho_2 \in (-1,1)$	$\tau_1, \tau_2 \in (-1,1)$
Treatment	Endpoints	$n_{trt}/(n_{trt}+n_{ctrl})$	$\tau \in (-1,1)$
Subpop	Endpoints	$n_{comm}/(n_1n_2)^{1/2}$	$\tau \in (-1,1)$
Non-Inf / Sup	Endpoints	1 or $(n_{PP}/n_{ITT})^{1/2}$	$\tau \in (-1,1)$
Non-Inf / Sup	Subpop	1 or $(n_{PP}/n_{ITT})^{1/2}$	$n_{comm}/(n_1n_2)^{1/2}$
Non-Inf/ Sup	Treatment	1 or $(n_{PP}/n_{ITT})^{1/2}$	$n_{trt}/(n_{trt}+n_{ctrl})$

# FWER for strategy $S_1$

Actual type I error of strategy  $S_1$  at nominal  $\alpha=0.025$   
(when  $H_{1,S}, H_{2,S}$  are both true, Bonf-Holm test for  $F_P$  and  $cF_S$ )



# Multiple primary and secondary hypotheses: Summary

---

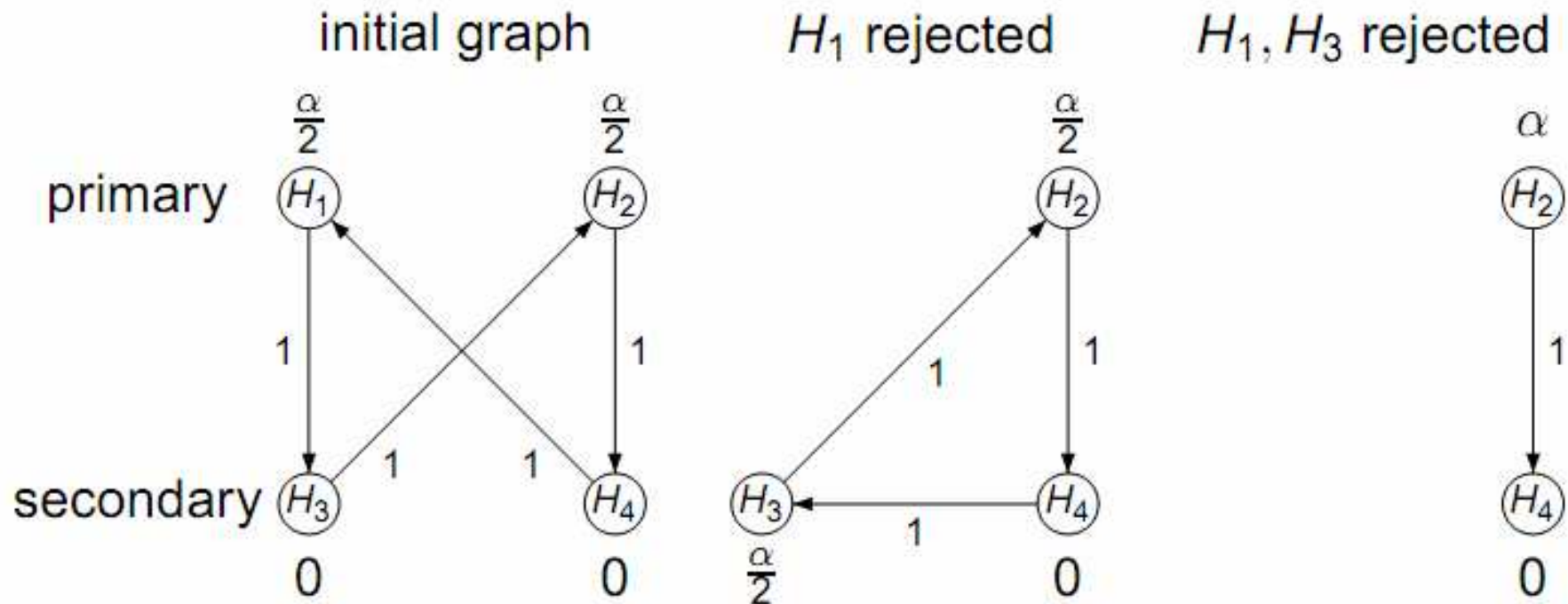
- Separate error control for primary and secondary hypotheses can be sensible
  - Additional consistency property avoids „illogical“ outcomes and reduces overall error rate (primary and secondary hyp. combined)
  - Allows to fully exploit the FWER level for primary hypotheses
  - Generalization to more than 2 treatment arms are possible
  - Buy in by regulators?
  - More on upper type I error bounds for consistent strategies (and on relation with group-sequential tests): Maurer, Glimm and Bretz (2010).

---

# Graphical methods controlling the FWER for the entire family $F_{PS}$

# Successive and consistent closed test procedures of level $\alpha$ defined by transition graphs

- Graphical representation of structural relation and of a Bonferroni based test procedure



# The transition graph test procedure: definition and sequential rejection algorithm

---

## ■ Vertices and initial levels:

- hypotheses  $H_i$  are represented by vertices of the graph with initial levels  $\alpha_i$  summing up to  $\alpha$ .
- Hypothesis  $H_i$  is rejected if associated p-value  $p_i < \alpha_i$  (Bonferroni)

## ■ Update of graph

- The level  $\alpha_i$  of a rejected hypothesis is distributed and added to the levels of the remaining hypotheses according to predefined weights on the directed edges between the vertices (transition weights)
- The transition weights between the remaining vertices representing candidates for further rejection are updated with a specific algorithm\*
- This process is repeated until no further hypothesis can be rejected.

\* Bretz, Maurer, Brannath, Posch (Stat. in Med. 2009)

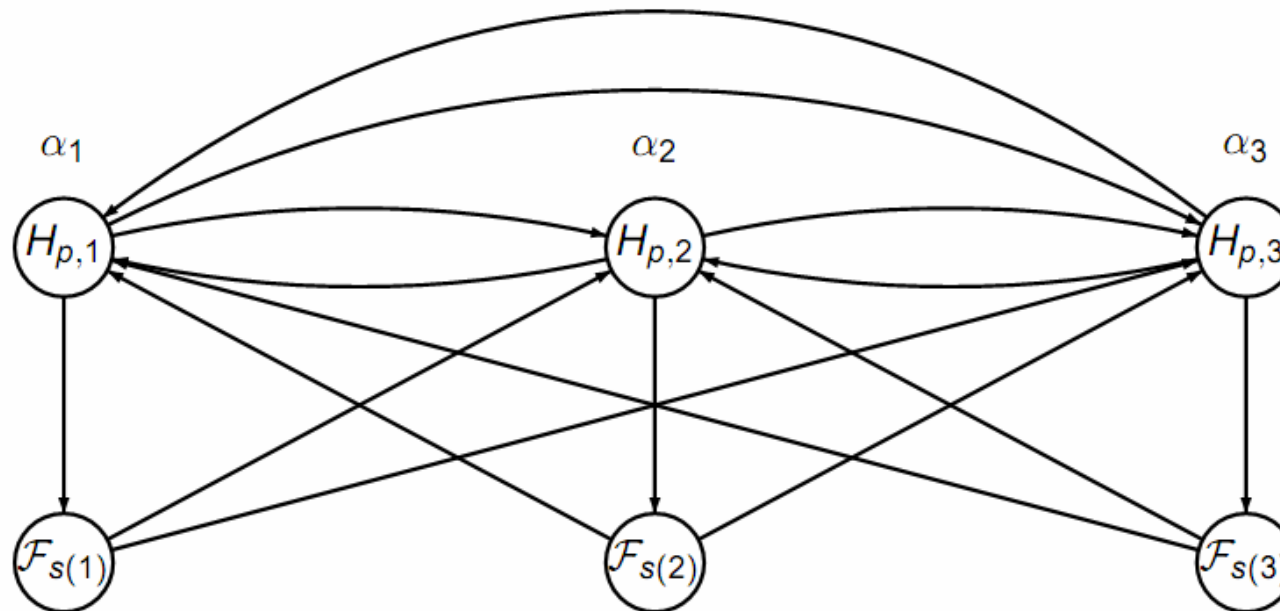
# The transition graph test procedure: Properties

---

- The transition graph test procedure
  - is equivalent to a unique closed test procedure based on Bonferroni level- $\alpha$  tests on each of the intersection hypotheses,
  - is independent of the rejection sequence and
  - is sequentially rejective;
    - the individual levels of not yet rejected hypotheses cannot decrease after rejection of a hypothesis (consonance of the test)
  - controls strongly the familywise type I error rate at level  $\alpha$ 
    - (i.e. is of multiple level  $\alpha$ )
  - covers most „classical“ Bonferroni based sequentially rejective procedures

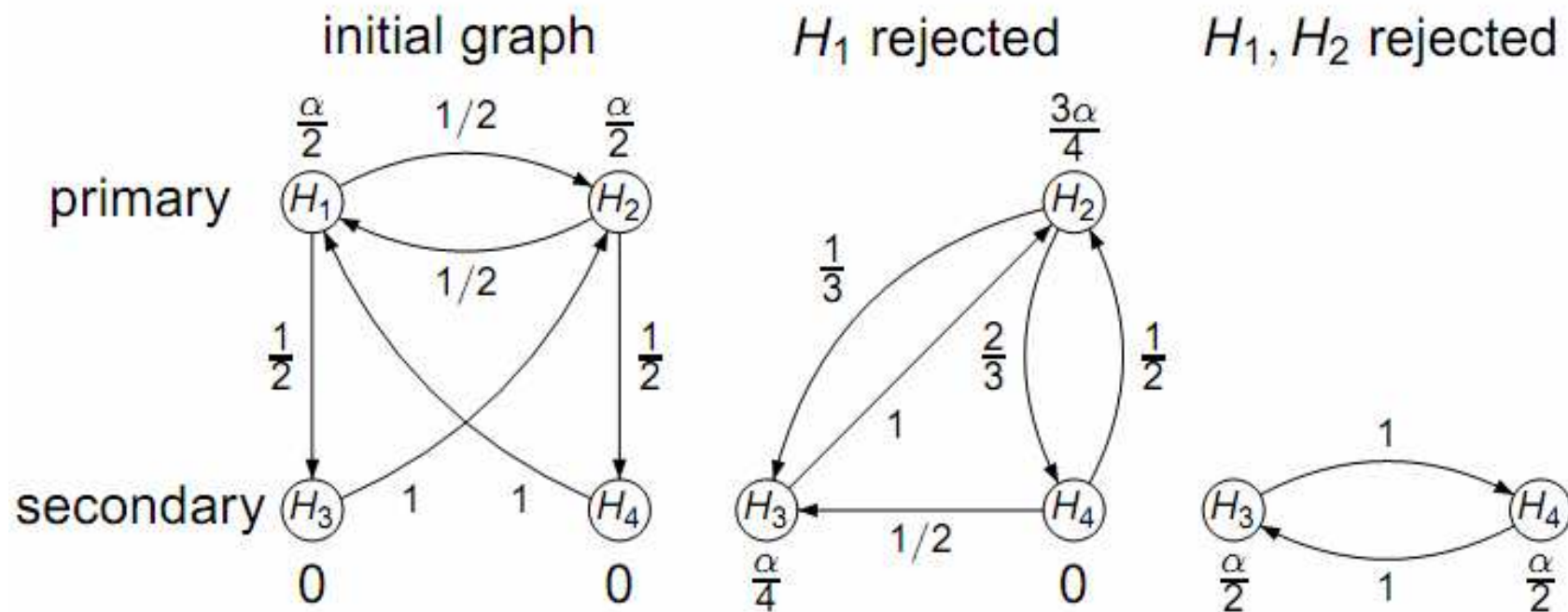
# Successive procedure for multiple primary hypotheses ...and each with a descendant family of secondary hypotheses

- A graph generates a successive procedure if
  - Initial weights are 0 on all secondary hypotheses
  - the only edges with positive weight leading into a secondary hypothesis are those originating at its parent primary hypotheses
  - no edges from a secondary hypothesis to another secondary hypothesis unless parents are the same.



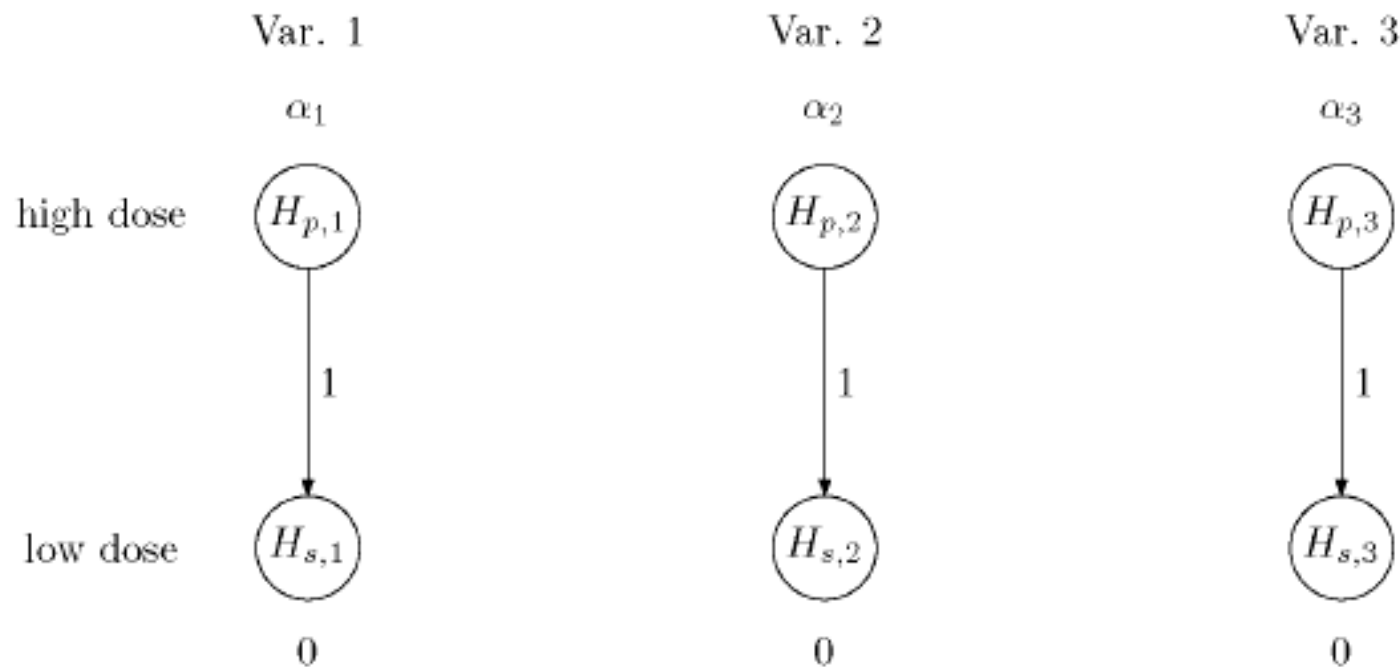
# Two primary hypotheses with one secondary descendant each

- An example where after rejection of one of the primary hypotheses its significance level is passed on equally to the other primary and the descendant secondary hypothesis



# Improving sequentially rejective procedures by weighted Simes Tests

- Graphical representation of a procedure proposed by Quan et al. (2009)
  - In addition to rejections possible by graph, reject **all** primary and secondary hypotheses if **all** are significant at level  $\alpha$



# Weighted Simes test

---

- Quan et al.'s procedure keeps the FWER under positive regression dependence, because it is a conservative simplification of the [Weighted Simes procedure](#) (Kling, 2005, has shown that this keeps the FWER).
- In general, the following multiple test protects the FWER at level  $\alpha$  (under positive regression dependence):
  - **Reject all hypotheses if all are significant at level  $\alpha$ .**
  - **Otherwise, reject those hypotheses that are rejected by a transition graph procedure.**

# Trimmed Simes test for two hypotheses

---

- Trimmed weighted Simes test for two hypotheses
  - $H_A$  and  $H_B$  with multivariate normal or t-distributed test statistics and associated univariate p-values  $p_A$  and  $p_B$
  - Given significance levels  $\alpha_A$  and  $\alpha_B$ ,  $\alpha_A + \alpha_B = \alpha$ ,
    - Reject  $H_A \cap H_B$  if
    - $p_A < \alpha_A$  and  $p_B < 1 - \alpha_B$  or
    - $p_B < \alpha_B$  and  $p_A < 1 - \alpha_A$  or
    - $\max(p_A, p_B) < \alpha$ .
  - The trimmed Simes test protects type I error rate at level  $\alpha$  for any correlation between the test statistics\*

\*Brannath, W., Bretz, F., Maurer, W., and Sarkar, S. (2009), "Trimmed Weighted Simes' Test for Two One-Sided Hypotheses With Arbitrarily Correlated Test Statistics," *Biometrical Journal*, 51, 885–898

# Properties of the trimmed 1-sided Simes test

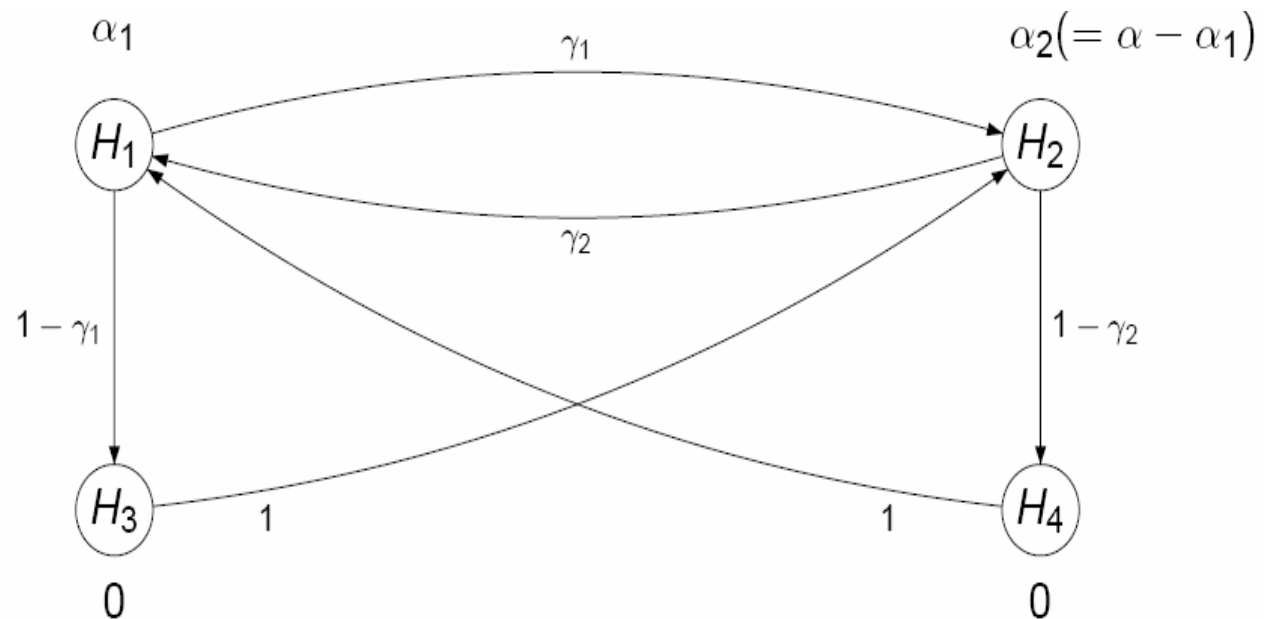
---

- For multivariate normal tests statistics level  $\alpha$  is protected irrespective of the correlation
  - for negatively correlated variables it is very conservative
  - for alternatives with positive effects in both variables and positive correlation, the power gain over Bonferroni is similar to Simes' test
- Test can be used in a closed test
- Strongly contradicting effects in the two test statistics do not result in a rejection of any of the null-hypotheses
  - Trimming is in the same spirit as the consistency requirement by Alosch and Huque\*, though somewhat weaker
    - \*Alosch, M. and Huque M. F. (2010), "A consistency-adjusted alpha adaptive strategy for sequential testing," *Statistics in Medicine*. 28.

# Application of trimmed Simes test to successive multiple tests with 2 primary hypotheses

- This multiple test protects the FWER at level  $\alpha$ 
  - i) **Retain all** four hypotheses if for **any**  $p_i > 1 - \alpha_i$
  - ii) **reject all** four hypotheses if **all**  $p_i < \alpha$
  - iii) **otherwise perform** a closed successive weighted Bonferroni-test, e.g. based on a successive **graphical approach** as below.

Reason: all intersections include never more than 2 hypotheses with positive weight



## Conclusions

---

- Multiple sources of multiplicity often induce partial order(s) of importance on the hypotheses;
- Test procedures consistent with such a partial order can be constructed (i.e. have succession property),
  - but need not control FWER at level  $\alpha$  for the combined family of primary and secondary hypotheses, even if they do so separately.
- Full FWER control can transparently be achieved by means of “successive” graphical procedures
- „Simes- and Dunnett-like“ considerations (=knowledge about correlations) can be combined with the graphical methods (Huque and Alosch, J Stat. Plan.& Inf. , 2008; Bretz et al., to appear)

# Backup slides

---

# Weighted Simes test

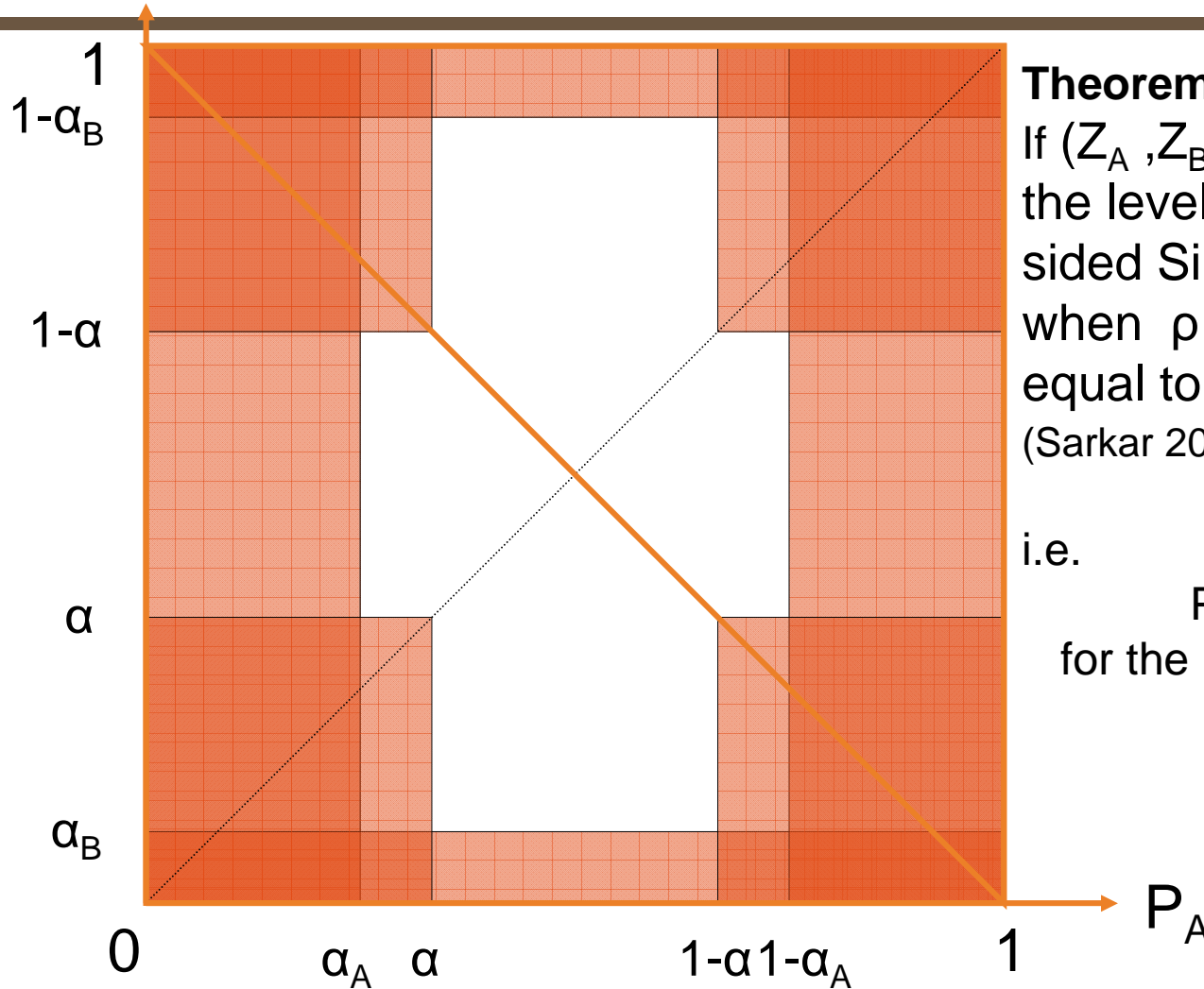
---

- **Weighted Simes:** Given hypotheses  $H_i$ ,  $i \in M = \{1, \dots, m\}$ , the intersection hypothesis  $H_M$  can be rejected if for some  $j \in M$

$$p_{(j)} \leq \sum_{i=1}^j \alpha_{(i)} :$$

- where  $p_{(j)}$  are the ordered p-values and  $\alpha_{(j)}$  are the corresponding local significance levels. **Kling (2005) showed that this global test controls the Type I error rate at level  $\alpha$  if the test statistics are positive regression dependent.**
- Assuming positive regression dependence one can show that the following multiple test protects the FWER at level  $\alpha$ :
- **In addition to the hypotheses rejected by a transition graph procedure of level  $\alpha$ , all hypotheses can be rejected if all are significant at level  $\alpha$ .**

# Unequal $\alpha$ -split, 2-sided Simes



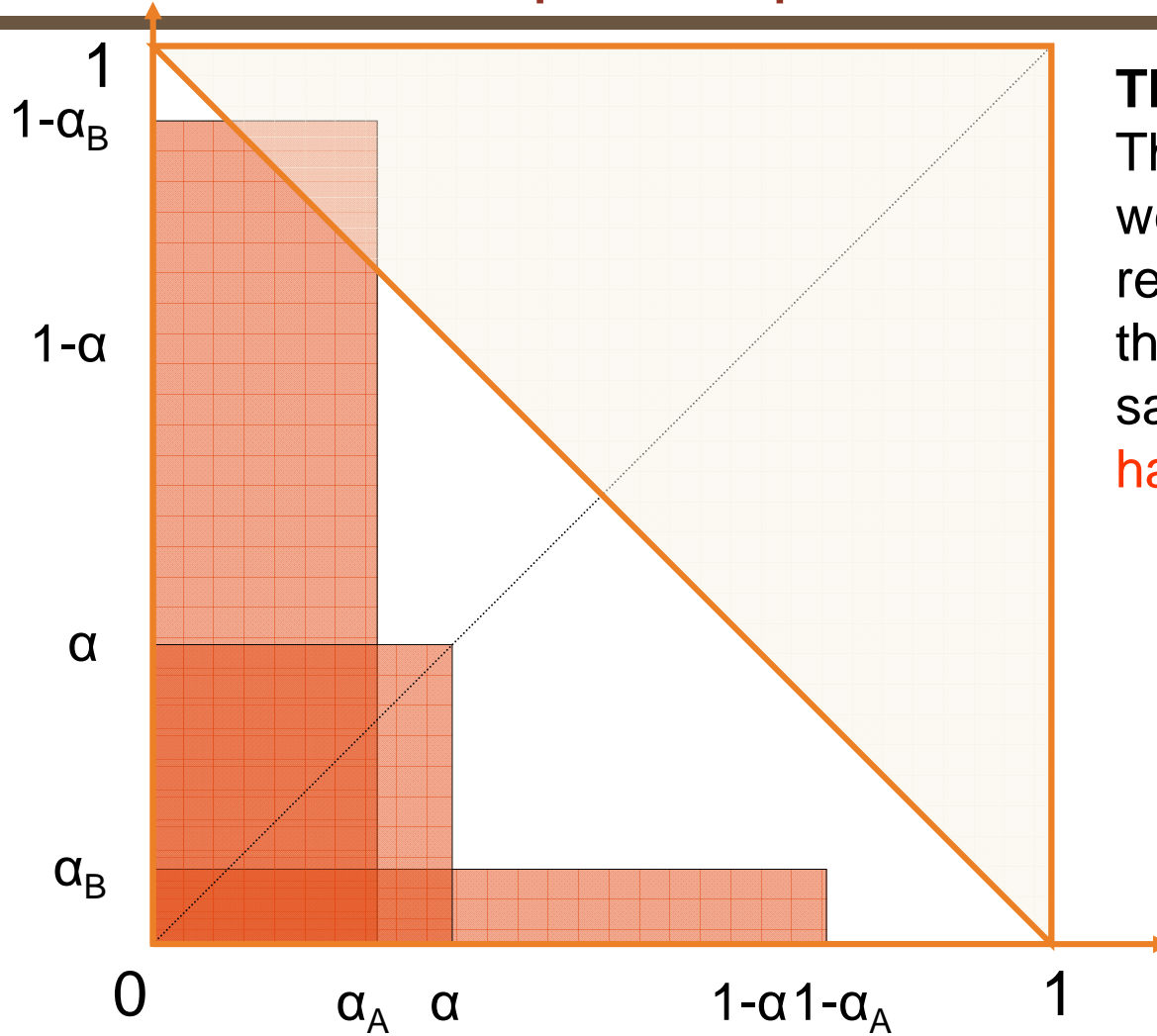
## Theorem 1:

If  $(Z_A, Z_B) \sim \text{BVN}(0, 0, \rho)$   
 the level of the weighted two-  
 sided Simes' test is equal to  $2\alpha$   
 when  $\rho = 0$  and less than or  
 equal to  $2\alpha$  when  $\rho \neq 0$   
 (Sarkar 2007).

i.e.

$P(\{\text{Reject } H_0\}) \leq 2\alpha$   
 for the rejection region to the left  
 for  $-1 \leq \rho \leq 1$ ,  
 with equality for  
 $\rho \in \{-1, 0, 1\}$

# Modified unequal $\alpha$ -split, 1-sided Simes



## Theorem 2

The modified one-sided weighted Simes test, which rejects  $H_{AB}$  if either one of the conditions below is satisfied,

has level  $\leq \alpha$  for any  $\rho$ .

$$P_A \leq \alpha \quad \text{and} \quad P_B \leq \alpha$$

or

$$P_A \leq \alpha_A \quad \text{and} \quad P_B \leq 1 - \alpha_B$$

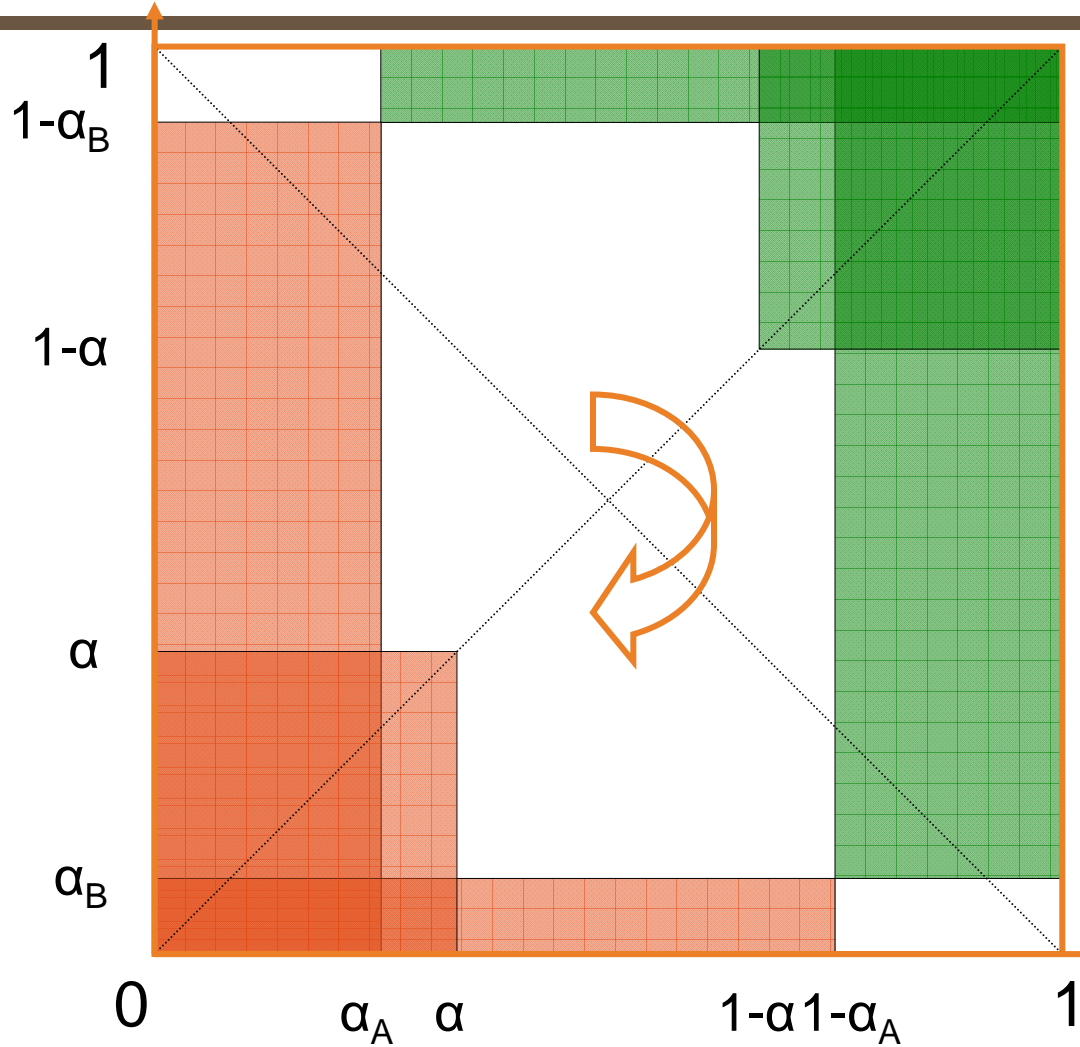
or

$$P_B \leq \alpha_B \quad \text{and} \quad P_A \leq 1 - \alpha_A$$

Why?

# Modified unequal $\alpha$ -split, 1-sided Simes

## An intuitive „proof“ of theorem 2



Turning the **orange modified 1-sided** rejection region  $\{O\}$  around the midpoint  $(0.5, 0.5)$  by  $180^\circ$  produces the **green** region  $\{G\}$ .

$P(\{O\}) = P(\{G\})$   
 when  $(Z_A, Z_B) \sim (Z_B, Z_A)$  and  
 $Z_i \sim -Z_i, i = A, B$ .

Since  $\{O\}$  and  $\{G\}$  are disjoint and  
 $\{O\} \cup \{G\} \subset \{\text{Rej. 2-sided}\}$ ,  
 we have:

$P(\{O\}) + P(\{G\}) \leq 2\alpha$   
 and hence  $P(\{O\}) \leq \alpha$