Adaptive graph-based multiple testing procedures

Koenig F.,Klinglmueller F., Posch M.,
Medical University of Vienna - Section of Medical Statistics
www.meduniwien.ac.at/user/franz.koenig

PSI and RSS Journal Club
2015-05-21

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013–602552.

"Modern" multiple testing procedures

- Recently many multiple testing procedures that address specific multiplicity issues in clinical trials have been proposed.
- Reflect the contextual relationships between hypotheses in the inference procedure (e.g. test secondary hypotheses only if primary hypotheses are rejected)
- Examples are Fixed Sequence Test, Fallback Test, Gatekeeping Tests, ...
- Especially graph-based multiple testing procedures:
  - can be easily tailored to the problem at hand,
  - make the guiding principle behind the procedures more transparent,
  - help to communicate the procedures to clinicians and regulators.

[Bretz, Maurer, Brannath, Posch (2009)]

Multiple testing in adaptive designs

- Methods for a limited number of standard situations:
  - treatment selection
  - multiple (co-primary) endpoints
  - subgroup selection

  (Bauer & Kieser 1999; Hommel 2001; Posch et al. (2005); ...)
- Reflect only very simple relations/hierarchies between hypotheses
- Difficult or impossible to tailor to more general multiple testing problems
- Closed test of adaptive combination tests does not allow to use the preplanned test if no adaptations are performed

Multiplicity in adaptive clinical trials

Review of scientific advice letters [Elsäßer et al. (2014)]

Review 59 scientific advise requests concerning adaptive trials. They identify that a large proportion concerns multiple testing issues and state: "However, even though a huge range of statistical methodology to avoid type I error inflation in adaptive clinical trials has been developed over the years, type I error control in adaptive clinical trials surprisingly is still a frequent major concern raised in the SA letters."

Multiplicity control in adaptive designs

- Multiple testing procedures for adaptive designs that reflect the contextual relations between hypotheses start to be addressed only very recently.
  - Adaptive design for primary and secondary hypotheses [Tamhane et al. (2012)]
  - Graph based partitioning algorithm that applies the graphical approach to adaptive combination tests [Sugitani et al. (2013)].
  - Adaptive graph-based multiple testing procedures based on the partial conditional error rate approach [Klinglmueller et al. (2014)]

Adaptive graph-based multiple testing procedures

A general framework to

1. address multiplicity in complex clinical trials
2. reflect the relative importances, contextual relations, logical restrictions of clinical hypotheses
3. **permit adaptive interim analysis, i.e., trial modifications based on unblinded trial data or external information**

Strict FWE control if adaptations are performed:

- To deal with multiplicity: Apply graphical approach `[(Bretz, Maurer, Brannath, Posch '09)]`
- To account for adaptivity: Use partial conditional errors `[(Posch, Futschik '08),(Posch, Maurer, Bretz '10)]`

Example: Multiple treatment arms, multiple endpoints

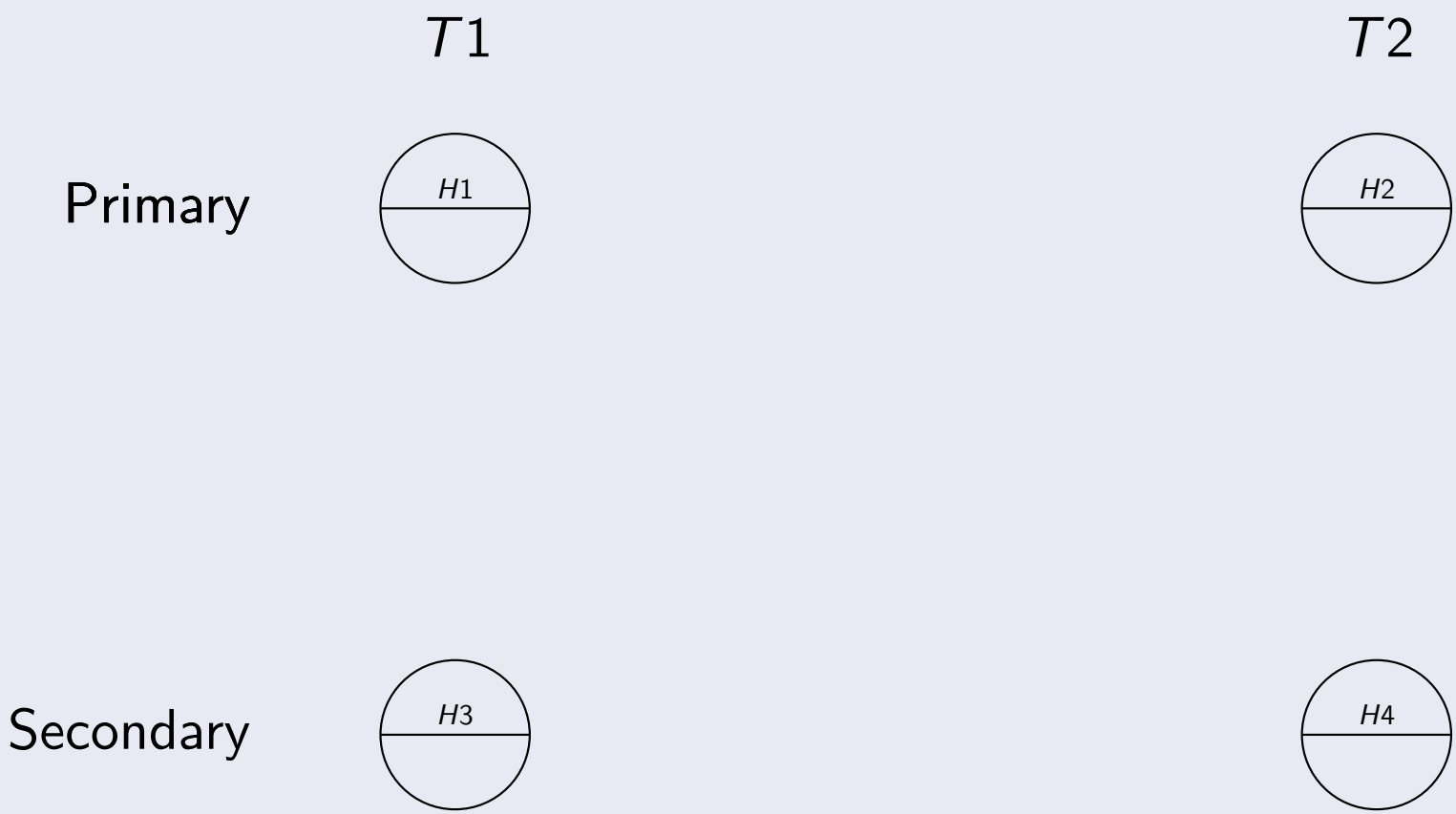
Late phase development of a new drug

- Two treatments, $T1$, $T2$, *e.g.*, high dose, low dose, are compared to placebo.
- Two endpoints, one primary and one secondary.
- There are 4 hypotheses to be tested

Desired properties for multiple test:

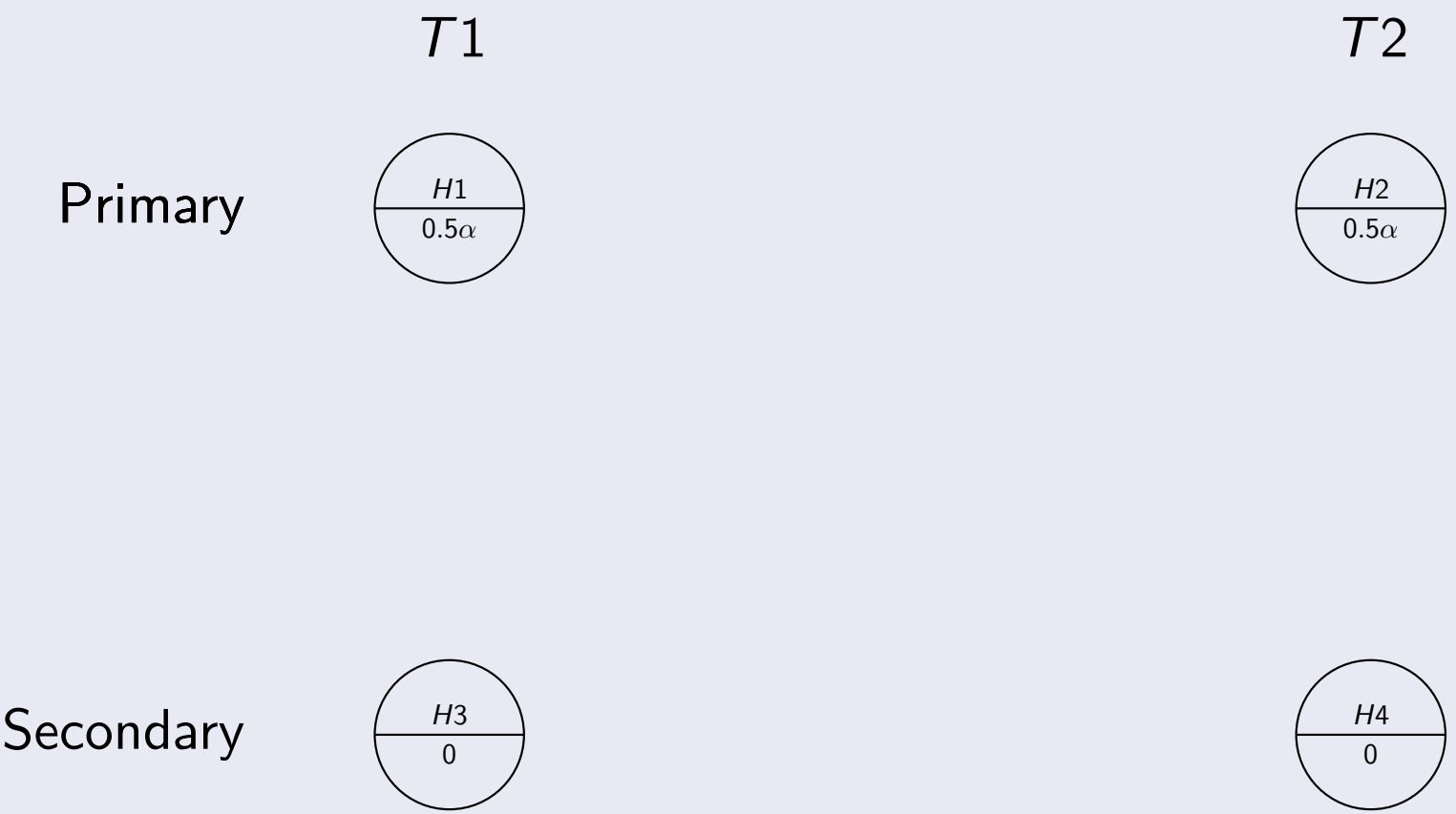
1. Family wise error control ($\alpha = .025$ one sided)
2. Assuming equal effect sizes both treatments have equal chances of a positive result.
3. Test secondary hypothesis only if the corresponding primary hypothesis is rejected.
4. Reject as many hypotheses as possible.

Two doses, two endpoints, $\alpha = .025$



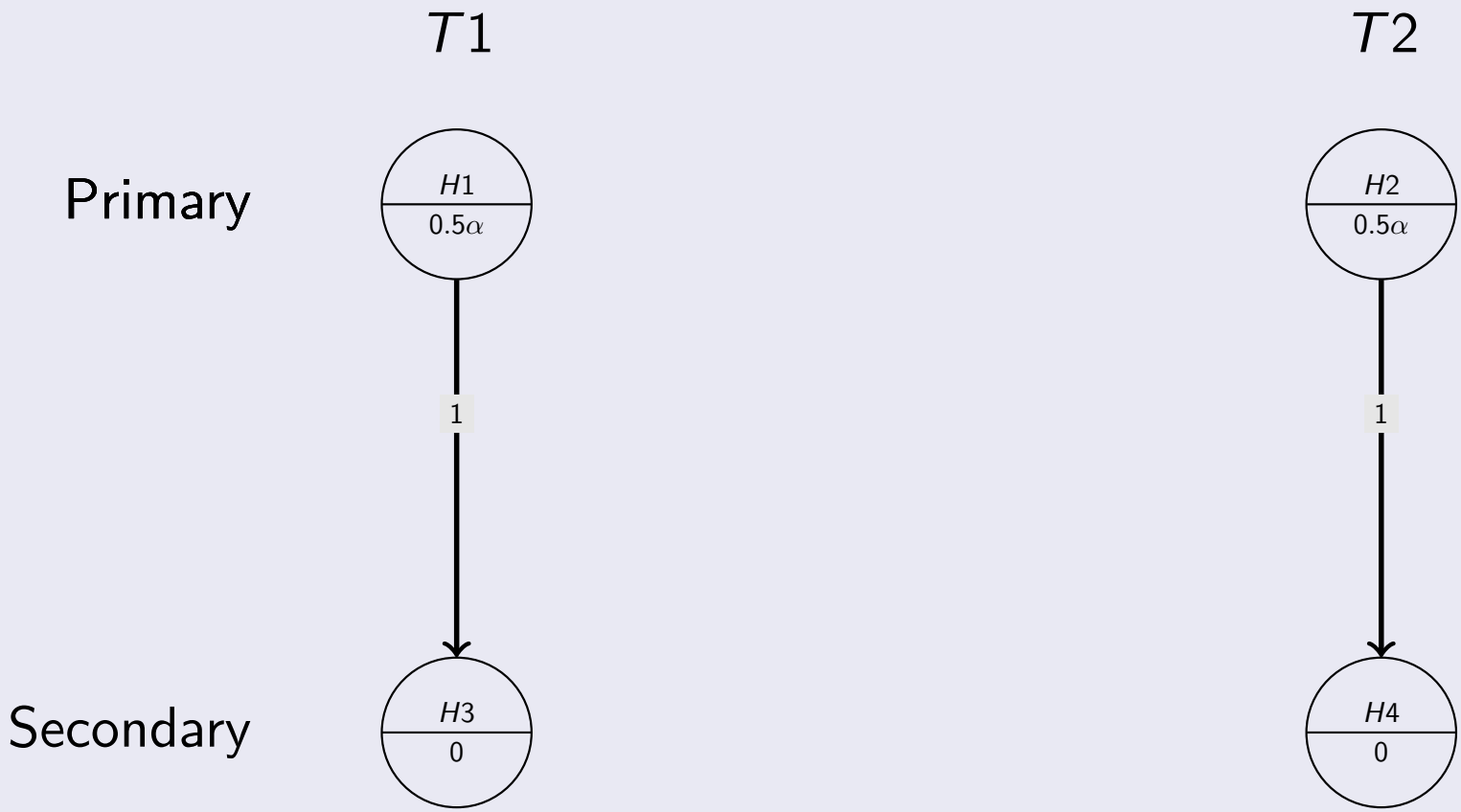
- 2. Split α using equal weights between doses

Two doses, two endpoints, $\alpha = .025$



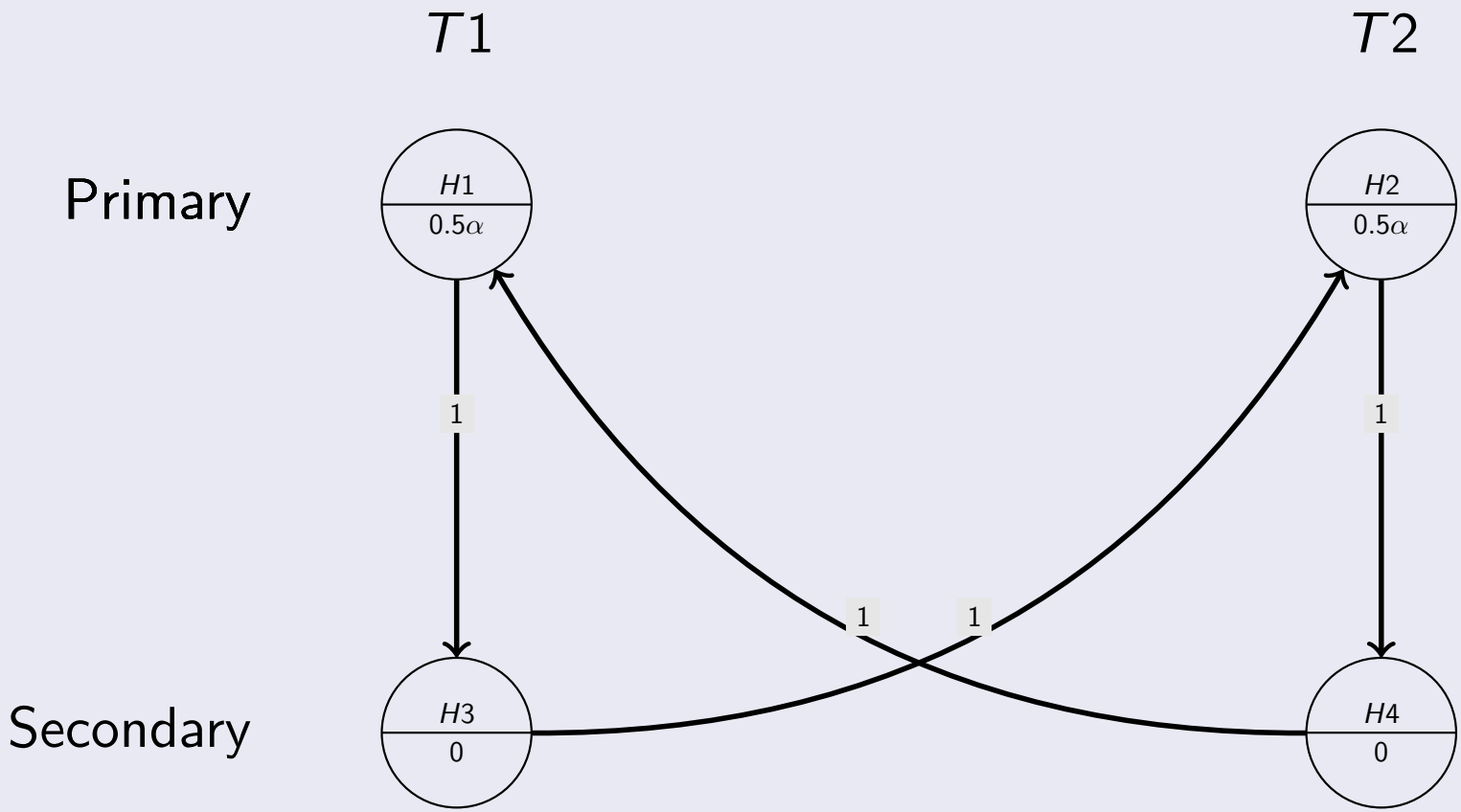
- 3. Initially give zero weight to secondary hypotheses

Two doses, two endpoints, $\alpha = .025$



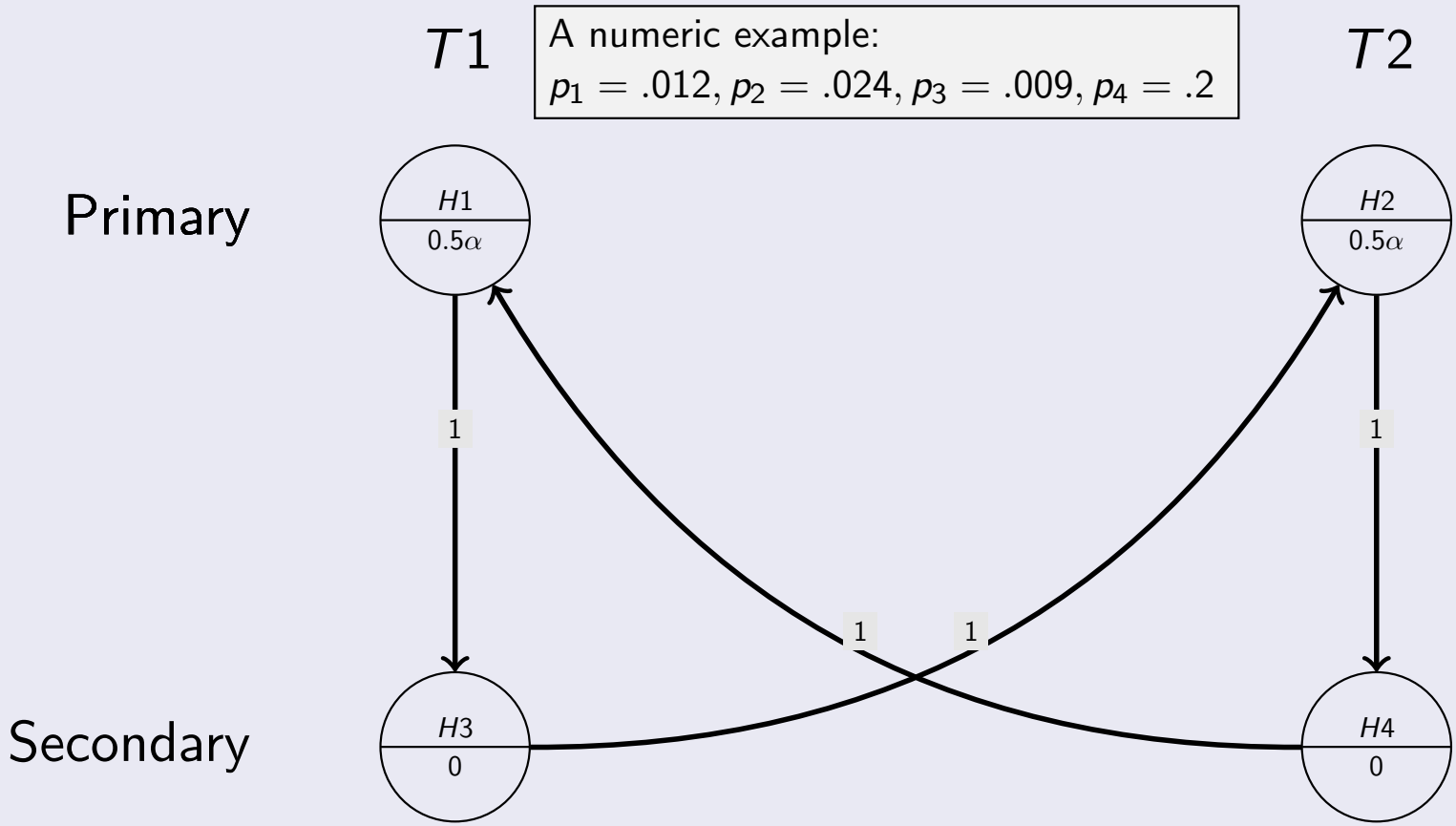
- 4. If rejected, reallocate weight from primary to secondary

Two doses, two endpoints, $\alpha = .025$



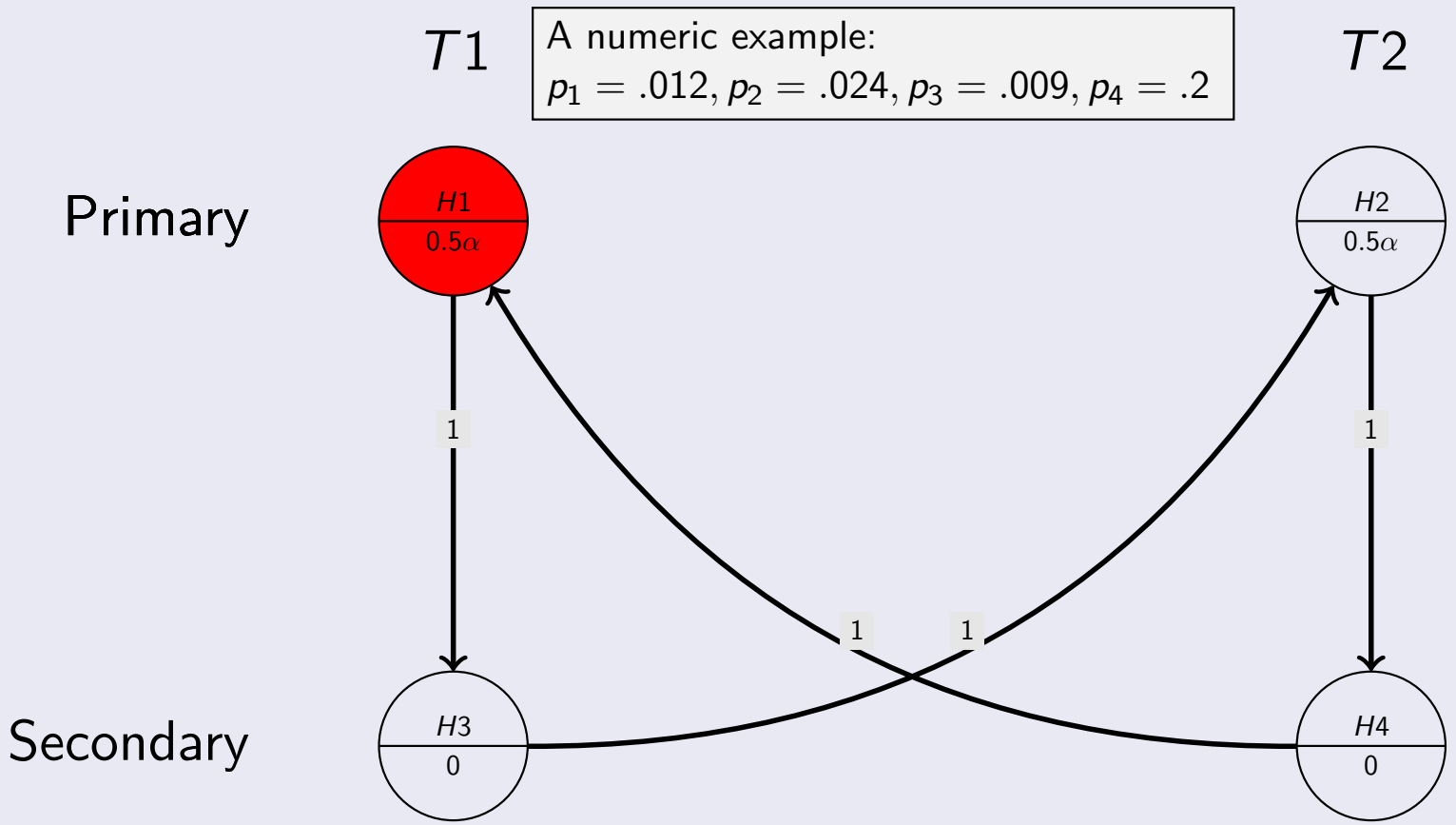
- 4. If both endpoints of a dose are rejected reallocate weight to other dose

Two doses, two endpoints, $\alpha = .025$



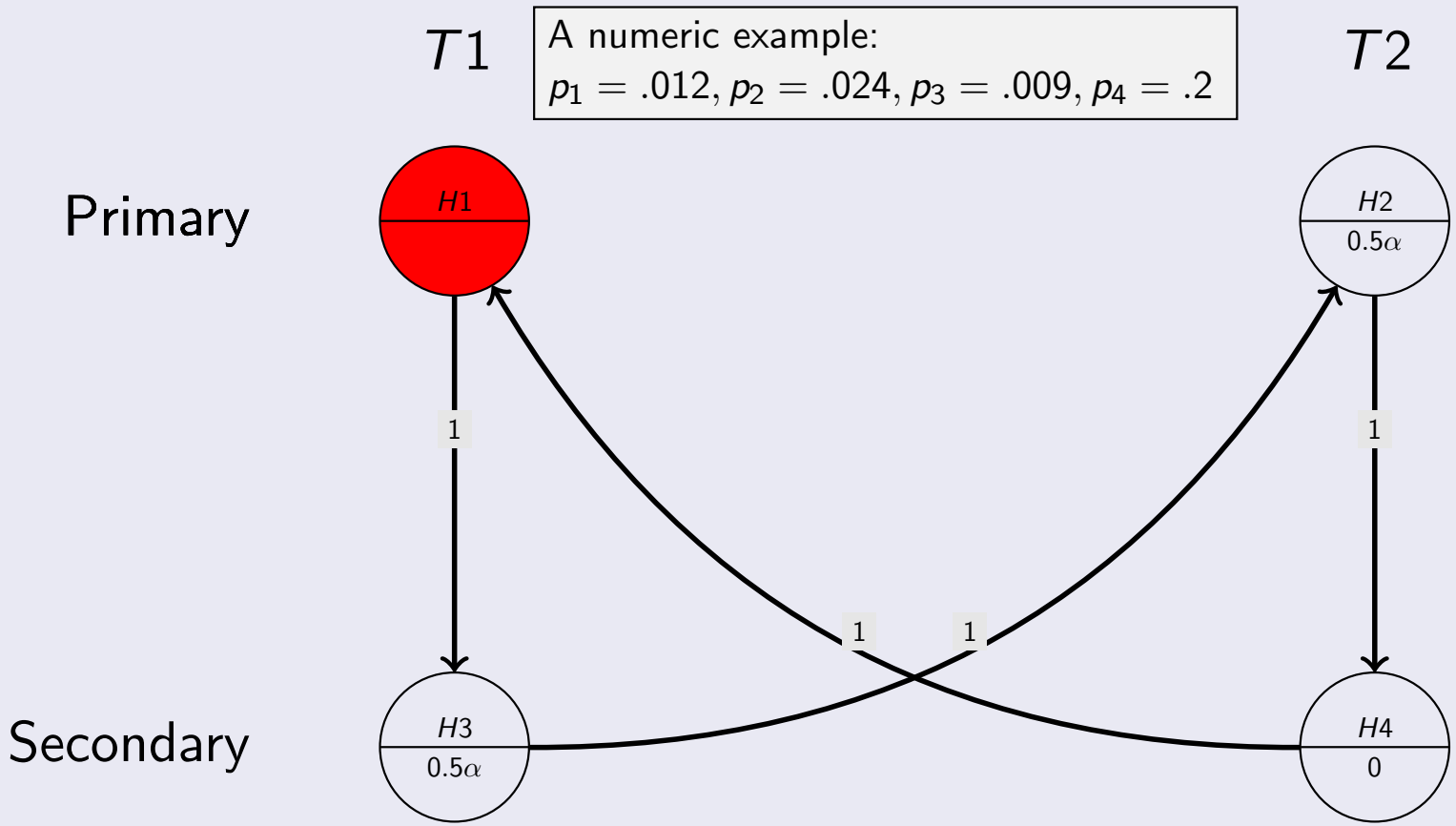
- Example: Reject H1 as $p_1 < 0.5\alpha$. Update graph by removing H_1

Two doses, two endpoints, $\alpha = .025$

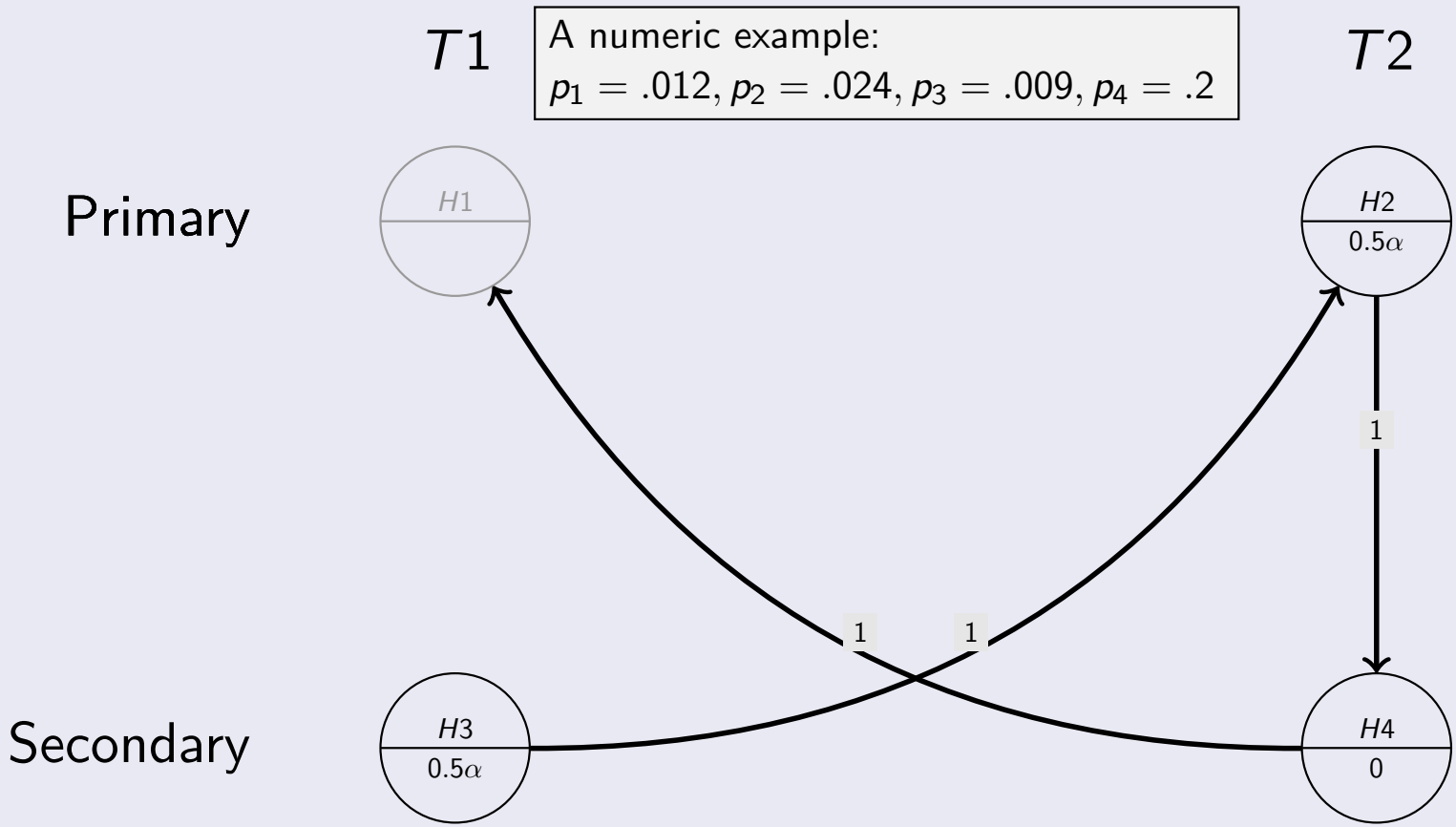


- Reallocate weight of removed hypothesis

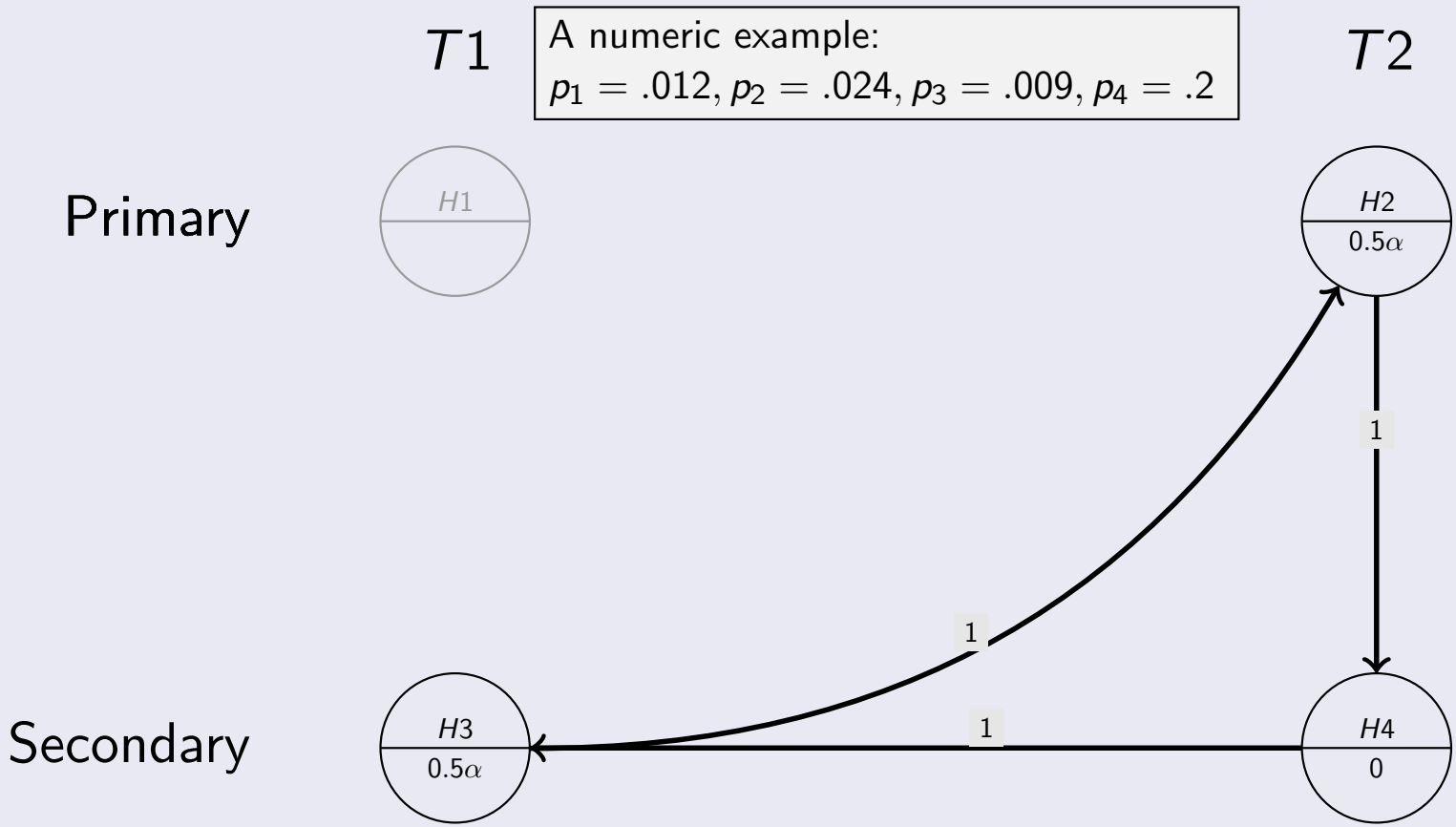
Two doses, two endpoints, $\alpha = .025$



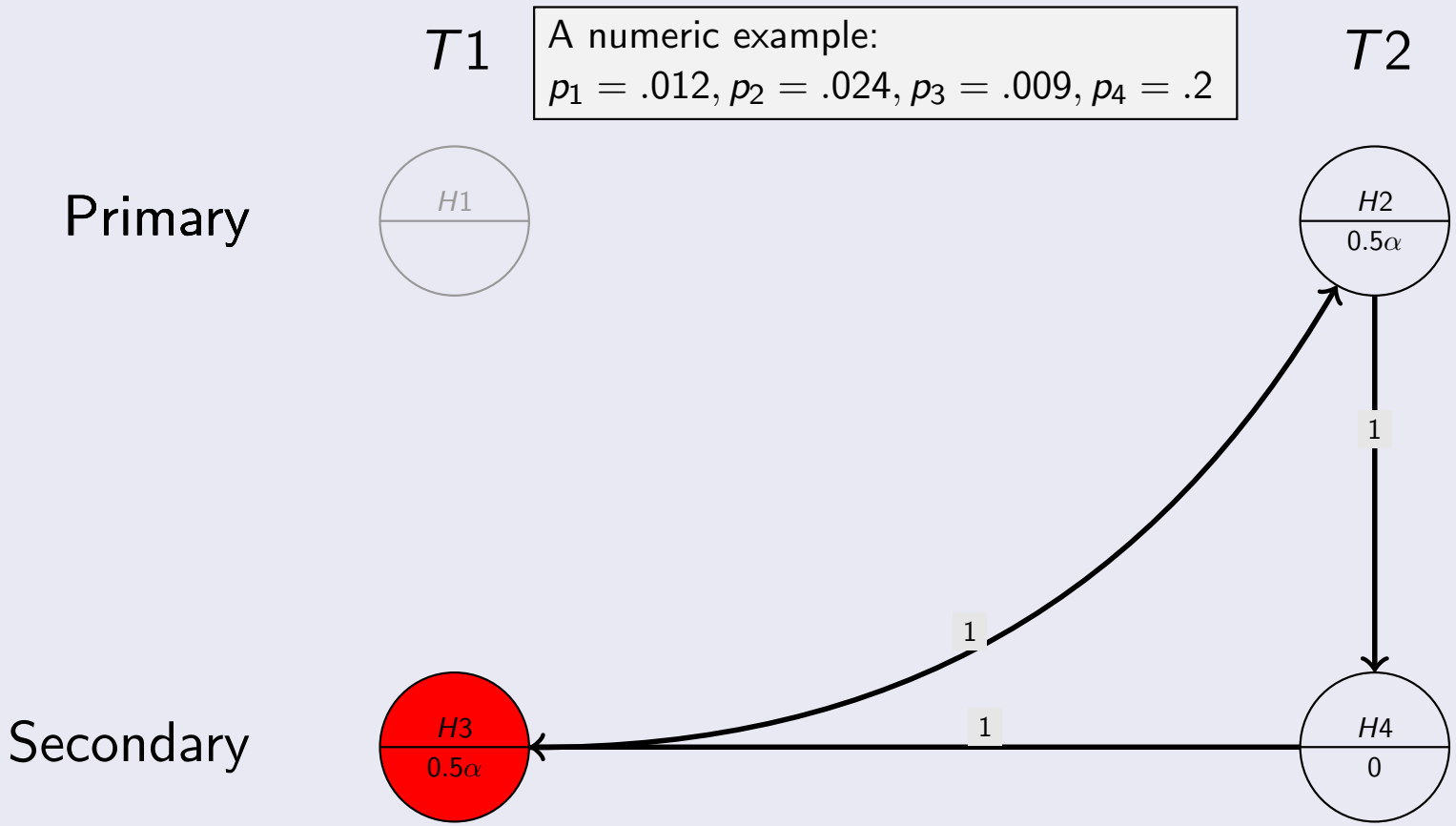
Two doses, two endpoints, $\alpha = .025$



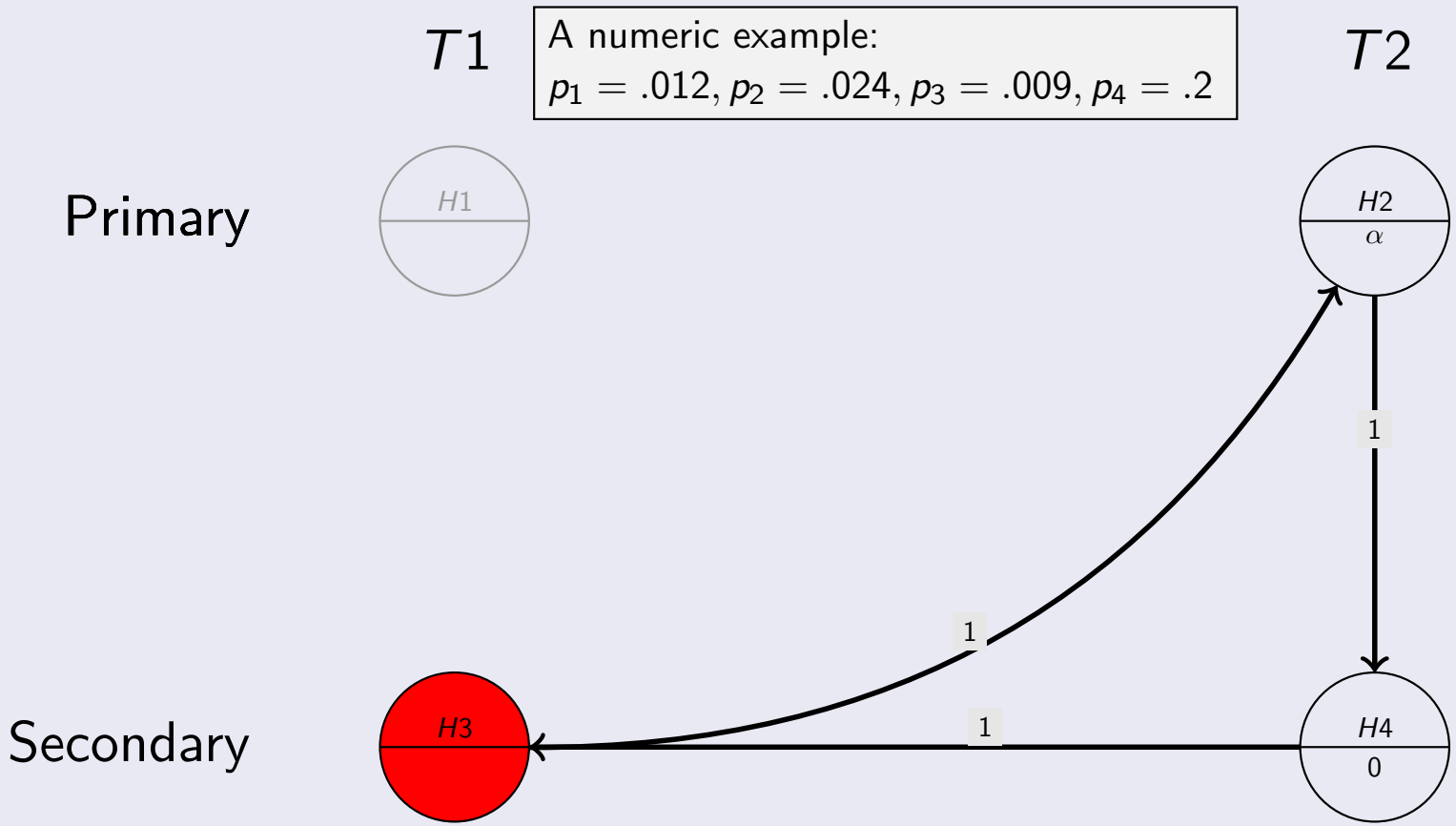
Two doses, two endpoints, $\alpha = .025$



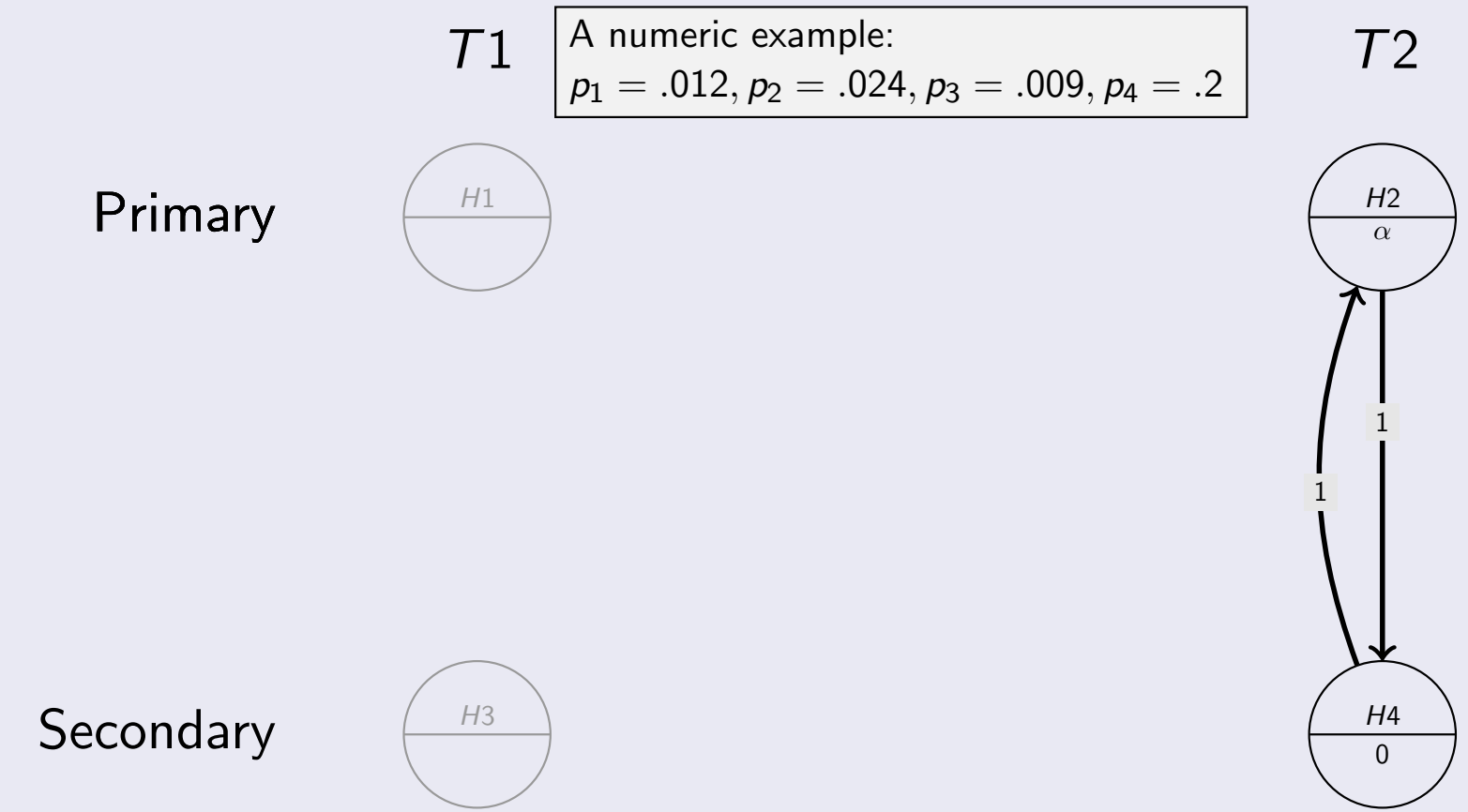
Two doses, two endpoints, $\alpha = .025$



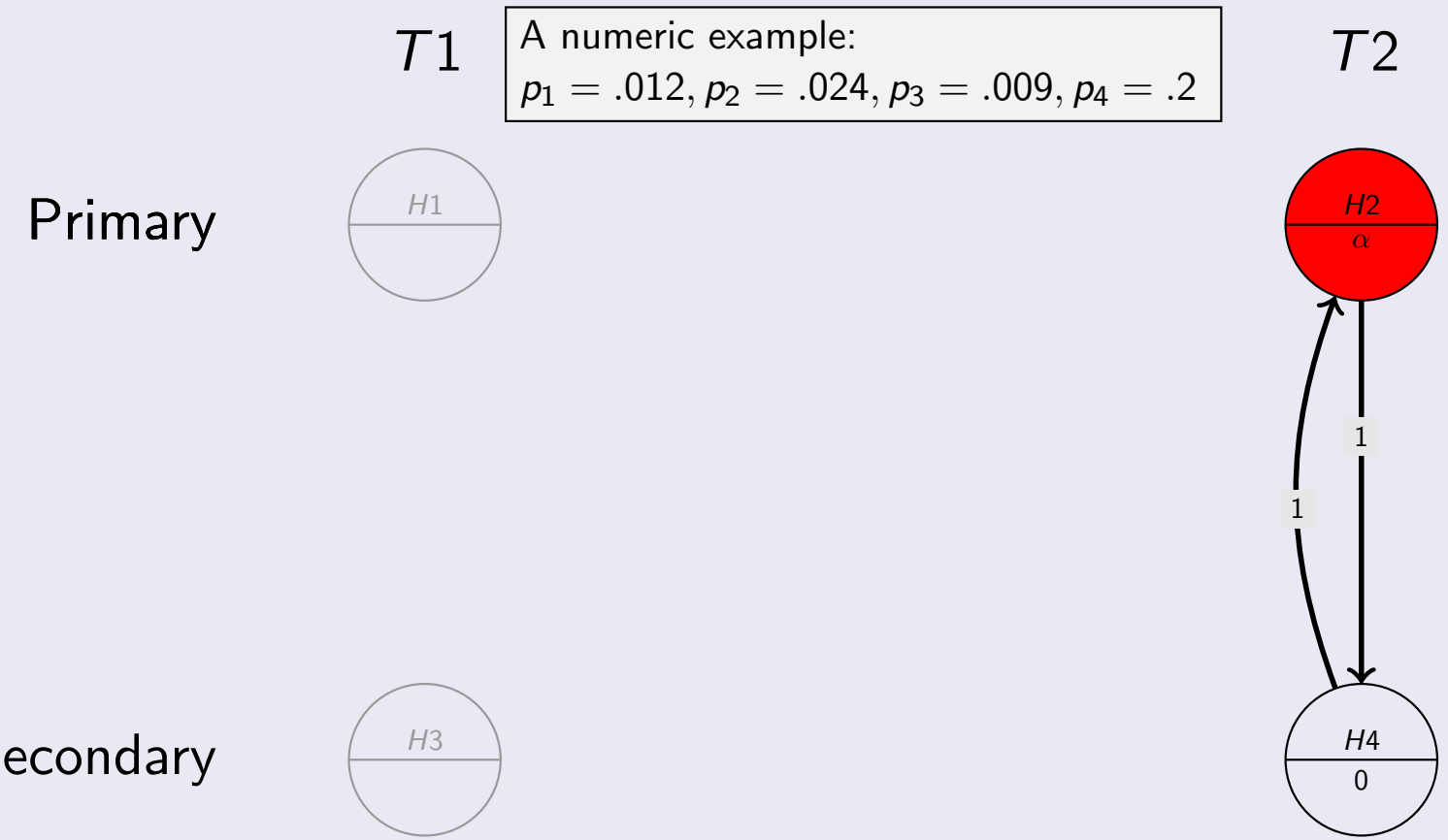
Two doses, two endpoints, $\alpha = .025$



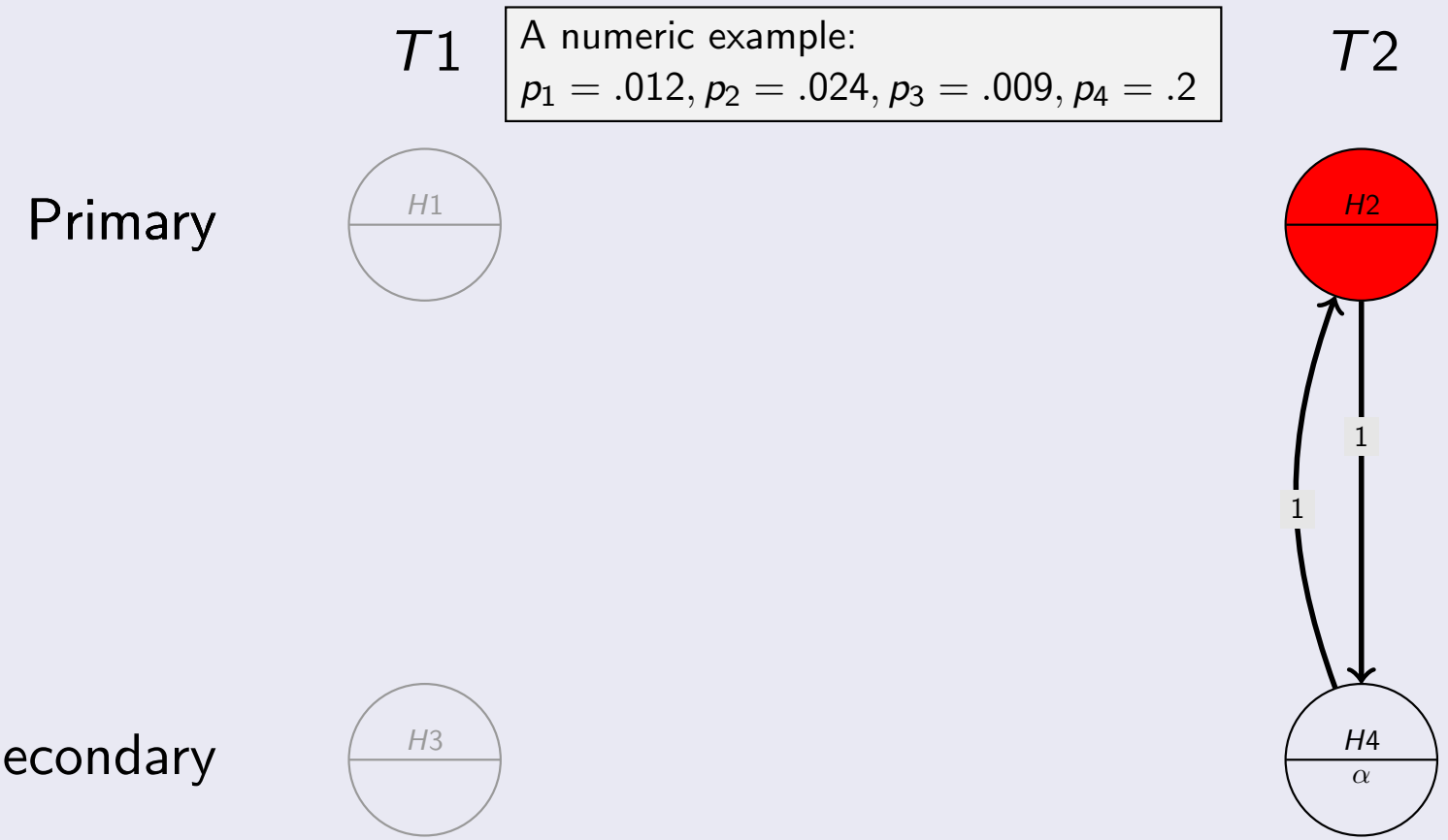
Two doses, two endpoints, $\alpha = .025$



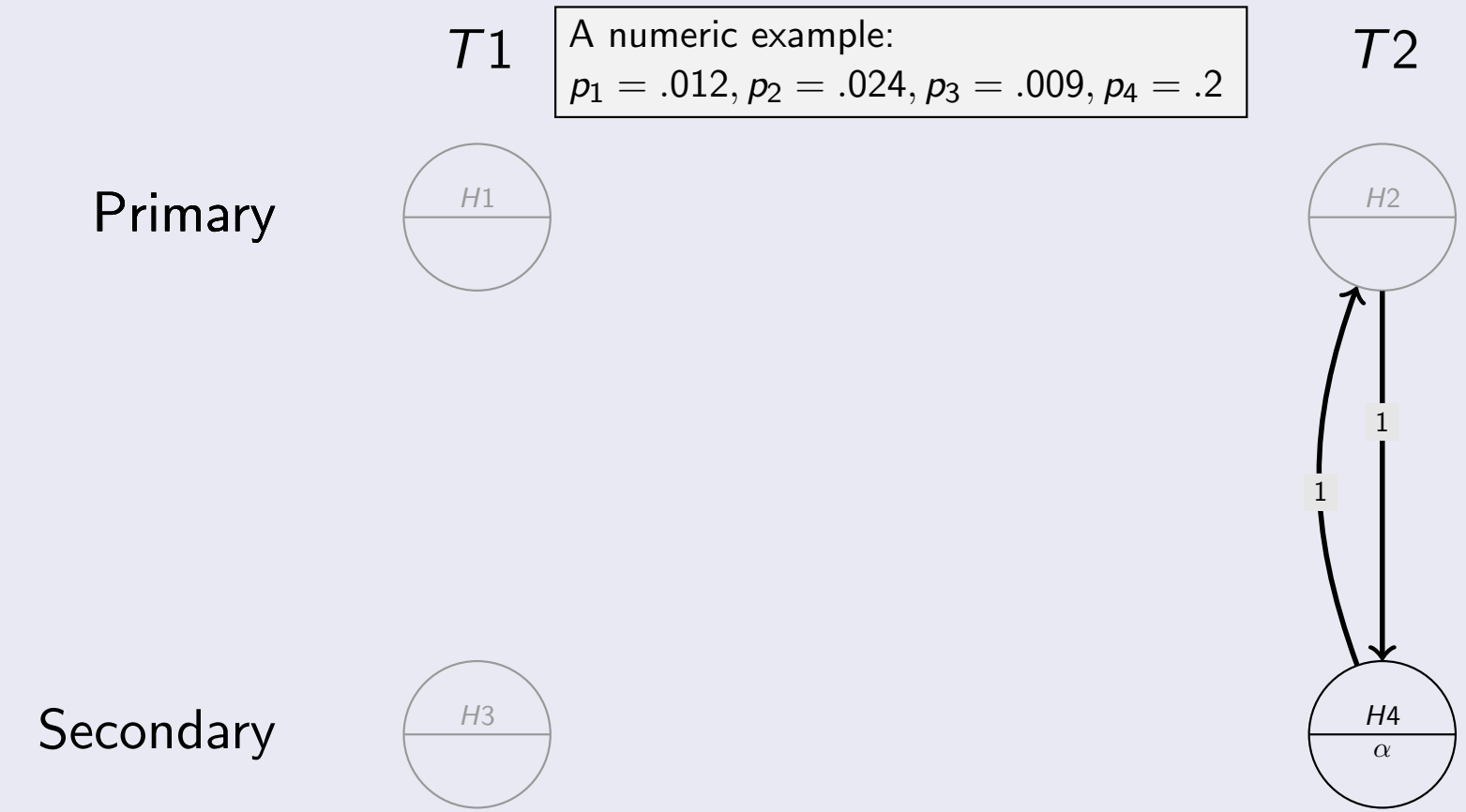
Two doses, two endpoints, $\alpha = .025$



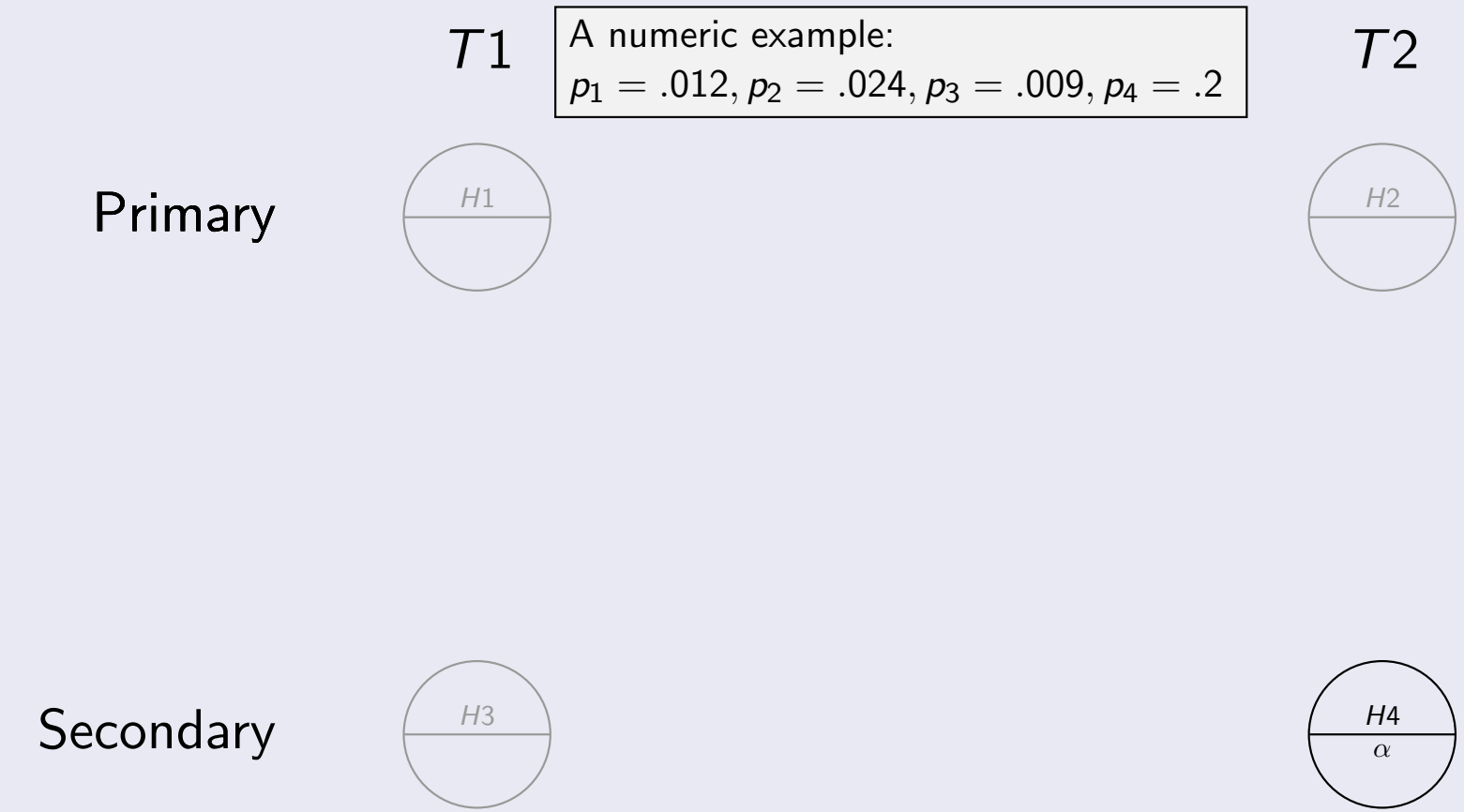
Two doses, two endpoints, $\alpha = .025$



Two doses, two endpoints, $\alpha = .025$



Two doses, two endpoints, $\alpha = .025$



Graphical approach to multiple testing procedures

- The graphical approach presents an intuitive way to construct and communicate multiple testing procedures that, reflect the relative importances, contextual relations, or logical restrictions between hypotheses.
- It provides strict control of the family wise error rate (FWER)
- *The graph defines a closed testing procedure of weighted Bonferroni tests.*
- The sequential rejection principle provides a shortcut to the closed test.

[Bretz et al. '09,'11]

The Closure Principle

- 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

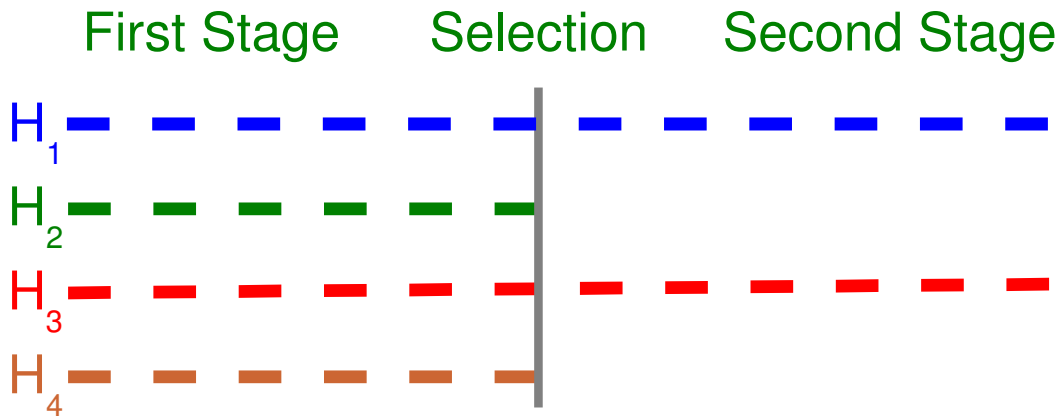
- The closed testing procedure controls the FWE at α in the strong sense.
- Requires $2^m - 1$ tests!
- The graph and algorithm implicitly define
  - weighted Bonferroni tests for all intersection hypotheses
  - a shortcut that reduces the number of tests: in each step, a large number of intersection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

Closure of weighted tests

Intersection H_0	Fixed Sample
$H_1 \cap H_2 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$
$H_1 \cap H_2 \cap H_3$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$
$H_1 \cap H_2 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$
$H_1 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_4 \leq 0.5\alpha$
$H_2 \cap H_3 \cap H_4$	$p_2 \leq 0.5\alpha \vee p_3 \leq 0.5\alpha$
$H_1 \cap H_2$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$
$H_1 \cap H_3$	$p_1 \leq \alpha$
$H_1 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_4 \leq 0.5\alpha$
$H_2 \cap H_3$	$p_2 \leq 0.5\alpha \vee p_3 \leq 0.5\alpha$
$H_2 \cap H_4$	$p_2 \leq \alpha$
$H_3 \cap H_4$	$p_3 \leq 0.5\alpha \vee p_4 \leq 0.5\alpha$
H_1	$p_1 \leq \alpha$
H_2	$p_2 \leq \alpha$
H_3	$p_3 \leq \alpha$
H_4	$p_4 \leq \alpha$

$2^4 - 1 = 15$ Intersection hypotheses!

Adaptive Designs with Selection of Hypotheses.

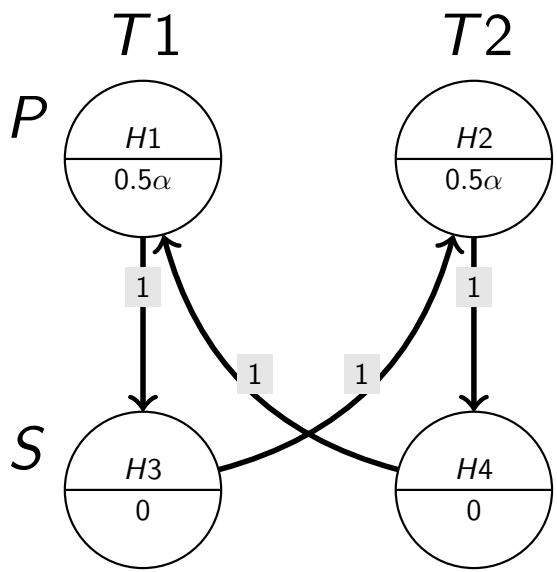


- Start with a graph specifying the multiple testing procedure for all m hypotheses.
- In an interim analysis some hypotheses are dropped.
- Only for the continued hypotheses further observations are collected.
- The data of both stages is used in the final test.
- Control of the FWE in the strong sense.

Applications: Treatment or subgroup selection in clinical trials, ...

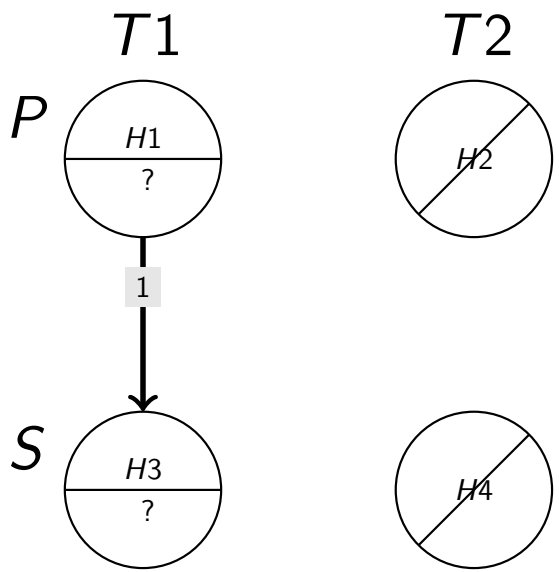
Dropping of a treatment at interim

Assume that midway throughout the trial (*e.g.*, following safety concerns) the trial data is unblinded and the decision is made to stop sampling for $T2$.



Dropping of a treatment at interim

Assume that midway throughout the trial (*e.g.*, following safety concerns) the trial data is unblinded and the decision is made to stop sampling for $T2$.

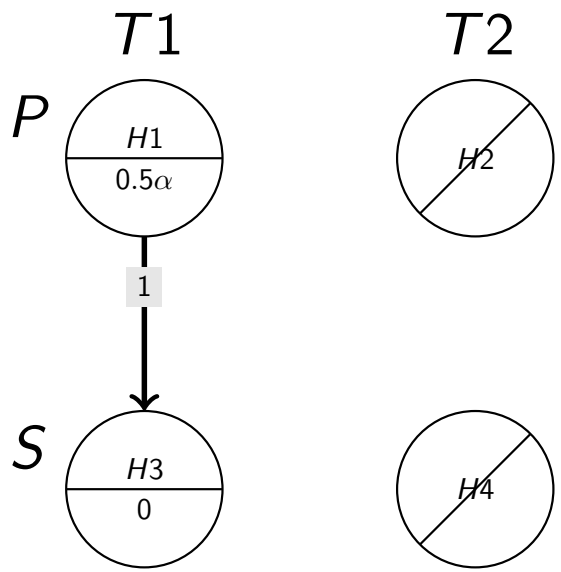


Dropping of a treatment at interim

Assume that midway throughout the trial (*e.g.*, following safety concerns) the trial data is unblinded and the decision is made to stop sampling for $T2$.

A simple strategy

- Having dropped $T2$ simply accept H_2 and H_4, discard reserved $\alpha/2$
- Test H_1 and H_3 using a fixed-sequence test at level $\alpha/2$
- Conservative - cannot use $\alpha/2$ foreseen for higher dose
- Does not allow further adaptations (e.g. sample size reassessment)
- Can we do better?

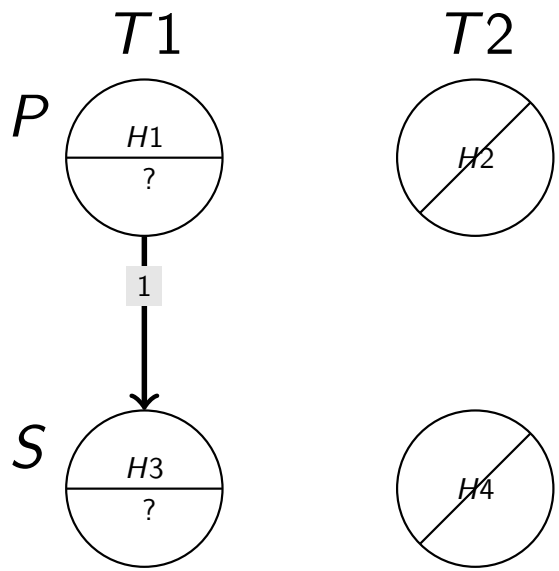


Dropping of a treatment at interim

Assume that midway throughout the trial (*e.g.*, following safety concerns) the trial data is unblinded and the decision is made to stop sampling for $T2$.

Desired strategy

- Use 1^{st}-stage data from $T2$.
- Optionally re-allocate samples from dropped treatment arm to $T1$ and control.
- Use an updated graph to define new test procedure!



Weighted closed test

Intersection H_0	Fixed Sample	Simple strategy	Adaptive
$H_1 \cap H_2 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_1 \cap H_2 \cap H_3$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_1 \cap H_2 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_1 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_4 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_2 \cap H_3 \cap H_4$	$p_2 \leq 0.5\alpha \vee p_3 \leq 0.5\alpha$	$p_3 \leq 0.5\alpha$	?
$H_1 \cap H_2$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_1 \cap H_3$	$p_1 \leq \alpha$	$p_1 \leq \alpha$	?
$H_1 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_4 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_2 \cap H_3$	$p_2 \leq 0.5\alpha \vee p_3 \leq 0.5\alpha$	$p_3 \leq 0.5\alpha$	?
$H_2 \cap H_4$	$p_2 \leq \alpha$		?
$H_3 \cap H_4$	$p_3 \leq 0.5\alpha \vee p_4 \leq 0.5\alpha$	$p_3 \leq 0.5\alpha$	?
H_1	$p_1 \leq \alpha$	$p_1 \leq \alpha$	?
H_2	$p_2 \leq \alpha$		?
H_3	$p_3 \leq \alpha$	$p_3 \leq \alpha$	?
H_4	$p_4 \leq \alpha$		?

Can we use the first stage data from the dropped treatment when testing corresponding intersections?

Weighted closed test

Intersection H_0	Fixed Sample	Simple strategy	Adaptive
$H_1 \cap H_2 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_1 \cap H_2 \cap H_3$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_1 \cap H_2 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_1 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_4 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?
$H_2 \cap H_3 \cap H_4$	$p_2 \leq 0.5\alpha \vee p_3 \leq 0.5\alpha$	$p_3 \leq 0.5\alpha$	?
$H_2 \cap H_4$	$p_2 \leq \alpha$		?
$H_3 \cap H_4$	$p_3 \leq 0.5\alpha \vee p_4 \leq 0.5\alpha$	$p_3 \leq 0.5\alpha$	?
H_1	$p_1 \leq \alpha$	$p_1 \leq \alpha$	?
H_2	$p_2 \leq \alpha$		?
H_3	$p_3 \leq \alpha$	$p_3 \leq \alpha$	?
H_4	$p_4 \leq \alpha$		?

Can we use the first stage data from the dropped treatment when testing corresponding intersections?

Solutions for $H_0 = H_1 \cap H_2 \cap H_3 \cap H_4$

Intersection H_0	Fixed Sample	Simple strategy	Adaptive
$H_1 \cap H_2 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?

Solutions for $H_0 = H_1 \cap H_2 \cap H_3 \cap H_4$

Intersection H_0	Fixed Sample	Simple strategy	Adaptive
$H_1 \cap H_2 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?

Conditional error [Müller, Schäfer, '04]

Second stage test, *i.e.*, using only independent second stage observations, at conditional level

$$A_{1234} = E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha\}} \middle| \text{First Stage Data}\right]$$

Solutions for $H_0 = H_1 \cap H_2 \cap H_3 \cap H_4$

Intersection H_0	Fixed Sample	Simple strategy	Adaptive
$H_1 \cap H_2 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?

Conditional error [Müller, Schäfer, '04]

Secon
observ

Caveat!

For multiple hypotheses one would need to know the joint conditional distribution of second stage statistics, which in general is unknown, *e.g.*, multiple endpoints. Parametric solutions exist, *e.g.*, for many-to-one comparisons [Koenig et al. '08]

Solutions for $H_0 = H_1 \cap H_2 \cap H_3 \cap H_4$

Intersection H_0	Fixed Sample	Simple strategy	Adaptive
$H_1 \cap H_2 \cap H_3 \cap H_4$	$p_1 \leq 0.5\alpha \vee p_2 \leq 0.5\alpha$	$p_1 \leq 0.5\alpha$	?

Partial conditional error [Posch et al. '08,'10]

Second stage test, *i.e.*, based on independent second stage observations, at conditional "level":

$$B_{1234} = E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \middle| \text{First Stage Data}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}} \middle| \text{First Stage Data}\right]$$

Conditional second stage levels
($\mathcal{X}$... First stage data)

Intersection H_0	Adaptive	
$H_1 \cap H_2 \cap H_3 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
$H_1 \cap H_2 \cap H_3$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
$H_1 \cap H_2 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
$H_1 \cap H_3 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_4 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
$H_2 \cap H_3 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
$H_1 \cap H_2$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
$H_1 \cap H_3$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq \alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq \alpha\}} \mid \mathcal{X}\right]$
$H_1 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_4 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
$H_2 \cap H_3$	$E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
$H_2 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_2 \leq \alpha\}} \mid \mathcal{X}\right]$	
$H_3 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}} \mid \mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_4 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}} \mid \mathcal{X}\right]$
H_1	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq \alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq \alpha\}} \mid \mathcal{X}\right]$
H_2	$E_{H_0}\left[\mathbf{1}_{\{p_2 \leq \alpha\}} \mid \mathcal{X}\right]$	
H_3	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq \alpha\}} \mid \mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq \alpha\}} \mid \mathcal{X}\right]$
H_4	$E_{H_0}\left[\mathbf{1}_{\{p_4 \leq \alpha\}} \mid \mathcal{X}\right]$	

Conditional second stage levels

$H_1 \cap H_2 \cap H_3 \cap H_4$ ($\mathcal{X}$... First stage data)

Intersection H_0	Adaptive	Simple strategy
$H_1 \cap H_2 \cap H_3 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$
$H_1 \cap H_2 \cap H_3$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$
$H_1 \cap H_2 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$
$H_1 \cap H_3 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_4 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$
$H_2 \cap H_3 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$
$H_1 \cap H_2$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$
$H_2 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_2 \leq \alpha\}}\middle\vert\mathcal{X}\right]$	
$H_3 \cap H_4$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_4 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$
H_1	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq \alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq \alpha\}}\middle\vert\mathcal{X}\right]$
H_2	$E_{H_0}\left[\mathbf{1}_{\{p_2 \leq \alpha\}}\middle\vert\mathcal{X}\right]$	
H_3	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq \alpha\}}\middle\vert\mathcal{X}\right]$	$E_{H_0}\left[\mathbf{1}_{\{p_3 \leq \alpha\}}\middle\vert\mathcal{X}\right]$
H_4	$E_{H_0}\left[\mathbf{1}_{\{p_4 \leq \alpha\}}\middle\vert\mathcal{X}\right]$	

Adaptive	Simple strategy
$E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right] + E_{H_0}\left[\mathbf{1}_{\{p_2 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$	$\geq E_{H_0}\left[\mathbf{1}_{\{p_1 \leq 0.5\alpha\}}\middle\vert\mathcal{X}\right]$

Adaptive graph-based multiple testing procedures

- Define a pre-planned test using the graphical approach
- At interim perform adaptations based on internal or external data, *e.g.*, dropping of treatments, sample size reassessment
- Flexibility: No specific selection rule nor the number of hypotheses to be selected needs to be pre-specified.
- Use an updated graph to derive suitable second stage multiple testing procedure, *e.g.*, remove nodes of dropped treatments
- If no adaptation is performed, the pre-planned sequentially rejective test can be applied (no price has to be paid!)
- Use of conditional error principle ensures family wise error rate control

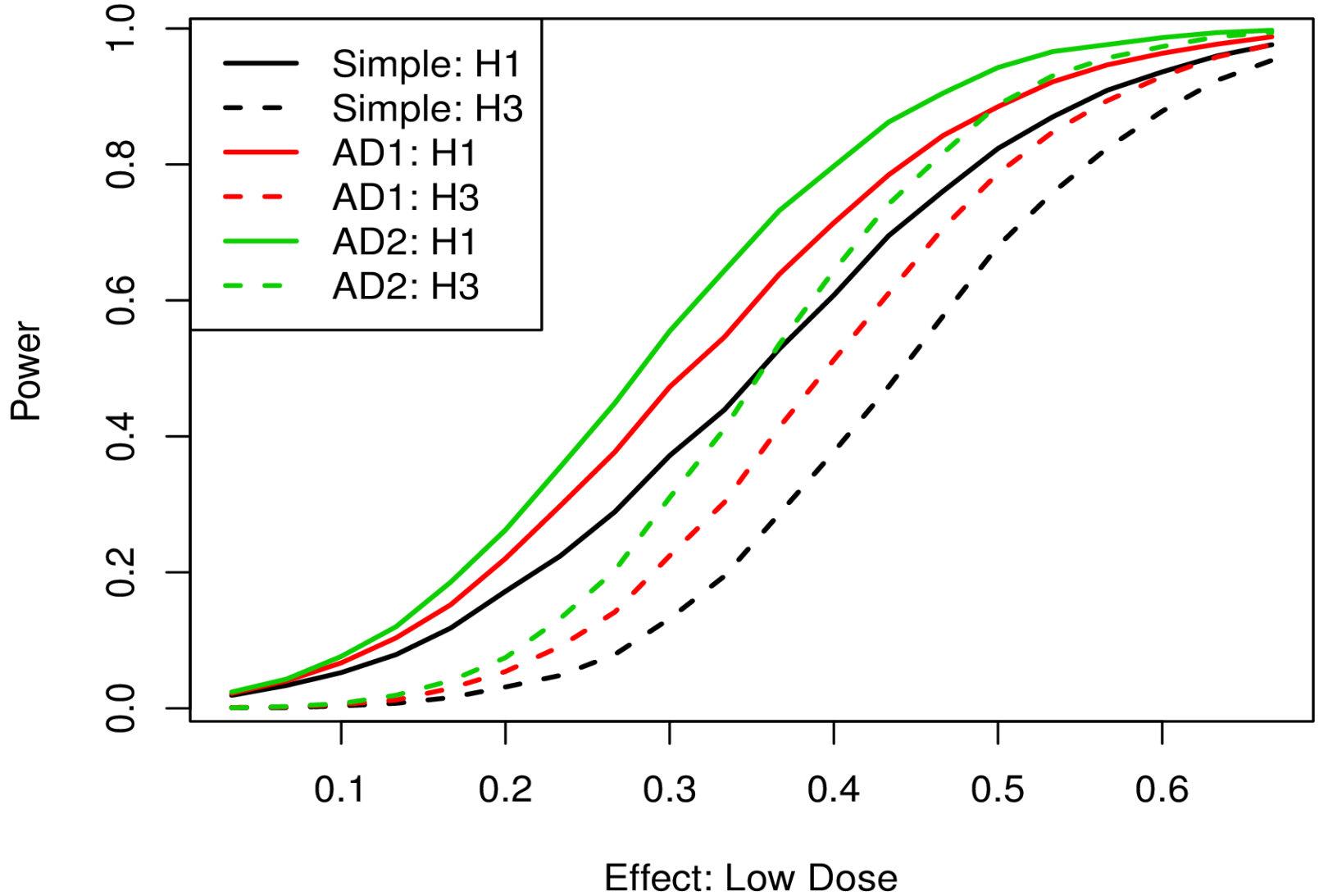
Simulation study

- Two treatments against control, two endpoints
- Common known standard deviation $\sigma = 1$
- $T1$ has effect $\frac{1}{2}\delta$, $T2$ (δ)
- Equal effect sizes in either endpoint
- Sample size $n = 80$ per treatment arm
- After half of the measurements have been collected, the T2 is dropped

Test procedures:

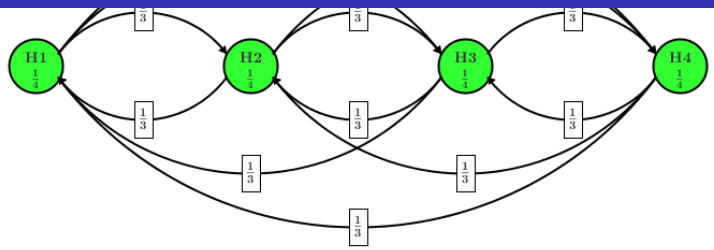
1. Simple: Retain H_2, H_4, test H_1, H_3 sequentially at level $\frac{\alpha}{2}$
2. AD1: Use adaptive test
3. AD2: Use adaptive test, and reallocate the 40 patients that would have received T2 to T1 and control

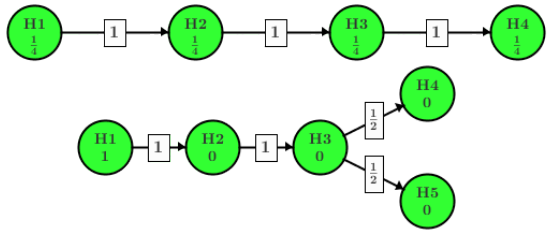
Simulation

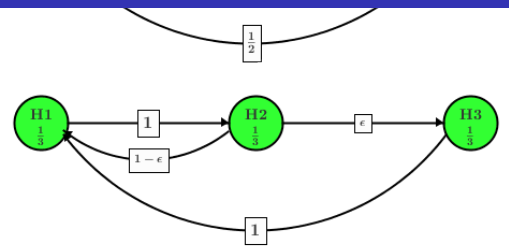

Simple: H1
Simple: H3
AD1: H1
AD1: H3
AD2: H1
AD2: H3
Power
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
0.6
Effect: Low Dose

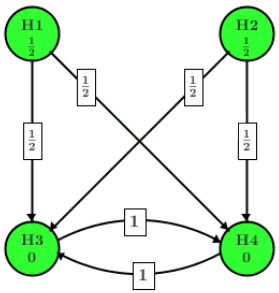
Our approach is suitable for

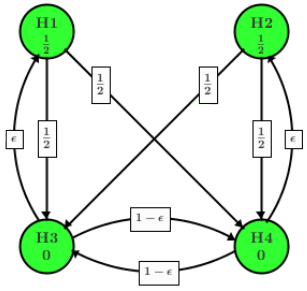
- General graphical procedures
- Treatment selection
- Subgroup selection
- Re-weighting (*e.g.*, change in priorities)
- Sample size reassessment
- Theoretically interim analyses may be unscheduled

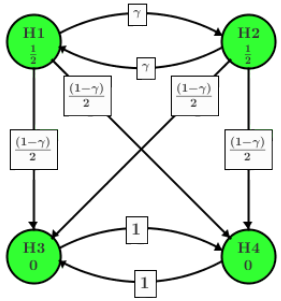

H1
H2
H3
H4

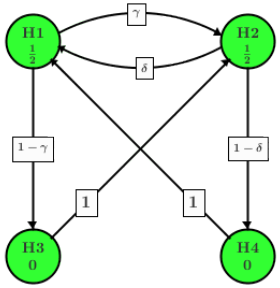

H1
H2
H3
H4
H1
H2
H3
H4
H5

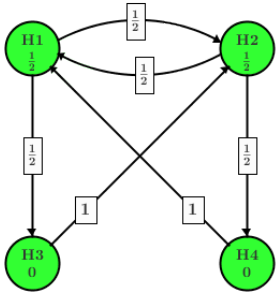

H1
H2
H3

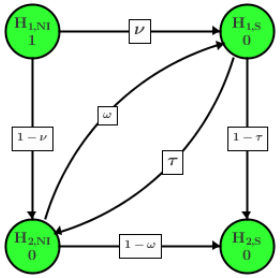

H1
H2
H3
H4

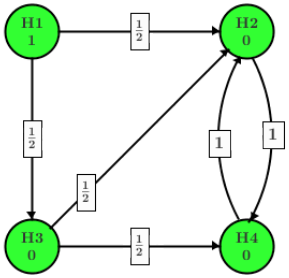

H1
H2
H3
H4

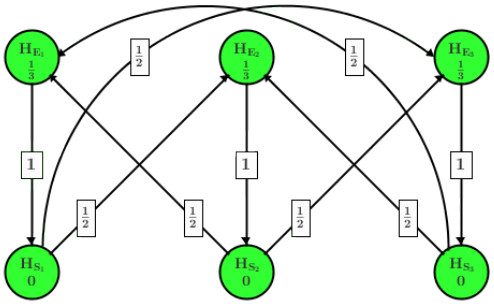

H1
H2
H3
H4

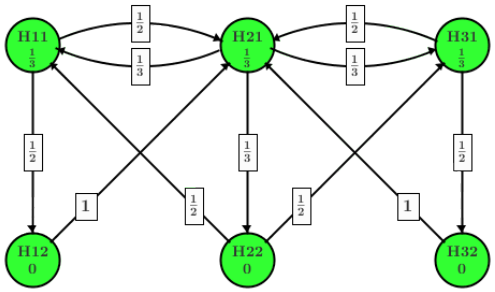

H1
H2
H3
H4

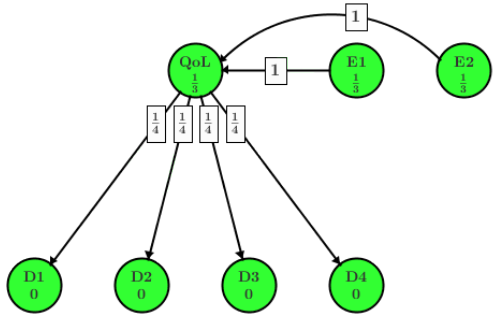

H1
H2
H3
H4

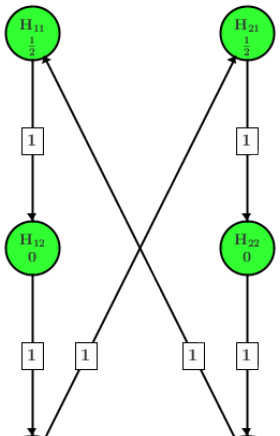

H1,NI
H1,S
H2,NI
H2,S

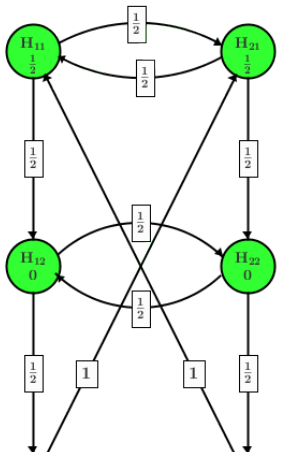

H1
H2
H3
H4

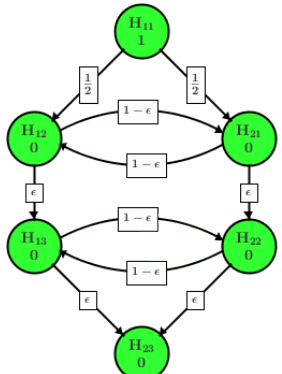

H1
H2
H3

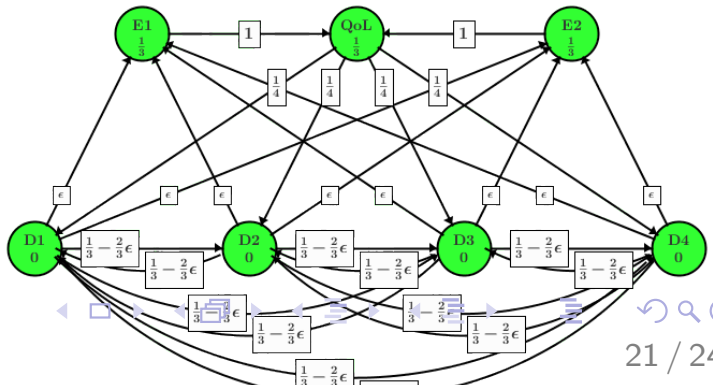

H11
H21
H31
H12
H22
H32


QoL
E1
E2
D1
D2
D3
D4


H11
H21
H12
H22


H11
H21
H12
H22


H11
H12
H21
H13
H22
H23

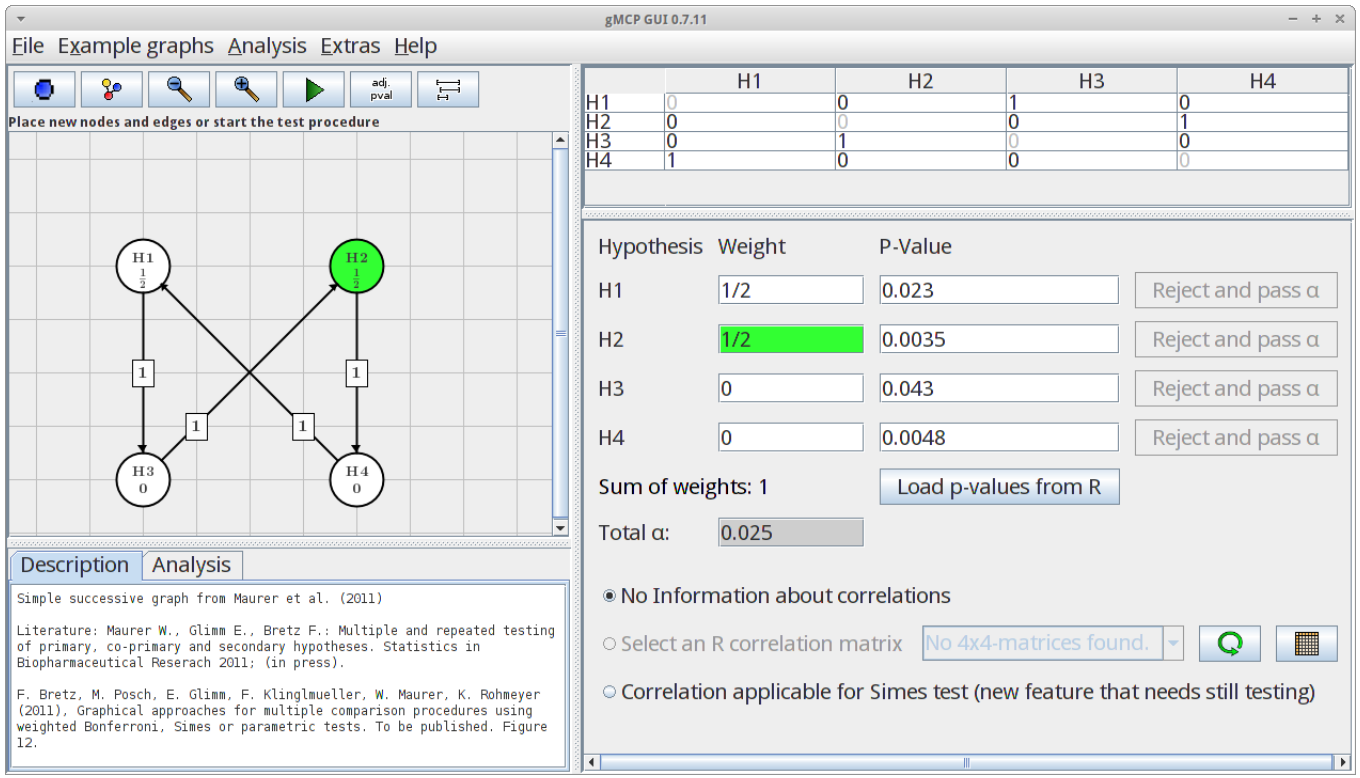

E1
QoL
E2
D1
D2
D3
D4

Extensions

- Simultaneous confidence intervals [(Magirr '12)]
- Mixed parametric procedures
- Fully sequential tests
- GNU R package gMCP

▸ To conditional error

▸ To 2^{nd}-stage levels



Discussion & Conclusion

- Intuitive way to design complex multiple testing procedures
- Covers a large class of multiple testing procedures
- The weighted directed graph completely defines the adaptive multiple testing procedure
- Flexibility to perform mid-trial adaptations based on internal or external information
- Adaptations are not restricted to dropping of hypotheses
- No assumptions on the joint distribution of test statistics across hypotheses.
- Without adaptation the pre-planned test can be performed
- Multiplicity from different sources can be adjusted for
- Strong control of the FWER

References

1. A. Elsaesser, J. Regnstrom, T. Vetter, F. Koenig, R. Hemmings, M. Greco, M. Papaluca-Amati, and M. Posch (2013). Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. *Trials*
2. Mueller H, Schaefer H, A general statistical principle for changing a design any time during the course of a trial *Stat Med* 23/16, 2497-508
3. Koenig F, Brannath W, Bretz F, Posch M, (2008) Adaptive Dunnett Tests for Treatment Selection *Stat Med* 27/10, 1612-25
4. Posch M, Futschik A (2008) A Uniform Improvement of Bonferroni-Type Tests by Sequential Tests *JASA* 103/481, 299-308
5. Bretz F, Maurer W, Brannat W, Posch M, (2009) A graphical approach to sequentially rejective multiple testing procedures. *Stat Med* 28/4, 586-604
6. Posch M, Maurer W, Bretz F (2010) Type I error rate control in adaptive designs for confirmatory clinical trials with treatment selection at interim *Pharm Stat*
7. Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K (2011) Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes or parametric tests. *Biom J* 53/6, 894-913
8. Magirr D, Jaki T, Posch M, Klinglmueller F (2013), Simultaneous confidence intervals that are compatible with closed testing in adaptive designs. *Biometrika*
9. Klinglmueller F, Posch M, Koenig F (2014), Adaptive graph-based multiple testing procedures. *Pharmaceutical Statistics*