Power Considerations for Clinical Trials in Presence of Multiplicity

Kaushik Patra, PhD

Alexion Pharmaceuticals

BASS XXVI

OCTOBER 21, 2019

Background

- Hypotheses associated with secondary endpoints are typically ranked in the protocol
  - Based on clinical importance or likelihood of success
- These hypotheses are tested only if the primary endpoint is significant
- Q: does the pre-specified order in the protocol always help?

FDA guidance

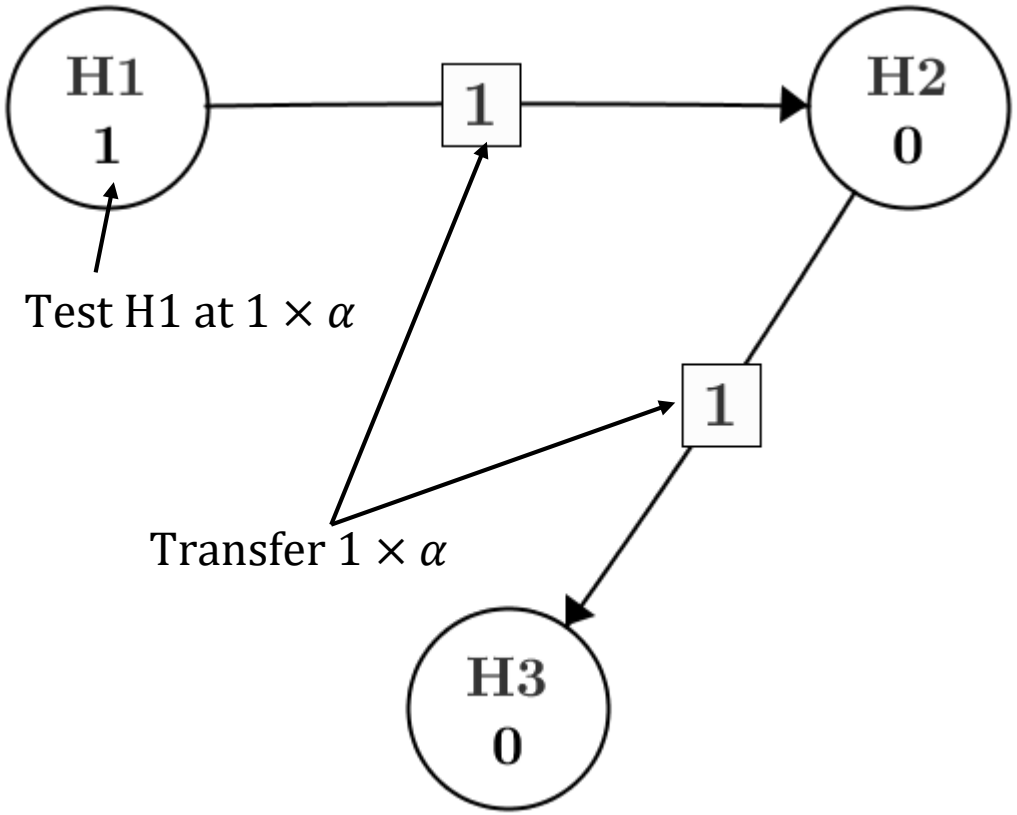
- Reference: FDA draft guidance (2017) “Multiple Endpoints in Clinical Trials”
- Multiplicity adjustment is related to evaluation of multiple hypotheses
- Regulatory requirement to protect against “false claims” i.e. claim for treatment effect when there is none.
- Control studywise Type I error rate at α = .05 (for example)

Example

Null Hypotheses	Nominal p-values (2-sided)
H1	p1 = .051
H2	p2 = .024
H3	p3 = .016

Overall alpha = 0.05

Fixed sequence testing



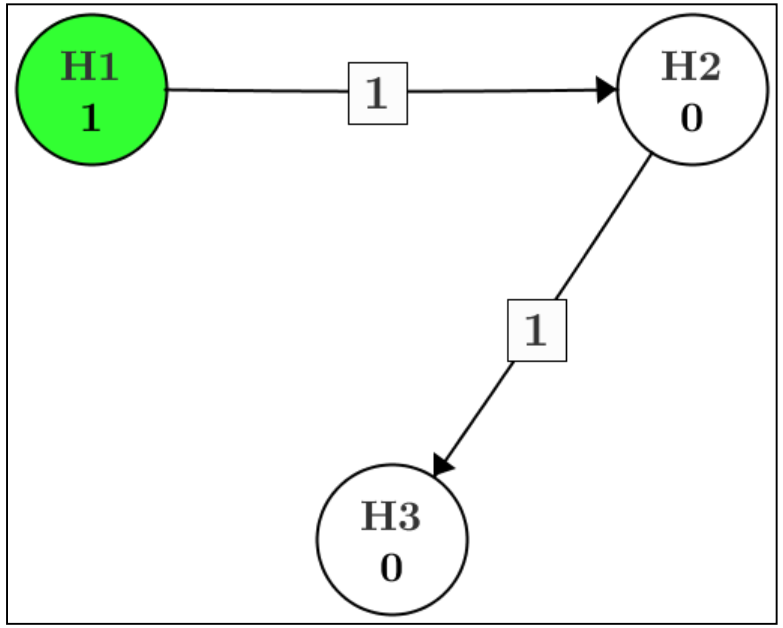
Testing:

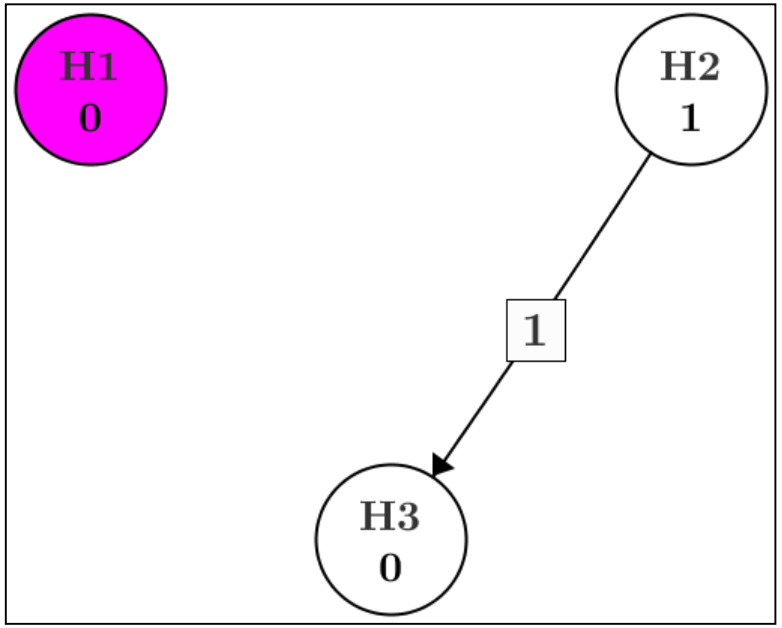
1. Test H1 first at full $\alpha = 0.05$
   a. If p1 < 0.05 i.e. H1 is rejected then test H2
   b. If accepted, then stop – no further testing should be performed.
2. Same rule for H2 and H3.

Order	H	P-values	Decision
1	H1	0.051	Not rejected
2	H2	0.024	Not rejected
3	H3	0.016	Not rejected

Fixed sequence testing

Order	H	P-values	Decision
1	H1	0.024	Rejected
2	H2	0.051	Not rejected
3	H3	0.016	Not rejected





Fixed sequence testing

1. Ranking the hypotheses can be a challenging task especially when prior experience is limited

2. Study success is heavily dependent on the order

Possible improvement

1. Using Fixed sequence testing, the first hypothesis always gets the full 'weight' i.e. it has to be tested at α = .05

2. Do we always need to test the first hypothesis at α = .05? -not necessarily

3. May use any fraction of α = .05 e.g. .04 or .03 or .01

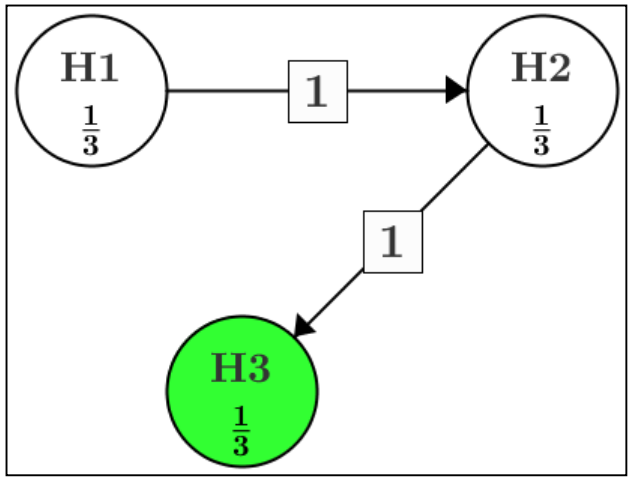
4. Fraction is determined by the relative importance (weight) of the hypothesis
5. Remaining fraction of α can be distributed among the rest of the hypotheses

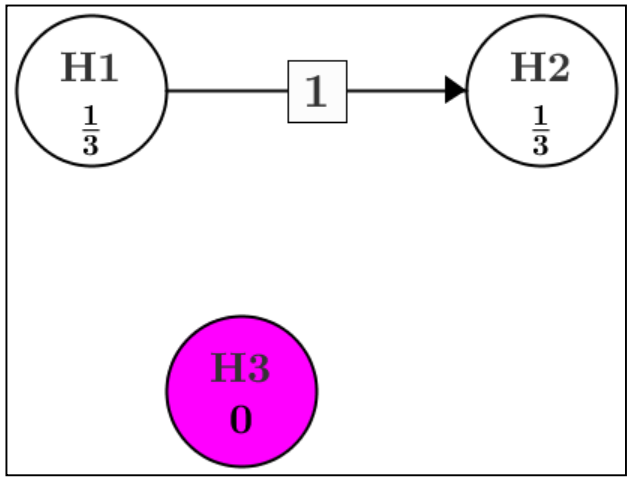
Fallback procedure

Testing:

1. Assign weights: 1/3, 1/3, 1/3
   i.e. distribute α as $\alpha/3$, $\alpha/3$, $\alpha/3$.
2. Test H1 at 0.05/3 = .0167
   a. If H1 is rejected then 'release' $\alpha/3$
      test H2 at $\alpha/3+ \alpha/3$ i.e. full $2\alpha/3$ = 0.033
   b. If H1 is not rejected then test H2 at $\alpha/3$ = 0.0167
3. If H2 is rejected, proceed to test H3 in similar fashion

Order	H	Weight	P-values	Decision
1	H1	1/3	0.051	Not rejected
2	H2	1/3	0.024	Not rejected
3	H3	1/3	0.016	Rejected



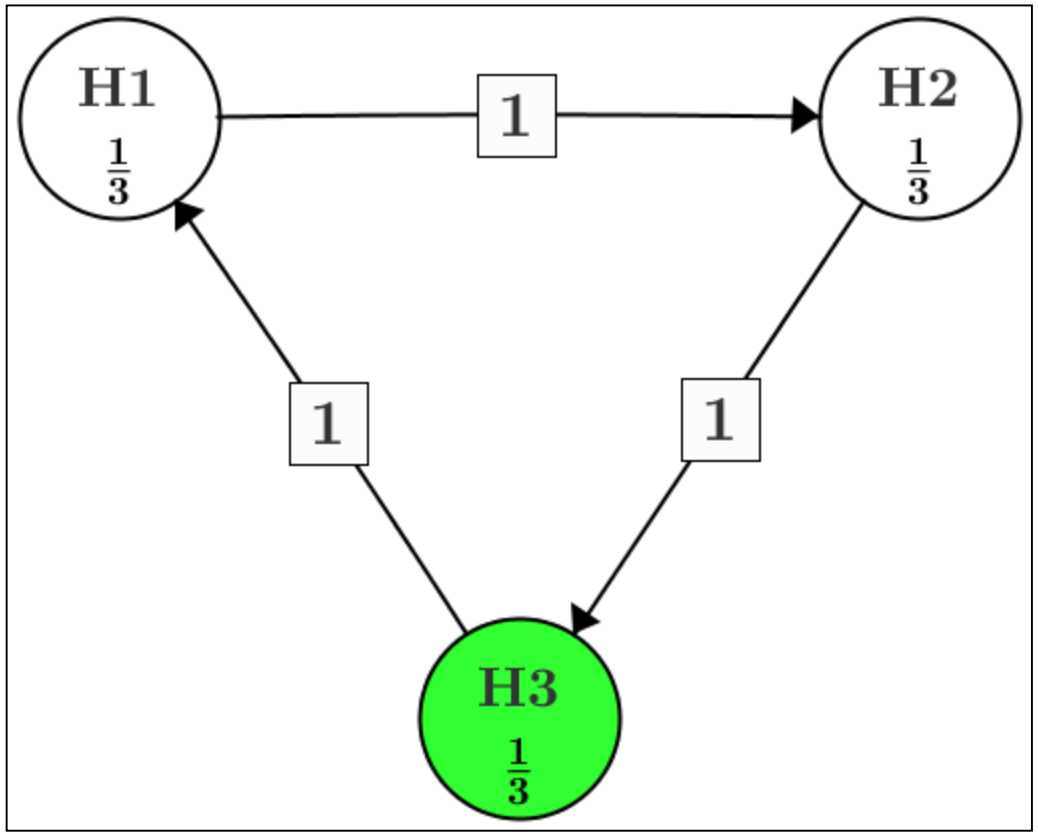


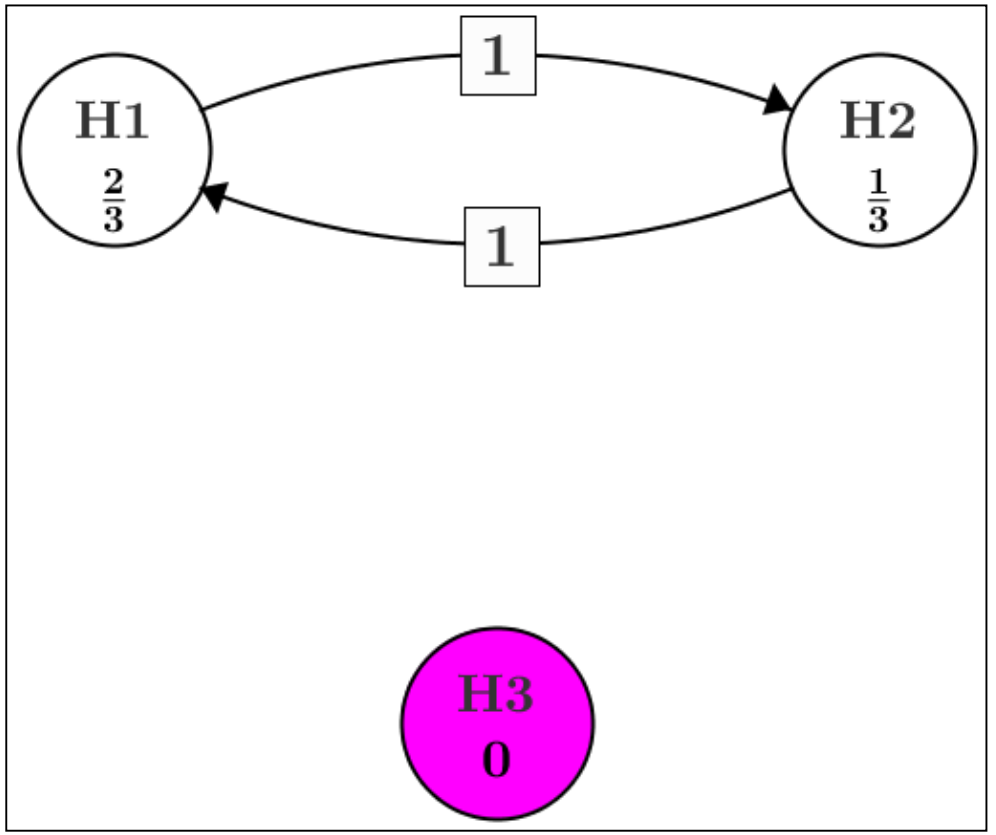
Fallback procedure

1. Because of the flexibility of splitting α, fallback procedure allowed for rejection of the most significant test

2. Note: Order of the hypotheses can not be changed

What if we could ‘pass’ the unused alpha back to first hypothesis only?

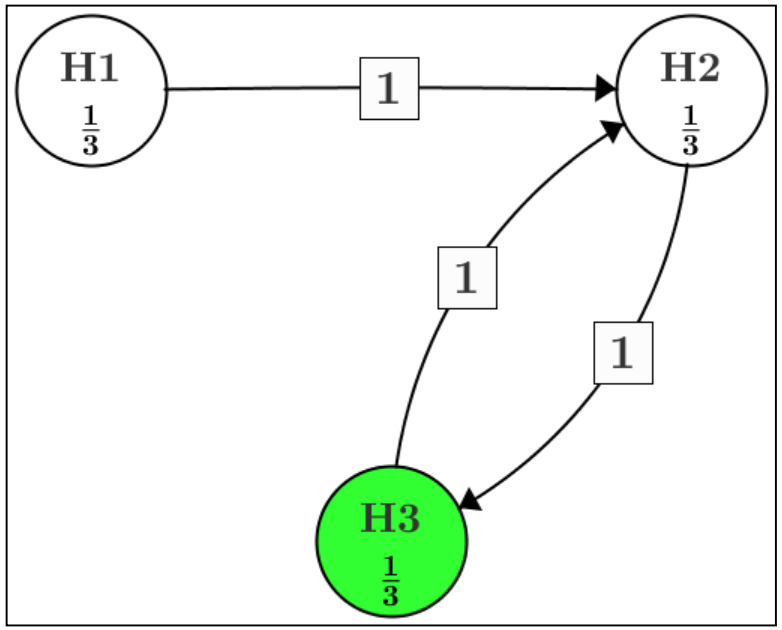


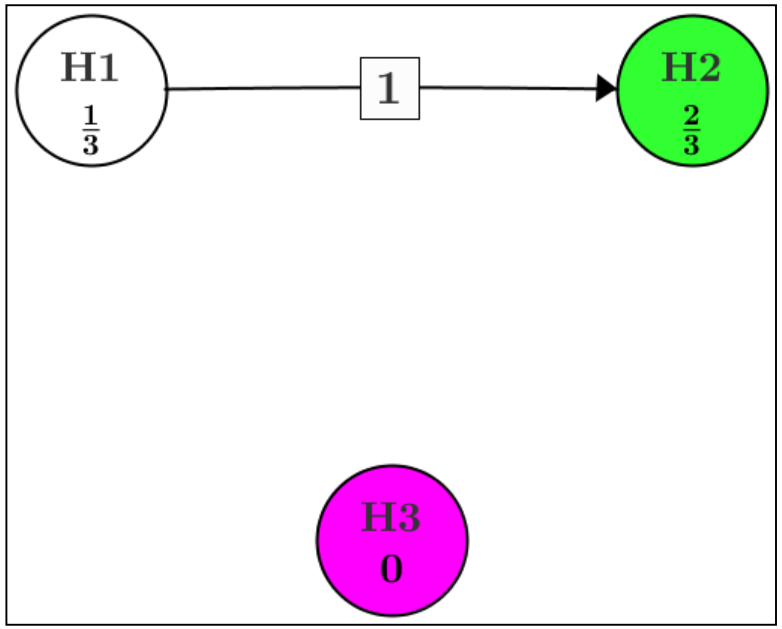


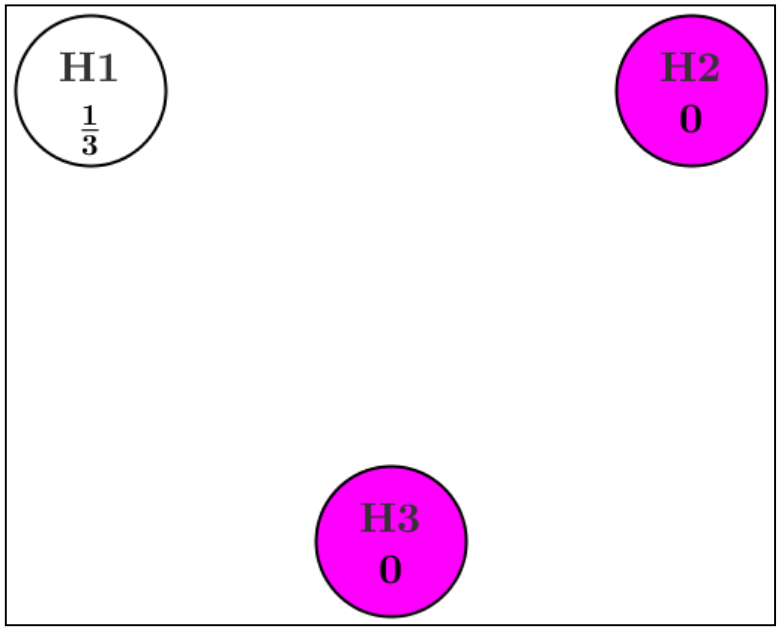
What if we could ‘pass’ the unused alpha back to Second hypothesis only?

Order	H	Weight	P-values	Decision
1	H1	1/3	0.051	Not rejected
2	H2	1/3	0.024	Rejected
3	H3	1/3	0.016	Rejected

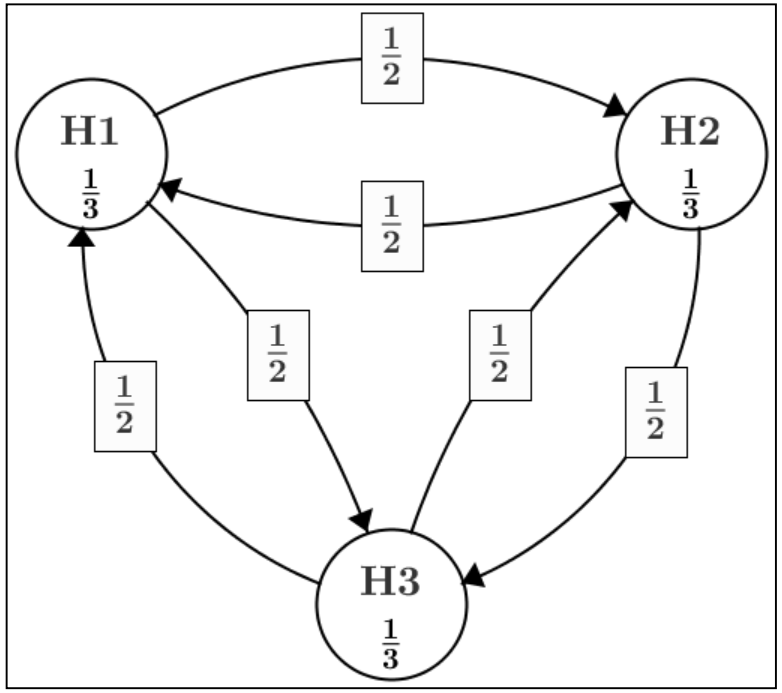
By relaxing the pre-specified order restriction, 2 out 3 hypotheses are rejected.

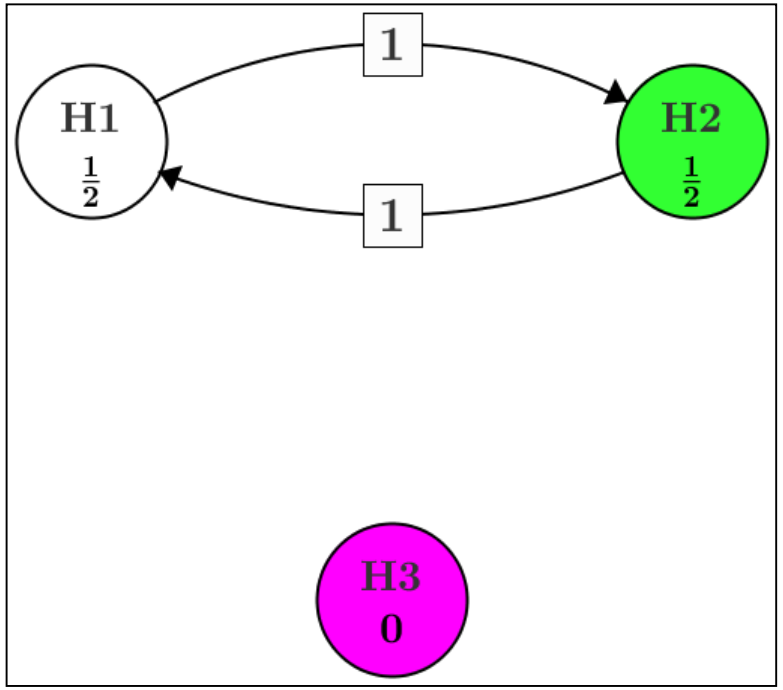


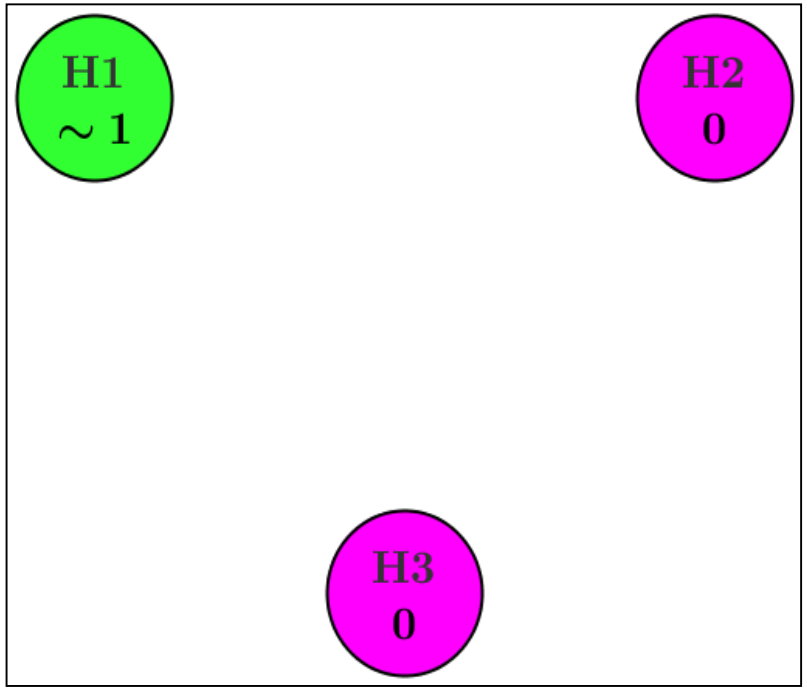




Chain procedure - general form







p1 = 0.051, p2 = 0.024, p3 = 0.016

Chain procedure - general form

1. Allows for assignment of weights to hypotheses according to clinical importance
2. Allows for 'unused' alpha to be passed along (propagated) to other hypotheses
3. Order of hypotheses is no longer a concern
4. Weights and passing rules are only initial specification – after at least one hypothesis is rejected, weights and passing rules are automatically updated

Contains Nonbinding Recommendations

Draft — Not for Implementation

APPENDIX: THE GRAPHICAL APPROACH

Bretz F, Hothorn T, Westfall P. *Multiple Comparisons Using R*, CRC Press (Taylor & Francis Group), Chapman and Hall, 2010.

FDA draft guidance (2017)

Case Study 1: Neuromyelistis Optica

- Primary endpoint: Time to first relapse
  - p-value < 0.0001
- 6 secondary endpoints
  - To be tested using a fixed sequence approach

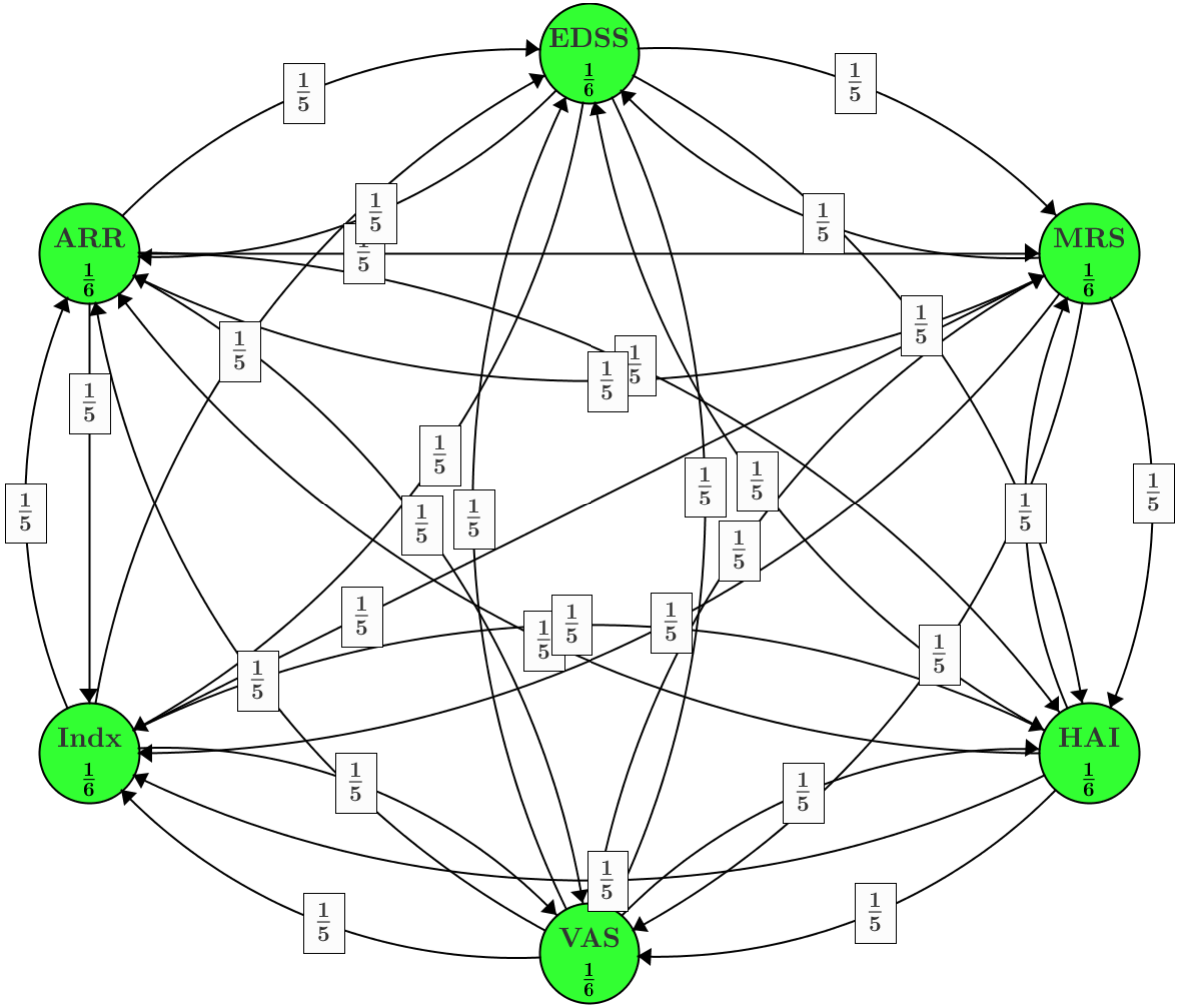
Pre-specified order	Endpoint	Nominal p-values	Decision
1	Annualized relapse rate (ARR)	0.0001	Rejected
2	EDSS	0.0597	Not rejected
3	Modified Rankin Scale (MRS)	0.0154	Not rejected
4	Hauser Ambulation Index (HAI)	0.0002	Not rejected
5	EQ-5D VAS	0.0309	Not rejected
6	EQ-5D Index	0.0077	Not rejected

Case Study 1: Implementing chain procedure

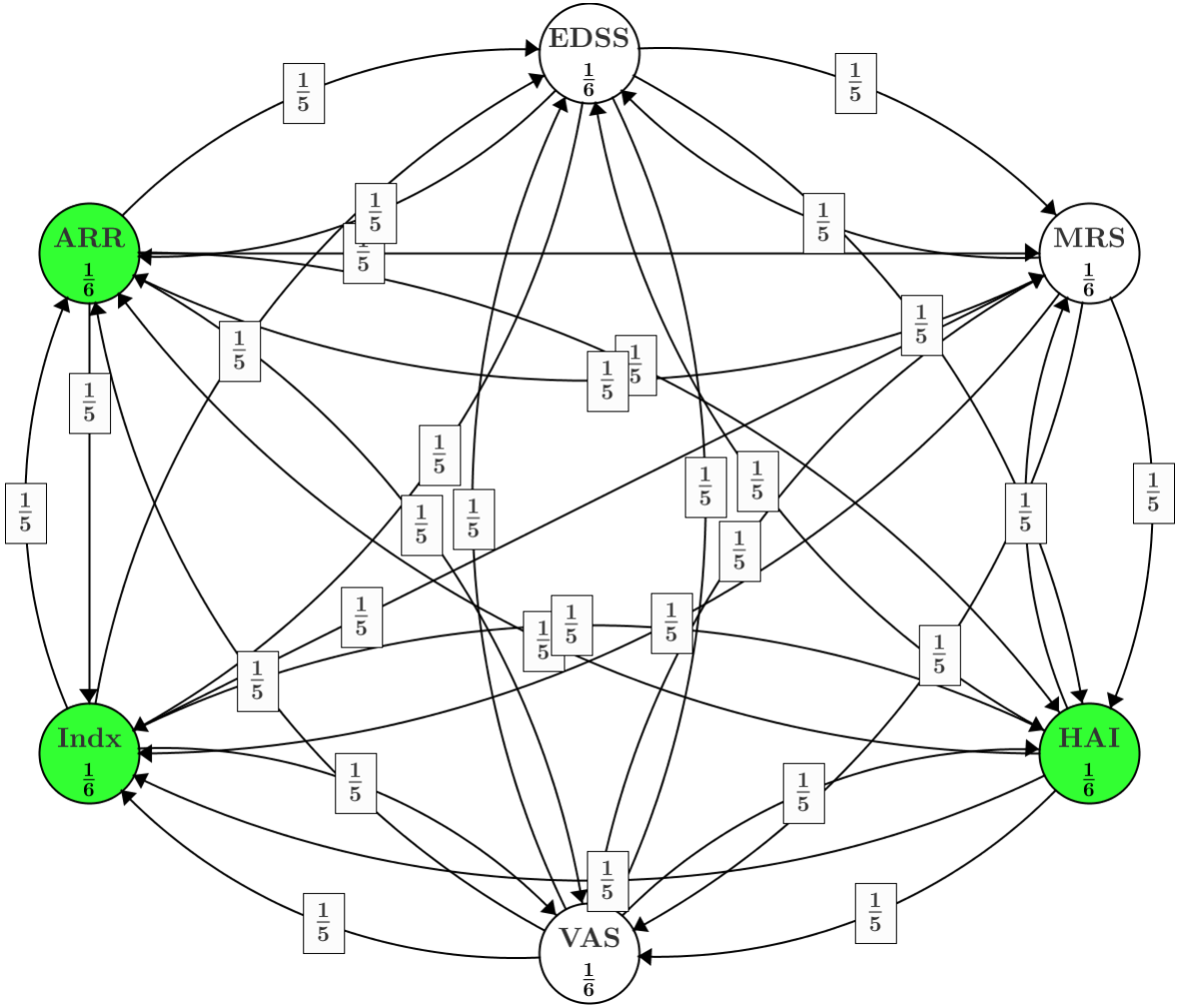
- Consider equal weight for each of the 6 hypotheses = 1/6
- Pass unused alpha equally to other hypotheses (propagation rule)

	ARR	EDSS	MRS	HAI	VAS	Indx
ARR	0	1/5	1/5	1/5	1/5	1/5
EDSS	1/5	0	1/5	1/5	1/5	1/5
MRS	1/5	1/5	0	1/5	1/5	1/5
HAI	1/5	1/5	1/5	0	1/5	1/5
VAS	1/5	1/5	1/5	1/5	0	1/5
Indx	1/5	1/5	1/5	1/5	1/5	0

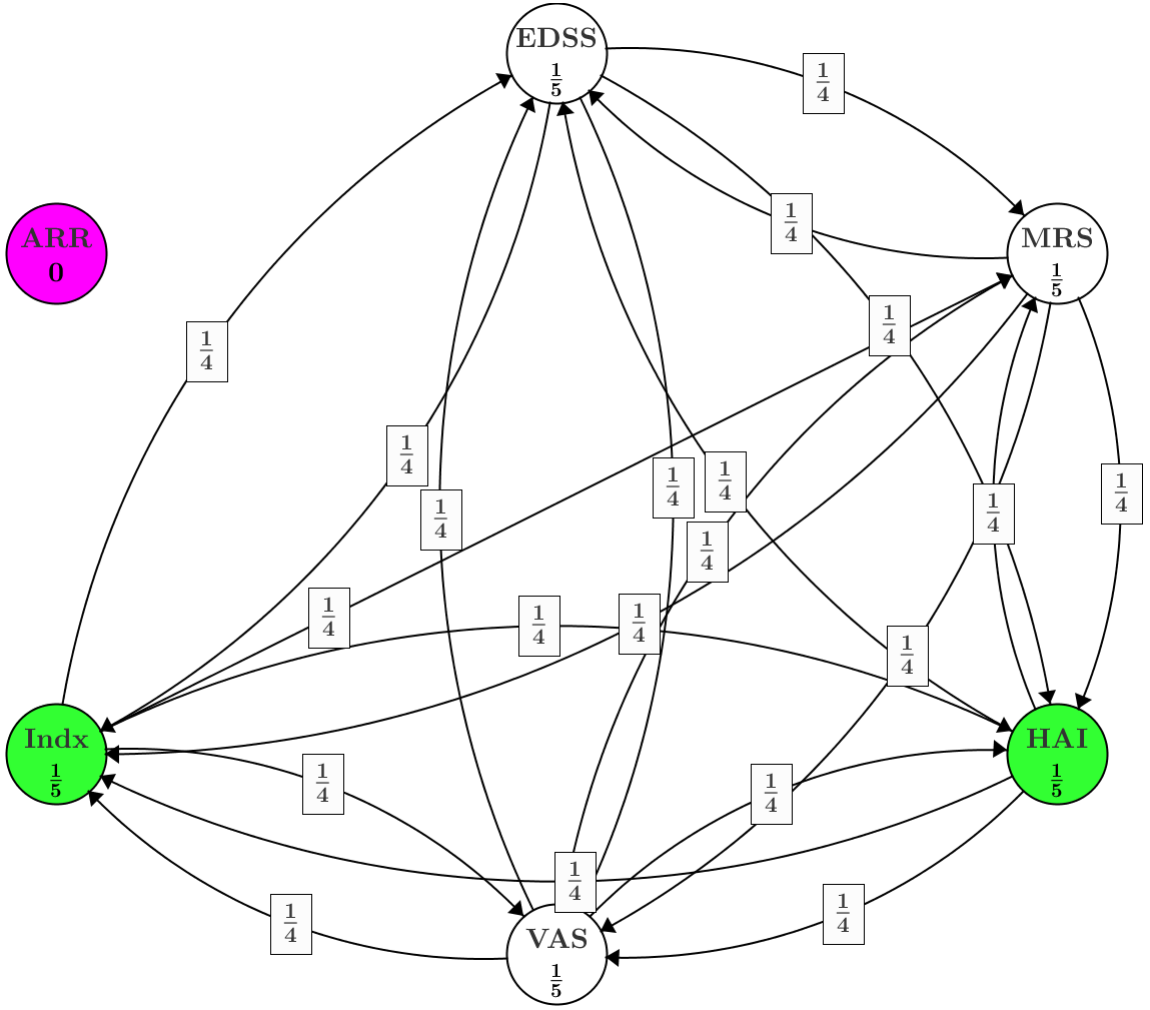
Case Study 1: Chain procedure



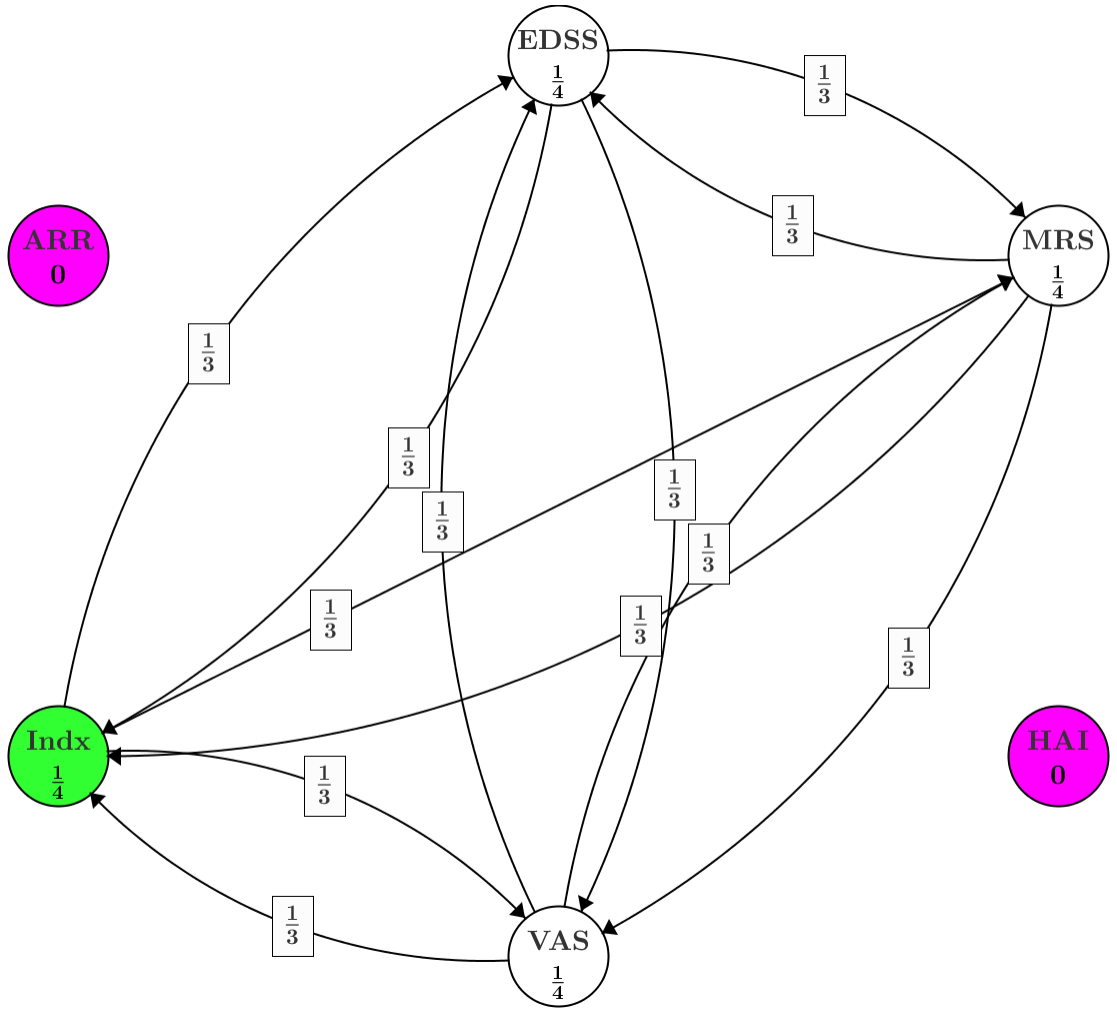
Case Study 1: Chain procedure



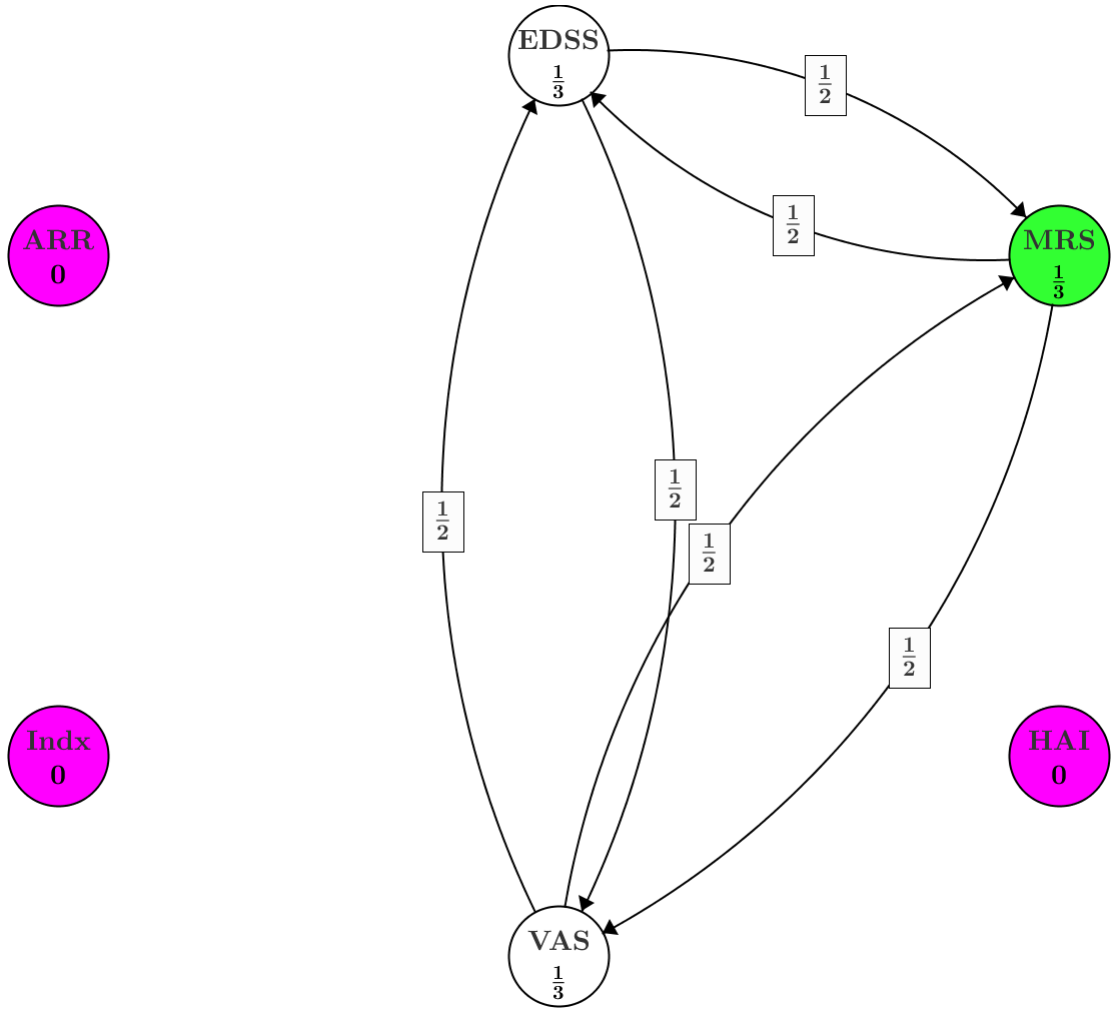
Case Study 1: Chain procedure



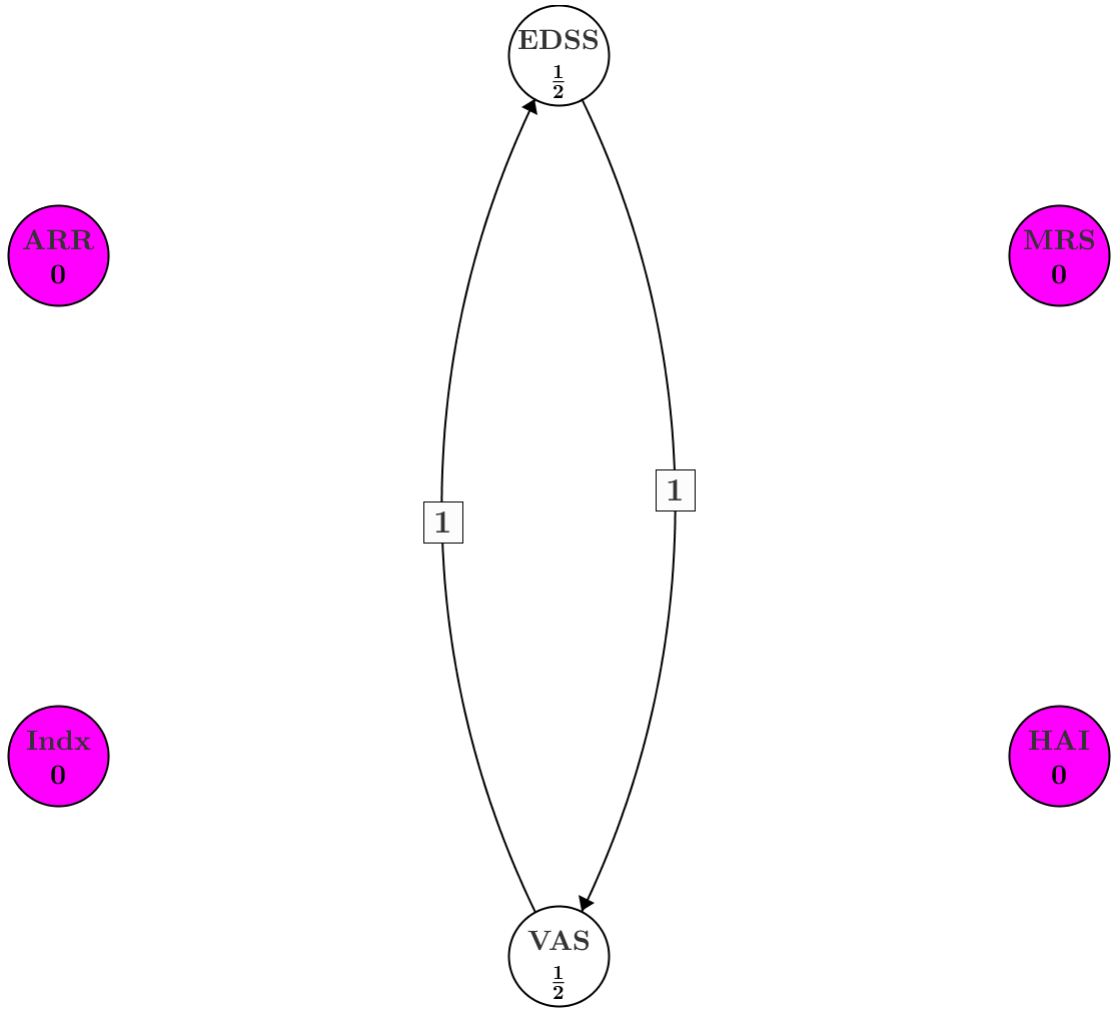
Case Study 1: Chain procedure



Case Study 1: Chain procedure



Case Study 1: Chain procedure



Fixed sequence vs chain procedure

Pre-specified order	Hypothesis	Raw p-values	Fixed sequence	Chain
1	Annualized relapse rate (ARR)	.0001	Rejected	Rejected
2	EDSS	.0597	Not rejected	Not rejected
3	Modified Rankin Scale (MRS)	.0154	Not rejected	Rejected
4	HAI	.0002	Not rejected	Rejected
5	EQ-5D VAS (VAS)	.0309	Not rejected	Not rejected
6	EQ-5D Index (Indx)	.0077	Not rejected	Rejected
Number of hypotheses rejected			**1**	**4**

Other approaches

H	RAW	BONFERRONI	HOLM	HOMMEL	HOCHBERG	FIXEDSEQ	FALLBACK	CHAIN
1	0.0001	0.0006	0.0006	0.0006	0.0006	0.0001	0.0006	0.0006
2	0.0597	0.3582	0.0618	0.0597	0.0597	0.0597	0.1791	0.0618
3	0.0154	0.0924	0.0462	0.0462	0.0462	0.0597	0.0924	0.0462
4	0.0002	0.0012	0.0010	0.0010	0.0010	0.0597	0.0012	0.0010
5	0.0309	0.1854	0.0618	0.0597	0.0597	0.0597	0.0924	0.0618
6	0.0077	0.0462	0.0308	0.0308	0.0308	0.0597	0.0462	0.0308

Power considerations

- The choice of the MCP method to use for a specific clinical trial will depend on the objectives and the design of the trial.
- The method should be decided upon prospectively.
- Sponsors should consider the variety of methods available and select the most powerful method that is suitable for the design and objective of the study and maintains Type I error rate control.

EMA guidance (EMA, 2017)

- “Significant effects in [secondary endpoints] can be considered for an additional claim only after the primary objective of the clinical trial has been achieved, and if they were part of the confirmatory strategy.”

Case Study 2: Myasthenia Gravis (MG)

Endpoint type	Endpoint	Active (n=62)	Placebo (n=63)	Treatment Effect	
		Mean (SE)	Mean (SE)	Difference (95% CI)	p-value
Primary	E_1	−4.1 (0.5)	−2.3 (0.5)	−1.8 (−3.2 , −0.5)	0.0077
Secondary	E_2	−4.6 (0.6)	−1.7 (0.6)	−2.9 (−4.6 , −1.2)	0.0007
	E_3	−7·9 (1·0)	−4·6 (1·0)	−3·3 (−5·9 , −0·6)	0·0168
	E_4	−13·8 (1·6)	−6·7 (1·6)	−7·1 (−11·3 , −3·0)	0·0009

E_1: Change from baseline in Myasthenia Gravis Activities of Daily Living

E_2: Change from baseline in Quantitative Myasthenia Gravis

E_3: Change from baseline in Myasthenia Gravis Composite.

E_4: Change from baseline in Myasthenia Gravis Quality of Life questionnaire

- E_2 is an important endpoint yet not elevated to the primary endpoint

Case Study 2: Planning for a new study in MG

- Objective: Superiority of a new experimental compound over placebo
- New compound has similar MOA as old compound
- Same primary (E_1) and secondary endpoints (E_2, E_3, E_4)
- Total sample size should not exceed previous study size by a huge margin
  - Positive experience from previous study
  - Recruitment challenge in rare disease
  - A moderate increase is allowed for expanded population

Case Study 2: Planning for a new study in MG

- Study success
  - Primary endpoint statistically significant

    +
  - Either (E_2 and E_3 significant) OR (E_2 and E_4 significant)
- Task: determine sample size to maximize the probability of success
  - Implicit determination of optimum multiplicity adjustment

Some important success criteria
(based on multiplicity adjustment)

- Marginal power: power for an endpoint after multiplicity adjustment
- Disjunctive power: at least one endpoint is significant
- Subset disjunctive power: at least one endpoint out of a subset of endpoints is significant
- Conjunctive power: all endpoints are significant
- Weighted power: weighted average of marginal powers
- Custom success criteria: a meaningful combinations of above

Case Study 2: Planning for a new study in MG

- $r_i = 1 \Leftrightarrow E_i$ significant, i=1, 2, 3, 4

Disjunctive power:	P(E_2 or E_3 or E_4 significant)	$= P(r_2 + r_3 + r_4 \geq 1)$
Conjunctive power:	P(E_2 and E_3 and E_4 significant)	$= P(r_2 + r_3 + r_4 = 3)$
Subset disjunctive power:	P(Either E_3 or E_4 significant)	$= P(r_3 + r_4 \geq 1)$
Custom success criteria:	**P(Either E_2 and E_3 OR E_2 and E_4 significant)**	$= \boldsymbol{P(r_2 r_3 + r_2 r_4 \geq 1)}$

- Success criteria are evaluated only if the primary endpoint is significant

Case Study 2: Planning for a new study in MG

- Chain procedure will be used
- Different choices of initial weights for α and propagation matrix will be considered
- Different choices of sample size will be used
  - N = 120 ~ comparable to the old study
  - N = 160 ~ 80% power for the primary endpoint
  - N = 210 ~ 90% power for the primary endpoint
- Simulations
  - Target treatment effect: similar to old study
  - Pessimistic treatment effect: 25% less treatment effect (sensitivity purpose)
    - Both scenarios to be replicated for zero and non-zero (0.5) correlation among endpoints

Clinical Scenario Evaluations (CSE)

Treatment effect	Correlation	N
Target	0	120
		160
		210
	0.5	120
		160
		210
Pessimistic	0	120
		160
		210
	0.5	120
		160
		210

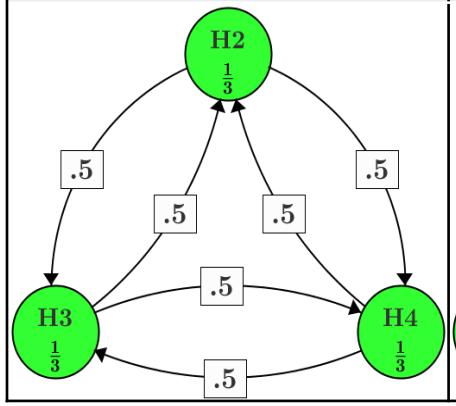
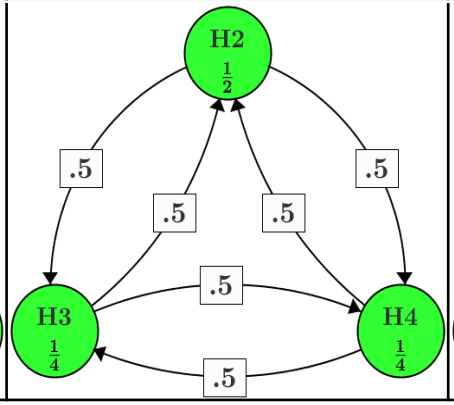
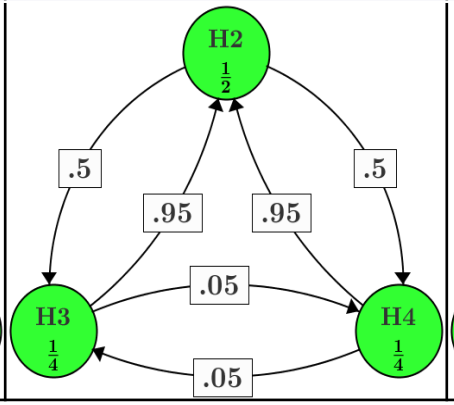
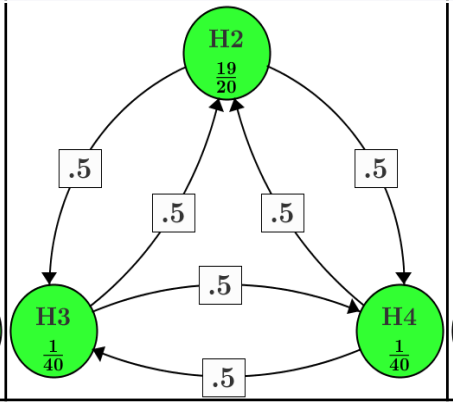
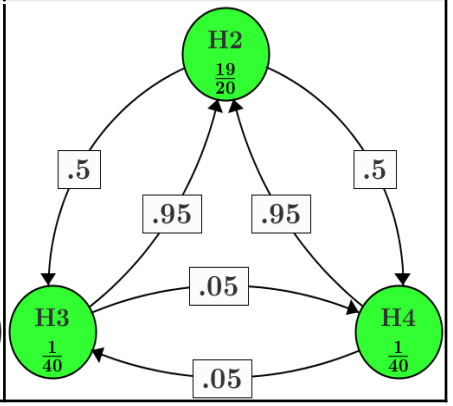
Optimum selection of initial weight and propagation matrix

Initial weights for alpha

	H2	H3	H4
Equal weight	1/3	1/3	1/3
Unequal weight	1/2	1/4	1/4
Extremely unequal weight	.95	.05	.05
	19/20	1/40	1/40

Propagation

	Symmetric			Asymmetric		
	H2	H3	H4	H2	H3	H4
H2	0	.5	.5	0	.5	.5
H3	.5	0	.5	.95	0	.05
H4	.5	.5	0	.95	.05	0

Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5
• Equal weight • Symmetric propagation	• Unequal weight • Symmetric propagation	• Unequal weight • Asymmetric propagation	• Extremely unequal weight • Symmetric propagation	• Extremely unequal weight • Asymmetric propagation
				

Evaluation criteria

- Select the strategy that maximizes Pr(custom success criteria)
  - Preference given to symmetric propagation over asymmetric propagation
    - Eg. Strategy 4 gets preference over strategy 5

Probability of success (Target treatment effect)

Strategy	N = 120	N = 160	N = 210
Correlation = 0			
1	0.51	0.72	0.88
2	0.52	0.73	0.88
3	0.52	0.73	0.88
4	0.53	0.73	0.88
5	0.53	0.73	0.88
Correlation = 0.5			
1	0.58	0.74	0.88
2	0.58	0.75	0.88
3	0.58	0.75	0.88
4	0.59	0.75	0.88
5	0.59	0.75	0.88

Probability of success (Pessimistic treatment effect)

Strategy	N = 120	N = 160	N = 210
Correlation = 0			
1	0.24	0.44	0.44
2	0.25	0.45	0.45
3	0.26	0.45	0.46
4	0.28	0.48	0.48
5	0.28	0.48	0.48
Correlation = 0.5			
1	0.35	0.53	0.53
2	0.36	0.53	0.53
3	0.37	0.54	0.54
4	0.39	0.56	0.56
5	0.39	0.56	0.56

Overview of success criteria

	N = 120	N = 160	N = 210
Correlation = 0			
Disjunctive power	0.64	0.79	0.90
Conjunctive power	0.30	0.52	0.73
Probability of success	0.53	0.73	0.88
Weighted power	0.76	0.87	0.94
Correlation = 0.5			
Disjunctive power	0.66	0.79	0.90
Conjunctive power	0.44	0.62	0.78
Probability of success	0.59	0.75	0.88
Weighted power	0.76	0.88	0.94

Conclusion

- Power consideration for secondary endpoints in pivotal studies is a non-trivial yet important consideration
- Study success criteria may not be unique but needs to be pre-specified
- Sample size determination should be evaluated using the CSE framework

Software

- SAS PROC MULTTEST
- SAS macros from multxpert.com
- SAS macros: Analysis of Clinical Trials Using SAS: A Practical Guide, Second Edition - by Alex Dmitrienko and Gary G. Koch
- R gMCP package
- Mediana package: https://cran.r-project.org/web/packages/Mediana/vignettes/mediana.html

References

- **Multiple testing procedures**
- Bretz, F., Maurer, W., Brannath, W., Posch, M. (2009). A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine. 28, 586-604.
- Dmitrienko, A., Tamhane, A.C., Bretz, F. (editors). Multiple Testing Problems in Pharmaceutical Statistics. Chapman and Hall/CRC Press, New York, 2009.
- FDA (U.S. Food and Drug Administration). Multiple endpoints in clinical trials: Guidance for Industry. 2017.
- **CSE**
  Benda, N., Branson, M., Maurer, W., Friede, T. Aspects of modernizing drug development using clinical scenario planning and evaluation. Drug Information Journal. 44:299-315, 2010.
- Dmitrienko, A., Pulkstenis, E. (editors). Clinical Trial Optimization Using R. Chapman and Hall/CRC Press, New York, 2017.

Thank you!

Case Study 2: Myasthenia Gravis (MG)

- Adjusted p-values

Test	Weight	Nominal p-value	Adjusted p-values						
			Bonferroni	Holm	Hommel	Hochberg	Fixed Sequence	Fallback	Chain
E_2	1/3	0.0007	0.0021	0.0021	0.0014	0.0018	0.0007	0.0021	0.0021
E_3	1/3	0.0168	0.0504	0.0168	0.0168	0.0168	0.0168	0.0252	0.0168
E_4	1/3	0.0009	0.0027	0.0021	0.0018	0.0018	0.0168	0.0027	0.0021