

## Lecture 2: Advanced multiplicity correction

*Adaptive Designs and Multiple Testing Procedures for Clinical Trials*

October 2019

# Outline

- 1 Hierarchical testing
- 2 Semi-parametric procedures
- 3 Closed testing
- 4 Graphical approaches
- 5 Multiplicity in practice

## Hierarchical testing

# Hierarchical test procedures

- Suppose hypotheses can be ordered into a *pre-specified* hierarchy  $H_1, \dots, H_K$ , before the data are observed
  - ▶ Clinical relevance
  - ▶ Dose concentration
  - ▶ Time sequence
- Hierarchical test procedures: tests the hypotheses in the pre-defined hierarchical order
  - ▶ Fixed sequence procedure
  - ▶ Fallback procedure

# Fixed sequence procedure

- Each hypothesis is tested in the pre-specified sequence at level  $\alpha$  until the first non-rejection

# Fixed sequence procedure

- Each hypothesis is tested in the pre-specified sequence at level  $\alpha$  until the first non-rejection
- *Rejection rule*
  - ▶ if  $p_1 \leq \alpha$ , reject  $H_1$  and continue; else stop
  - ▶ if  $p_2 \leq \alpha$  reject  $H_2$  and continue; else stop
  - ▶ ...
  - ▶ if  $p_k \leq \alpha$  reject  $H_k$  and continue; else stop

# Fixed sequence procedure

- Each hypothesis is tested in the pre-specified sequence at level  $\alpha$  until the first non-rejection
- *Rejection rule*
  - ▶ if  $p_1 \leq \alpha$ , reject  $H_1$  and continue; else stop
  - ▶ if  $p_2 \leq \alpha$  reject  $H_2$  and continue; else stop
  - ▶ ...
  - ▶ if  $p_k \leq \alpha$  reject  $H_k$  and continue; else stop
- As soon as a hypothesis  $H_i$  cannot be rejected,  $p_i > \alpha$ , the procedure stops and all remaining hypotheses  $H_{i+1}, \dots, H_K$  are not rejected

# Fixed sequence procedure

## Advantages

- Simple procedure
- Optimal (maximises power) if previous hypotheses rejected

## Disadvantages

- Ordering of testing sequence is critically important, and may be based on subjective information
- Minimises power if a previous hypothesis is not rejected
- Once a hypothesis is not rejected, no further testing is allowed



# Fallback procedure

- Again test each hypothesis in the pre-specified sequence, but split the  $\alpha$  between hypotheses
- Assign  $\alpha_i$  to hypothesis  $H_i$ , where  $\sum_i^k \alpha_i = \alpha$
- $H_1$  is tested at level  $\alpha'_1 = \alpha_1$
- For  $i \geq 2$ ,  $H_i$  is tested at level  $\alpha'_i$ , where

$$\alpha'_i = \begin{cases} \alpha_i & \text{if } H_{i-1} \text{ is not rejected} \\ \alpha_i + \alpha'_{i-1} & \text{if } H_{i-1} \text{ is rejected} \end{cases}$$

- Test all hypotheses even if initial hypotheses are not rejected
- Fallback procedure is implemented in the `multxpert` R package

# Example

- $\alpha = 0.025$
- p-values  $p_1 = 0.03$ ,  $p_2 = 0.004$ ,  $p_3 = 0.01$
- For fallback procedure, suppose  $\alpha$  split equally:  
 $\alpha_1 = \alpha_2 = \alpha_3 = 0.025/3$

p-value	Fixed sequence	Fallback procedure
0.03	$\alpha = 0.025$	$\alpha'_1 = 0.025/3$

# Example

- $\alpha = 0.025$
- p-values  $p_1 = 0.03$ ,  $p_2 = 0.004$ ,  $p_3 = 0.01$
- For fallback procedure, suppose  $\alpha$  split equally:  
 $\alpha_1 = \alpha_2 = \alpha_3 = 0.025/3$

p-value	Fixed sequence	Fallback procedure
0.03	$\alpha = 0.025$	$\alpha'_1 = 0.025/3$
0.004	Not tested	$\alpha'_2 = 0.025/3$

# Example

- $\alpha = 0.025$
- p-values  $p_1 = 0.03$ ,  $p_2 = 0.004$ ,  $p_3 = 0.01$
- For fallback procedure, suppose  $\alpha$  split equally:  
 $\alpha_1 = \alpha_2 = \alpha_3 = 0.025/3$

p-value	Fixed sequence	Fallback procedure
0.03	$\alpha = 0.025$	$\alpha'_1 = 0.025/3$
0.004	Not tested	$\alpha'_2 = 0.025/3$
0.01	Not tested	$\alpha'_3 = 0.05/3$

## Semi-parametric procedures

# Šidák procedure

- Hypotheses  $H_1, \dots, H_K$
- Aim to control the FWER at level  $\alpha$

# Šidák procedure

- Hypotheses  $H_1, \dots, H_K$
- Aim to control the FWER at level  $\alpha$
- Šidák (also known as Dunn–Šidák) procedure uses the adjusted significance level

$$\alpha_S = 1 - (1 - \alpha)^{1/K}$$

- More powerful than Bonferroni for  $K > 1$
- However, only guaranteed to control the FWER for **independent** or **positively correlated** test statistics
  - ▶ e.g. appropriate for **many-to-one comparisons**, where a positive correlation is induced by the control group

# Hochberg

- **Ordered**  $p$ -values  $p_{(1)} < \dots < p_{(K)}$  with corresponding hypotheses  $H_{(1)}, \dots, H_{(K)}$
- Uses same significance thresholds as the Holm procedure, but reversed (step-up rather than step-down)
- *Rejection rule*
  - ▶ If  $p_{(K)} \leq \alpha$ , reject  $H_{(1)}, \dots, H_{(K)}$  and stop; else continue
  - ▶ If  $p_{(K-1)} \leq \alpha/2$ , reject  $H_{(1)}, \dots, H_{(K-1)}$  and stop; else continue
  - ▶ ...
  - ▶ If  $p_{(1)} \leq \alpha/K$ , reject  $H_{(1)}$  and stop
- Find largest  $i$  such that  $p_{(i)} \leq \alpha/(K - i + 1)$  and reject all hypotheses  $H_{(i)}, H_{(i+1)}, \dots, H_{(1)}$



# Hochberg

- More powerful than Bonferroni and Holm
- Again, only guaranteed to control FWER under certain **correlation assumptions**
  - ▶ e.g. When test statistics are independent or positively correlated

# Hommel

- *Rejection rule*

- ▶ Let  $j$  be the largest integer for which

$$p_{(K-j+i)} > \frac{i\alpha}{j}$$

for all  $i = 1, \dots, j$

- ▶ If no such  $j$  exists, reject all hypotheses
- ▶ Otherwise, reject all  $H_{(i)}$  with  $p_{(i)} \leq \alpha/j$
- More powerful than Hochberg, but needs same distributional assumptions
- Hommel and Hochberg procedures are implemented in the `multxpt` R package

## Closed testing

# Closed test procedures

- General methodology to construct multiple testing procedures which strongly control the FWER
- Includes many well-known procedures as special cases

# Closed test procedures

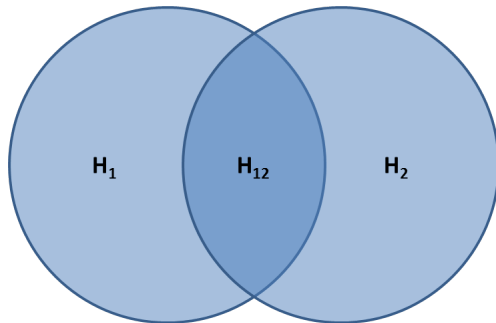
- General methodology to construct multiple testing procedures which strongly control the FWER
- Includes many well-known procedures as special cases
- Closed test procedures consider all *intersection hypotheses*

$$H_J = \bigcap_{i \in J} H_i, \quad J \subseteq \{1, \dots, K\}$$

- **Closure principle**: An individual hypothesis  $H_i$  is rejected at familywise level  $\alpha$  only if every intersection hypothesis  $H_J$  with  $i \in J$  is rejected at local level  $\alpha$

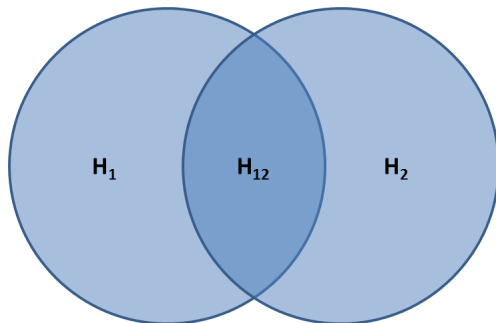
# Closure principle

Venn diagram for  $K = 2$  hypotheses



# Closure principle

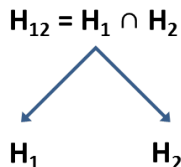
Venn diagram for  $K = 2$  hypotheses



- Test  $H_{12} = H_1 \cap H_2$  using Bonferroni or Dunnett etc. at level  $\alpha$
- Test  $H_1$  and  $H_2$  using a level  $\alpha$  test

# Closure principle

$K = 2$  hypotheses



- Reject  $H_1$  overall if  $H_{12}$  and  $H_1$  are rejected locally at level  $\alpha$
- If  $K > 2$ , several intersection hypotheses have to be tested
- Different tests can be chosen for each (intersection) hypothesis

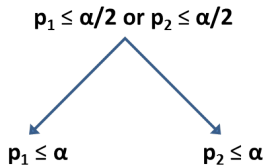


# Closure principle

Holm

*Holm's procedure* is the closure principle applied to Bonferroni:

$H_{12}$  is rejected if either  $p_1 \leq \alpha/2$  **or**  $p_2 \leq \alpha/2$

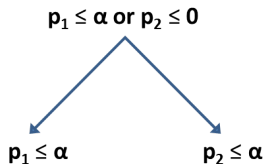


# Closure principle

## Fixed sequence procedure

Bonferroni could also be applied with unequal splitting of the significance level into  $\alpha_1$  and  $\alpha_2$ , where  $\alpha_1 + \alpha_2 = \alpha$

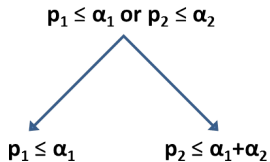
Setting  $\alpha_1 = \alpha$  and  $\alpha_2 = 0$  gives the *fixed sequence procedure*:



# Closure principle

## Fallback procedure

In general, we recover the *fallback procedure*:



# Closure principle

$k = 3$  hypotheses

$$H_1 \cap H_2 \cap H_3$$

$$H_1 \cap H_2$$

$$H_1 \cap H_3$$

$$H_2 \cap H_3$$

$$H_1$$

$$H_2$$

$$H_3$$

# Closure principle

**Fixed sequence procedure** with a-priori fixed order

$$H_1 \rightarrow H_2 \rightarrow H_3$$

$$H_1 \cap H_2 \cap H_3$$

$$p_1 \leq \alpha$$

$$H_1 \cap H_2$$

$$p_1 \leq \alpha$$

$$H_1 \cap H_3$$

$$p_1 \leq \alpha$$

$$H_2 \cap H_3$$

$$p_2 \leq \alpha$$

$$H_1$$

$$p_1 \leq \alpha$$

$$H_2$$

$$p_2 \leq \alpha$$

$$H_3$$

$$p_3 \leq \alpha$$

# Closure principle

## Holm

$$\begin{array}{c} H_1 \cap H_2 \cap H_3 \\ p_1 \leq \alpha/3 \text{ or } p_2 \leq \alpha/3 \text{ or } p_3 \leq \alpha/3 \end{array}$$

$$\begin{array}{c} H_1 \cap H_2 \\ p_1 \leq \alpha/2 \text{ or } p_2 \leq \alpha/2 \end{array}$$

$$\begin{array}{c} H_1 \cap H_3 \\ p_1 \leq \alpha/2 \text{ or } p_3 \leq \alpha/2 \end{array}$$

$$\begin{array}{c} H_2 \cap H_3 \\ p_2 \leq \alpha/2 \text{ or } p_3 \leq \alpha/2 \end{array}$$

$$\begin{array}{c} H_1 \\ p_1 \leq \alpha \end{array}$$

$$\begin{array}{c} H_2 \\ p_2 \leq \alpha \end{array}$$

$$\begin{array}{c} H_3 \\ p_3 \leq \alpha \end{array}$$

# Closed test procedures

## *Advantages*

- Includes many well-known procedures as special cases
- Closed test procedures are more powerful than the procedures they are derived from

# Closed test procedures

## Advantages

- Includes many well-known procedures as special cases
- Closed test procedures are more powerful than the procedures they are derived from
- By construction, it is *coherent*: if null hypothesis  $H_I$  is rejected, all subsets  $H_J \subseteq H_I$  are rejected as well



# Closed test procedures

## Advantages

- Includes many well-known procedures as special cases
- Closed test procedures are more powerful than the procedures they are derived from
- By construction, it is *coherent*: if null hypothesis  $H_I$  is rejected, all subsets  $H_J \subseteq H_I$  are rejected as well
- Any non-coherent multiple testing procedure can be replaced by a coherent one that is at least as powerful
- Any coherent multiple test controlling FWER is a closed test

# Closed test procedures

## Advantages

- Includes many well-known procedures as special cases
- Closed test procedures are more powerful than the procedures they are derived from
- By construction, it is *coherent*: if null hypothesis  $H_I$  is rejected, all subsets  $H_J \subseteq H_I$  are rejected as well
- Any non-coherent multiple testing procedure can be replaced by a coherent one that is at least as powerful
- Any coherent multiple test controlling FWER is a closed test

## Disadvantages

- No natural point estimates or confidence intervals
- Can be a very large number of intersection hypotheses to test as  $k$  increases: worst-case is  $2^K - 1$ . However, shortcuts exist

## Graphical approaches

# Motivating example: Diabetes trial

- Trial compares two doses  $D_1$  or  $D_2$  against placebo in diabetes patients for two endpoints
  - ▶ Primary endpoint: HbA1c
  - ▶ Secondary endpoint: Body weight
- There is a natural order: a primary endpoint is more important than a secondary endpoint
  - ▶ We test the primary null hypothesis first; only if this is rejected do we test the secondary hypothesis

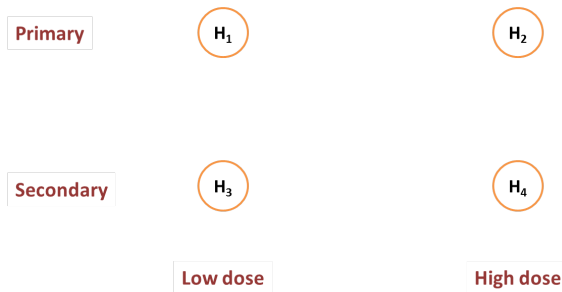
# Motivating example: Diabetes trial

- Trial compares two doses  $D_1$  or  $D_2$  against placebo in diabetes patients for two endpoints
  - ▶ Primary endpoint: HbA1c
  - ▶ Secondary endpoint: Body weight
- There is a natural order: a primary endpoint is more important than a secondary endpoint
  - ▶ We test the primary null hypothesis first; only if this is rejected do we test the secondary hypothesis
- Both doses are equally important

# Motivating example: Diabetes trial

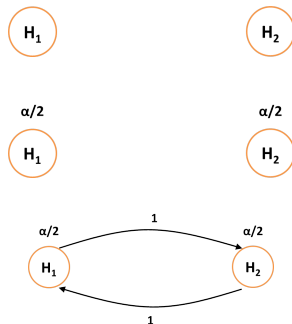
- *Objective*: test all four hypotheses under strong FWER while reflecting clinical objectives
- Standard multiple testing procedures do not reflect the relative importance of the two endpoints or the underlying structure
- In general, need test procedures that can deal with complex trial objectives and multiple structured hypotheses: **Graphical approaches** (also known as chain or sequentially rejective procedures)

# Diabetes example



# Graphical approaches

- 1 Hypotheses  $H_1, \dots, H_K$  represented as nodes
- 2 Split of significance level  $\alpha$  into  $(\alpha_1, \dots, \alpha_K)$
- 3 “ $\alpha$  propagation” through weighted, directed edges





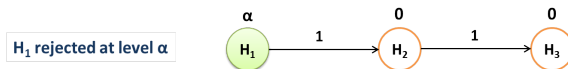
# Graphical approaches

- Technical basis of the graphical approach: the graph defines a closed testing procedure with weighted Bonferroni tests for each intersection hypothesis
- Equivalent formulation of split of  $\alpha$  using **weights**  $(w_1, \dots, w_K)$  where  $\sum_{i=1}^K w_i = 1$  and  $\alpha_i = \alpha w_i$ .
- $\alpha$ -propagation: If a hypothesis  $H_i$  can be rejected at level  $\alpha_i$  (i.e.  $p_i \leq \alpha_i$ ), propagate its level  $\alpha_i$  to the remaining (not yet tested) hypotheses, according to a prefixed rule, and continue testing with the updated  $\alpha$  levels.

# Examples

## Fixed sequence procedure

- Assume  $H_1 \rightarrow H_2 \rightarrow H_3$
- Example fixed sequence procedure:

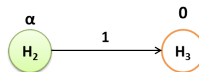


# Examples

## Fixed sequence procedure

- Assume  $H_1 \rightarrow H_2 \rightarrow H_3$
- Example fixed sequence procedure:

$H_2$  rejected at level  $\alpha$



# Examples

## Fixed sequence procedure

- Assume  $H_1 \rightarrow H_2 \rightarrow H_3$
- Example fixed sequence procedure:

$H_3$  not rejected at level  $\alpha$  (stop)



# Examples

- *Bonferroni*: no  $\alpha$ -propagation (no edges between nodes)

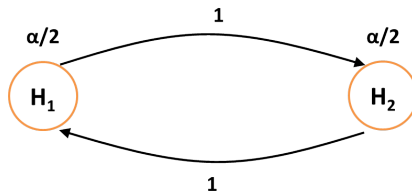


# Examples

- *Bonferroni*: no  $\alpha$ -propagation (no edges between nodes)

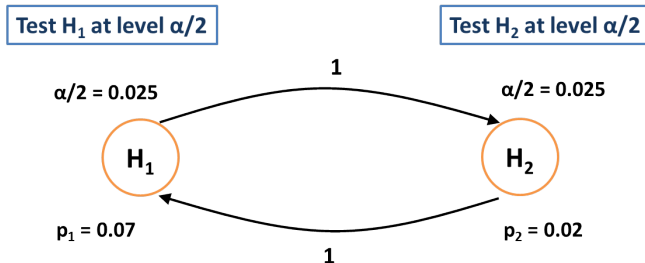


- *Holm*: includes  $\alpha$ -propagation  $\rightarrow$  more powerful



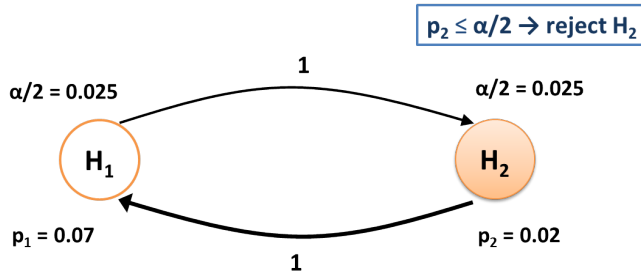
# Examples

**Holm procedure** with  $\alpha = 0.05$



# Examples

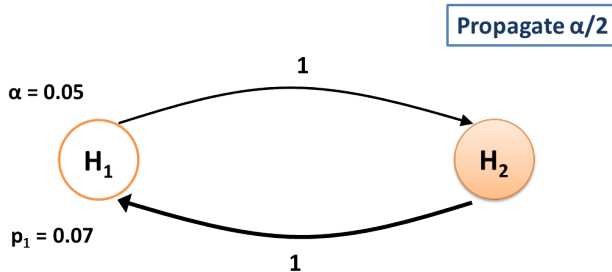
Holm procedure with  $\alpha = 0.05$





# Examples

Holm procedure with  $\alpha = 0.05$



# Examples

**Holm procedure** with  $\alpha = 0.05$

Remove node for  $H_2$

$\alpha = 0.05$



$p_1 = 0.07$

# Examples

**Holm procedure** with  $\alpha = 0.05$

Test  $H_1$  at level  $\alpha$   
 $p_1 > \alpha \rightarrow$  do not reject  $H_1$

$\alpha = 0.05$

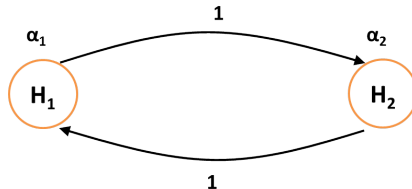


$p_1 = 0.07$

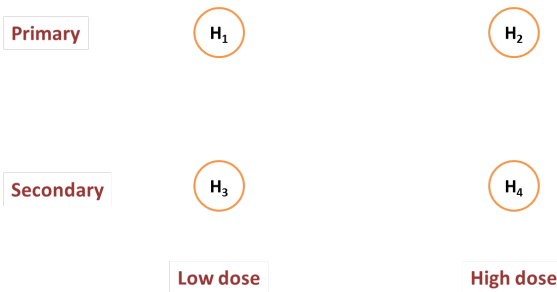
# Examples

## Weighted Holm procedure

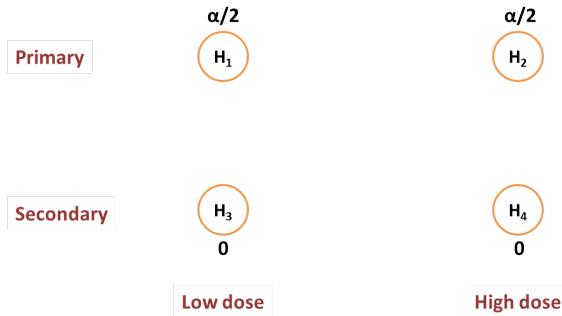
Use  $\alpha_1, \alpha_2$  (where  $\alpha_1 + \alpha_2 = \alpha$ ) instead of  $\alpha_1 = \alpha_2 = \alpha/2$



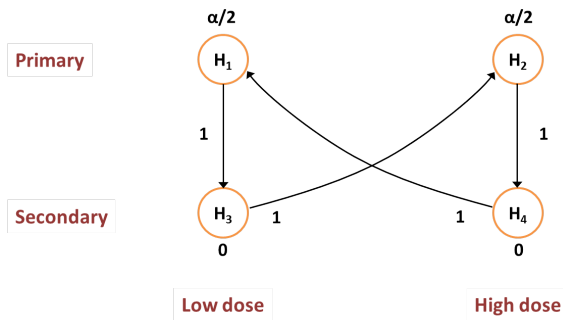
# Diabetes example



# Diabetes example



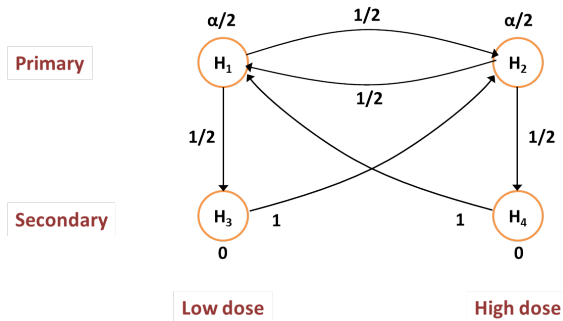
# Diabetes example



$$\alpha = \begin{pmatrix} \frac{\alpha}{2} & \frac{\alpha}{2} & 0 & 0 \end{pmatrix} \quad \mathbf{G} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

(Based on the graphs in Maurer et al., 2011)

# Diabetes example



$$\alpha = \begin{pmatrix} \frac{\alpha}{2} & \frac{\alpha}{2} & 0 & 0 \end{pmatrix} \quad \mathbf{G} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

(Based on the graphs in Maurer et al., 2011)



# Update algorithm

Transition matrix  $\mathbf{G} = (g_{ij})$ , where  $g_{ij}$  is the fraction of  $\alpha_i$  allocated to  $H_j$  if  $H_i$  is rejected.

Require  $0 \leq g_{ij} \leq 1$ ,  $g_{ii} = 0$  and  $\sum_{k=1}^K g_{ik} = 1$  for  $i, j = 1, \dots, K$ .

- 0 Set  $J = \{1, \dots, K\}$
- 1 Select a  $j \in J$  such that  $p_j \leq \alpha_j$  and reject  $H_j$ ; otherwise stop
- 2 Update the graph:

$$J \rightarrow J \setminus \{j\}$$

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

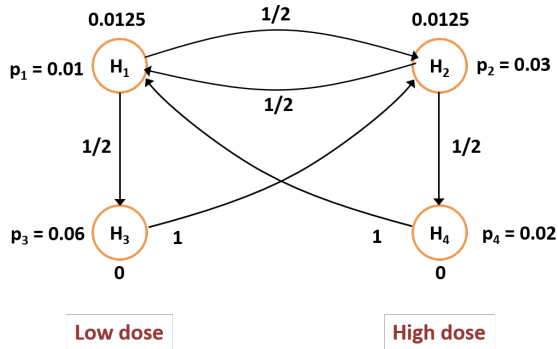
$$g_{\ell k} \rightarrow \begin{cases} \frac{g_{\ell k} + g_{\ell j} g_{jk}}{1 - g_{\ell j} g_{j\ell}} & \text{for } \ell, k \in J, \ell \neq k, g_{\ell j} g_{j\ell} < 1 \\ 0 & \text{otherwise} \end{cases}$$

- 3 If  $|J| \geq 1$ , go to Step 1; otherwise stop

(Bretz et al., 2009)

# Diabetes example

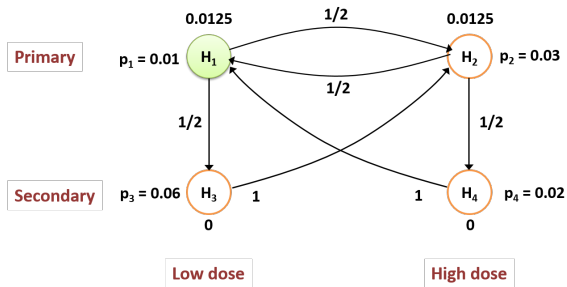
$$\alpha = 0.025$$



(Based on the graphs in Maurer et al., 2011)

# Diabetes example

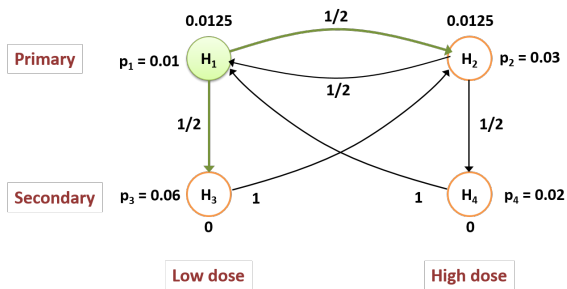
$$\alpha = 0.025$$



(Based on the graphs in Maurer et al., 2011)

# Diabetes example

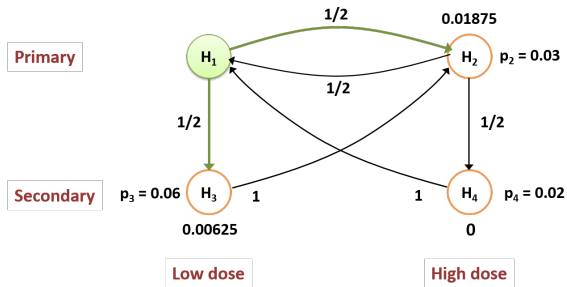
$$\alpha = 0.025$$



(Based on the graphs in Maurer et al., 2011)

# Diabetes example

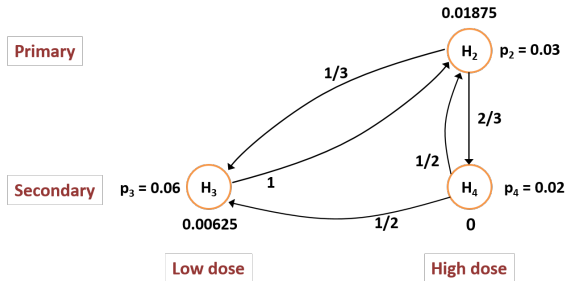
$$\alpha = 0.025$$



(Based on the graphs in Maurer et al., 2011)

# Diabetes example

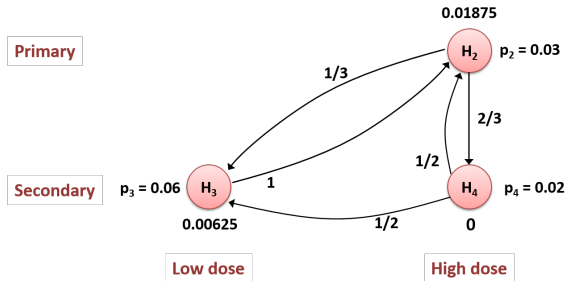
$$\alpha = 0.025$$



(Based on the graphs in Maurer et al., 2011)

# Diabetes example

$$\alpha = 0.025$$



(Based on the graphs in Maurer et al., 2011)

# gMCP package

gMCP GUI 0.8.10

File Example graphs Analysis Extras Help

Place new nodes and edges or start the test procedure

Graph

```

graph LR
    H1((H1  
1/2)) -- 1 --> H2((H2  
1/2))
    H2 -- 1 --> H1
  
```

Transition Matrix

	H1	H2
H1	0	1
H2	1	0

Hypothesis      Weights      P-Value

H1      0.5      0.015     

H2      0.5      0.097     

Sum of weights: 1     

Total  $\alpha$ :      0.025

☒ No information about correlations (Bonferroni based weighted tests)

☐ Select an R correlation matrix                 

☐ Use Simes test

Description      Analysis

Graph representing the (unweighted) Bonferroni-Holm-Procedure

The graph is a complete graph, where all nodes have the same weights and each edge weight is  $1/(n-1)$ .

Literature: Holm, S. (1979). A simple sequentially



# Summary

- Graphical approach allows the following (Bretz et al. 2015):
  - ▶ Tailor advanced multiple test procedures to a structured families of hypotheses
  - ▶ Visualise complex decision strategies in an efficient and easily communicable way
  - ▶ Ensure strong FWER control
- Approach covers many common multiple test procedures as special cases
- Many possible extensions as well

## Multiplicity in practice

# Choosing a multiple testing procedure

A lot of advice available!

- Dmitrienko and D'Agostino (2013) Traditional multiplicity adjustment methods in clinical trials. *Stat Med* 32:5172-5218.
- Dmitrienko and D'Agostino (2018) Multiplicity Considerations in Clinical Trials. *NEJM* 378:2115-2122.
- Howard *et al.* (2018) Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *SMMR* 27(5):1513-1530.
- Li *et al.* (2017) An introduction to multiplicity issues in clinical trials: the what, why, when and how. *Int J Epi* 46(2):746-55
- Wang *et al.* (2011) Regulatory Perspectives on Multiplicity in Adaptive Design Clinical Trials throughout a Drug Development Program. *J Biopharm Stat* 21:846-59.
- Wason *et al.* (2014) Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? *Trials* 15:364.

# Choosing a multiple testing procedure

- Need to identify the adjustment method aligned with the structure of the clinical objectives, that provides highest statistical power.
- Use all available clinical and statistical information to arrive at the most appropriate and efficient method.
- Often need to do a simulation-based assessment of operating characteristics of procedures under trial-specific assumptions.
- Helpful to classify methods (see Dmitrienko and D'Agostino, 2013)
  - ▶ Logical restrictions
  - ▶ Distributional information

# Choosing a multiple testing procedure

## Logical restrictions

- Group procedures into three classes:
  - ▶ Single-step that test all hypotheses simultaneously (e.g. Bonferroni)
  - ▶ Stepwise that rely on data-driven ordering (e.g. Holm and Hochberg)
  - ▶ Stepwise that rely on pre-specified ordering. (e.g. Fixed-sequence or Fallback)
- Single-step most basic. Stepwise more powerful; easier to tailor.

# Choosing a multiple testing procedure

## Distributional information

- Again group into three classes:
  - ▶ Non-parametric procedures: without distributional assumptions (e.g. Bonferroni, Holm, Fixed-sequence, Fallback)
  - ▶ Semi-parametric methods: control under e.g. independence or positive correlation, but not full specification of a joint distribution (e.g. Hochberg and Hommel)
  - ▶ Parametric methods: control FWER only when joint distribution of test statistics fully specified (e.g. Dunnett)
- In general, power increases as more distributional assumptions are added

# Regulatory view

- So far, we have been assuming that we have to adjust for multiplicity.
- Often we will need to, but not always.
- Latest (draft) regulatory guidance from the FDA and EMA:
  - ▶ EMA: *Guideline on multiplicity issues in clinical trials* (2016)
  - ▶ FDA: *Multiple Endpoints in Clinical Trials* (2017)

# Software

## R

- **multxpert** package for common adjustments in single family (non-parametric, semi-parametric and parametric procedures) and multiple families (gatekeeping procedures).
  - ▶ <http://multxpert.com/wiki/Software>
  - ▶ “Multiple Testing Problems in Pharmaceutical Statistics” by Dmitrienko *et al.* (2009)
- **gMCP** for graph based multiple test procedures.
- **multcomp** package addresses multiplicity issues in general linear and non-linear models.
  - ▶ Adjustments for ANOVA, ANCOVA, regression models, and more.
  - ▶ “Multiple Comparisons Using R” by Bretz *et al.* (2016)



# Software

## Stata

- Several commands support correcting for multiple comparisons.
- oneway (one-way ANOVA) commands
- Powerful test command for testing linear hypotheses from regress, stcox, logit, svy has an mtest option that supports Bonferroni, Holm, etc.

# Summary

- There is a lot of advice available on how to choose a multiple testing procedure.
- But no golden rule: have to tailor to your trial.
- Important as often required by regulatory authorities.
- Much software out there to do this in practice.

# References I

M. Alosh, F. Bretz and M. Huque

Advanced multiplicity adjustment methods in clinical trials

*Statistics in Medicine*, 33:693–713, 2011.

A. Dmitrienko, A. C. Tamhane and F. Bretz

*Multiple testing problems in pharmaceutical statistics*, CRC Press, 2009.

F. Bretz et al.

Confirmatory Seamless Phase II/III Clinical Trials with Hypotheses Selection at Interim: General Concepts

*Biometrical Journal*, 48:623–634, 2006.

F. Bretz et al.

Adaptive designs for confirmatory clinical trials

*Statistics in Medicine*, 28:1181–1217, 2009.

F. Bretz, T. Hothorn, P. Westfall

*Multiple comparisons using R*, CRC Press, 2016.

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

A graphical approach to sequentially rejective multiple test procedures

*Statistics in Medicine*, 28:586–604, 2009.

# References II

F. Bretz, W. Maurer and J. Maca

Graphical approaches to multiple testing

in *Clinical Trial Biostatistics and Biopharmaceutical Applications*, CRC Press, 2015.

W. Maurer, E. Glimm and F. Bretz

Multiple and repeated testing of primary, coprimary, and secondary hypotheses

*Statistics in Biopharmaceutical Research*, 3(2):336–352, 2011.

H. Schmidli et al.

Confirmatory Seamless Phase II/III Clinical Trials with Hypotheses Selection at Interim: Applications and Practical Considerations

*Biometrical Journal*, 48:635–643, 2006.

J.P. Shaffer

Multiple Hypothesis Testing

*Annual Review Psychology*, 46:561-84, 1995.