Visualizing multiple objectives in flexible and group sequential trials

Florian Klinglmueller[1,2]

http://float.lefant.net

FWF P23167 & J3721

Acknowledgment: Martin Posch[1]**, Franz Koenig**[1]

[1] Medical University of Vienna, Austria

[2] University of Padova, Italy

Adaptive Design & Multiple Testing Workshop, Koeln, Deutschland - 2015

“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)

Graph-based multiple testing procedures

Notation

- $H_1, \ldots, H_m$: m null hypotheses.
- $p_1, \ldots, p_m$: m elementary p-values
- Initial significance level $\alpha_1, ..., \alpha_m$, $\alpha = \sum_{i=1}^{m} \alpha_i$
- Weighted directed graph with transition matrix $\mathbf{G} = (g_{ij})_{i,j \in \{1,...,m\}}$

Graph-based multiple testing

1. Reject any hypothesis for which $p_i \leq \alpha_i$
2. Reallocate significance levels $\alpha_j = \alpha_j + g_{ij}\alpha_i$
3. Update the graph
4. Goto 1

Control of the FWE

Theorem BRETZ ET. AL ('09)

The initial levels α, the graph and algorithm define a unique multiple testing procedure controlling strongly the FWE at level α.

Proof:

- The graph and algorithm define weighted Bonferroni tests for all intersection hypotheses $H_J = \bigcap_{j \in J} H_j$, $J \subset \{1, ..., m\}$ with local significance levels $\alpha_i(J)$ with $\alpha \geq \sum_{i=1}^{k} \alpha_i(J)$.
- The algorithm is a short cut for the resulting closed test.
- Provides strong control of the family wise error rate
  MARCUS '76

Extensions

- Instead of performing a weighted Bonferroni test to each intersection hypothesis other weighted tests may be used:
  1. Simes or parametric tests* BRETZ ET AL ('11)
  2. Group sequential tests MAURER & BRETZ ('13)
  3. Adaptive combination tests SUGITANI ET AL. ('13)
  4. Adaptive tests based on partial conditional error rates* KLINGLMUELLER ET AL. ('14)
- * If the intersection tests are not consonant no general short-cut exists and the whole closed test must be performed!
- Here we extend adaptive graph-based multiple testing procedures (4) to also permit early rejection for success.

Extensions

- Instead of performing a weighted Bonferroni test to each intersection hypothesis other weighted tests may be used:
  1. Simes or parametric tests* BRETZ ET AL ('11)
  2. Group sequential tests MAURER & BRETZ ('13)
  3. Adaptive combination tests SUGITANI ET AL. ('13)
  4. Adaptive tests based on partial conditional error rates* KLINGLMUELLER ET AL. ('14)
- * If the intersection tests are not consonant no general short-cut exists and the whole closed test must be performed!
- Here we extend adaptive graph-based multiple testing procedures (4) to also permit early rejection for success.

Extensions

- Instead of performing a weighted Bonferroni test to each intersection hypothesis other weighted tests may be used:
  1. Simes or parametric tests* BRETZ ET AL ('11)
  2. Group sequential tests MAURER & BRETZ ('13)
  3. Adaptive combination tests SUGITANI ET AL. ('13)
  4. Adaptive tests based on partial conditional error rates* KLINGLMUELLER ET AL. ('14)
- * If the intersection tests are not consonant no general short-cut exists and the whole closed test must be performed!
- Here we extend adaptive graph-based multiple testing procedures (4) to also permit early rejection for success.

Example: 2 primary, 2 secondary hypotheses

New drug for the treatment of multiple sclerosis

- Two active treatment arms (high dose given once per day, low dose given 3 times per day), one placebo control arm
- Primary endpoint annualized relapse rate: H_1, H_2
- Secondary endpoint number of lesions in the brain: H_3, H_4

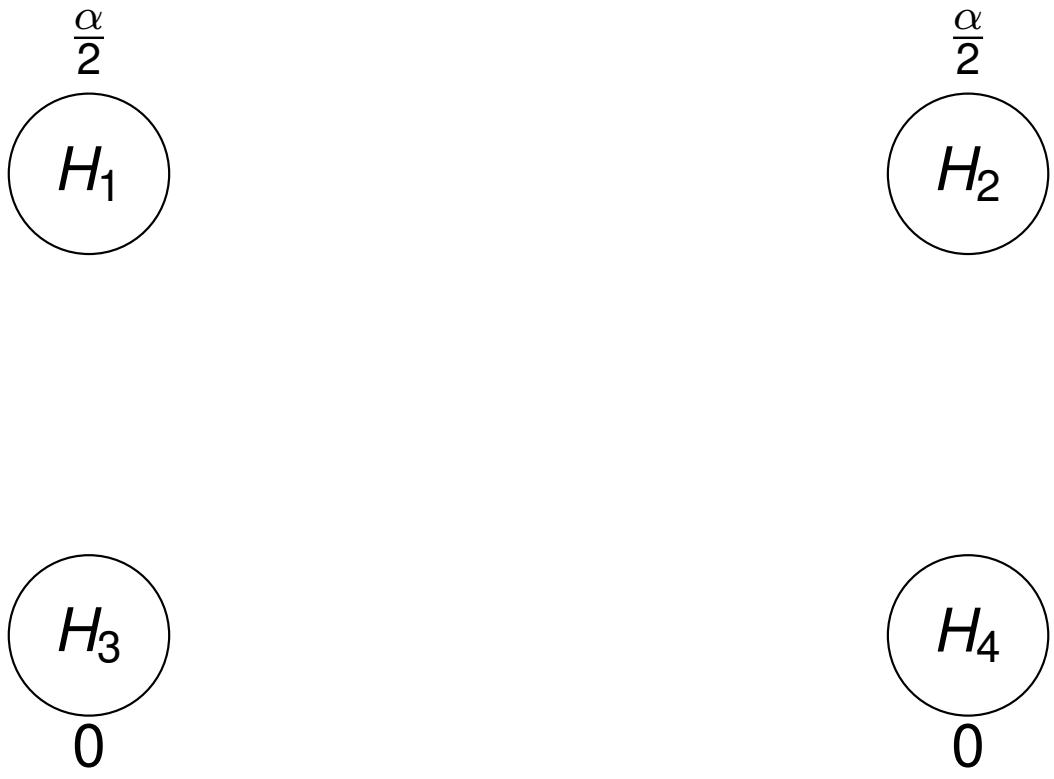
Testing Strategy

- Rejection of secondary hypotheses is only of interest if at least one of the primary hypotheses can be rejected
- Assuming equal efficacy the two treatments should have same probability of success.

Example: Tailoring the procedure

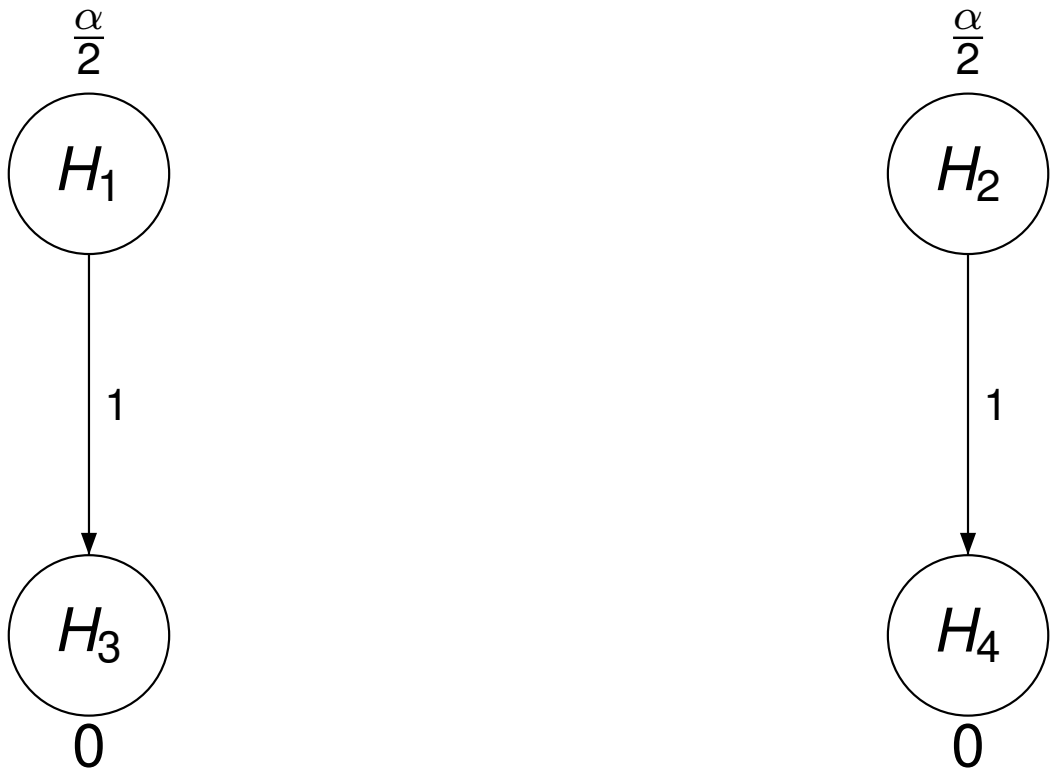
1. Split α equally between primary hypotheses
2. Give no α to secondary hypotheses
3. Reallocate significance levels to secondary hypotheses
4. Reallocate significance levels between treatment arms

Example: Tailoring the procedure



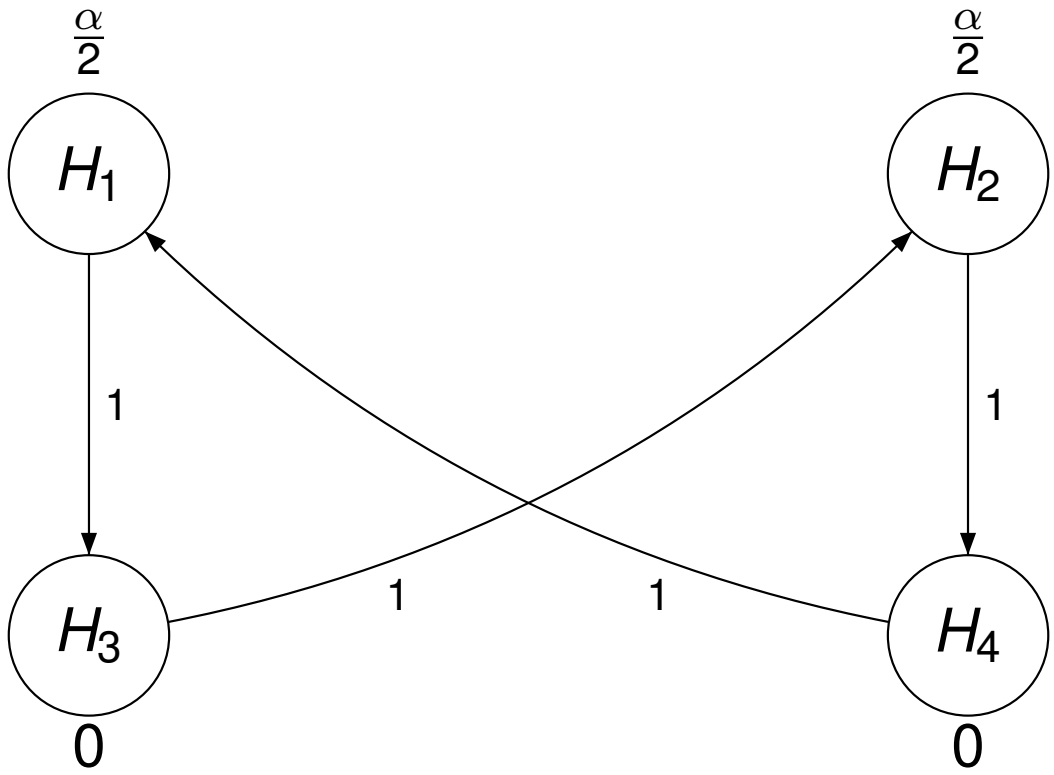
1. Split α equally between primary hypotheses
2. Give no α to secondary hypotheses
3. Reallocate significance levels to secondary hypotheses
4. Reallocate significance levels between treatment arms

Example: Tailoring the procedure



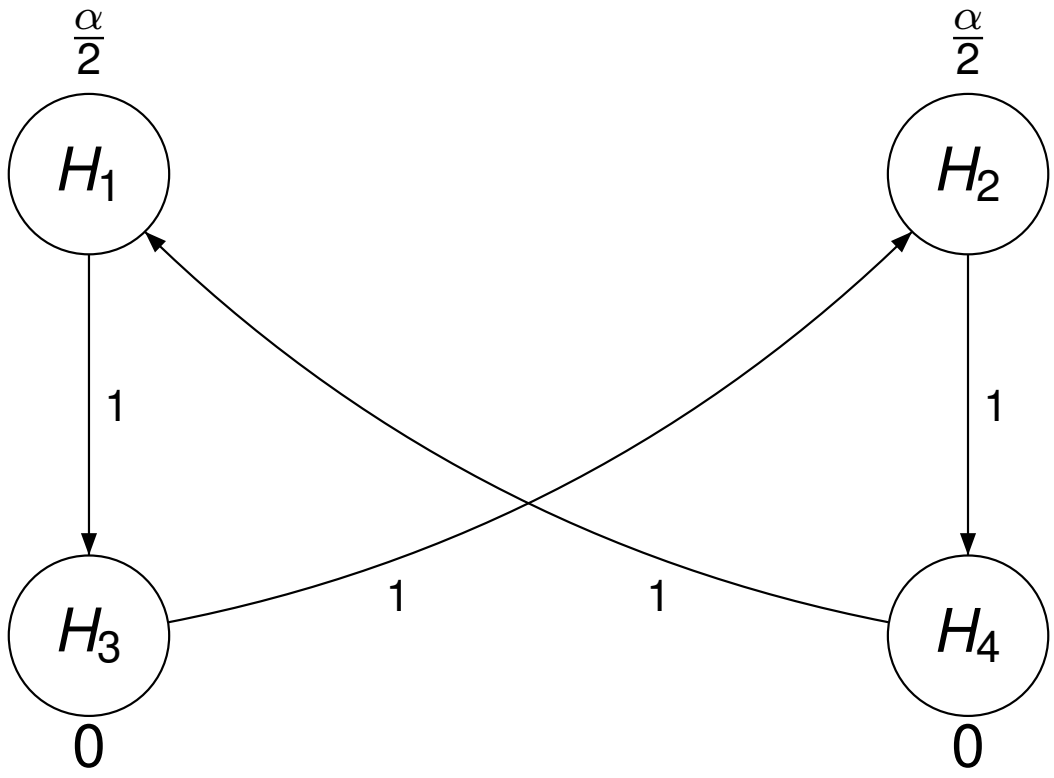
1. Split α equally between primary hypotheses
2. Give no α to secondary hypotheses
3. Reallocate significance levels to secondary hypotheses
4. Reallocate significance levels between treatment arms

Example: Tailoring the procedure



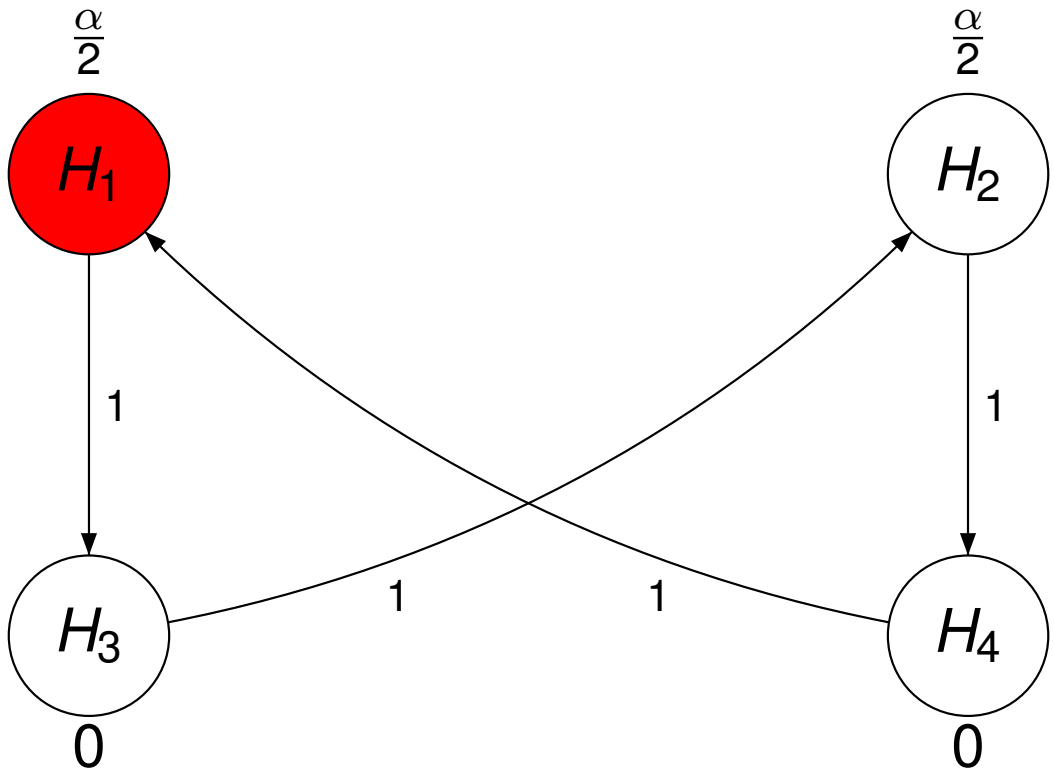
1. Split α equally between primary hypotheses
2. Give no α to secondary hypotheses
3. Reallocate significance levels to secondary hypotheses
4. Reallocate significance levels between treatment arms

Example: fixed sample test, $\alpha = 0.025$



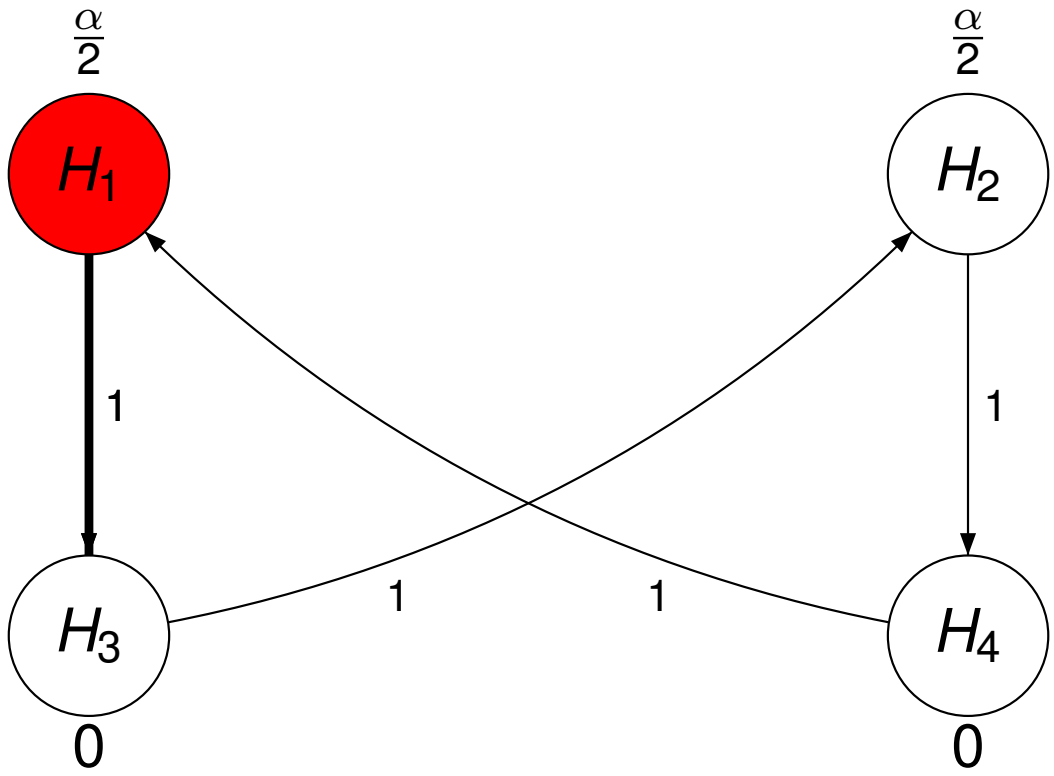
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1,

Example: fixed sample test, $\alpha = 0.025$



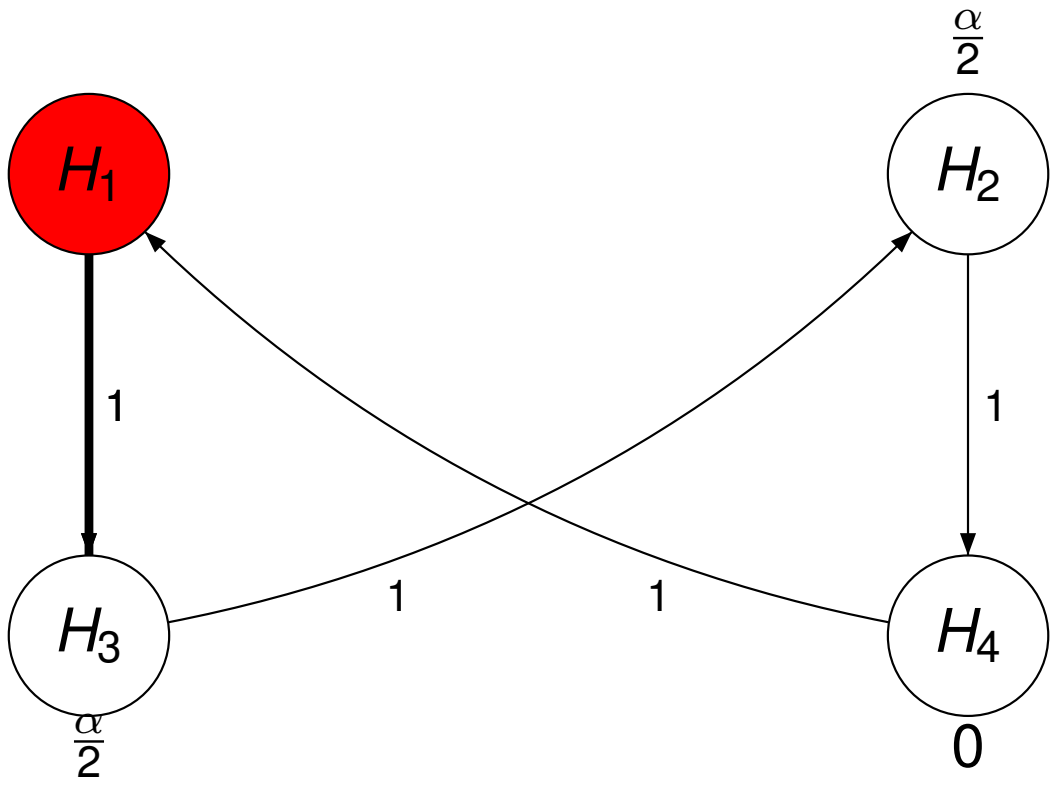
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1,

Example: fixed sample test, $\alpha = 0.025$



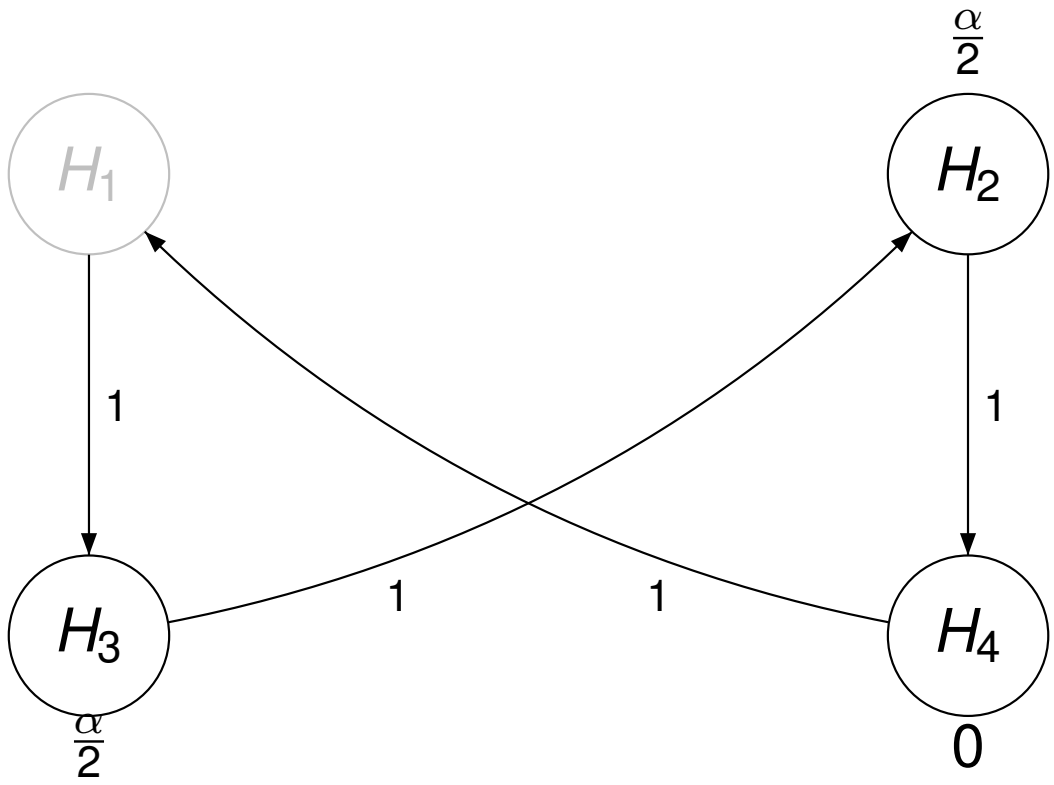
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1,

Example: fixed sample test, $\alpha = 0.025$



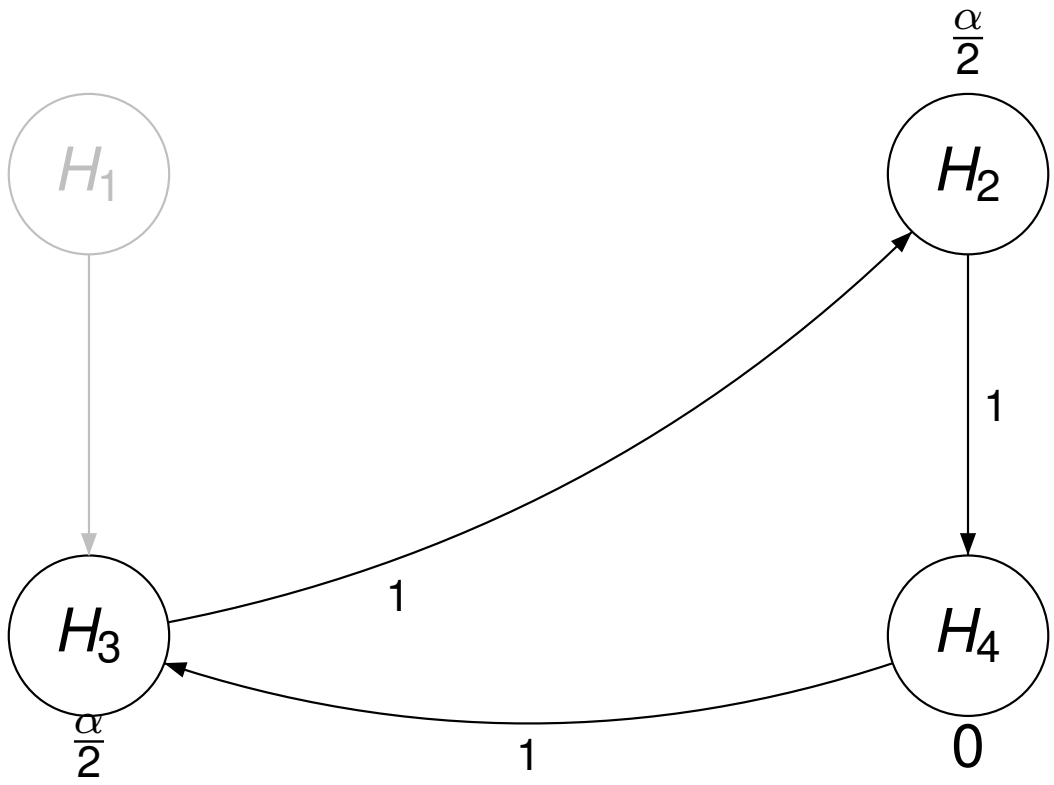
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1,

Example: fixed sample test, $\alpha = 0.025$



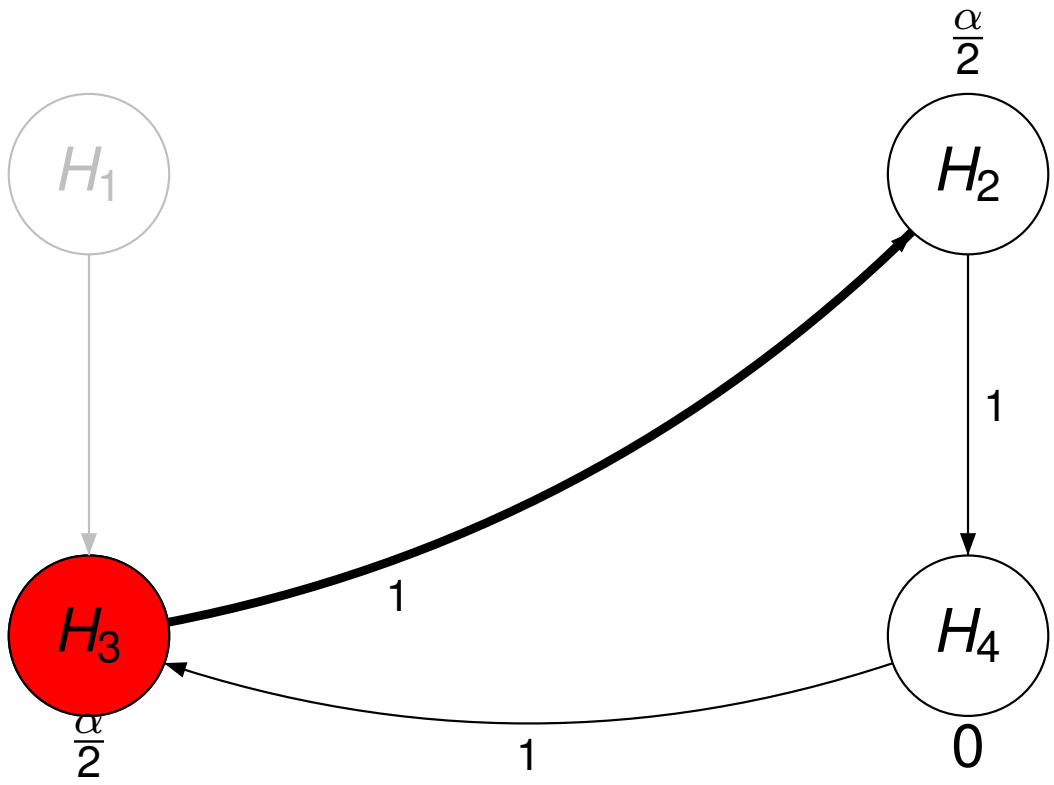
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1,

Example: fixed sample test, $\alpha = 0.025$



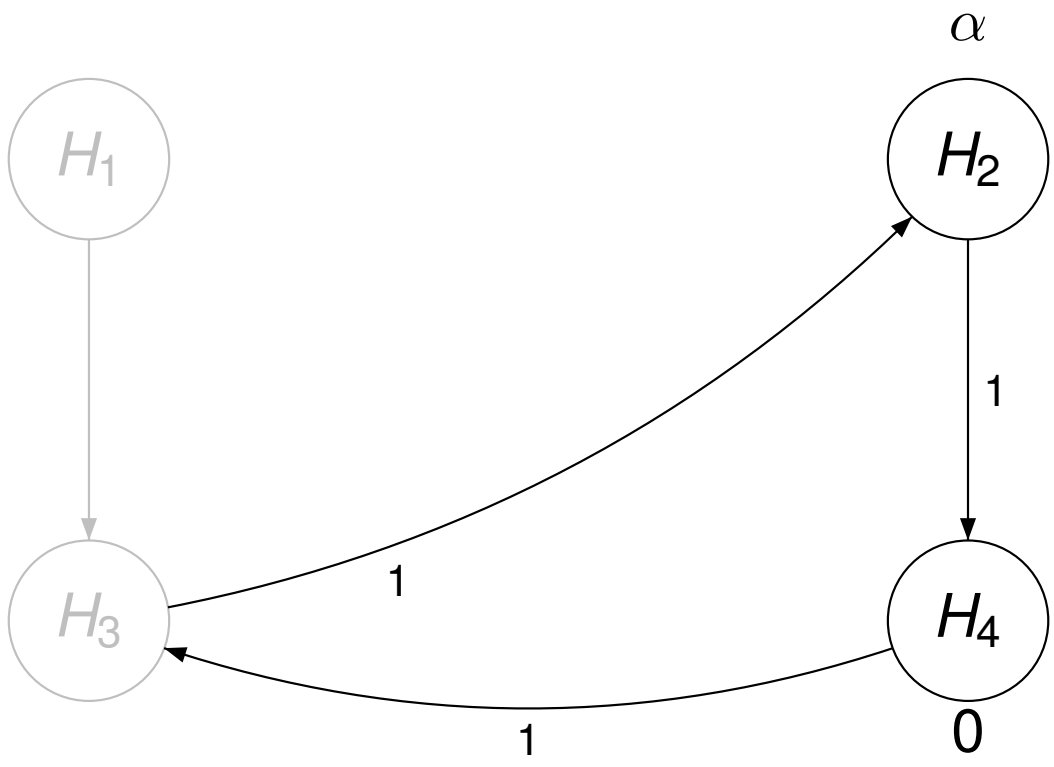
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1,

Example: fixed sample test, $\alpha = 0.025$



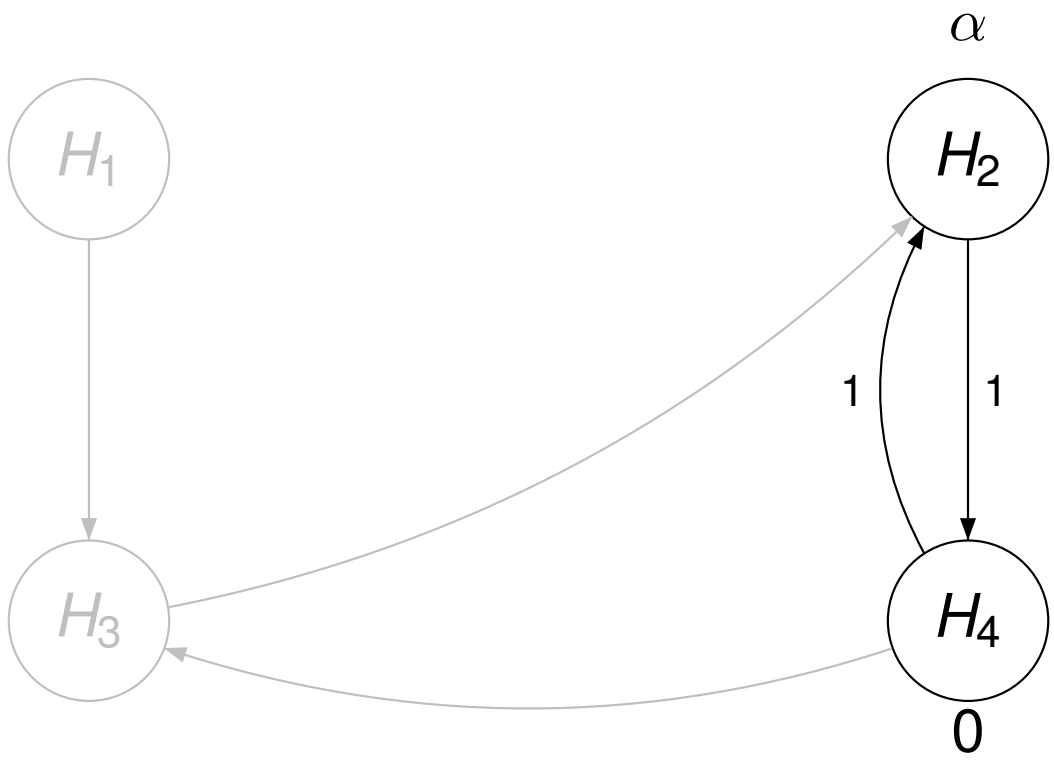
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1, then H_3,

Example: fixed sample test, $\alpha = 0.025$



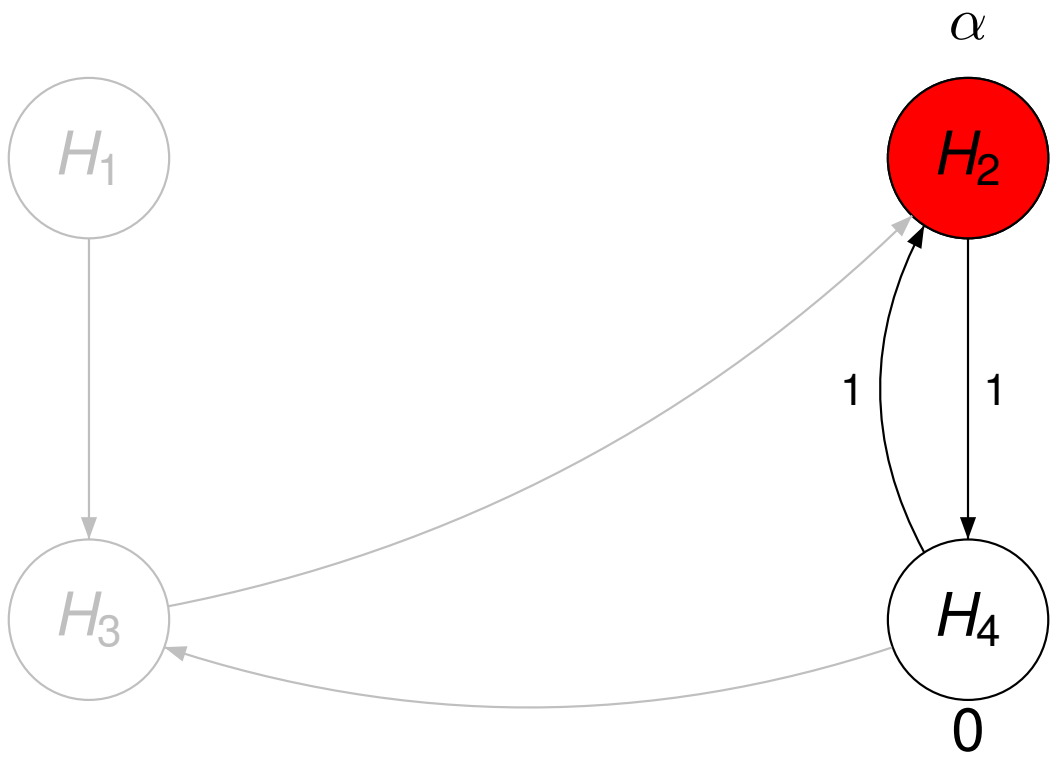
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1, then H_3,

Example: fixed sample test, $\alpha = 0.025$



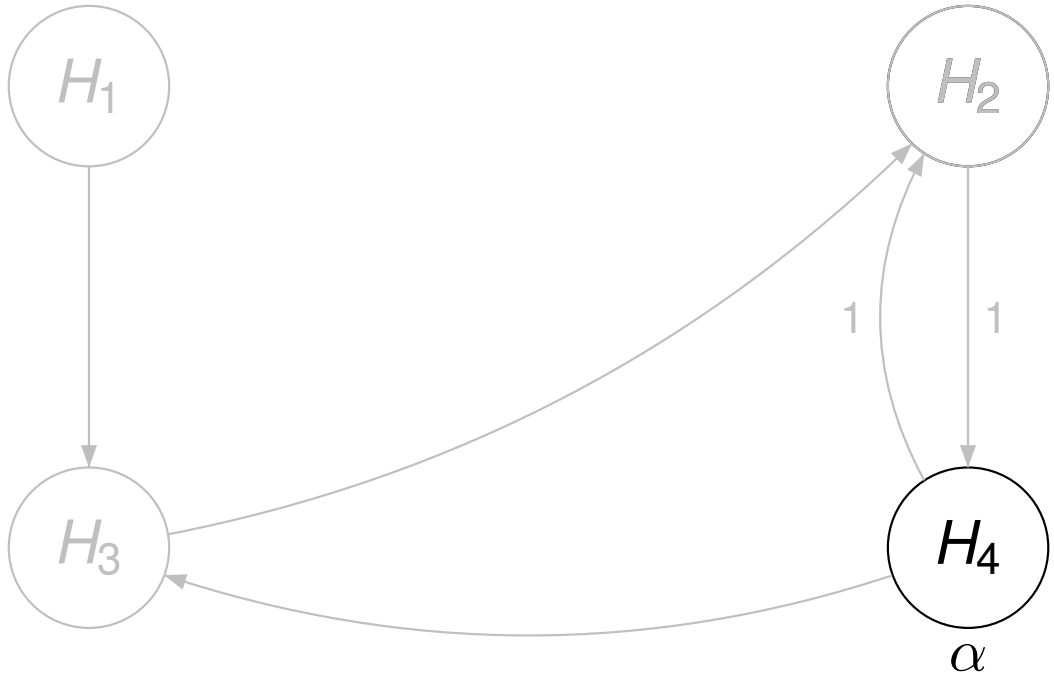
1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1, then H_3,

Example: fixed sample test, $\alpha = 0.025$



1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1, then H_3, finally H_2,

Example: fixed sample test, $\alpha = 0.025$



1. Perform the trial with a fixed sample size of N patients/arm
2. Control FWE at $\alpha = .025$ one-sided
3. Observe: $p_1 = 0.004$, $p_2 = 0.017$, $p_3 = 0.011$, $p_4 = 0.032$
4. Reject H_1, then H_3, finally H_2, but not H_4.

Group sequential design

- Instead of performing the trial with a fixed sample size, we perform two pre-planned interim analyses.
- After $1/3 * N$ and $2/3 * N$ observations per group have been collected, data will be unblinded.
- Hypotheses are rejected early if interim test statistics (cumulative p-values $p_i^{(k)}$) cross an early rejection boundary ($\text{gsB}_i^{(k)}(\alpha_i)$).
- Use O'Brien Fleming type boundaries (O'BF) for the primary hypotheses, Pocock (Poc) type boundaries for the secondary hypotheses. (GLIMM ET AL. (2010))
- Using these types of boundaries we can apply the graphical approach at each interim analysis. MAURER & BRETZ (2013).

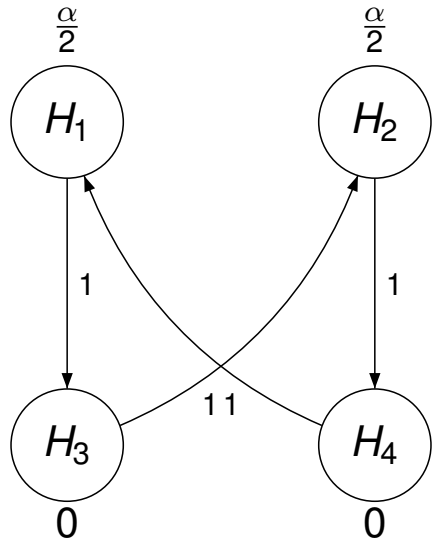
Group sequential boundaries

- We compute boundaries for all different local significance levels that can appear in the graph to get critical values (adjusted significance levels) for each interim/final analysis.
- Boundaries are computed based on Poc/O'BF type-spending function using R-package `ldbounds`

Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011

Example: group sequential design - Stage 1

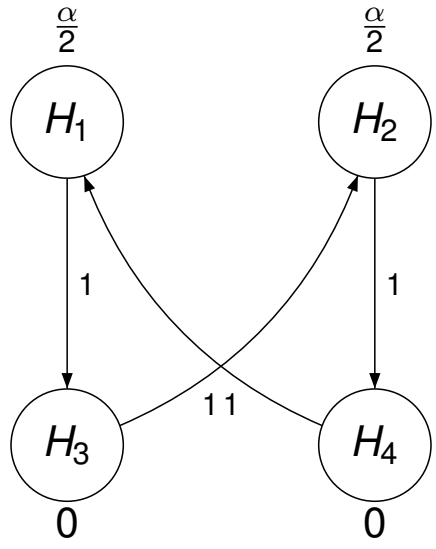
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



1. Trial with three equal sized stages and at most N patients per group.
2. After Stage 1 (Stage 2) we have $N/3$ ($2N/3$) observations per arm.
3. Interim 1 observe: $p_1^{(1)} = 0.09$, $p_2^{(1)} = 0.17$, $p_3^{(1)} = 0.13$, $p_4^{(1)} = 0.04$
4. Interim 2 observe: $p_1^{(2)} = 0.001$, $p_2^{(2)} = 0.07$, $p_3^{(2)} = 0.021$, $p_4^{(2)} = 0.22$

Example: group sequential design - Stage 2

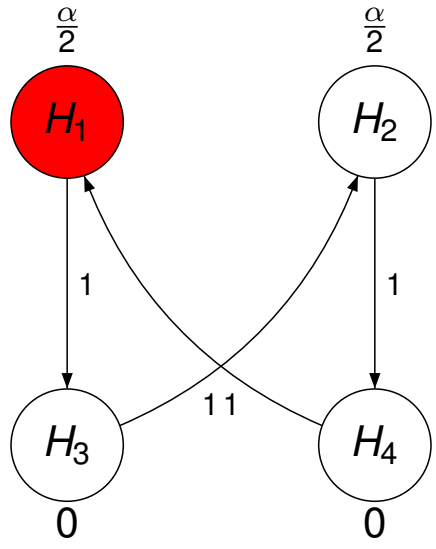
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



1. Trial with three equal sized stages and at most N patients per group.
2. After Stage 1 (Stage 2) we have $N/3$ ($2N/3$) observations per arm.
3. Interim 1 observe: $p_1^{(1)} = 0.09$, $p_2^{(1)} = 0.17$, $p_3^{(1)} = 0.13$, $p_4^{(1)} = 0.04$
4. Interim 2 observe: $p_1^{(2)} = 0.001$, $p_2^{(2)} = 0.07$, $p_3^{(2)} = 0.021$, $p_4^{(2)} = 0.22$

Example: group sequential design - Stage 2

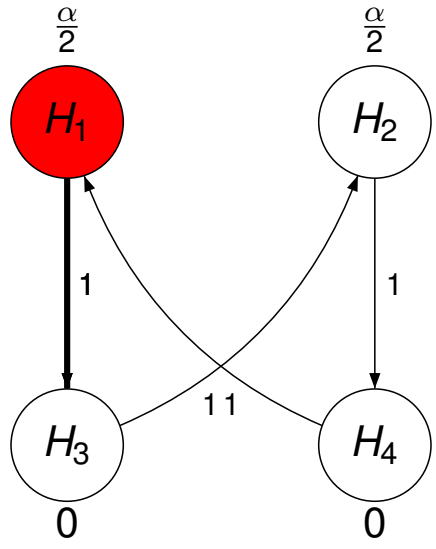
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



1. Trial with three equal sized stages and at most N patients per group.
2. After Stage 1 (Stage 2) we have $N/3$ ($2N/3$) observations per arm.
3. Interim 1 observe: $p_1^{(1)} = 0.09$, $p_2^{(1)} = 0.17$, $p_3^{(1)} = 0.13$, $p_4^{(1)} = 0.04$
4. Interim 2 observe: $p_1^{(2)} = 0.001$, $p_2^{(2)} = 0.07$, $p_3^{(2)} = 0.021$, $p_4^{(2)} = 0.22$

Example: group sequential design - Stage 2

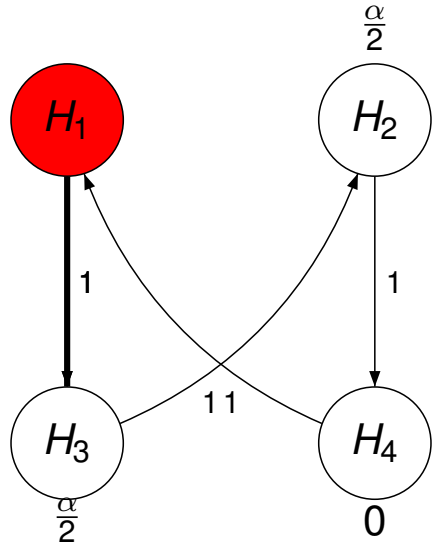
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



1. Trial with three equal sized stages and at most N patients per group.
2. After Stage 1 (Stage 2) we have $N/3$ ($2N/3$) observations per arm.
3. Interim 1 observe: $p_1^{(1)} = 0.09$, $p_2^{(1)} = 0.17$, $p_3^{(1)} = 0.13$, $p_4^{(1)} = 0.04$
4. Interim 2 observe: $p_1^{(2)} = 0.001$, $p_2^{(2)} = 0.07$, $p_3^{(2)} = 0.021$, $p_4^{(2)} = 0.22$

Example: group sequential design - Stage 2

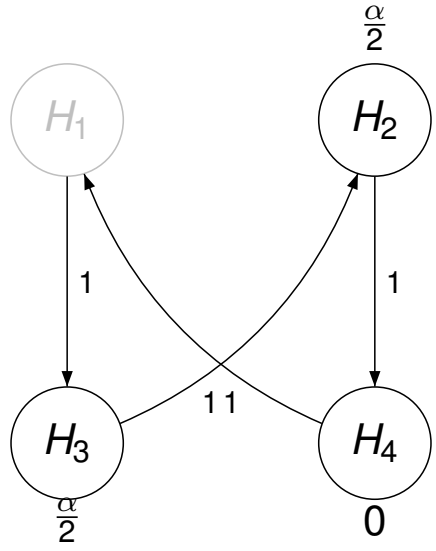
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



1. Trial with three equal sized stages and at most N patients per group.
2. After Stage 1 (Stage 2) we have $N/3$ ($2N/3$) observations per arm.
3. Interim 1 observe: $p_1^{(1)} = 0.09$, $p_2^{(1)} = 0.17$, $p_3^{(1)} = 0.13$, $p_4^{(1)} = 0.04$
4. Interim 2 observe: $p_1^{(2)} = 0.001$, $p_2^{(2)} = 0.07$, $p_3^{(2)} = 0.021$, $p_4^{(2)} = 0.22$

Example: group sequential design - Stage 2

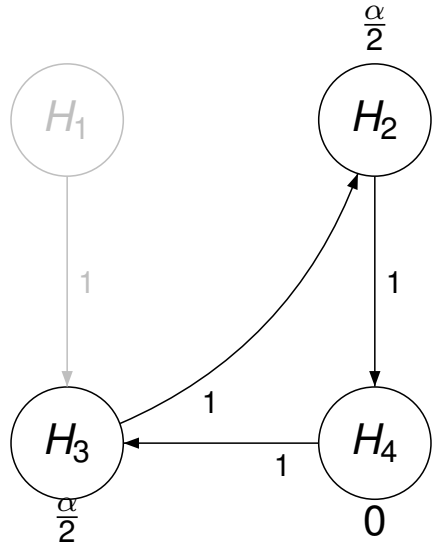
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



1. Trial with three equal sized stages and at most N patients per group.
2. After Stage 1 (Stage 2) we have $N/3$ ($2N/3$) observations per arm.
3. Interim 1 observe: $p_1^{(1)} = 0.09$, $p_2^{(1)} = 0.17$, $p_3^{(1)} = 0.13$, $p_4^{(1)} = 0.04$
4. Interim 2 observe: $p_1^{(2)} = 0.001$, $p_2^{(2)} = 0.07$, $p_3^{(2)} = 0.021$, $p_4^{(2)} = 0.22$

Example: group sequential design - Stage 2

Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



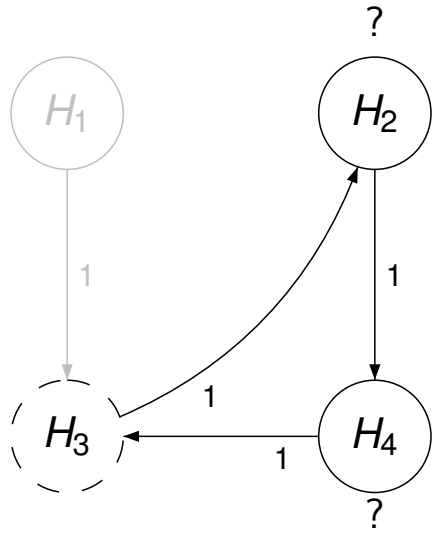
1. Trial with three equal sized stages and at most N patients per group.
2. After Stage 1 (Stage 2) we have $N/3$ ($2N/3$) observations per arm.
3. Interim 1 observe: $p_1^{(1)} = 0.09$, $p_2^{(1)} = 0.17$, $p_3^{(1)} = 0.13$, $p_4^{(1)} = 0.04$
4. Interim 2 observe: $p_1^{(2)} = 0.001$, $p_2^{(2)} = 0.07$, $p_3^{(2)} = 0.021$, $p_4^{(2)} = 0.22$

Comments on group sequential design

- Having rejected the primary hypothesis H_1 early, we can either:
  1. Continue recruiting to the corresponding treatment arm;
  2. Stop recruiting to the corresponding treatment arm;
- First option means there are little savings from early rejection as the full number of patients has to be recruited
- First option also means that one could gather second stage observations for the primary endpoint - which may not be consistent with the first stage outcome
- Second option abandons any chance to reject the secondary hypothesis H_3
- Second option also means that local significance levels of H_1 and H_3 cannot be allocated to H_2 and subsequently H_4

Example: final analysis

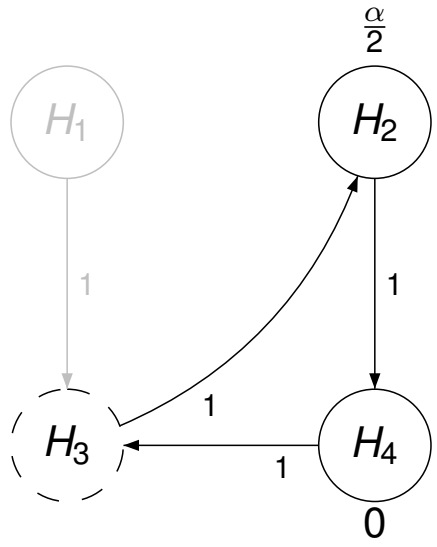
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



- If we knew that we can not reject H_3, at the end we had to test H_2 using O'BF boundary at level 0.0125 (i.e. 0.0118).
- If we knew that we can reject H_3 we could reallocate the α and test H_2 using O'BF boundary at level 0.025.
- We extend adaptive graph-based multiple testing procedures (KLINGLMUELLER ET AL. (2014)) to offer an intermediate solution

Example: final analysis

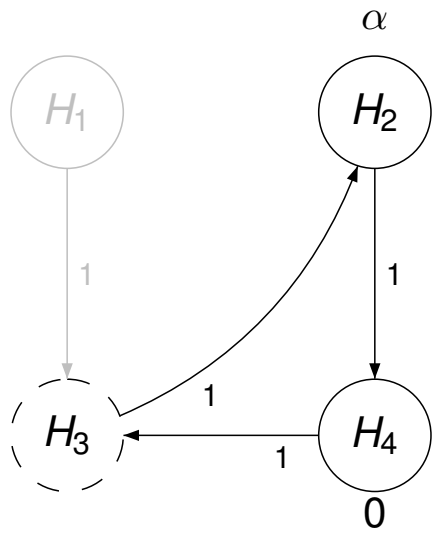
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



- If we knew that we can not reject H_3, at the end we had to test H_2 using O'BF boundary at level 0.0125 (i.e. 0.0118).
- If we knew that we can reject H_3 we could reallocate the α and test H_2 using O'BF boundary at level 0.025.
- We extend adaptive graph-based multiple testing procedures (KLINGLMUELLER ET AL. (2014)) to offer an intermediate solution

Example: final analysis

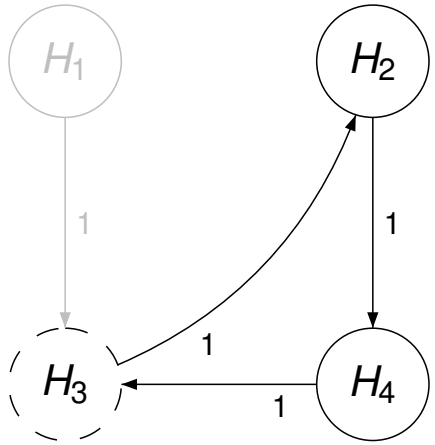
Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



- If we knew that we can not reject H_3, at the end we had to test H_2 using O'BF boundary at level 0.0125 (i.e. 0.0118).
- If we knew that we can reject H_3 we could reallocate the α and test H_2 using O'BF boundary at level 0.025.
- We extend adaptive graph-based multiple testing procedures (KLINGLMUELLER ET AL. (2014)) to offer an intermediate solution

Example: final analysis

Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.0125	0.00002	0.0022	0.0118
Pocock	0.0125	0.0057	0.0052	0.0050
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011



- If we knew that we can not reject H_3, at the end we had to test H_2 using O'BF boundary at level 0.0125 (i.e. 0.0118).
- If we knew that we can reject H_3 we could reallocate the α and test H_2 using O'BF boundary at level 0.025.
- We extend adaptive graph-based multiple testing procedures (KLINGLMUELLER ET AL. (2014)) to offer an intermediate solution

Adaptive graph-based multiple testing (ctd.)

- Assume we stop recruiting for Treatment 1
- We use the Stage 1 and 2 observations to compute *partial conditional error rates*.

$$A_j(J) = P_{H_J}\left[p_j^{(3)} \leq \text{gsB}_i^{(k)}\left(\alpha_i(J)\right) | \text{First-,Secondstage Data}\right]$$

- The sum $B_J = \sum_{j \in J} A_j(J)$ provides a conditional 'level' for an adapted test of H_J based on (independent) third stage data (KLINGLMUELLER ET AL. (2014)).
- In the final analysis apply a weighted Bonferroni test at level B_J to H_J based on Stage 3 data alone.
- We can also modify the Stage 3 design - e.g. reallocate patients of the dropped treatment arm to Treatment 2.
- Caveat: In general we need to construct adapted tests for all $2^m - 1$ intersection hypotheses $H_J = \bigcap_{i \in J} H_i$, $J \subseteq \{1, ..., m\}$

Adaptive graph-based multiple testing (ctd.)

- Assume we stop recruiting for Treatment 1
- We use the Stage 1 and 2 observations to compute *partial conditional error rates*.

$$A_j(J) = P_{H_J}\left[p_j^{(3)} \leq \text{gsB}_i^{(k)}\left(\alpha_i(J)\right) | \text{First-,Secondstage Data}\right]$$

- The sum $B_J = \sum_{j \in J} A_j(J)$ provides a conditional 'level' for an adapted test of H_J based on (independent) third stage data (KLINGLMUELLER ET AL. (2014)).
- In the final analysis apply a weighted Bonferroni test at level B_J to H_J based on Stage 3 data alone.
- We can also modify the Stage 3 design - e.g. reallocate patients of the dropped treatment arm to Treatment 2.
- Caveat: In general we need to construct adapted tests for all $2^m - 1$ intersection hypotheses $H_J = \bigcap_{i \in J} H_i$, $J \subseteq \{1, ..., m\}$

Adaptive graph-based multiple testing (ctd.)

- Assume we stop recruiting for Treatment 1
- We use the Stage 1 and 2 observations to compute *partial conditional error rates*.

$$A_j(J) = P_{H_J}\left[p_j^{(3)} \leq \text{gsB}_i^{(k)}\left(\alpha_i(J)\right) | \text{First-,Secondstage Data}\right]$$

- The sum $B_J = \sum_{j \in J} A_j(J)$ provides a conditional 'level' for an adapted test of H_J based on (independent) third stage data (KLINGLMUELLER ET AL. (2014)).
- In the final analysis apply a weighted Bonferroni test at level B_J to H_J based on Stage 3 data alone.
- We can also modify the Stage 3 design - e.g. reallocate patients of the dropped treatment arm to Treatment 2.
- Caveat: In general we need to construct adapted tests for all $2^m - 1$ intersection hypotheses $H_J = \bigcap_{i \in J} H_i$, $J \subseteq \{1, ..., m\}$

Local significance levels for final closed test

	$A_1(J)$	$A_2(J)$	$A_3(J)$	$A_4(J)$	B_J
$H_{\{4\}}$	0	0	0	0.002	0.002
$H_{\{3\}}$	0	0	0.135	0	0.135
$H_{\{3,4\}}$	0	0	0.057	0.0003	0.057
$H_{\{2\}}$	0	0.086	0	0	0.086
$H_{\{2,4\}}$	0	0.086	0	0	0.086
$H_{\{2,3\}}$	0	0.033	0.057	0	0.09
$H_{\{2,3,4\}}$	0	0.033	0.057	0	0.09
$H_{\{1\}}$	1	0	0	0	1
$H_{\{1,4\}}$	1	0	0	0.0003	1
$H_{\{1,3\}}$	1	0	0	0	1
$H_{\{1,3,4\}}$	1	0	0	0.0003	1
$H_{\{1,2\}}$	1	0.033	0	0	1.033
$H_{\{1,2,4\}}$	1	0.033	0	0	1.033
$H_{\{1,2,3\}}$	1	0.033	0	0	1.033
$H_{\{1,2,3,4\}}$	1	0.033	0	0	1.033

- Partial conditional error rates $A_j(J)$ and sums B_J for all intersection hypotheses

Local significance levels for final closed test

	$A_1(J)$	$A_2(J)$	$A_3(J)$	$A_4(J)$	B_J
$H_{\{4\}}$	0	0	0	0.002	0.002
$H_{\{3\}}$	0	0	0.135	0	0.135
$H_{\{3,4\}}$	0	0	0.057	0.0003	0.057
$H_{\{2\}}$	0	0.086	0	0	0.086
$H_{\{2,4\}}$	0	0.086	0	0	0.086
$H_{\{2,3\}}$	0	0.033	0.057	0	0.09
$H_{\{2,3,4\}}$	0	0.033	0.057	0	0.09
$H_{\{1\}}$	1	0	0	0	1
$H_{\{1,4\}}$	1	0	0	0.0003	1
$H_{\{1,3\}}$	1	0	0	0	1
$H_{\{1,3,4\}}$	1	0	0	0.0003	1
$H_{\{1,2\}}$	1	0.033	0	0	1.033
$H_{\{1,2,4\}}$	1	0.033	0	0	1.033
$H_{\{1,2,3\}}$	1	0.033	0	0	1.033
$H_{\{1,2,3,4\}}$	1	0.033	0	0	1.033

- Partial conditional error rates $A_j(J)$ and sums B_J for all intersection hypotheses
- All intersections with H_1 have been rejected at interim

Local significance levels for final closed test

	$A_1(J)$	$A_2(J)$	$A_3(J)$	$A_4(J)$	B_J
$H_{\{4\}}$	0	0	0	0.002	0.002
$H_{\{3\}}$	0	0	0.135	0	0.135
$H_{\{3,4\}}$	0	0	0.057	0.0003	0.057
$H_{\{2\}}$	0	0.086	0	0	0.086
$H_{\{2,4\}}$	0	0.086	0	0	0.086
$H_{\{2,3\}}$	0	0.033	0.057	0	0.09
$H_{\{2,3,4\}}$	0	0.033	0.057	0	0.09

- Partial conditional error rates $A_j(J)$ and sums B_J for all intersection hypotheses
- All intersections with H_1 have been rejected at interim

Local significance levels for final closed test

	$A_1(J)$	$A_2(J)$	$A_3(J)$	$A_4(J)$	B_J
$H_{\{4\}}$	0	0	0	0.002	0.002
$H_{\{3\}}$	0	0	0.135	0	0.135
$H_{\{3,4\}}$	0	0	0.057	0.0003	0.057
$H_{\{2\}}$	0	0.086	0	0	0.086
$H_{\{2,4\}}$	0	0.086	0	0	0.086
$H_{\{2,3\}}$	0	0.033	0.057	0	0.09
$H_{\{2,3,4\}}$	0	0.033	0.057	0	0.09

- Partial conditional error rates $A_j(J)$ and sums B_J for all intersection hypotheses
- All intersections with H_1 have been rejected at interim
- We have no Stage 3 data for H_3

Local significance levels for final closed test

	$A_1(J)$	$A_2(J)$	$A_3(J)$	$A_4(J)$	B_J
$H_{\{4\}}$	0	0	0	0.002	0.002
$H_{\{3,4\}}$	0	0	0.057	0.0003	0.057
$H_{\{2\}}$	0	0.086	0	0	0.086
$H_{\{2,4\}}$	0	0.086	0	0	0.086
$H_{\{2,3\}}$	0	0.033	0.057	0	0.09
$H_{\{2,3,4\}}$	0	0.033	0.057	0	0.09

- Partial conditional error rates $A_j(J)$ and sums B_J for all intersection hypotheses
- All intersections with H_1 have been rejected at interim
- We have no Stage 3 data for H_3
- We allocate partial conditional error rate of H_3 to H_2 and test the corresponding intersections at level B_J
- Note that since $B_{\{2,3,4\}} > B_{\{2\}}$ there is no penalty for not rejecting H_3
- Also test $H_{\{3,4\}}$ at full level $B_{\{3,4\}}$

Local significance levels for final closed test

	$A_1(J)$	$A_2(J)$	$A_3(J)$	$A_4(J)$	B_J
$H_{\{4\}}$	0	0	0	0.002	0.002
$H_{\{3,4\}}$	0	0	0.057	0.0003	0.057
$H_{\{2\}}$	0	0.086	0	0	0.086
$H_{\{2,4\}}$	0	0.086	0	0	0.086
$H_{\{2,3\}}$	0	0.033	0.057	0	0.09
$H_{\{2,3,4\}}$	0	0.033	0.057	0	0.09

- Partial conditional error rates $A_j(J)$ and sums B_J for all intersection hypotheses
- All intersections with H_1 have been rejected at interim
- We have no Stage 3 data for H_3
- We allocate partial conditional error rate of H_3 to H_2 and test the corresponding intersections at level B_J
- Note that since $B_{\{2,3,4\}} > B_{\{2\}}$ there is no penalty for not rejecting H_3
- Also test $H_{\{3,4\}}$ at full level $B_{\{3,4\}}$

Example: final analysis (ctd.)

- Reject H_2 if the p-value of an adapted test using only Stage 3 data falls below 0.087
- Reject H_4 if the p-value of an adapted test using only Stage 3 data false below 0.002
- Assuming standard normal observations and no sample size increase this corresponds to critical boundaries 0.0231 for $p_2^{(3)}$ and 0.011 for $p_4^{(3)}$ - which are just the O'BF and POC boundaries for Stage 3 at level α

Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011

- Assume we observe $p_2^{(3)} = 0.021$ and $p_4^{(3)} = 0.047$
- Then we reject H_2 but not H_4.

Example: final analysis (ctd.)

- Reject H_2 if the p-value of an adapted test using only Stage 3 data falls below 0.087
- Reject H_4 if the p-value of an adapted test using only Stage 3 data false below 0.002
- Assuming standard normal observations and no sample size increase this corresponds to critical boundaries 0.0231 for $p_2^{(3)}$ and 0.011 for $p_4^{(3)}$ - which are just the O'BF and POC boundaries for Stage 3 at level α

Type	local-α	Stage 1	Stage 2	Stage 3
O'BF	0.025	0.0001	0.006	0.0231
Pocock	0.025	0.011	0.011	0.011

- Assume we observe $p_2^{(3)} = 0.021$ and $p_4^{(3)} = 0.047$
- Then we reject H_2 but not H_4.

Simulations: Setup

- Two treatment arms in comparison to control; 4 elementary hypotheses
- Graph-based procedure as in the case study; one-sided tests; level α
- Standard normal observations, mean vector $(m1, m2, m3, m4)$.
- Correlation between treatment arms $1/2$; between endpoints $1/3$.
- Only 1 pre-planned interim analysis after $1/2$ of observations have been collected, 100 patients per group and stage

Simulation: Tests and selection procedures

- Apply group sequential test (gsGMCP) MAURER ('13)
- Apply adaptive group sequential test (agsGMCP)
- Test each hypothesis using the group sequential test (Stage2) - stop treatment arms only if the hypotheses for both endpoints can be rejected
- Stop treatment arm if at least the primary hypothesis is rejected (tapered)
- Reallocate sample size to remaining treatment and control if a treatment is dropped (SSR)
- Use O'BF for primary and Poc for secondary hypotheses, or O'BF for both.

Simulations: Results

Bounds	Method	m1	m2	m3	m4	Stage1 P1	Stage2 P1	Stage2 PB	tapered PB	Stage2 PA	tapered PA	ASN
O'BF-Poc	gsGMCP	0.3	0	0.3	0	11.1	77.8	2.3	2	60.6	58.5	588
	agsGMCP					11.2	77.8	2.3	2.3	60.6	58.5	588
	SSR								2.3		58.6	
	gsGMCP	0.3	0.3	0.3	0.3	18.2	89.7	72.1	71.5	76.7	73.2	549
	agsGMCP					18.6	89.7	72.1	72.2	76.7	75.9	548
	SSR								72.7		76.4	
O'BF-O'BF	gsGMCP	0.3	0	0.3	0	11.1	77.7	2.3	1.9	65.4	58.9	589
	agsGMCP					11.1	77.7	2.3	2.3	65.4	59	589
	SSR								2.3		59	
	gsGMCP	0.3	0.3	0.3	0.3	18	89.5	73.1	72.5	81.2	72.1	558
	agsGMCP					18.1	89.5	73.1	73.1	81.2	78.2	556
	SSR								73.6		79	

- More early rejections due to $B_J \geq 1$ (Stage1)
- Power to reject any primary hypothesis is barely affected
- When a treatment is dropped: higher power to reject both (PB) treatments and both hypotheses in a treatment arm (PA)
- Effect of sample size reallocation small (SSR)
- Choice of spending function not clear O'BF-O'BF outperforms O'BF-Poc

Discussion

- Using this approach we may perform adaptive design modifications, e.g. sample size reassessment, adding/dropping of hypotheses, adding/dropping of interim analyses
- Adaptive interim analyses are not limited to the second to last stage - however, designing the remaining test (adapted weights, adapted spending functions, ...) may become cumbersome
- Software will be available in an upcoming version of R-package `gMCP`

References

- F. Bretz, W. Maurer, W. Brannath, and M. Posch. A graphical approach to sequentially rejective multiple test procedures. *StatMed*, 28:586–604, 2009.
- F. Bretz, M. Posch, E. Glimm, F. Klinglmueller, W. Maurer, and K. Rohmeyer. Graphical approaches for multiple endpoint problems using weighted bonferroni, simes or parametric tests. *BiomJ* 53: 894-913, 2011.
- Klinglmueller F, Posch M, and Koenig F. Adaptive graph-based multiple testing procedures. *Pharm Stat* 13.6 (2014): 345-356.
- Sugitani, T., Hamasaki, T., and Hamada, C. (2013). Partition testing in confirmatory adaptive designs with structured objectives. *Biom J*, 55(3), 341-359.
- Klinglmueller F. (2013). Flexible tests for clinical trials testing multiple hypotheses. PhD Thesis: Vienna University of Technology
- Tamhane, A. C., Wu, Y., and Mehta, C. R. (2012). Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (II): sample size re-estimation. *Stat Med*, 31(19), 2041-2054.
- Maurer, Willi, and Frank Bretz. "Multiple testing in group sequential trials using graphical approaches." *Stat Biopharm Res* 5.4 (2013): 311-320.