Adaptive, graph based multiple testing procedures and a uniform improvement of Bonferroni type tests.

Martin Posch
Center for Medical Statistics, Informatics and Intelligent Systems
Medical University of Vienna

Padua, 2015

Acknowledgements

Frank Bretz, Andreas Futschik, Florian Klinglmüller, Franz König, Willi Maurer

This work was supported by the Austrian Science Fund FWF Project 23167-B13

A graphical approach to multiple testing in clinical trials

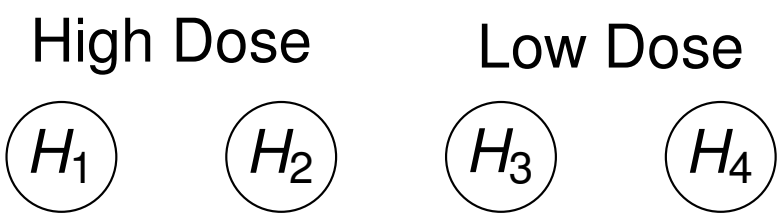
- In clinical trials often multiple hypotheses are tested simultaneously,
  - Primary/secondary endpoints.
  - Comparison of multiple treatment arms
- Multiplicity adjustment is important to control the overall probability of a false positive result, the family wise error rate (FWE).
- Research questions are typically not symmetric but differ in importance and there are certain relationships between them.
- Using directed weighted graphs the structure of the research problem can be mapped onto the multiple testing procedure.

BRETZ, MAURER, BRANNATH & POSCH (2008)

Example: Multiple Sclerosis Trial

Late phase development of a new drug for the indication of multiple sclerosis

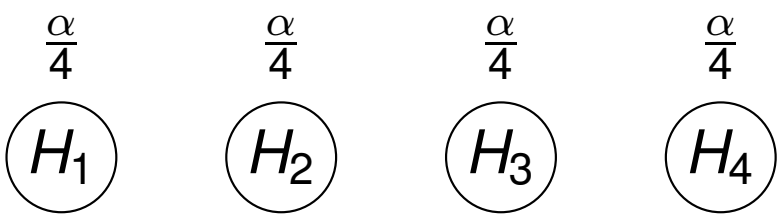
- Two dose levels (High, Low) are compared to placebo.
- Two endpoints:
  1. Annualized relapse rate
  2. Number of lesions in the brain
- Four elementary hypotheses



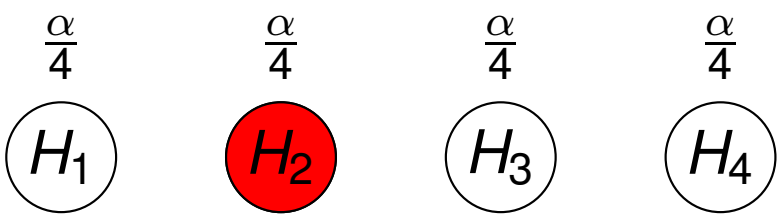
The Bonferroni Adjustment

Distributing the level between the hypotheses

- Test of m hypotheses $H_1, \ldots, H_m$.
- For each hypothesis a test with p-value p_i, $i = 1, \ldots, m$ is defined.
- Bonferroni adjustment: Test each hypothesis at level α/m.
- Example:

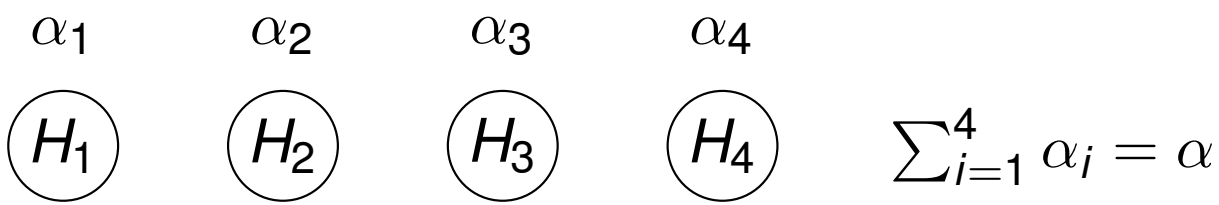


- Start with the Bonferroni test with levels α/m.
- If a hypothesis can be rejected, test the remaining with a Bonferroni test for $m-1$ hypotheses.
- If a hypothesis can be rejected, test the remaining with a Bonferroni test for $m-2$ hypotheses....
- Stop as soon as no further hypothesis can be rejected.
- Example:



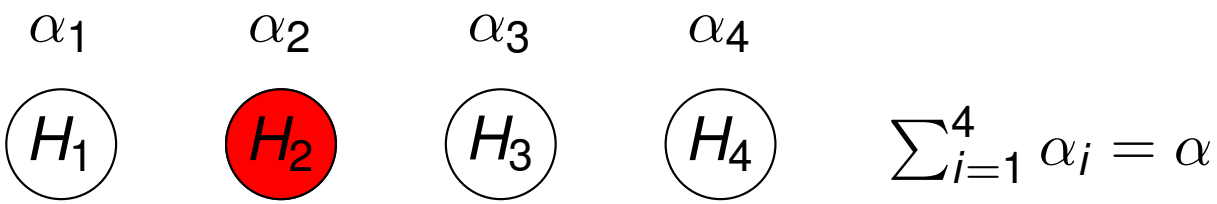
The Weighted Bonferroni Test

- Split the level α across the m hypotheses such that $\sum_{i=1}^{m} \alpha_i = \alpha$.
  E.g., for $m = 4$



- for $\alpha_i = \alpha/m$ the conventional Bonferroni test results.
- The levels α_i are predefined.

Weighted Sequentially Rejective Bonferroni Tests

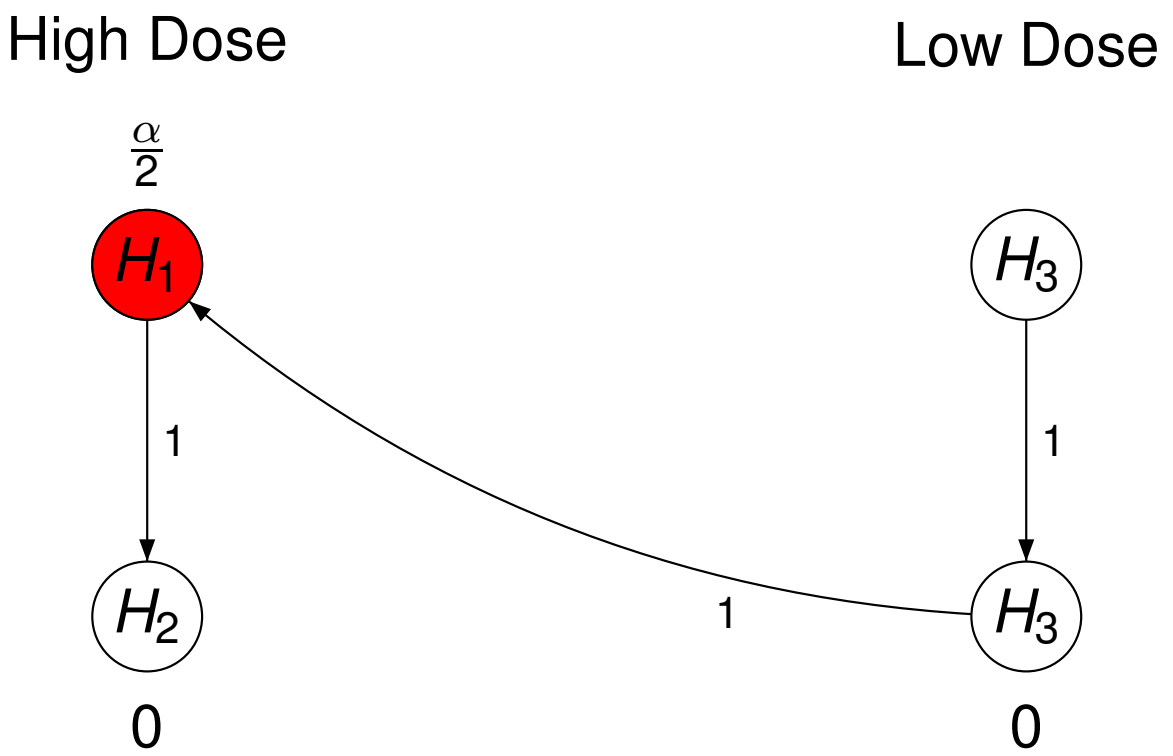


- The procedure controls the FWE if the level for each hypothesis never decreases in the stepwise procedure (e.g., $\alpha_i^{(134)} \geq \alpha_i, i = 1, 3, 4$).

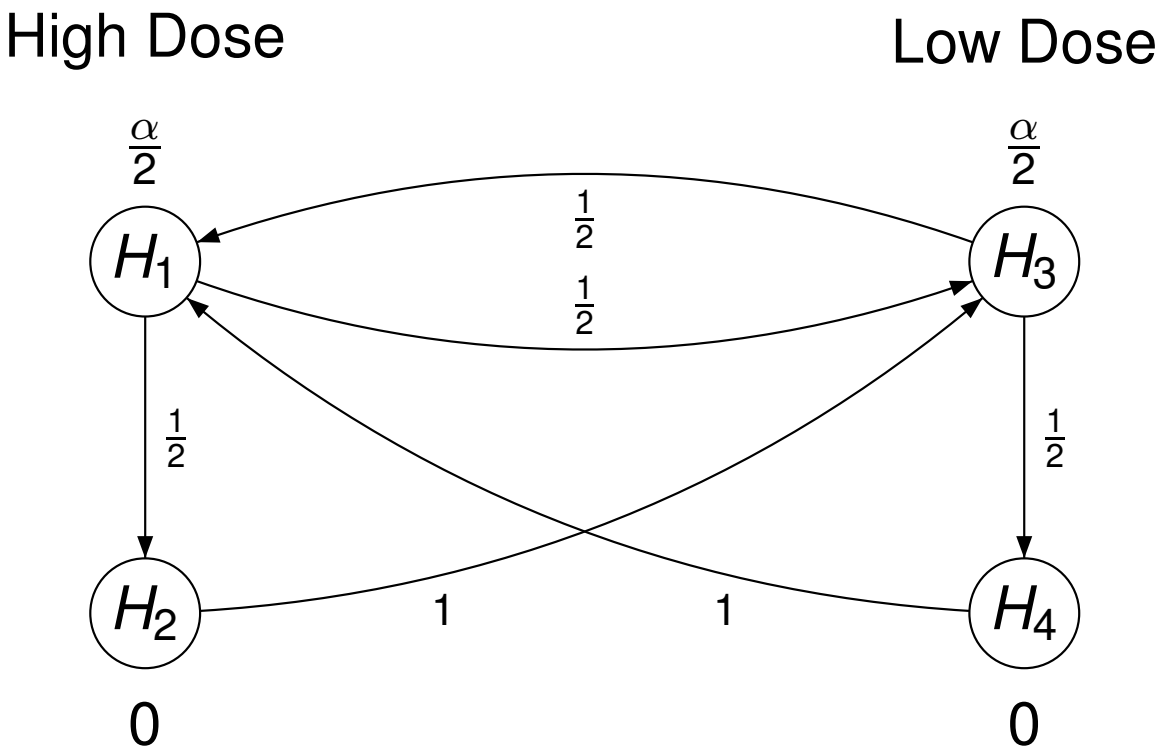
HOMMEL, MAURER, BRETZ (2007), GOEMAN & SOLARI (2010)

- We need to define weighted Bonferroni tests for each of $2^m - 1$ subsets of hypotheses!

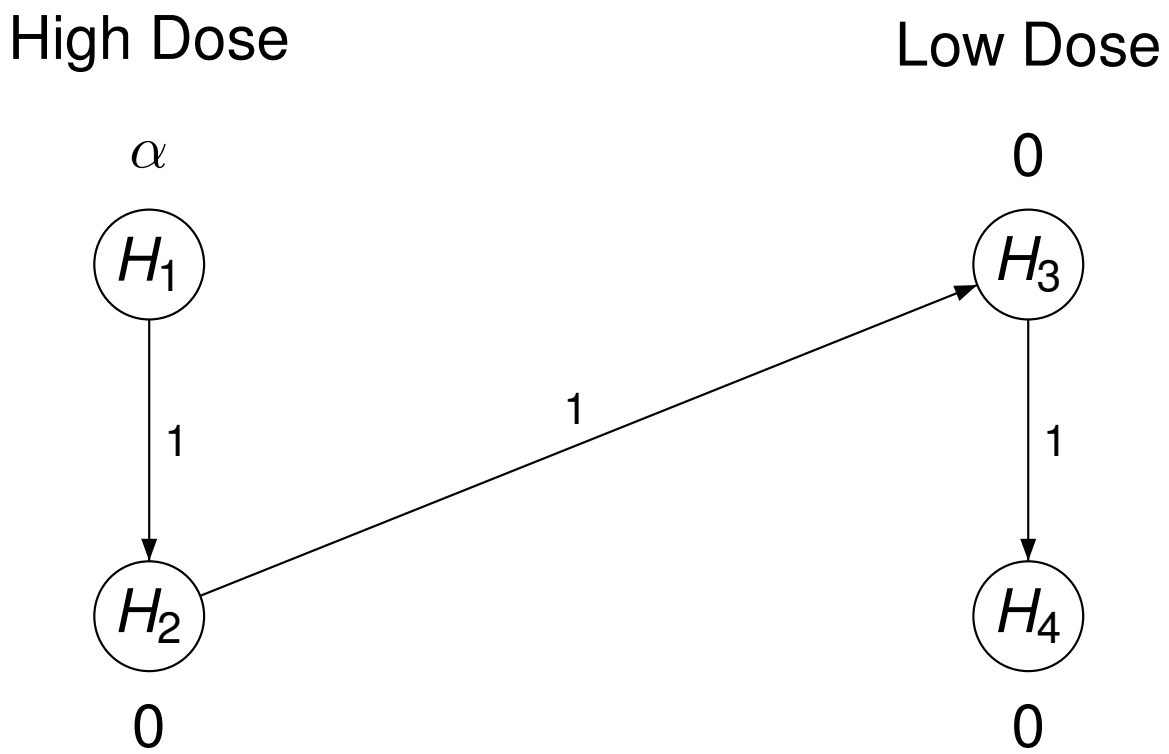
A Graph Based Test



Giving more Weight to Primary Hypotheses



Fixed Sequence Test



Defining a Multiple Test Procedure with Graphs

General definition of the multiple test

- $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, $\sum_{i=1}^{m}, \alpha_i = \alpha$, initial levels
- $\mathbf{G} = (g_{ij}) : m \times m$ transition matrix
  g_{ij} with $0 \leq g_{ij} \leq 1$, $g_{ii} = 0$ and $\sum_{j=1}^{m} g_{ij} \leq 1$ for all $i = 1, \ldots, m$.

- g_{ij}...fraction of the level of H_i that is allocated to H_j.
- $\mathbf{G}$ and α determine the graph and the multiple test.

The Testing Procedure

Set $J = \{1, \ldots, m\}$.

1. Select a j such that $p_j \leq \alpha_j$.
   If no such j exists, stop, otherwise reject H_j.
2. Update the graph:

$$J \to J/\{j\}$$

$$\alpha_\ell \to \begin{cases} \alpha_\ell + \alpha_j g_{j\ell}, & \ell \in J \\ 0, & \text{otherwise} \end{cases}$$

$$g_{\ell k} \to \begin{cases} \frac{g_{\ell k} + g_{\ell j} g_{jk}}{1 - g_{\ell j} g_{j\ell}}, & \ell, k \in J, \ell \neq k \\ 0, & \text{otherwise} \end{cases}$$

3. Go to step 1.

Control of the FWE

Theorem

The initial levels α, the transition matrix **G** and the algorithm define a unique multiple testing procedure controlling strongly the FWER at level α.

BRETZ, MAURER, BRANNATH & POSCH '08

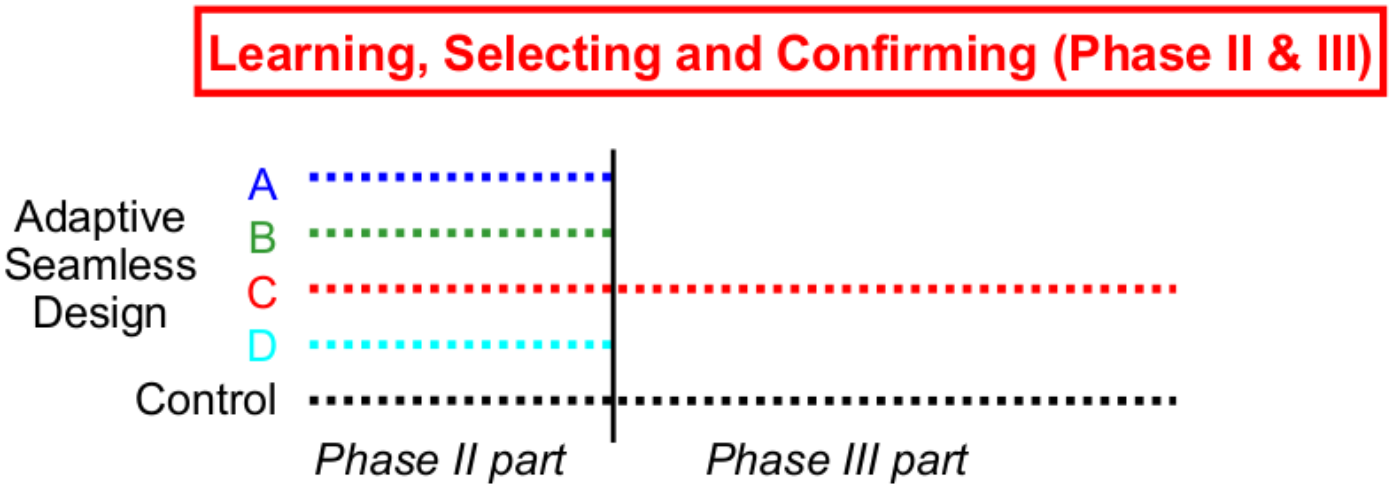
Behind the Scenes:
The graph based procedure is a shortcut for a closed tests

The Closure Principle (Marcus et al. 1976)

- For $J \subseteq \{1, \ldots, m\}$ let $H_J = \bigcap_{i \in J} H_i$.
- For each H_J define a level α test.
- Reject H_i if all H_J for which $i \in J$ can be rejected at level α

- The closed testing procedure controls the FWE at α in the strong sense.
- Requires $2^m - 1$ tests!
- The graph and algorithm implicitly define
  - weighted Bonferroni tests for all intersection hypotheses
  - a shortcut that reduces the number of tests: in each step, a large number of intersection hypotheses are tested implicitly.

Adaptive Designs with Treatment Selection



- In a first stage start with several treatments.
- Based on interim results stop for futility or select one or more treatment arms.
- Inference should be based on data from both stages.
- Increase in efficiency compared to separate designs

THALL ET AL. '88, KIESER & BAUER '99, HOMMEL '01, STALLARD AND TODD '03, KÖNIG ET AL. '06, POSCH ET AL. 10.

A Graphical Approach for Treatment Selection Designs

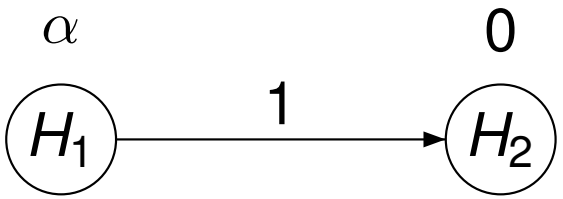
- Start out with the test of m hypotheses and a graph specifying the multiple testing procedure. ▸ Graph
- In an unblinded interim analysis a subset $S \subset \{1, \ldots, m\}$ is selected.
- Only for the selected hypotheses further observations are collected.
- The data of both stages is used in the final test.
- Control of the familywise error rate (FWE) in the strong sense.

A Simple Conservative Adaptive Test

Set p-values of dropped hypotheses equal to 1. Perform the sequentially rejective test specified by the graph.

Example (continued): Assume the high dose is dropped. Set $p_1 = 1, p_2 = 1$ and compute p_3, p_4 as pre-planned. Then apply the test based on the graphical approach.

- This procedure controls the FWE.
- However, e.g., for fixed sequence tests not always a satisfactory solution



- No further adaptations (e.g. sample size reassessment) is possible.

Can we do better?

Adaptive Tests Defined by Graphs
Application of partial conditional error rates

Define an adaptive test for each intersection hypothesis $H_J, J \subseteq \{1, \ldots, m\}$:

- The graph defines levels $\alpha_i^{(J)}$ for each H_J.
- Compute Partial Conditional Error Rates (Posch, Maurer, Bretz '10)

$$A_i = P(p_i \leq \alpha_i^{(J)} | \text{First Stage Data}), i \in J,$$

  where p_i denote p-values of the fixed sample tests.
- Plan a second stage test for H_J at (conditional) level

$$A_J = \sum_{i \in J} A_i$$

  based only on data corresponding to selected hypotheses.
- Apply the closed testing principle: Reject all individual hypotheses H_i if all intersection hypothesis H_J with $i \in J$ can be rejected.

Properties of the Adaptive Test

- Flexibility: No specific selection rule nor the number of hypotheses to be selected needs to be pre-specified.
- Strong control of the FWE.
- There is no simple sequential algorithm as for the fixed sample, graph based test. (Only for the special case of the Bonferroni Holm test a shortcut is available, see Scherag et al. ’09).
- However, the testing procedure is fully specified by the graph.

Simulation Study

- Comparison of two dose groups to a control
- Two normally distributed endpoints
- $H_1 : \mu_{C1} \leq \mu_1 \quad H_3 : \mu_{C3} \leq \mu_3$
  $H_2 : \mu_{C2} \leq \mu_2 \quad H_4 : \mu_{C4} \leq \mu_4$
- z-tests for each of the 4 comparisons
- "Per dose fixed sequence test" ▸ Graph
- Interim analysis after half of the preplanned 100 patients per group

Results: Select the most promising treatment
(According to the interim estimate of the primary endpoint)

Con
Strictly conservative procedure, setting p-values of dropped hypotheses to 1.

CER
Conditional error rate based procedure

Method	Effect sizes				Power				
	θ_1	θ_2	θ_3	θ_4	P_1	P_2	P_3	P_4	P_{any}
Con	0.5	0.4	0.3	0.2	0.78	0.56	0.11	0.02	**0.89**
CER					0.79	0.59	0.12	0.03	**0.90**
Con	0.5	0.4	0.4	0.3	0.65	0.47	0.26	0.11	**0.91**
CER					0.66	0.50	0.26	0.14	**0.92**
Con	0.5	0.5	0.5	0.5	0.47	0.43	0.48	0.43	**0.95**
CER					0.48	0.45	0.48	0.45	**0.96**

Selecting the lower dose at interim

Method	Effect sizes				Power				
	θ_1	θ_2	θ_3	θ_4	P_1	P_2	P_3	P_4	P_{any}
Con	0.5	0.4	0.4	0.3	0	0	0.72	0.32	**0.72**
CER					0	0	0.79	0.42	**0.79**

Example: Fixed Sequence Procedure

Method	Effect sizes				Power				
	θ_1	θ_2	θ_3	θ_4	P_1	P_2	P_3	P_4	P_{any}
Con	0.5	0.4	0.4	0.3	0	0	0	0	**0**
CER					0	0	0.68	0.347	**0.68**

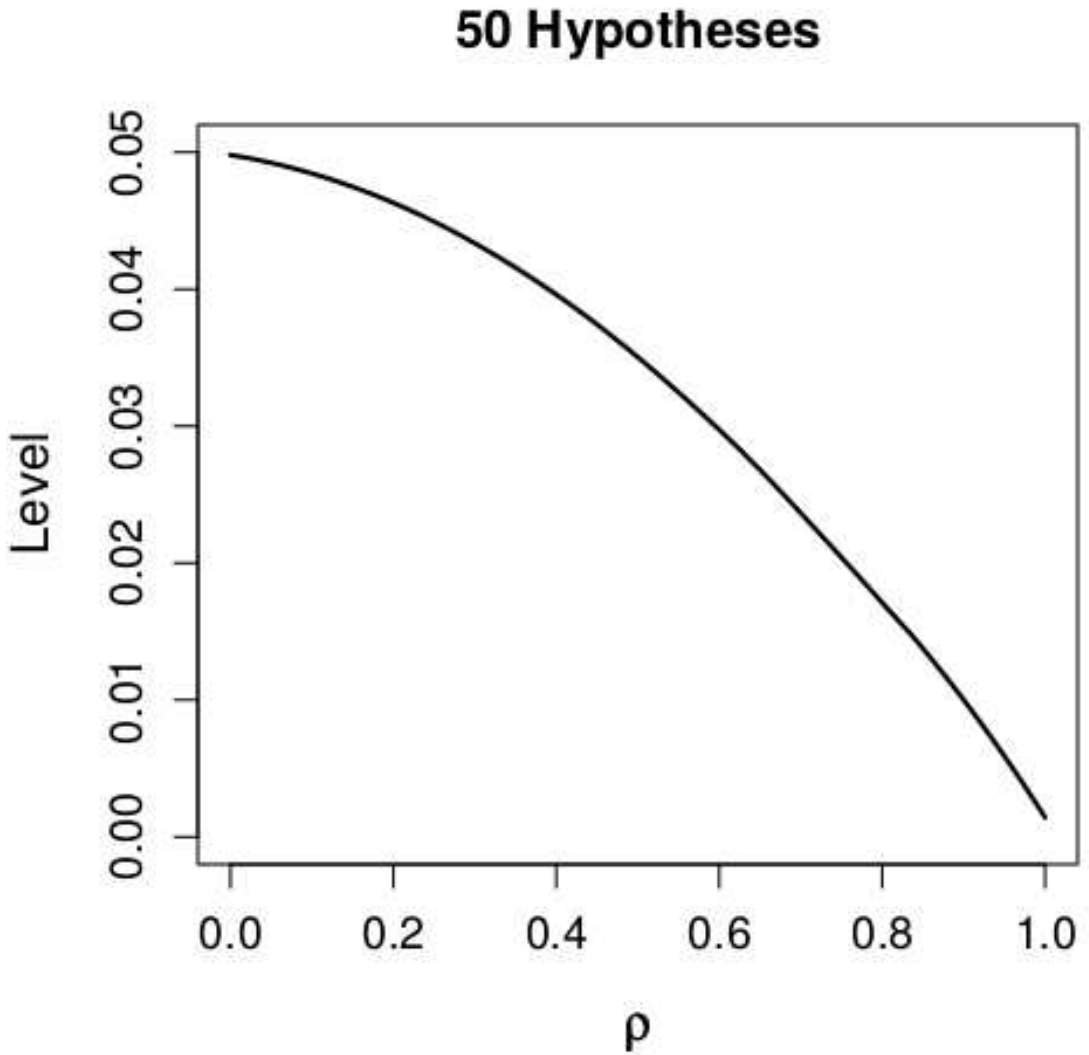
Summary

- Graphs are a simple tool to
  - specify intersection hypotheses tests for closed test procedures based on weighted Bonferroni tests
  - implement short cuts
  - communicate the testing strategy
- Further Research
  - Use weighted tests that make use of stochastic dependency of test statistics.
  - Data dependent weights.
  - Exploiting logical relations.

Multiple Testing on the Beach

Actual Level of the Bonferroni Test

One sample t-tests for equi-correlated multivariate normal data (n=200)



A uniform improvement of Bonferroni-Type tests

. . . by a sequential test

- Consider a fixed sample experiment with sample size n
- The fixed sample (weighted) Bonferroni test rejects H_J if

$$p_i \leq \alpha_i^{(J)}$$

for some $i \in J$.

- After each observation t compute for all $i \in J$ the partial conditional error rate

$$A_{i,t}^{(J)} = P(p_i \leq \alpha_i^{(J)} | \mathbf{X}_t),$$

where $\mathbf{X}_t$ denotes the observations from the first t observational units.

- The improved sequential test additionally rejects H_J, already after t observations if

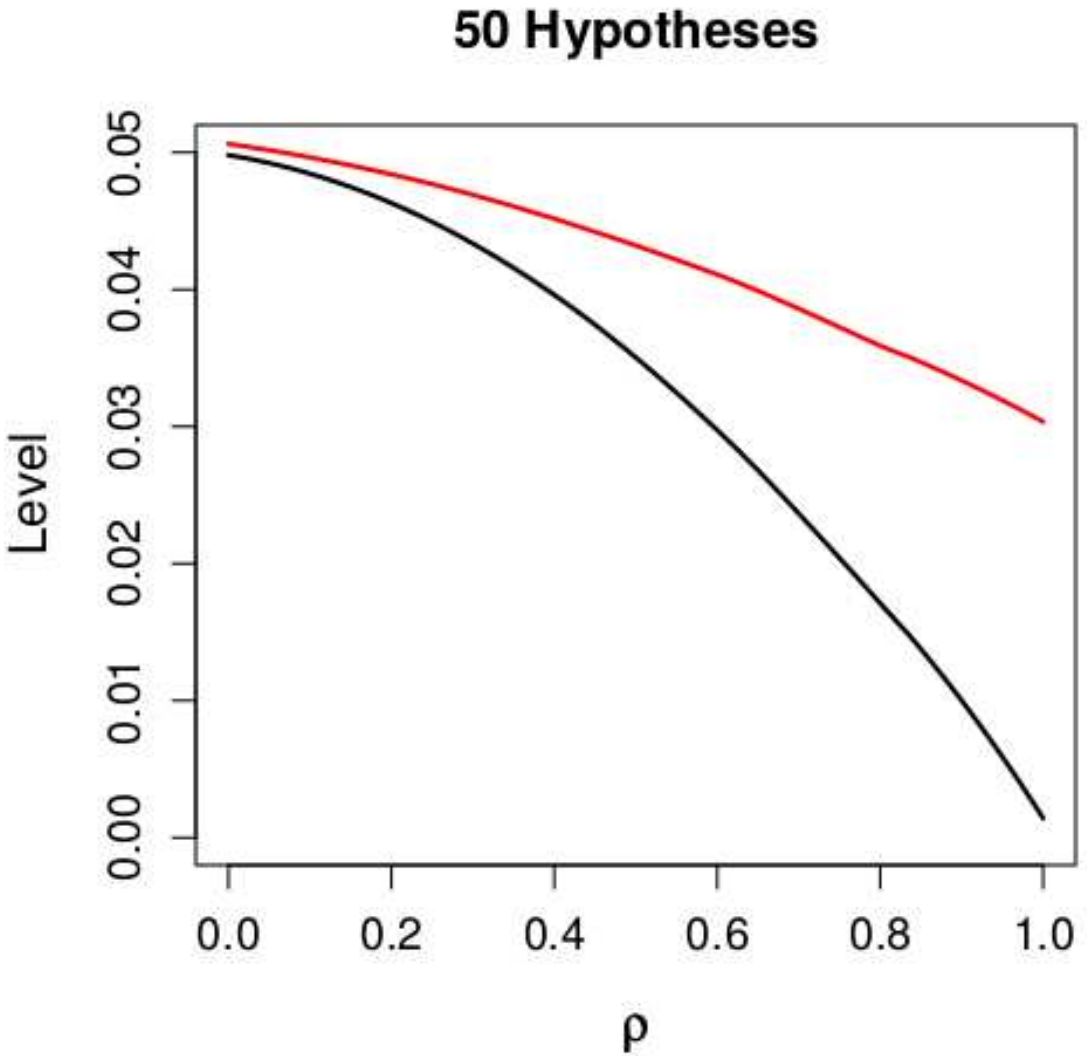
$$\sum_{i \in J} A_{i,t}^{(J)} \geq 1$$

for some $t = 1, \ldots, n$ and has level α for all dependence structures.

M. Posch, A. Futschik, *JASA*, **103**:299-308, 2008.

Level of the Sequential Bonferroni Test

t-tests for equi-correlated multivariate normal data ($n = 200$)

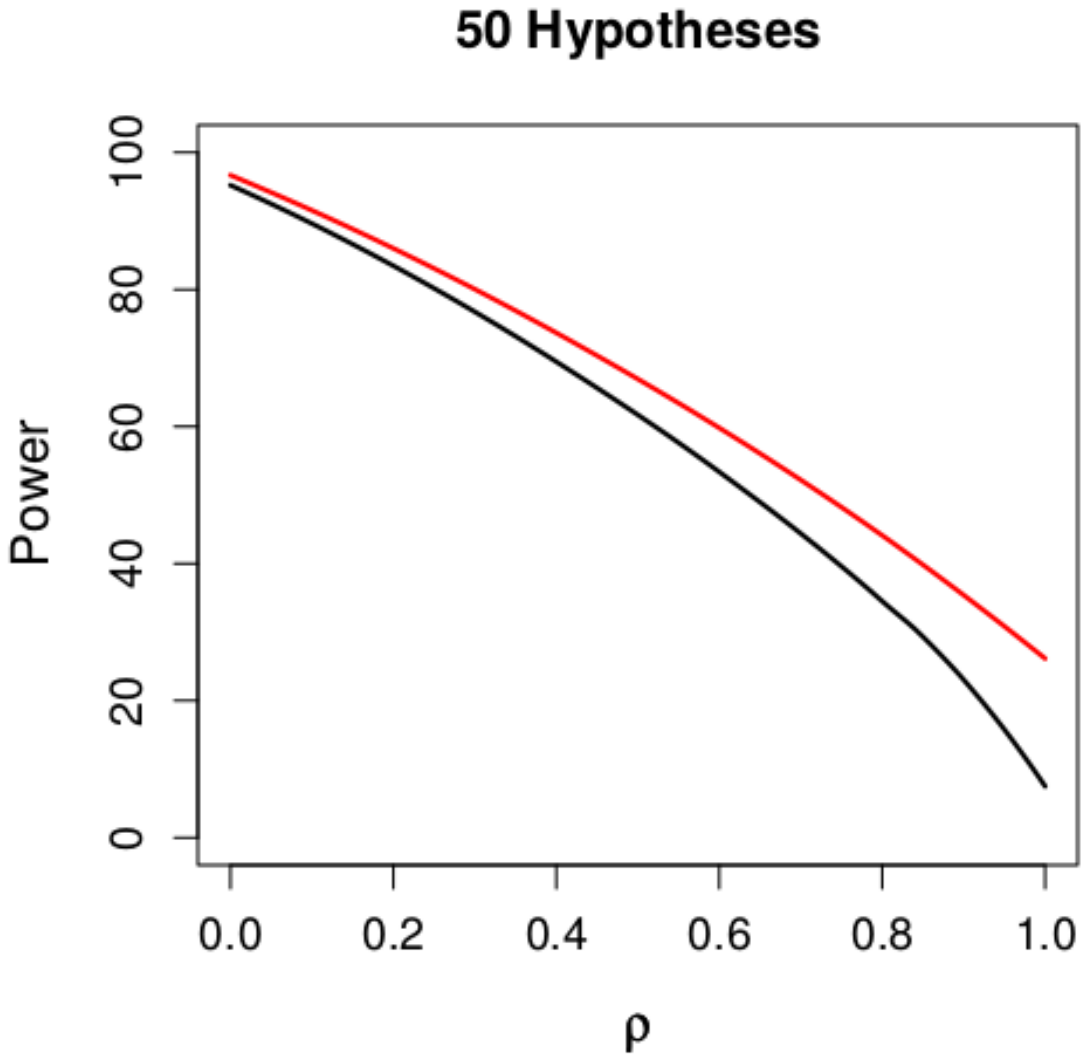


– sequential test

– fixed sample test

Power of the Sequential Bonferroni Test

t-tests for equi-correlated multivariate normal data ($n = 200$, $\theta = \sigma 2/\sqrt{200}$)

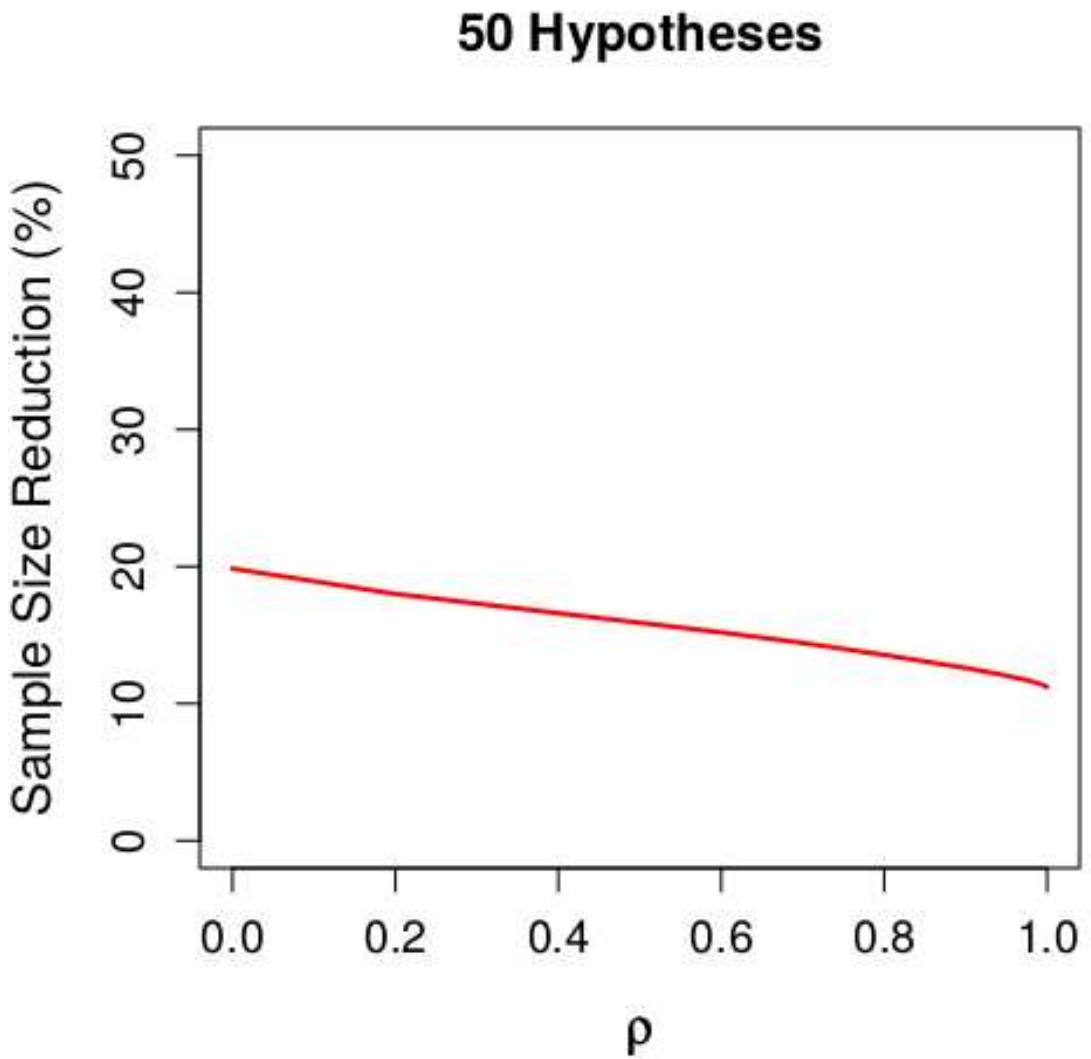


– sequential test

– fixed sample test

Sample Size Reduction of the Seq. Bonferroni Test

t-tests for equi-correlated multivariate normal data ($n = 200$, $\theta = \sigma 2/\sqrt{200}$)



– sequential test

– fixed sample test

- For $n \to \infty$ the sequential test has asymptotic level α
- For finite n the level decreases with increasing positive correlation
- For finite n the level decreases also with K
- What if $n \to \infty$ and $K \to \infty$?

Proposition

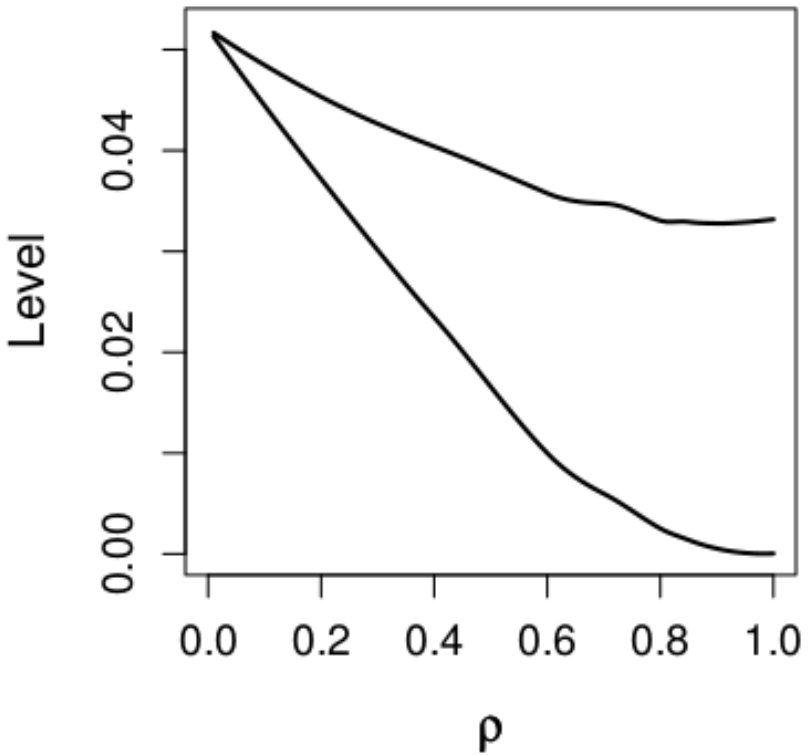
Assume that the $p_{k,n}$ are p-values for one-sided z-tests and are perfectly positively correlated ($\rho = 1$). Then the sequential Bonferroni test has asymptotic level α if

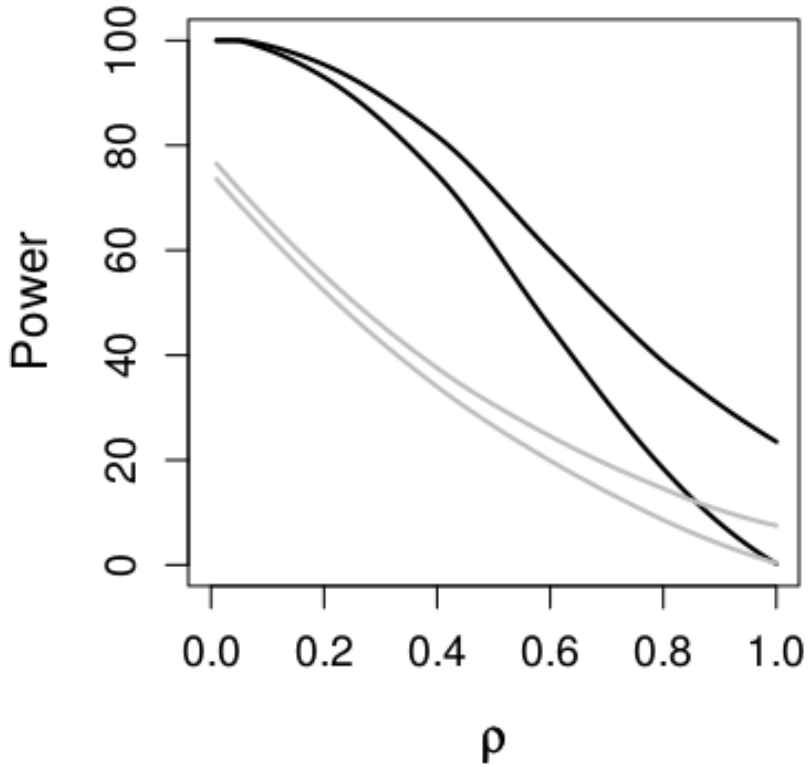
$$\frac{K}{n^{1-\beta}} \to 0$$

for some $0 < \beta < 1$.

Example: K=10000, n=25

Lower lines: Fixed sample Bonferroni test, upper lines: sequential test





Back lines: for all hypotheses the alternative $\theta_l = \sigma\, 2/\sqrt{25}$ holds

Gray lines: for 5% of the hypoheses the alternative holds.

Summary & Extensions

The sequential test

- uniformly improves the fixed sample Bonferroni test
- needs no additional assumptions on the joint distribution of test statistics
- asymptotically exhaust the level α
- is very flexible: different test statistics can be used for the elementary hypotheses

Extensions

- Tests for elementary hypotheses with the closed testing procedure:
- Weighted Bonferroni tests & graph based procedures

F. Bretz, W. Maurer, W. Brannath, and M. Posch.

A graphical approach to sequentially rejective multiple test procedures.

Statistics in Medicine, 28:586–604, 2008.

F. Bretz, W. Maurer, and G. Hommel.

Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures.

Statistics in Medicine, 2011.

F. Bretz, M. Posch, E. Glimm, F. Klinglmueller, W. Maurer, and K. Rohmeyer.

Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests.

Biometrical Journal, 53(6):894–913, 2011.

F. Klinglmueller, M. Posch, and F. Koenig.

Adaptive graph-based multiple testing procedures.

Pharmaceutical Statistics, 13:345–356, 2014.


M. Posch and F. Futschik.

A uniform improvement of Bonferroni type tests by sequential tests.

JASA, 103:299–308, 2008.


M. Posch, F. König, M. Branson, W. Brannath, C. Dunger-Baldauf, and P. Bauer.

Testing and estimation in flexible group sequential designs with adaptive treatment selection.

Statistics in Medicine, 24:3697–3714, 2005.


M. Posch, W. Maurer, and F. Bretz.

Type I error rate control in adaptive designs for confirmatory clinical trials with treatment selection at interim.

Pharmaceutical Statistics, 10:96–104, 2010.