Key Multiplicity Issues in Clinical Trials (Part II)

Alex Dmitrienko (Mediana Inc)
admitrienko@medianainc.com

Introduction

Multiplicity in clinical trials

Complex objectives in clinical trials

Complex sets of objectives are introduced to improve the efficiency of clinical trials or development programs

Multiple clinical endpoints, multiple doses or regimens, multiple patient populations, multiple decision points (interim and final analyses)

Multiplicity issues

Complex objectives induce multiplicity and lead to inflated false-positive rates

Multiplicity issues in clinical trials

FDA guidance

Draft guidance on multiple endpoints in clinical trials (January 2017)

EMA guidance

Draft guideline on multiplicity issues in clinical trials (April 2017)

Revision of *Points to consider on multiplicity issues in clinical trials* (September 2002)

General outline

Part I

Traditional multiplicity problems

Clinical trials with a single source of multiplicity or single multiplicity component

Part II

Advanced multiplicity problems

Clinical trials with several sources of multiplicity or several multiplicity components

Part I: Traditional multiplicity problems

Single family of null hypotheses

$$H_1, \ldots, H_m$$

Trials with a single source of multiplicity

Part I: Outline

Module A

Clinical trial examples

Module B

Key concepts (inferential goals, error rate definitions, classification of multiple testing procedures)

Module C

Nonparametric and semiparametric multiple testing procedures I (data-driven hypothesis ordering)

Part I: Outline

Module D

Nonparametric multiple testing procedures II (pre-specified hypothesis ordering)

Module E

Parametric multiple testing procedures

Module F

Simultaneous confidence intervals

Module G

Power calculations

Part II: Advanced multiplicity problems

Multiple families of null hypotheses

Family 1 $H_1, \ldots, H_{k_1}$

$\ldots$

Family m $H_{k_{m-1}+1}, \ldots, H_{k_m}$

Trials with multiple sources of multiplicity

Several sources of multiplicity

Multiple endpoints and multiple dose-control comparisons

First source: Multiple endpoints (primary and secondary endpoints)

Second source: Evaluate efficacy at different dose levels

Multiple endpoints and multiple patient populations

First source: Multiple endpoints

Second source: Evaluate efficacy in two or more pre-defined patient populations

Part II: Outline

Module A

Clinical trial examples

Module B

Introduction to gatekeeping strategies

Module C

Parallel gatekeeping strategies

Part II: Outline

Module D

General gatekeeping strategies

Module E

Group-sequential trials with multiple objectives

Module F

General adaptive trials with multiple objectives

Module G

Power calculations

Books

Analysis of Clinical Trials Using SAS (Second Edition)

Edited by Alex Dmitrienko (Mediana) and Gary Koch (UNC-Chapel Hill)

SAS Press (2017)

Chapter 5: Multiplicity adjustment methods

Introduction to multiplicity problems arising in clinical trials, popular multiple testing procedures as well as gatekeeping procedures for problems with several sources of multiplicity

Books

Multiple Testing Problems in Pharmaceutical Statistics

Edited by Alex Dmitrienko (Eli Lilly), Ajit Tamhane (Northwestern University), Frank Bretz (Novartis, Hannover Medical School)

Chapman and Hall/CRC Press (2009)

Comprehensive summary of methodological, regulatory and practical issues related to multiplicity problems in clinical trials, including problems with several sources of multiplicity (Chapter 5)

Review papers

Recent review papers and tutorials

Dmitrienko, D'Agostino and Huque. (2013). Key multiplicity issues in clinical drug development.

Dmitrienko and D'Agostino. (2013). Tutorial in Biostatistics: Traditional multiplicity adjustment methods in clinical trials.

Alosh, Bretz and Huque (2014). Advanced multiplicity adjustment methods in clinical trials.

Web site

Instant Training web site

http://sprmm.com

Supplementary materials

Presentation slides

SAS and R code

References

Module A
Clinical Trial Examples

Module A outline

A1. Clinical trial examples

Examples of confirmatory clinical trials with "multivariate" multiplicity problems

Combination of several multiplicity components such as multiple endpoints, multiple doses or regimens, multiple patient populations and multiple decision points

A1. Clinical trial examples

Example 1

Prostate cancer trial: Multiple endpoints (primary and secondary endpoints)

Example 2

Type 2 diabetes mellitus trial: Multiple dose-placebo comparisons

Example 3

Schizophrenia trial: Multiple endpoints and multiple dose-placebo comparisons

A1. Clinical trial examples

Example 4

Prostate cancer trial: Multiple endpoints and multiple decision points

Example 5

Breast cancer trial: Multiple patient populations and multiple decision points

Example 1

Multiple endpoints

Prostate cancer trial

ALSYMPCA trial (Sartor et al., 2014)

Objective

Evaluate the effects of radium-223 dichloride on symptomatic skeletal events in patients with castration-resistant prostate cancer

Trial arms

Experimental treatment versus placebo

Multiple endpoints

Primary endpoint

Endpoint P: Overall survival

Secondary endpoints

Endpoint S1: Time to first symptomatic skeletal event

Endpoint S2: Occurrence of new symptomatic pathological bone fractures

Endpoint S3: Occurrence of spinal cord compression

Role of secondary endpoints

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."

Example 2

Multiple doses

Type 2 diabetes mellitus trial

Saxagliptin Phase III trial (Frederich et al., 2012)

Objective

Evaluate the efficacy and safety of saxagliptin in patients with Type 2 diabetes mellitus

Primary endpoint

Change from baseline to Week 24 in HbA1c

Multiple doses

Trial arms

Several treatment regimens (Saxagliptin 2.5 mg q.A.M., 5 mg q.A.M. and 2.5/5 mg q.A.M.) versus placebo

Families of dose-placebo comparisons

Primary family: Saxagliptin 2.5 mg q.A.M. versus placebo, Saxagliptin 5 mg q.A.M. versus placebo

Secondary family: Saxagliptin 2.5/5 mg q.A.M. versus placebo

Example 3

Multiple endpoints and multiple doses

Schizophrenia trial

Lurasidone Phase III trial (Meltzer et al., 2011)

Objective

Evaluate the short-term efficacy and safety of lurasidone in the treatment of acute schizophrenia

Trial arms

Two doses of lurasidone (40 mg and 120 mg) versus placebo

Multiple endpoints and multiple doses

Primary endpoint

Endpoint P: Change from baseline to Week 6 in PANSS (Positive and Negative Syndrome Scale) total score

Secondary endpoint

Endpoint S: Change from baseline to Week 6 in CGI-S (Clinical Global Impressions – Severity) score

Multiple endpoints and multiple doses

Schizophrenia trial

Pomaglumetad methionil Phase III trial (Downing et al., 2014)

Multiple endpoints

Primary endpoint: PANSS total score

Secondary endpoint: PSP (Personal and Social Performance Scale) score

Multiple dose-placebo tests

Two doses of experimental treatment (40 mg and 80 mg) versus placebo

Multiple endpoints and multiple doses

Multiple patient populations

Overall trial population

Pre-defined subpopulation which excluded non-Hispanic white patients with the A/A genotype at the serotonin 2A receptor single nucleotide polymorphism (rs7330461)

Example 4

Multiple endpoints and multiple decision points

Prostate cancer trial

SPARC trial (Sternberg et al., 2009)

Objective

Evaluate the efficacy and safety of satraplatin (in combination with prednisone) in patients with castrate-refractory prostate cancer

Trial arms

Satraplatin plus prednisone versus placebo plus prednisone

Multiple endpoints and multiple decision points

Multiple endpoints

Primary endpoint: Progression-free survival

Secondary endpoint: Overall survival

Multiple decision points

Single interim analysis and final analysis

Example 5

Multiple populations and multiple decision points

Breast cancer trial

BELLE-3 trial in patients with locally advanced or metastatic breast cancer

Objective

Evaluate the efficacy and safety of buparlisib (in combination with fulvestrant)

Primary endpoint

Progression-free survival

Multiple populations and multiple decision points

Trial arms

Buparlisib plus fulvestrant versus placebo plus fulvestrant

Multiple patient populations

Overall trial population

Pre-specified subpopulation (patients with PIK3CA mutations)

Multiple decision points

Single interim analysis and final analysis

Multiple doses and multiple decision points

COPD trial

Seamless Phase II/Phase III trial (Barnes et al., 2010)

Objective

Evaluate the efficacy of indacaterol in patients with chronic obstructive pulmonary disease

Multiple doses and multiple decision points

Primary endpoint

Forced expiratory volume in 1 second (FEV1) at 14 days

Trial arms

Indacaterol 75 mcg, 150 mcg, 300 mcg and 600 mcg versus placebo

Multiple decision points

Single interim analysis and final analysis

Multiplicity in adaptive trials

EMA guideline (EMA, 2017)

"There are further areas concerning multiplicity in clinical trials which, according to the above list of issues, are not the focus of this document. For example, there is a rapid advance in methodological richness and complexity regarding interim analyses, with the possibility to stop early either for futility or with a claim for efficacy, or stepwise designed studies, with the possibility for adaptive changes in the trial's next steps."

Module B
Introduction to Gatekeeping Strategies

Module B outline

B1. Gatekeeping strategies

General regulatory considerations

Classification of gatekeeping strategies

B2. Other regulatory considerations

Secondary analyses in confirmatory trials

Section B1
Gatekeeping Strategies

Advanced multiplicity problems

Hierarchy of multiple objectives

Primary objectives

Secondary objectives

Exploratory objectives

Multiple objectives in clinical trials

Primary objectives

Directly related to the trial's outcome and presented in product label using inferential statements

Example: p-values and/or confidence intervals

Secondary objectives (Key secondary objectives)

Provide key supportive evidence and are presented in product label using inferential statements

Example: p-values and/or confidence intervals

Multiple objectives in clinical trials

Exploratory objectives (Other secondary objectives)

Play a general supportive role and are presented in product label using descriptive statements

Example: Descriptive statistics or plots (survival curves)

P-values or confidence intervals may not be used

Multiple objectives in clinical trials

FDA guidance (FDA, 2017)

"Endpoints essential to establish effectiveness for approval are called *primary endpoints*. *Secondary endpoints* may be used to support the primary endpoint(s) and/or demonstrate additional effects."

"Positive results on the secondary endpoints can be interpreted *only* if there is first a demonstration of a treatment effect on the primary endpoint family"

Advanced multiplicity problems

Hierarchy of multiple objectives

Secondary objectives

FDA guidance (FDA, 2017)

"Presenting p-values from descriptive analyses (that is, from analyses that were not prespecified and for which appropriate multiplicity adjustments were not applied) is inappropriate because doing so would imply a statistically rigorous conclusion and convey a level of certainty about the effects that is not supported by that trial."

Gatekeeping procedures

Gatekeeping procedures

Definition

Procedures for multiple families of null hypotheses

Global FWER control

Control global familywise error rate over multiple families

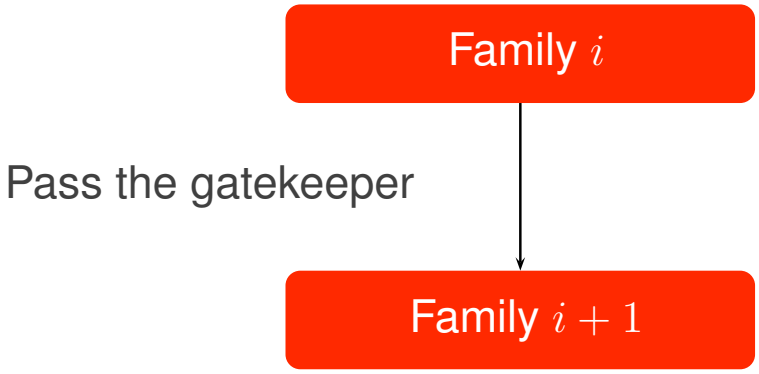
Help provide important information on secondary objectives for prescribing physicians, patients, etc

Optimal distribution of power

Maximize power by accounting for hierarchical structure of multiple families, e.g., more power for more important objectives

Gatekeeper

Family i ($i = 1, \ldots, m-1$) serves as a gatekeeper for Family $i+1$



Gatekeeping strategies

Example 3: Schizophrenia trial

Lurasidone Phase III development program in schizophrenia

Importance of using gatekeeping procedures has been emphasized in the clinical publications (Meltzer et al., 2011; Nasrallah et al., 2013)

Other examples

Gatekeeping procedures have been successfully developed in multiple Phase III programs, e.g., insomnia (Herring et al., 2016) or Type 2 diabetes programs (Frederich et al., 2012)

Gatekeeping strategies

Industry-wide guidelines

Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (Turk et al., 2008)

"To decrease the probability of a Type I error... in pain clinical trials, the use of gatekeeping procedures and other methods that correct for multiple analyses is recommended when a single primary endpoint does not adequately reflect the overall benefits of treatment"

Classification of gatekeepers

Classification of gatekeepers

Serial gatekeepers

Maurer, Hothorn and Lehmacher (1995), Westfall and Krishen (2001)

Parallel gatekeepers

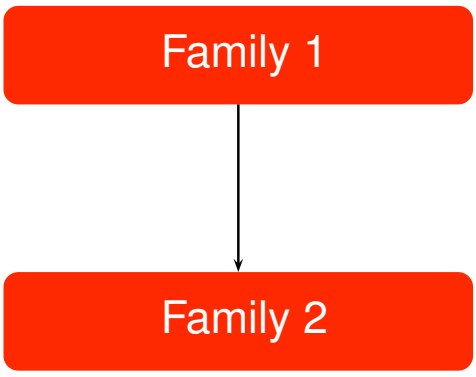
Dmitrienko, Offen and Westfall (2003), Dmitrienko and Tamhane (2009)

General gatekeepers

Dmitrienko and Tamhane (2011, 2013)

Classification of gatekeeping procedures

Sequential testing

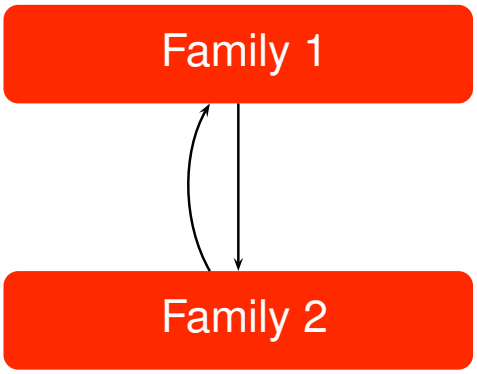


Families of hypotheses are tested sequentially starting with Family 1

Error rate is transferred along the sequence

Classification of gatekeeping procedures

Sequential testing with re-testing

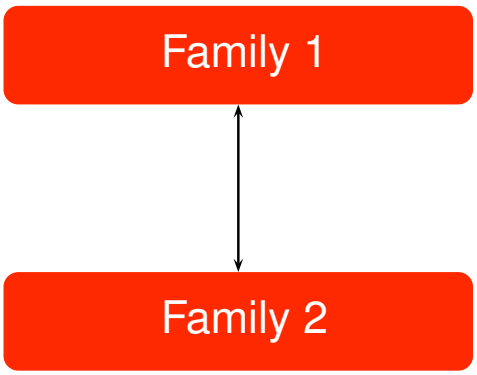


Families of hypotheses are tested sequentially starting with Family 1 with a re-testing loop

Error rate is transferred along the sequence and then back to Family 1

Classification of gatekeeping procedures

Simultaneous testing with re-testing



Families of hypotheses are tested simultaneously with potential re-testing

Error rate is transferred among families

Serial gatekeepers

Two families of null hypotheses

Serial gatekeeping

Family 1 $H_1, \ldots, H_k$

Family 2 $H_{k+1}, \ldots, H_{2k}$

Family 1 is a serial gatekeeper for Family 2: All null hypotheses must be rejected in Family 1 to proceed to Family 2

Serial gatekeeping procedures

Statement of problem

F_i is a serial gatekeeper for F_{i+1}, i.e., all null hypotheses must be rejected in F_i to proceed to F_{i+1}, $i = 1, \ldots, m-1$

Gatekeeping procedures

All tests in F_i are carried out at α level, $i = 1, \ldots, m-1$, and any procedure can be used in F_m

Serial gatekeeping procedure controls global FWER at α

Example 1: Prostate cancer trial

Primary hypothesis

H_1 (Endpoint P: Overall survival)

Secondary hypotheses

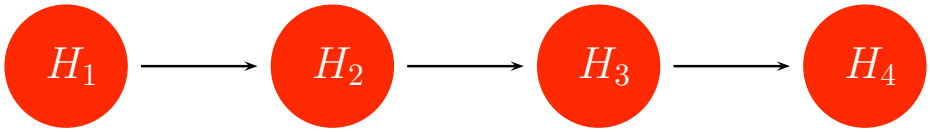
H_2 (Endpoint S1: Time to first symptomatic skeletal event)

H_3 (Endpoint S2: Occurrence of new symptomatic pathological bone fractures)

H_4 (Endpoint S3: Occurrence of spinal cord compression)

Example 1: Prostate cancer trial

Testing strategy



Sequential testing approach: Each hypothesis is tested at α provided all preceding hypotheses are rejected

Parallel gatekeepers

Two families of null hypotheses

Parallel gatekeeping

Family 1 $H_1, \ldots, H_k$

Family 2 $H_{k+1}, \ldots, H_{2k}$

Family 1 is a parallel gatekeeper for Family 2: One or more null hypotheses must be rejected in Family 1 to proceed to Family 2

Example 2: Type 2 diabetes mellitus trial

Primary comparisons

H_1 (Saxagliptin 2.5 mg q.A.M. versus placebo)

H_2 (Saxagliptin 5 mg q.A.M. versus placebo)

Secondary comparison

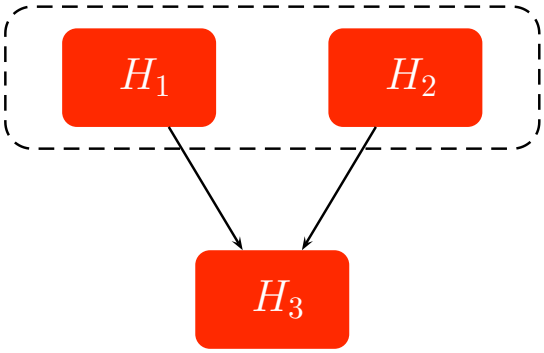
H_3 (Saxagliptin 2.5/5 mg q.A.M. versus placebo)

Example 2: Type 2 diabetes mellitus trial

Testing strategy

Primary family

Secondary family



Family 1 is a parallel gatekeeper for Family 2

General gatekeepers

Example 3: Schizophrenia trial

Primary endpoint tests

H_1 (40 mg versus placebo on the primary endpoint)

H_2 (120 mg versus placebo on the primary endpoint)

Secondary endpoint tests

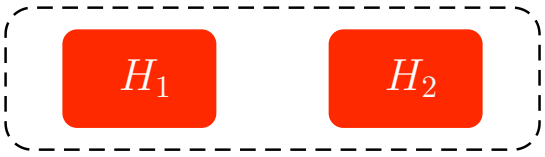
H_3 (40 mg versus placebo on the secondary endpoint)

H_4 (120 mg versus placebo on the secondary endpoint)

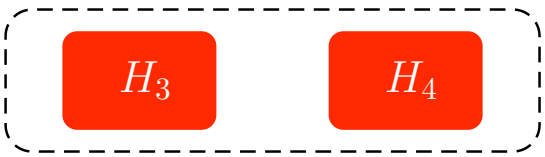
Example 3: Schizophrenia trial

Families of hypotheses

Family 1



Family 2



What are the most meaningful logical relationships among the hypotheses in Family 1 and Family 2?

Example 3: Schizophrenia trial

Logical relationships

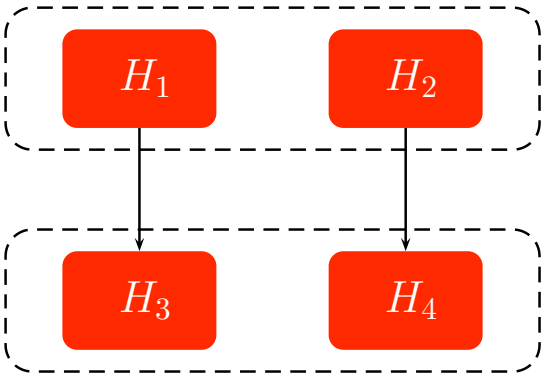
A dose-placebo test on the key secondary endpoint is relevant only if a significant treatment effect was established for this dose on the primary endpoint (Hung and Wang, 2009)

Example 3: Schizophrenia trial

Meaningful logical relationships

Family 1

Family 2

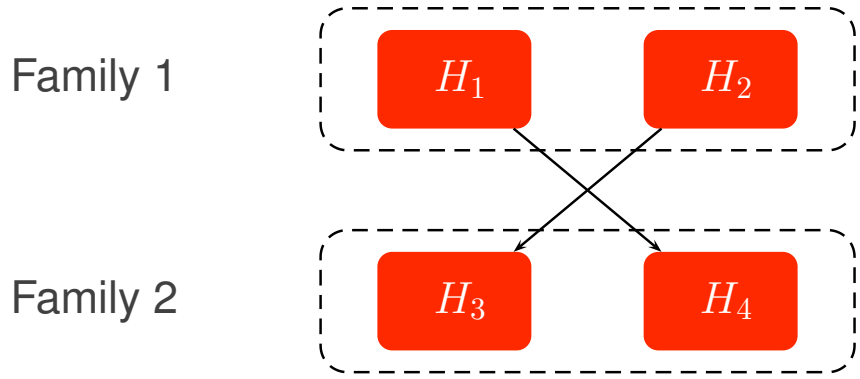


H_3 can be tested only if H_1 is rejected

H_4 can be tested only if H_2 is rejected

Example 3: Schizophrenia trial

These logical relationships are not clinically meaningful



Family 1 is not a parallel gatekeeper for Family 2

Family 1 is a general gatekeeper for Family 2

Section B2
Other Regulatory Considerations

Other regulatory considerations

Topics

Pseudospecificity

Sample size calculations in clinical trials with secondary analyses

Pooled analysis of secondary endpoints

Pseudospecificity

Multiple objectives in clinical trials

Pseudospecificity

FDA's restrictions on secondary objectives in product labels

Key secondary objectives (secondary endpoints) should provide additional information on the treatment's efficacy

Secondary objectives should not be clinically related to the primary objective

Major depressive disorder

Primary endpoint

Montgomery-Asberg Depression Rating Scale total score (MADRS)

Key secondary endpoints

Sheehan Disability Scale Global Functional Impairment score

MADRS-based remission status (MADRS total score is below 6)

Prevention of episodic migraine

Phase III development program

EVOLVE-1 and EVOLVE-2 trials

Primary endpoint

Mean change from baseline in the number of monthly migraine headache days

Key secondary endpoints

Proportion of participants with reduction from baseline $\geq 50\%$ and $\geq 75\%$ in monthly migraine headache days

Pseudospecificity

FDA guidance (FDA, 2017)

"Another example is a secondary endpoint of the percentage of patients whose symptoms are "very improved," when the primary endpoint is the percentage of patients with any amount of improvement for the same symptoms. Adjustment for multiplicity is necessary to demonstrate these additional effects."

Pseudospecificity

FDA guidance (FDA, 2017)

"An example of the second type of component analysis might be found in trials of antipsychotic drugs, in which positive and negative symptoms are domains collected in the Positive and Negative Syndrome Scale (PANSS) and often analyzed separately in addition to the overall scale. Interpretation of analyses of any subscale domain, however, is dependent on that subscale domain having been previously evaluated and determined to be valid as a stand-alone clinical measure."

Sample Size Calculations

Clinical trials with secondary endpoints

Sample size calculations

Sample size calculations are typically performed based on the primary endpoint

Should key secondary endpoints play any role in power/sample size calculations?

Sample size calculations

FDA guidance (FDA, 2017)

"If success on the secondary endpoints is important, the secondary endpoints should be considered when determining study design (e.g., sample size)."

Case study

FREEDOM trial

Phase III clinical trial for prevention of fractures in women with osteoporosis (Cummings et al., 2009)

Trial design

Novel treatment (denosumab) versus control

Endpoints

Primary endpoint and two key secondary endpoints

FREEDOM trial

Primary endpoint

Endpoint P. Risk of developing new vertebral fractures

Key secondary endpoints

Endpoint S1. Risk of developing new nonvertebral fractures (including hip fractures)

Endpoint S2. Risk of developing new hip fractures

Important considerations

Important to establish a beneficial treatment effect on very rare hip fractures

FREEDOM trial

Sample size

7868 patients enrolled in the trial

Power calculations

Endpoint P (new vertebral fractures) was powered at 99%

Endpoint S2 (new hip fractures) was powered at 91%

Sample size was driven by a secondary endpoint

Clinical trials with secondary endpoints

Analytical approach

Closed-form expressions for power/sample size calculations are not available in trials with complex objectives

Simulation-based approach

Simulation-based power/sample size calculations are recommended (Module G)

Pooled Analysis Of Secondary Endpoints

Analysis of secondary endpoints

Primary endpoint

Strict consistency criterion is applied to primary endpoints (beneficial effect must be demonstrated in each trial within a development program)

Secondary endpoints

Option 1: Analysis of secondary endpoints can be performed using the same consistency criterion (significance in each trial)

Option 2: Pooled analysis across trials within a program can be performed (Wang et al., 2013)

Analysis of secondary endpoints

Secondary endpoints

Option 1: Overly conservative approach

Option 2: Advantages of the pooled analysis include improved power of secondary analyses and lower likelihood of observing inconsistent results across the trials

Case study

Cystic fibrosis

Orkambi (lumacaftor/ivacaftor): Cystic fibrosis in patients who are homozygous for the F508del mutation in the CFTR gene (Wainwright et al., 2015)

TRAFFIC and TRANSPORT trials

Phase III program

Two trials (TRAFFIC and TRANSPORT trials)

Trial arms

Two treatment regimens (Regimen 1 and Regimen 2) versus placebo

Primary endpoint

Endpoint P. Absolute change from baseline in percent predicted FEV1 (ppFEV1)

TRAFFIC and TRANSPORT trials

Key secondary endpoints

Endpoint S1. Relative change in ppFEV1

Endpoint S2. Absolute change BMI

Endpoint S3. Absolute change in CFQ-R respiratory domain

Endpoint S4. Patients with $\geq 5\%$ increase in relative change in ppFEV1

Endpoint S5. Number of pulmonary exacerbations

TRAFFIC and TRANSPORT trials

Summary of trial results (one-sided p-values)

Endpoint	Pooled trials	
	Regimen 1	Regimen 2
S1	0.0001	0.0001
S2	0.0001	0.0004
S3	0.0071	0.0512
S4	0.0001	0.0001
S5	0.0014	0.0001

Significant results for all secondary endpoints on Regimen 1 and for Endpoints S1 and S2 on Regimen 2

TRAFFIC and TRANSPORT trials

Results presented in product label

Product label presents the key secondary results only within each trial

Endpoint	Product label	
	Regimen 1	Regimen 2
S1	Significant	Significant
S2	Non-significant	Significant
S3	Non-significant	Non-significant
S4	Non-significant	Non-significant
S5	Non-significant	Non-significant

Ordering of secondary endpoints

EMA guideline (EMA, 2017)

"The ranking of endpoints in a hierarchy can be a source of controversy. In principle, the planning and assessment of a clinical trial should prioritise those endpoints of greatest interest from a clinical perspective, but it has become common practice to rank endpoints based on the likelihood that the individual null hypothesis can be rejected. "

Clinical trial optimization

Optimal selection of endpoint ordering

Optimization approach based on a clinically relevant criterion

Publications

Review of practical approaches to clinical trial optimization with applications to optimal selection of multiplicity adjustments in Phase III trials (*Clinical Trial Optimization Using R* edited by Dmitrienko and Pulkstenis, 2017)

Module C
Parallel Gatekeeping Strategies

Module C outline

C1. Nonparametric parallel gatekeeping procedures

Simple strategies based on nonparametric procedures (chain procedures/graphical methods) in a parallel gatekeeping setting

C2. Semiparametric parallel gatekeeping procedures

More powerful strategies based on semiparametric procedures in a parallel gatekeeping setting

Section C1
Nonparametric Parallel Gatekeeping Procedures

Parallel gatekeeping procedures

Simple gatekeeping strategies

Chain procedures/graphical methods can be applied to define simple gatekeeping strategies in problems with parallel gatekeepers (Bretz et al., 2009; Burman et., 2009)

Derived from nonparametric procedures (Bonferroni and Holm) and generally less powerful than semiparametric gatekeeping procedures

Example 2: Type 2 diabetes mellitus trial

Primary comparisons

H_1 (Saxagliptin 2.5 mg q.A.M. versus placebo)

H_2 (Saxagliptin 5 mg q.A.M. versus placebo)

Secondary comparison

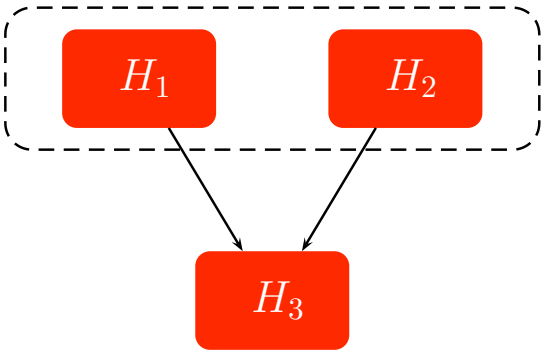
H_3 (Saxagliptin 2.5/5 mg q.A.M. versus placebo)

Example 2: Type 2 diabetes mellitus trial

Testing strategy

Primary family

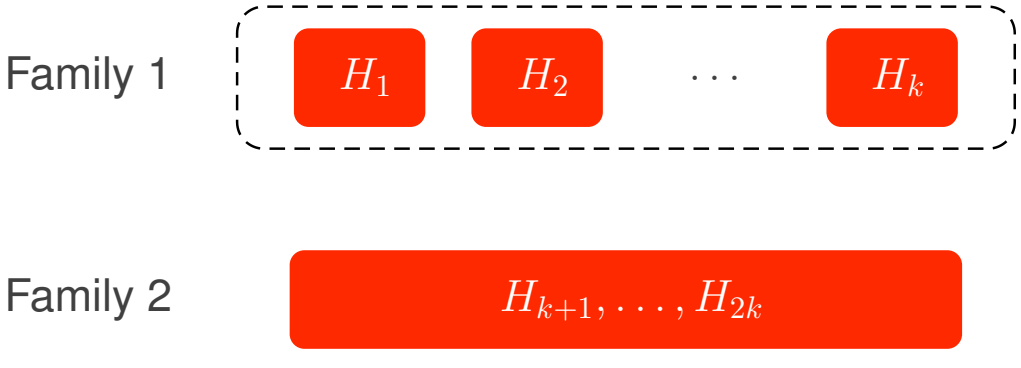
Secondary family



Family 1 is a parallel gatekeeper for Family 2

Parallel gatekeeping procedures

Trickle-down principle



Procedure 1 is applied in Family 1 at α

Procedure 2 is applied in Family 2 at $\alpha_2 \leq \alpha$ (α_2 depends on the number of null hypotheses rejected in Family 1)

Parallel gatekeeping procedures

Chain-based gatekeeping procedure

Family 1

Hypotheses are tested using Bonferroni procedure

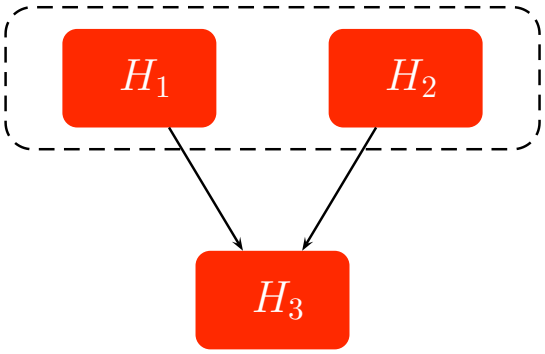
Family 2

Hypotheses are tested using an appropriate multiple testing procedure

Example 2: Type 2 diabetes mellitus trial

Chain-based gatekeeping procedure

Bonferroni procedure



Univariate test

Chain-based gatekeeping procedure

α **allocation rule**

w_1, w_2, w_3, Hypothesis weights

α **propagation rule**

Transition parameters:

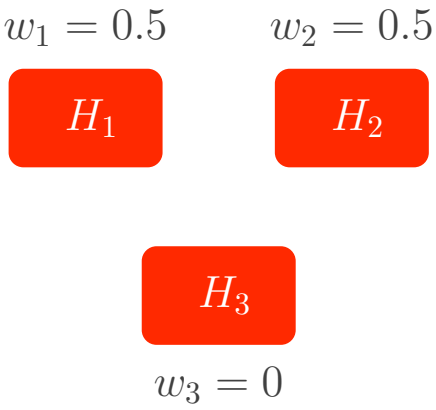
g_{12}, fraction of α carried forward from H_1 to H_2

g_{13}, fraction of α carried forward from H_1 to H_3

g_{23}, fraction of α carried forward from H_2 to H_3

Chain-based gatekeeping procedure

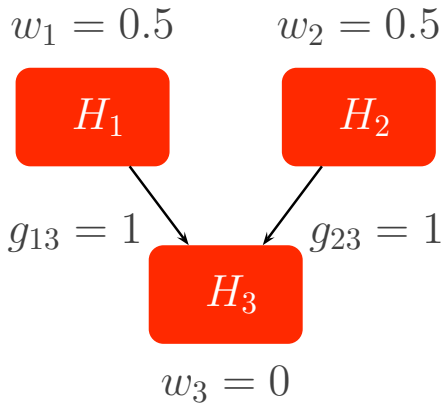
α allocation rule



Non-negative weights are assigned to H_1 and H_2 and a zero weight to H_3

Chain-based gatekeeping procedure

α propagation rule



Rules for transferring the error rate from H_1 and H_2 to H_3

Chain-based gatekeeping procedure

α allocation rule

Vector of hypothesis weights

$$W = (0.5, 0.5, 0)$$

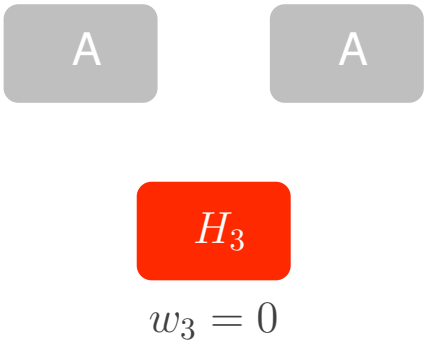
α propagation rule

Matrix of transition parameters

$$G = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Chain-based gatekeeping procedure

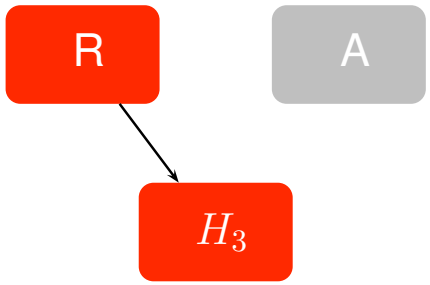
Outcomes in Family 1



If no hypotheses are rejected in Family 1, the weight of H_3 is set to 0, i.e., this hypothesis is accepted without testing

Chain-based gatekeeping procedure

Outcomes in Family 1

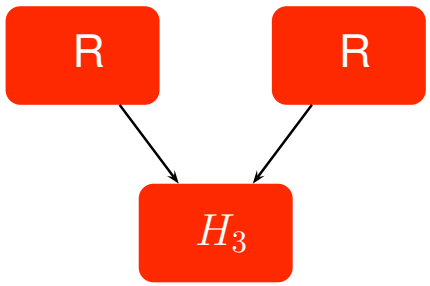


$w_3 = 0.5$

If only one hypothesis is rejected in Family 1 (e.g., H_1), the weight of H_3 is set to 0.5, i.e., this hypothesis is tested at $\alpha/2$

Chain-based gatekeeping procedure

Outcomes in Family 1



$w_3 = 1$

If both hypotheses are rejected in Family 1, the weight of H_3 is set to 1, i.e., this hypothesis is tested at the full α

Nonparametric gatekeeping strategies

Chain procedures

Memoryless procedures, which makes it difficult to deal with several families of hypotheses or account for complex logical relationships among null hypotheses, e.g., general gatekeepers

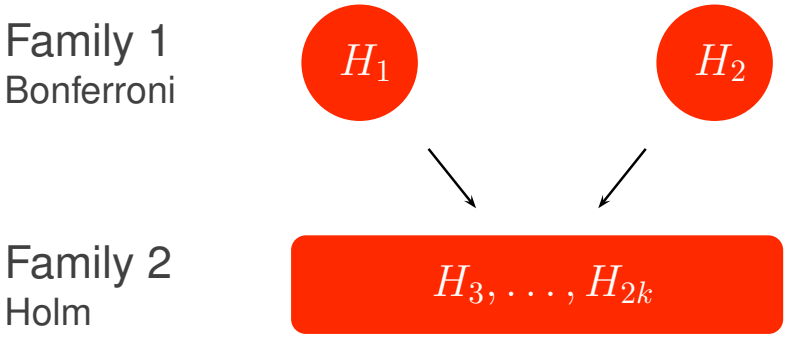
Advanced chain procedures

Chain procedures based on entangled graphs (Maurer and Bretz, 2013) support more complex types of logical relationships

Section C2
Semiparametric Parallel Gatekeeping Procedures

Nonparametric gatekeeping procedure

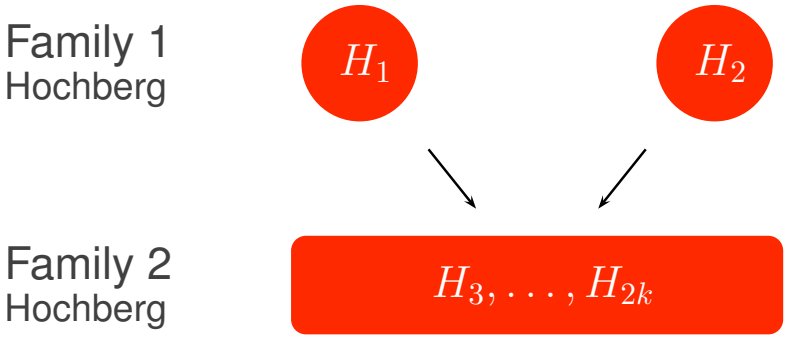
Chain-based gatekeeping procedure



α-propagation rule: If H_1 is rejected, $\alpha/2$ is transferred to Family 2 and, if H_2 is rejected, $\alpha/2$ is transferred to Family 2

Semiparametric gatekeeping procedure

More powerful Hochberg-based procedure



Parallel gatekeeping procedure with Hochberg-based components

Semiparametric gatekeeping procedures

Generous component procedure (Bonferroni) in Family 1

A fraction of α can be transferred to Family 2 even if some null hypotheses are accepted in Family 1

Greedy component procedure (Hochberg) in Family 1

All α is used up and nothing can be transferred to Family 2 if some null hypotheses are accepted in Family 1

Semiparametric gatekeeping procedures

Methodology

General method for building multistage parallel gatekeeping procedures (Dmitrienko, Tamhane and Wiens, 2008)

Key features

Based on a simple algorithm which streamlines implementation and communication of results

Defined using component procedures and have a flexible structure

Semiparametric gatekeeping procedures

Family 1

$F_1 = \{H_1, \ldots, H_k\}$, null hypotheses

$N_1 = \{1, \ldots, k\}$, index set

Procedure 1: Separable (generous) procedure with local FWER control within F_1

Family 2

$F_2 = \{H_{k+1}, \ldots, H_{2k}\}$, null hypotheses

$N_2 = \{k+1, \ldots, 2k\}$, index set

Procedure 2: Any procedure with local FWER control within F_2

α propagation rule

Family 1

Procedure 1 is applied at $\alpha_1 = \alpha$ level

$A_1 \subseteq N_1$, index set of null hypotheses accepted in F_1

Family 2

Procedure 2 is applied at α_2 level

$\alpha_2 = \alpha_1 - e(A_1)$, where $e(I)$ is the error rate function of Procedure 1

Property

Global FWER is controlled at α

Error rate function

Definition

Select an subset of the null hypotheses and assume they are true, i.e., H_i, $i \in I$, and $I \subseteq N_1$

Error rate function is the error rate for this subset of null hypothesis, i.e., probability of rejecting at least one true null hypothesis in the subset

$$e(I) = P\left\{ \bigcup_{i \in I} (\text{Reject } H_i) \middle| H_I \right\},$$

Error rate function

Properties

$e(\emptyset) = 0$ (no accepted hypotheses)

$e(N_1) = \alpha$ (all hypotheses are accepted)

Example

Error rate function of Bonferroni procedure is $e(I) = \alpha|I|/k$, where $|I|$ is number of elements in index set I

Properties of gatekeeping procedure

All null hypotheses are rejected in Family 1

A_1 is empty

$\alpha_2 = \alpha_1 - e(\emptyset) = \alpha_1 - 0 = \alpha$

Null hypotheses in Family 2 are tested at full α level

No null hypotheses are rejected in Family 1

$A_1 = N_1$

$\alpha_2 = \alpha_1 - e(N_1) = \alpha_1 - \alpha_1 = 0$

Null hypotheses in Family 2 are not tested

Separable procedures

Separability condition

Procedure 1 is separable (generous) if $e(I) < \alpha$ provided I is a proper subset of N_1

Implication

If a separable procedure is used in Family 1, a fraction of α can be carried over to Family 2 if one or more null hypotheses are rejected in Family 1

Bonferroni procedure

Three null hypotheses in Family 1

Index set I	Error rate function $e(I)$
$N_1 = \{1, 2, 3\}$	α
$\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$	$2\alpha/3$
$\{1\}$, $\{2\}$, $\{3\}$	$\alpha/3$
Empty	0

Error rate function of Bonferroni procedure is $e(I) = \alpha|I|/k$

Bonferroni procedure is separable (generous) because $e(I) < \alpha$ if I is a proper subset of N_1

Hochberg procedure

Three null hypotheses in Family 1

Index set I	Error rate function $e(I)$
$N_1 = \{1, 2, 3\}$	α
$\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$	α
$\{1\}$, $\{2\}$, $\{3\}$	α
Empty	0

Error rate function of Hochberg procedure is $e(I) = \alpha$ unless I is empty

Hochberg procedure is not separable (greedy)

Separability condition

Separability procedures

Most powerful procedures (Holm, Hochberg and Hommel procedures) are not separable

Truncated procedures

Separable versions of popular procedures (known as truncated procedures) can be constructed

Truncated procedure is based on a convex combination between a multiple procedure and Bonferroni procedure

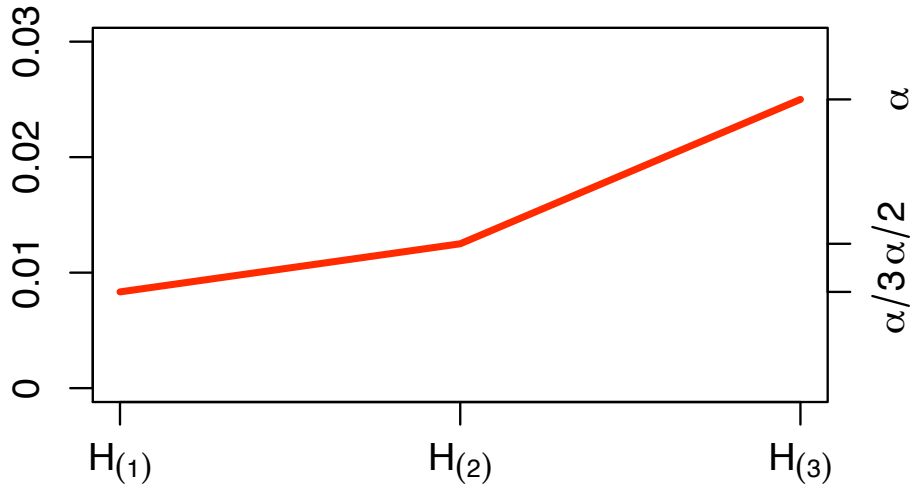
Truncated procedures

Truncated Hochberg procedure

Truncated Hochberg procedure is defined using a convex combination between Bonferroni and Hochberg procedures

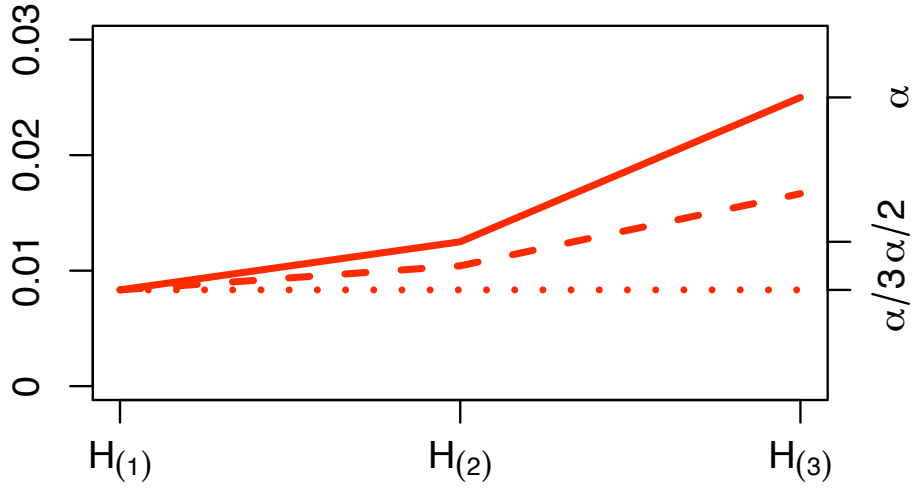
Regular Hochberg procedure

Three null hypotheses in Family 1 ($\alpha = 0.025$)



Truncated Hochberg procedure

Three null hypotheses in Family 1 ($\alpha = 0.025$)



—, Regular Hochberg procedure; - - -, Truncated Hochberg procedure; · · ·, Bonferroni procedure

Truncated Hochberg procedure

Significance levels for $H_{(1)}$, $H_{(2)}$ and $H_{(3)}$

Bonferroni procedure	Truncated Hochberg procedure	Hochberg procedure
$\alpha/3$	$(1-\gamma)\alpha/3+\gamma\alpha/3$	$\alpha/3$
$\alpha/3$	$(1-\gamma)\alpha/3+\gamma\alpha/2$	$\alpha/2$
$\alpha/3$	$(1-\gamma)\alpha/3+\gamma\alpha/1$	$\alpha/1$

$0 \le \gamma \le 1$, truncation parameter (greediness parameter)

Truncated Hochberg procedure

Decision rules

Reject $H_{(1)}$, $H_{(2)}$ and $H_{(3)}$ if

$$p_{(3)} \le (1-\gamma)\alpha/3+\gamma\alpha$$

Otherwise reject $H_{(1)}$ and $H_{(2)}$ if

$$p_{(2)} \le (1-\gamma)\alpha/3+\gamma\alpha/2$$

Otherwise reject $H_{(1)}$ if

$$p_{(1)} \le (1-\gamma)\alpha/3+\gamma\alpha/3=\alpha/3$$

Truncated Hochberg procedure

Special cases

Truncated Hochberg procedure simplifies to Bonferroni procedure if $\gamma = 0$ (generous/ separable)

Truncated Hochberg procedure simplifies to regular Hochberg procedure if $\gamma = 1$ (greedy/ non-separable)

Separability

Truncated Hochberg procedure is separable and can be used in parallel gatekeeping procedures if $0 \leq \gamma < 1$

Comparison of nonparametric and semiparametric gatekeeping procedures

Example 2: Type 2 diabetes mellitus trial

Three hypothesis tests

Family	Hypothesis	Raw p-value
Family 1	H_1	0.0082
	H_2	0.0174
Family 2	H_3	0.0202

Global familywise error rate: One-sided $\alpha = 0.025$

Example 2: Type 2 diabetes mellitus trial

Nonparametric parallel gatekeeping procedure

Family 1: Bonferroni procedure

Family 2: Univariate test

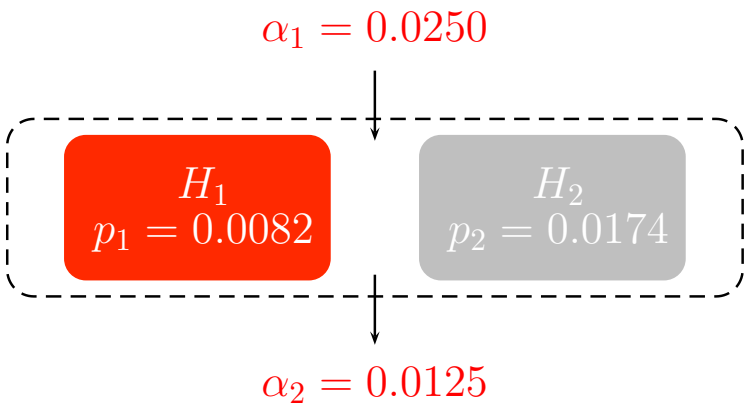
Semiparametric parallel gatekeeping procedure

Family 1: Truncated Hochberg procedure with $\gamma = 0.7$

Family 2: Univariate test

Nonparametric gatekeeping procedure

Bonferroni procedure



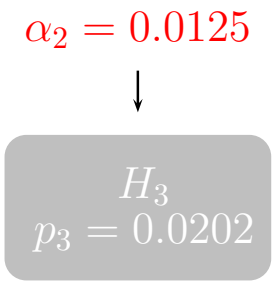
Test H_1 and H_2 at $\alpha_1/2 = 0.0125$

H_1 is rejected and H_2 is accepted

$\alpha_2 = \alpha_1/2 = 0.0125$ is carried over to Family 2

Nonparametric gatekeeping procedure

Univariate test



Test H_3 at $\alpha_2 = 0.0125$

H_3 is accepted

Semiparametric gatekeeping procedure

Family 1

Truncated Hochberg procedure with $0 \leq \gamma < 1$ at $\alpha_1 = \alpha$

Decision rules

$p_{(1)} < p_{(2)}$, Ordered p-values

Reject $H_{(1)}$ and $H_{(2)}$ if $p_{(2)} \leq (1-\gamma)\alpha/2 + \gamma\alpha = (1+\gamma)\alpha/2$

Reject $H_{(1)}$ only if $p_{(1)} \leq \alpha/2$ and $p_{(2)} > (1+\gamma)\alpha/2$

Semiparametric gatekeeping procedure

Family 2

Regular Hochberg procedure at $\alpha_2 = \alpha_1 - e(A_1)$

$e(I)$, error rate function of truncated Hochberg procedure

$$e(I) = [\gamma + (1-\gamma)|I|/2]\alpha$$

α propagation rule in Family 2

No null hypotheses are rejected in Family 1

$A_1 = \{1, 2\}$

$\alpha_2 = \alpha - e(A_1) = \alpha - \alpha = 0$

One null hypothesis is rejected in Family 1

$A_1 = \{1\}$ or $\{2\}$

$\alpha_2 = \alpha - e(A_1) = \alpha - [\gamma + (1-\gamma)/2]\alpha = (1-\gamma)\alpha/2$

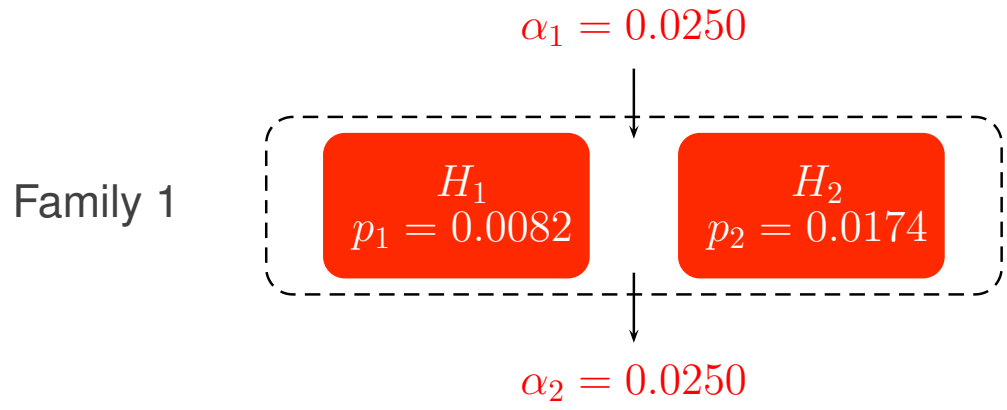
Two null hypotheses are rejected in Family 1

A_1 is empty

$\alpha_2 = \alpha - e(A_1) = \alpha - 0 = \alpha$

Semiparametric gatekeeping procedure

Truncated Hochberg procedure ($\gamma = 0.7$)



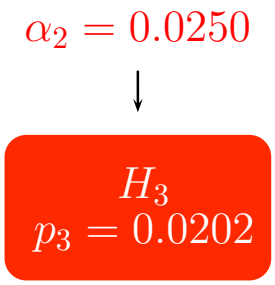
H_1 and H_2 are rejected since $p_{(2)} = p_2 \leq (1+\gamma)\alpha_1/2 = 0.02125$

$\alpha_2 = \alpha_1 = 0.025$ is carried over to Family 2

Semiparametric gatekeeping procedure

Univariate test

Family 2



Test H_3 at $\alpha_2 = 0.0250$

H_3 is rejected

Comparison of gatekeeping procedures

Nonparametric parallel gatekeeping procedure

One null hypothesis is rejected

Semiparametric parallel gatekeeping procedure

Three null hypotheses are rejected

More powerful component is used in Family 1 (truncated Hochberg is uniformly more powerful than Bonferroni)

In general, a more powerful component can be used in Family 2

Comparison of gatekeeping procedures

Selection of truncation parameter γ

Optimal value of γ depends on the effect sizes in Families 1 and 2 (Dmitrienko et al., 2011)

Clinical trial optimization methods can be applied to select truncation parameters in semiparametric gatekeeping procedures

Review of general approaches to clinical trial optimization and applications to trials with multiple objectives (*Clinical Trial Optimization Using R* edited by Dmitrienko and Pulkstenis, 2017)

Summary

Nonparametric gatekeeping strategies

Derived from nonparametric procedures, e.g., Bonferroni- or Holm-based gatekeeping procedures, which results in power loss

Semiparametric gatekeeping strategies

Derived from semiparametric procedures, e.g., Hochberg- or Hommel-based gatekeeping procedures, which leads to improved power

Rely on straightforward testing algorithms and are easy to implement

Software implementation

Software implementation in SAS

Custom macros

Semiparametric parallel gatekeeping procedures: MIXGATE macro

For more information and examples, see Brechenmacher and Dmitrienko (2017)

Semiparametric gatekeeping procedure

Example 2: Type 2 diabetes mellitus trial

Family 1: Truncated Hochberg procedure with $\gamma = 0.7$

Family 2: Univariate test

P-values

Family 1: $p_1 = 0.0082$ and $p_2 = 0.0174$

Family 2: $p_3 = 0.0202$

Semiparametric gatekeeping procedure in SAS

MIXGATE macro

```
data ex2;
    input hyp $ family parallel $ serial $ rawp;
    datalines;
    H1 1 000 000 0.0082
    H2 1 000 000 0.0174
    H3 2 110 000 0.0202
run;

%MixGate(indata=ex2, method=Standard,
         test=Hochberg, gamma=0.7,
         adjpout=adjp);

proc print data=adjp noobs label;
    var adj_p1-adj_p3;
run;
```

Semiparametric gatekeeping procedure in SAS

MIXGATE macro: Output

```
0.0164 0.0205 0.0205
```


Software implementation in R

Mediana package

AdjustPvalues function: Nonparametric and semiparametric gatekeeping procedures

Web sites

http://biopharmnet.com/mediana

https://cran.r-project.org/web/packages/Mediana/

Online manual

http://gpaux.github.io/Mediana/

Semiparametric gatekeeping procedure in R

AdjustPvalues function

```
rawp=c(0.0082,0.0174,0.0202)
families=families(family1=c(1, 2),
                  family2=c(3))
component.procedures=
              families(family1 ="HochbergAdj",
                       family2="HochbergAdj")
gamma=families(family1=0.7,
               family2=1)
adjp=AdjustPvalues(rawp,
          proc="ParallelGatekeepingAdj",
          par=parameters(family=families,
                         proc=component.procedures,
                         gamma=gamma))
round(adjp, 4)
```


Semiparametric gatekeeping procedure in R

AdjustPvalues function: Output

```
0.0164 0.0205 0.0205
```

Module D
General Gatekeeping Strategies

Module D outline

D1. General gatekeeping procedures

Powerful strategies based on semiparametric procedures in multiplicity problems with general gatekeepers

Section D1
General Gatekeeping Strategies

Example 3: Schizophrenia trial

Primary endpoint tests

H_1 (Dose L versus placebo on the primary endpoint)

H_2 (Dose H versus placebo on the primary endpoint)

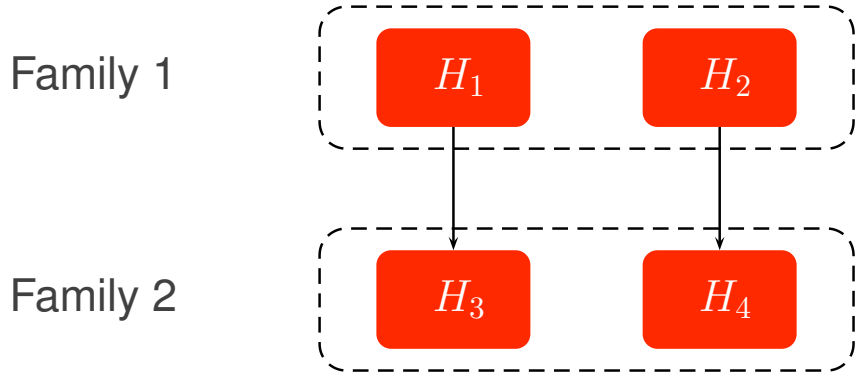
Secondary endpoint tests

H_3 (Dose L versus placebo on the secondary endpoint)

H_4 (Dose H versus placebo on the secondary endpoint)

Example 3: Schizophrenia trial

Meaningful logical relationships



H_3 can be tested only if H_1 is rejected

H_4 can be tested only if H_2 is rejected

Family 1 is a general gatekeeper for Family 2

General logical relationships

Mixture-based methodology

Mixture-based method for constructing gatekeeping procedures with general logical relationships (Dmitrienko and Tamhane, 2011, 2013)

Can be used with any combination of multiple testing procedures (nonparametric, semiparametric and fully parametric procedures)

Two-family testing problem

Family 1

$F_1 = \{H_1, \ldots, H_k\}$, null hypotheses

$N_1 = \{1, \ldots, k\}$, index set

Procedure 1: Separable procedure with local FWER control within Family 1, e.g., truncated Hochberg

Family 2

$F_2 = \{H_{k+1}, \ldots, H_{2k}\}$, null hypotheses

$N_2 = \{k+1, \ldots, 2k\}$, index set

Procedure 2: Any procedure with local FWER control within Family 2, e.g., regular Hochberg

General logical relationships

Goal

Logical relationships need to be defined at the null hypothesis level rather than at the family level

Simple logical relationships among multiple families

Serial gatekeepers: All null hypotheses in Family 2 are testable if all null hypotheses in Family 1 are rejected

Parallel gatekeepers: All null hypotheses in Family 2 are testable if one or more null hypotheses in Family 1 are rejected

Restriction functions

Logical restrictions

Logical relationships among null hypotheses in Families 1 and 2 are specified using a family of binary restriction functions

$L_i(I_1) = 0$ or 1, where $i \in N_2$ and $I_1 \subseteq N_1$

Testable null hypotheses

Null hypothesis in Family 2, e.g., H_i, $i \in N_2$, is testable if and only if $L_i(R_1) = 1$

R_1, index set of null hypotheses rejected in Family 1

Example 3: Schizophrenia trial

Restriction functions

Set of rejected null hypotheses	Null hypothesis	Testable
H_1, H_2	H_3	Yes
H_1	H_3	Yes
H_2	H_3	No
Empty	H_3	No
H_1, H_2	H_4	Yes
H_1	H_4	No
H_2	H_4	Yes
Empty	H_4	No

Mixture procedures

Three ingredients

Component procedures:

- Procedure 1: Local FWER control within F_1
- Procedure 2: Local FWER control within F_2

Mixing function

Mixture procedure

Component procedures and mixing function define a mixture procedure which takes into account logical relationships and controls global FWER in Families 1 and 2

Example 3: Schizophrenia trial

Procedure 1 (Family 1)

Bonferroni procedure (nonparametric procedure) ignores within-family correlation

Procedure 2 (Family 2)

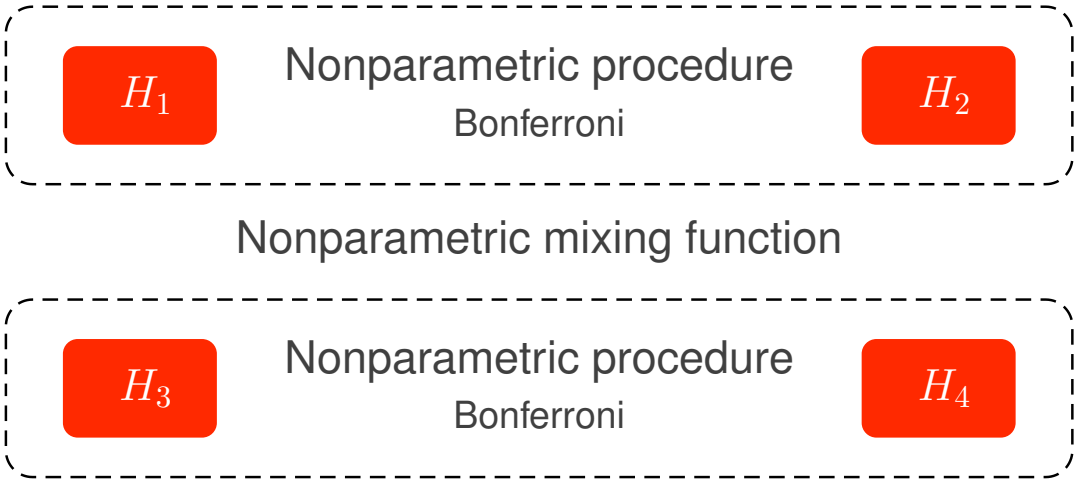
Bonferroni procedure (nonparametric procedure) ignores for within-family correlation

Mixing function

Nonparametric mixing function ignores between-family correlation

Example 3: Schizophrenia trial

Distributional relationships



Bonferroni-based gatekeeping procedure that ignores within-family and between-family correlations

Example 3: Schizophrenia trial

Procedure 1 (Family 1)

Dunnett procedure (parametric procedure) to account for within-family correlation

Procedure 2 (Family 2)

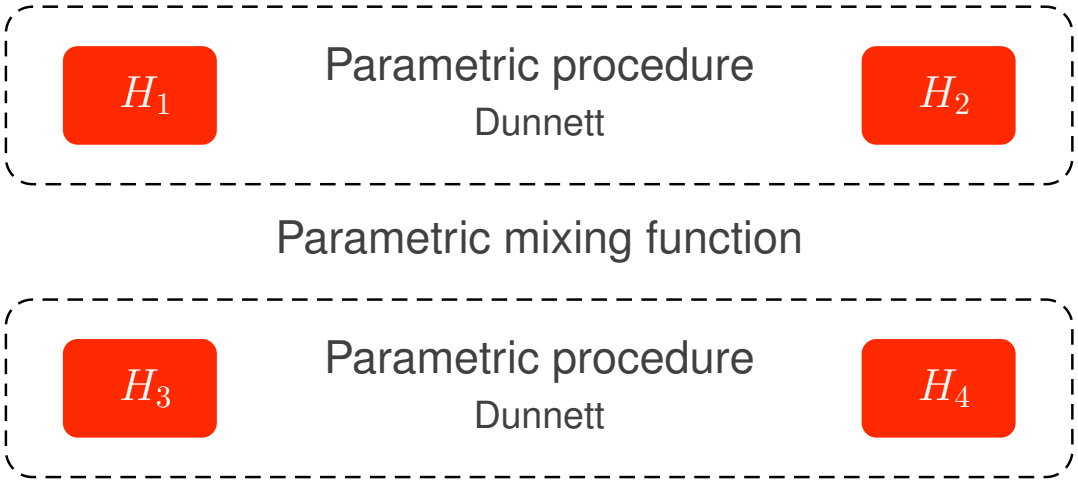
Dunnett procedure (parametric procedure) to account for within-family correlation

Mixing function

Parametric mixing function to account for between-family correlation

Example 3: Schizophrenia trial

Distributional relationships



Parametric gatekeeping procedure that accounts for within-family and between-family correlations

Example 3: Schizophrenia trial

Procedure 1 (Family 1)

Truncated Hochberg procedure (semiparametric procedure) to account for within-family correlation

Procedure 2 (Family 2)

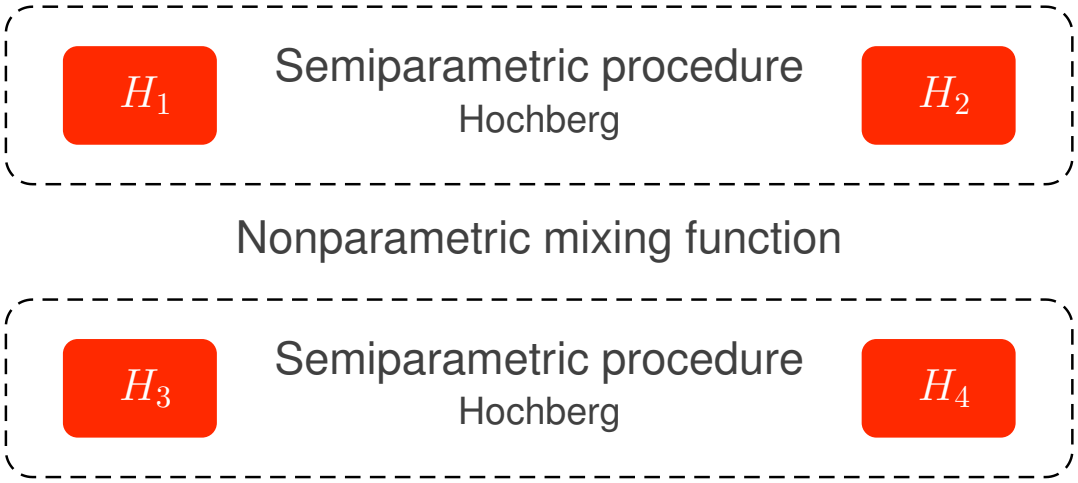
Hochberg procedure (semiparametric procedure) to account for within-family correlation

Mixing function

Nonparametric mixing function to account for between-family correlation

Example 3: Schizophrenia trial

Distributional relationships



Semiparametric gatekeeping procedure that accounts for within-family correlations only

Example 3: Schizophrenia trial

Procedure 1 (Family 1)

Truncated Hochberg procedure takes advantage of positive within-family correlation

Procedure 2 (Family 2)

Regular Hochberg to account takes advantage of positive within-family correlation

Mixing function

Nonparametric mixing function (between-family correlation is unknown)

Mixture-based gatekeeping procedures

Mixture-based gatekeeping procedures

Various extensions of original mixture-based methodology

Modified mixture-based gatekeeping procedures (Dmitrienko, Kordzakhia and Brechenmacher, 2016) are more powerful than standard mixture-based gatekeeping procedures and support stepwise testing algorithms

Enhanced mixture-based gatekeeping procedures (Kordzakhia et al., 2018) are more powerful than modified mixture-based gatekeeping procedures

Example 3: Schizophrenia trial

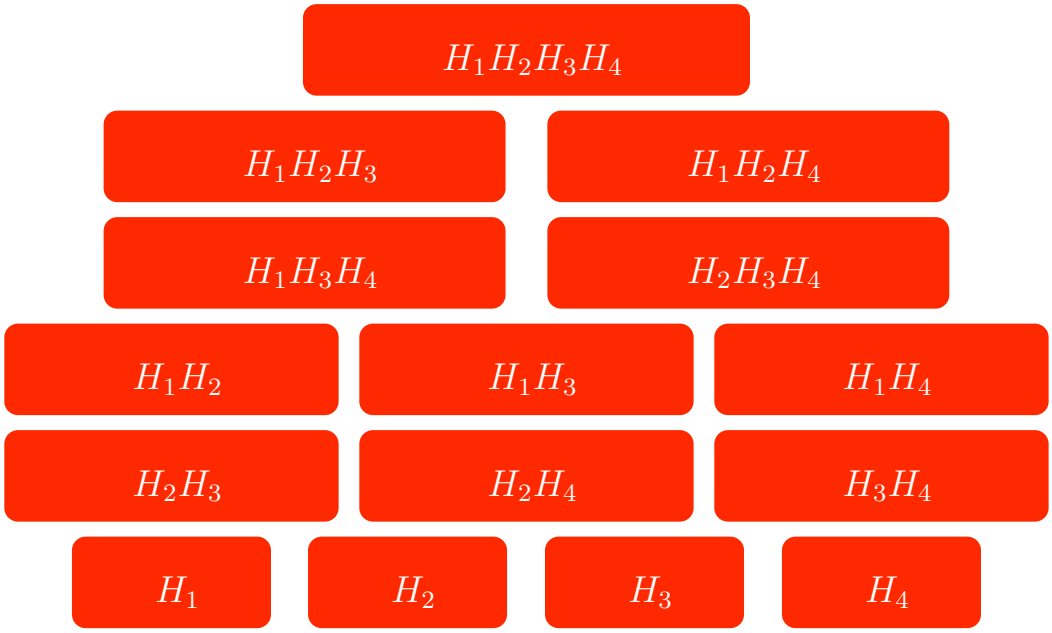
Closed testing procedure

Step 1: Define closed family which includes all possible intersections of H_1, H_2, H_3 and H_4

Step 2: Define α-level local tests for all intersection hypotheses

Step 3: Define decision rules: Reject a null hypothesis if all intersection hypotheses containing this null hypothesis are rejected by local tests

Step 1: Define closed family



Closed family associated with H_1, H_2, H_3 and H_4

Step 2: Define local tests

Example: Bonferroni-based gatekeeping procedure

Local test for each intersection hypothesis is based on minimum-p-value tests

By Bonferroni inequality, each local test is an α-level test and thus Bonferroni procedure controls FWER at α

Note

This procedure does not account for logical relationships and is used only for illustration purposes

Step 2: Define local tests

$H_1H_2H_3H_4$: $\min(p_1, p_2, p_3, p_4) \leq \alpha/4$

$H_1H_2H_3$: $\min(p_1, p_2, p_3) \leq \alpha/4$

$H_1H_2H_4$: $\min(p_1, p_2, p_4) \leq \alpha/4$

$H_1H_3H_4$: $\min(p_1, p_3, p_4) \leq \alpha/4$

$H_2H_3H_4$: $\min(p_2, p_3, p_4) \leq \alpha/4$

H_1H_2: $\min(p_1, p_2) \leq \alpha/4$

H_1H_3: $\min(p_1, p_3) \leq \alpha/4$

H_1H_4: $\min(p_1, p_4) \leq \alpha/4$

H_2H_3: $\min(p_2, p_3) \leq \alpha/4$

H_2H_4: $\min(p_2, p_4) \leq \alpha/4$

H_3H_4: $\min(p_3, p_4) \leq \alpha/4$

H_1: $p_1 \leq \alpha/4$

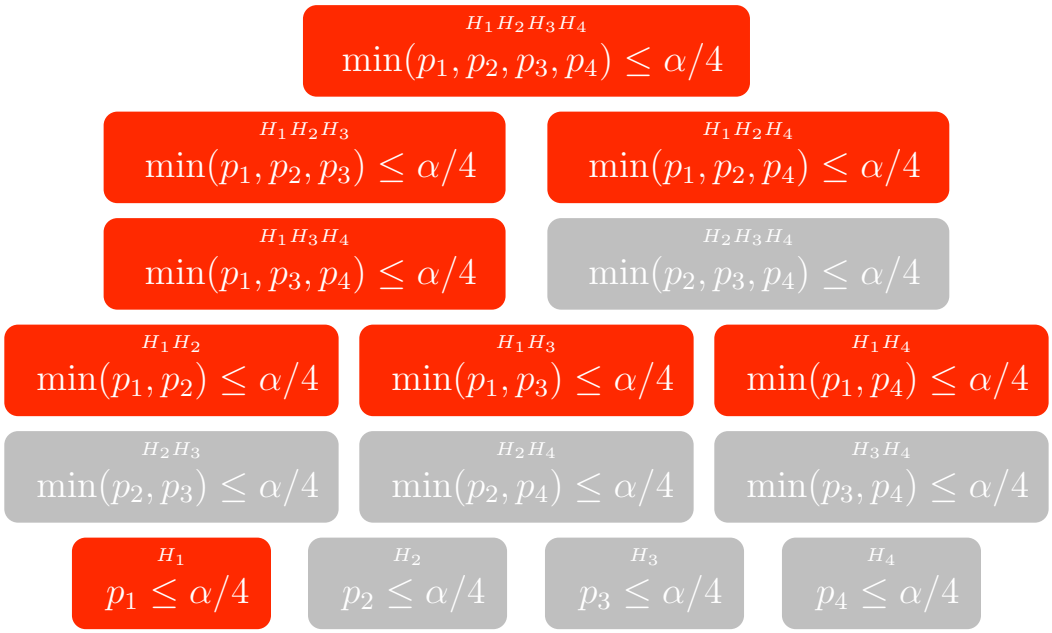
H_2: $p_2 \leq \alpha/4$

H_3: $p_3 \leq \alpha/4$

H_4: $p_4 \leq \alpha/4$

Each local test is an α-level test

Step 3: Define decision rules



H_1 is rejected if all highlighted intersections are rejected

Logical restrictions

Account for logical restrictions among null hypotheses

Local tests are modified to ensure that the gatekeeping procedure is consistent with logical restrictions

Restricted local tests are defined based on pre-specified restrictions

Example 3: Schizophrenia trial

Logical restrictions

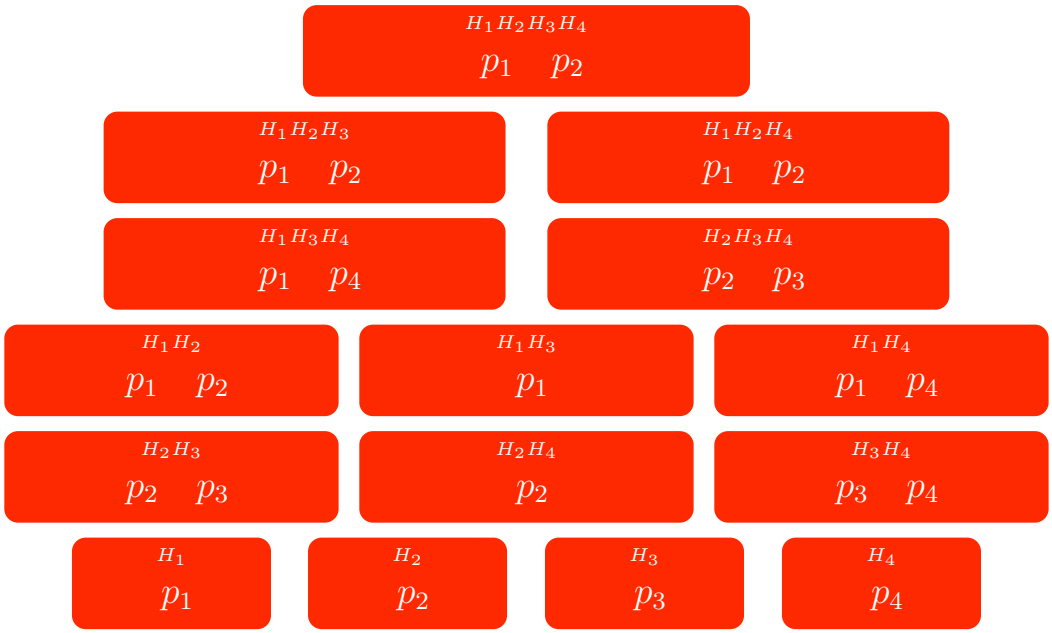
H_3 can be tested only if H_1 is rejected and H_4 can be tested only if H_2 is rejected

Restricted local tests

If an intersection contains H_1 and H_3, p_3 will be removed from the local test

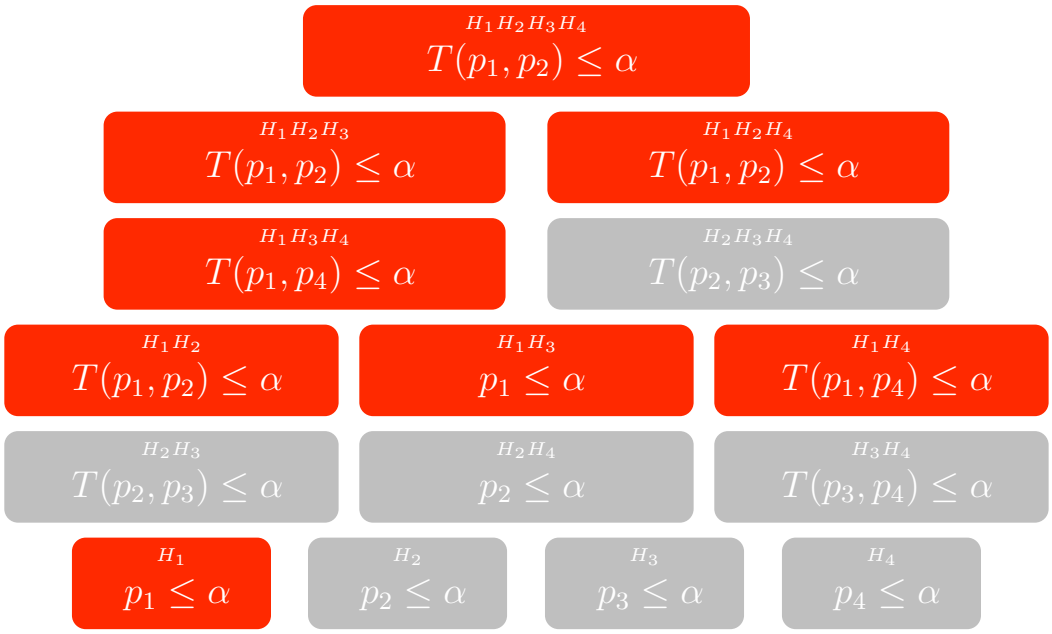
If an intersection contains H_2 and H_4, p_4 will be removed from the local test

Step 3: Restricted local tests



P-values used in each restricted local test

Step 3: Decision rules



H_1 is rejected if all highlighted intersections are rejected

Stepwise testing algorithm

Mixture-based gatekeeping procedures

Closed testing algorithm

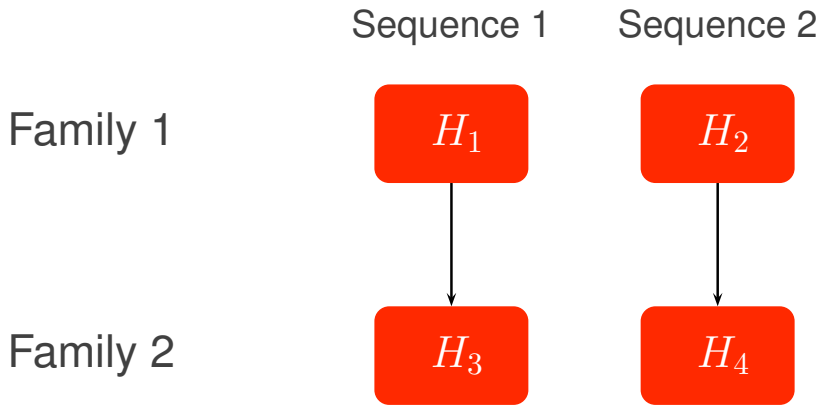
Multiplicity-adjusted p-values can be computed using the closure principle

Stepwise testing algorithm

Modified mixture-based gatekeeping procedures support a convenient stepwise testing algorithm in problems with two sequences of hypotheses (Dmitrienko, Kordzakhia and Brechenmacher, 2016)

Example 3: Schizophrenia trial

Two sequences of null hypotheses



Sequence 1: H_1 followed by H_3

Sequence 2: H_2 followed by H_4

Example 3: Schizophrenia trial

Four hypothesis tests

Family	Hypothesis	Raw p-value
Family 1	H_1	0.0101
	H_2	0.0233
Family 2	H_3	0.0022
	H_4	0.0167

Global familywise error rate: One-sided $\alpha = 0.025$

Example 3: Schizophrenia trial

Modified mixture-based gatekeeping procedure with a stepwise testing algorithm

Family 1: Truncated Hochberg procedure with $\gamma = 0.7$

Family 2: Regular Hochberg procedure if both hypotheses are rejected in Family 1 or univariate test if only one hypothesis is rejected in Family 1 (due to logical restrictions)

General gatekeeping procedure

Family 1

Truncated Hochberg procedure with $0 \leq \gamma < 1$ at $\alpha_1 = \alpha$

Decision rules

$p_{(1)} < p_{(2)}$, Ordered p-values

Reject $H_{(1)}$ and $H_{(2)}$ if $p_{(2)} \leq (1-\gamma)\alpha/2 + \gamma\alpha = (1+\gamma)\alpha/2$

Reject $H_{(1)}$ only if $p_{(1)} \leq \alpha/2$ and $p_{(2)} > (1+\gamma)\alpha/2$

α propagation rule in Family 2

No null hypotheses are rejected in Family 1

$A_1 = \{1, 2\}$

$\alpha_2 = \alpha - e(A_1) = \alpha - \alpha = 0$

One null hypothesis is rejected in Family 1

$A_1 = \{1\}$ or $\{2\}$

$\alpha_2 = \alpha - e(A_1) = \alpha - [\gamma + (1-\gamma)/2]\alpha = (1-\gamma)\alpha/2$

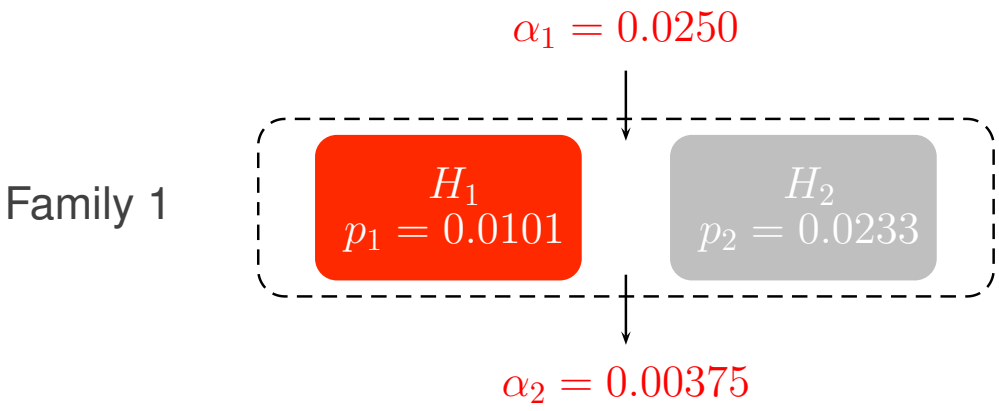
Two null hypotheses are rejected in Family 1

A_1 is empty

$\alpha_2 = \alpha - e(A_1) = \alpha - 0 = \alpha$

General gatekeeping procedure

Truncated Hochberg procedure ($\gamma = 0.7$)

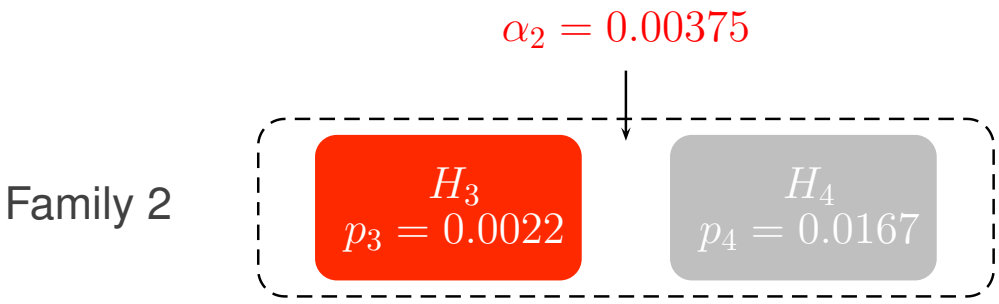


H_1 is rejected and H_2 is accepted since $p_{(2)} = p_2 > (1+\gamma)\alpha_1/2 = 0.02125$ but $p_{(1)} = p_1 \leq \alpha_1/2 = 0.0125$

$\alpha_2 = (1-\gamma)\alpha/2 = 0.00375$ is carried over to Family 2

General gatekeeping procedure

Univariate test



Test H_3 at $\alpha_2 = 0.00375$

H_3 is rejected and H_4 is not tested because this hypothesis is no longer relevant

Software implementation

Software implementation in SAS

Custom macros

General mixture-based gatekeeping procedures: MIXGATE macro

General gatekeeping procedure

Example 3: Schizophrenia trial

Family 1: Truncated Hochberg procedure with $\gamma = 0.7$

Family 2: Hochberg procedure

P-values

Family 1: $p_1 = 0.0101$ and $p_2 = 0.0233$

Family 2: $p_3 = 0.0022$ and $p_4 = 0.0167$

General gatekeeping procedure in SAS

MIXGATE macro

```
data ex3;
    input hyp $ family parallel $ serial $ rawp;
    datalines;
    H1 1 0000 0000 0.0101
    H2 1 0000 0000 0.0233
    H3 2 0000 1000 0.0022
    H4 2 0000 0100 0.0167
run;

%MixGate(indata=ex3, method=Standard,
         test=Hochberg, gamma=0.7,
         adjpout=adjp);

proc print data=adjp noobs label;
    var adj_p1-adj_p4;
run;
```

General gatekeeping procedure in SAS

MIXGATE macro: Output

```
0.0202 0.0274 0.0202 0.0274
```


Software implementation in R

Mediana package

AdjustPvalues function: General mixture-based gatekeeping procedures

General gatekeeping procedure in R

AdjustPvalues function

```
rawp=c(0.0101,0.0233,0.0022,0.0167)
families=families(family1=c(1, 2),
                  family2=c(3, 4))
component.procedures=
              families(family1="HochbergAdj",
                       family2="HochbergAdj")
gamma=families(family1=0.7,
               family2=1)
adjp=AdjustPvalues(rawp,
          proc="MultipleSequenceGatekeepingAdj",
          par=parameters(family=families,
                         proc=component.procedures,
                         gamma=gamma))
round(adjp, 4)
```


General gatekeeping procedure in R

AdjustPvalues function: Output

```
0.0202 0.0274 0.0202 0.0274
```

Summary

Summary

Gatekeeping procedures

Multiple testing procedures for several families of null hypotheses

Control global familywise error rate across families

Account for hierarchical structure of multiplicity problems (logical relationships among families of null hypotheses)

Enable clinical trial sponsors to enrich product labels by including key secondary findings

Module E
Group-sequential Trials with Multiple Objectives

Module E outline

E1. Introduction to group-sequential trials

Traditional group-sequential trials

E2. Group-sequential trials with multiple objectives

Group-sequential trials with several multiplicity components, e.g., multiple endpoints and multiple doses, etc

Section E1
Introduction to
Group-sequential Trials

Interim decision making

Sequential monitoring of safety and efficacy data

Interim analyses have become integral part of confirmatory Phase III trials

Ethical requirements

Imperative to ensure that patients are not exposed to inferior or harmful therapies

Financial considerations

Early stopping due to futility is desirable if a trial is unlikely to achieve its objectives

Decision rules at interim analyses

Early stopping

Early stopping due to futility or superior benefit

Trials may be stopped as soon as enough information is accumulated to reach a conclusion (positive and negative)

Design modification

Trial design may be revised to incorporate new information available at an interim look

Group-sequential designs

General setting

Clinical trial with a single objective

Single endpoint

Single patient population

Single dose compared to control

Two decision points

Interim analysis (early stopping due to futility or superior benefit)

Final analysis

Group-sequential designs

Recommended books

Jennison and Turnbull. (2000). *Group Sequential Methods with Applications to Clinical Trials*

Proschan, Lan and Wittes. (2006). *Statistical Monitoring of Clinical Trials: A Unified Approach*

Wassmer and Brannath. (2016). *Group Sequential and Confirmatory Adaptive Designs in Clinical Trials*

General setting

Hypothesis testing problem

θ, True treatment difference

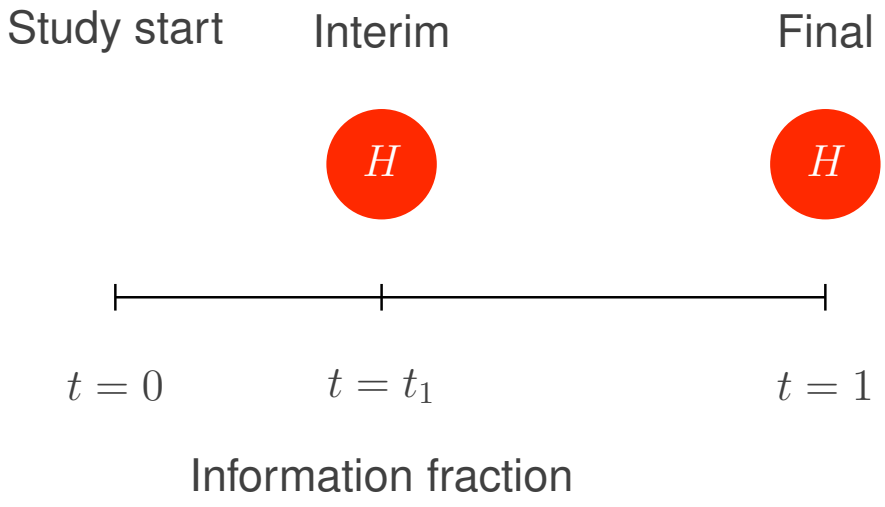
$H: \ \theta = 0$, Null hypothesis of no effect

$K: \ \theta = \theta_1 > 0$, One-sided alternative

Type I error rate

α, Type I error rate, e.g., $\alpha = 0.025$ (one-sided)

Trial diagram



Information fraction t $(0 \leq t \leq 1)$ is based on the number of patients or number of events

General setting

Test statistics and p-values

Z_1 and $p_1 = 1 - \Phi(Z_1)$, Interim analysis

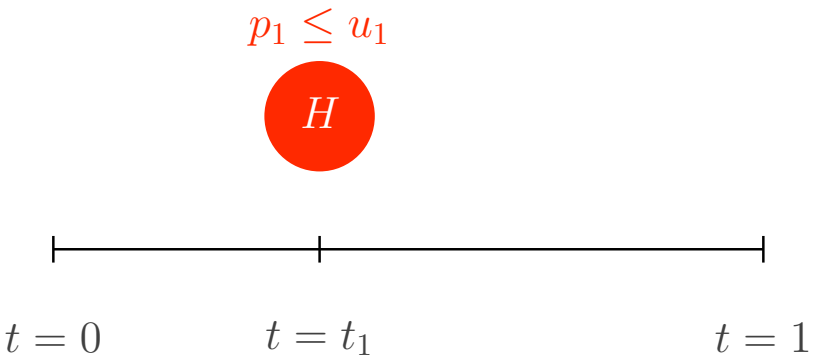
Z_2 and $p_2 = 1 - \Phi(Z_2)$, Final analysis

Stopping boundaries

U_1 and U_2, Stopping boundary for test statistics

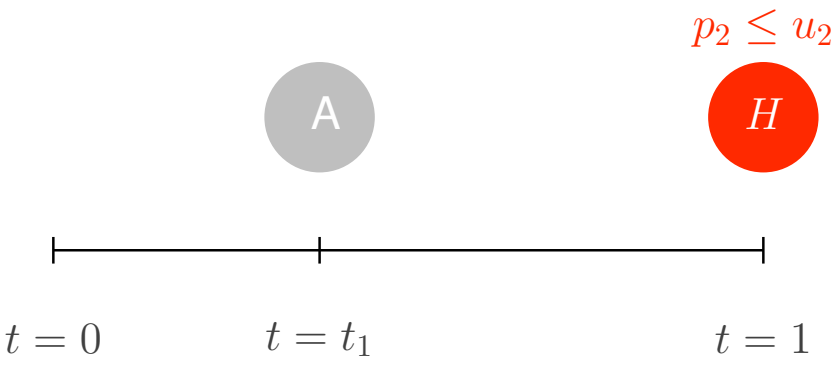
$u_1 = 1 - \Phi(U_1)$ and $u_2 = 1 - \Phi(U_2)$, Stopping boundary for p-values

Decision rule at the interim analysis



Reject H (stop early due to superior efficacy) if $p_1 \leq u_1$; continue to the final analysis if $p_1 > u_1$

Decision rule at the final analysis



Reject H (significant treatment effect) if $p_2 \leq u_2$; negative trial outcome if $p_2 > u_2$

Early stopping due to superior benefit

Type I error spending function (α spending function)

$\alpha(t)$, Non-decreasing function defined over $0 \leq t \leq 1$

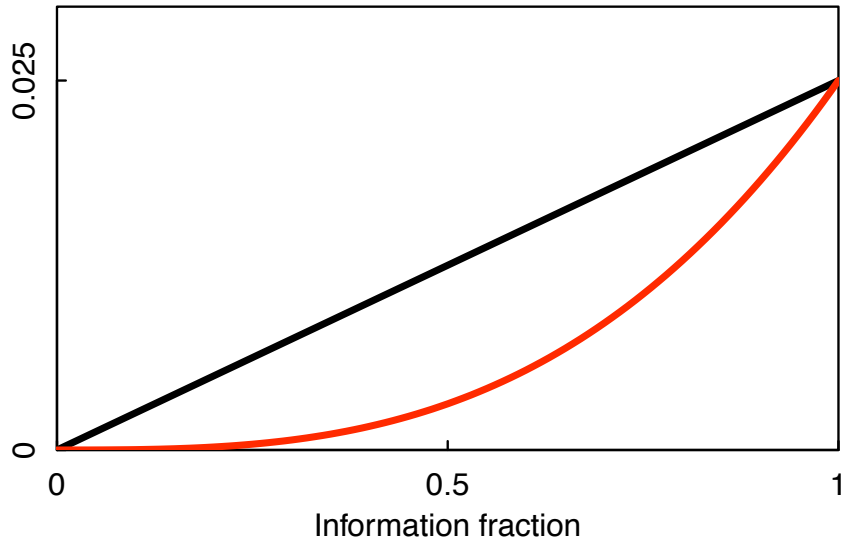
$\alpha(0) = 0$ and $\alpha(1) = \alpha$

Commonly used α spending functions

$\alpha(t) = \alpha t$, Approximation to Pocock boundary

$\alpha(t) = \alpha t^3$, Approximation to O'Brien-Fleming boundary

α spending functions



Black curve: Pocock α-spending function
Red curve: O'Brien-Fleming α-spending function

Comparison of α spending functions

Pocock α-spending function

α is spent uniformly over $0 \leq t \leq 1$

Higher probability of early stopping (generally undesirable)

O'Brien-Fleming α-spending function

Very little α is spent early in the trial

Lower probability of early stopping (generally desirable)

O'Brien-Fleming α-spending function is commonly used in clinical trials

Early stopping due to superior benefit

Interim analysis ($t = t_1$)

Compute U_1 from $P(Z_1 \geq U_1) = \alpha(t_1)$

Final analysis ($t = 1$)

Compute U_2 from $P(Z_1 < U_1,\ Z_2 \geq U_2) = \alpha(1) - \alpha(t_1) = \alpha - \alpha(t_1)$

Null distribution

Calculations are performed under the null distribution of no effect (H) to control Type I error rate

Stopping boundary

Design with a single interim analysis

Analysis	Stopping boundary	
Interim	$U_1 = 2.80$	$u_1 = 0.0026$
Final	$U_2 = 1.98$	$u_2 = 0.0240$

Interim analysis with a 50% information fraction ($t_1 = 0.5$)

O'Brien-Fleming spending function with $\alpha = 0.025$ (one-sided)

Section E2
Group-sequential Trials with Multiple Objectives

Example 4: Prostate cancer trial

Trial arms

Experimental treatment versus control

Two sources of multiplicity

Two decision points

Two key endpoints

Multiple endpoints

Primary endpoint: Progression-free survival (PFS)

Secondary endpoint: Overall survival (OS)

Example 4: Prostate cancer trial

Two decision points

Interim analysis and final analysis

Decision rule at the interim analysis

Early stopping due to superior efficacy

Example 4: Prostate cancer trial

Null hypotheses

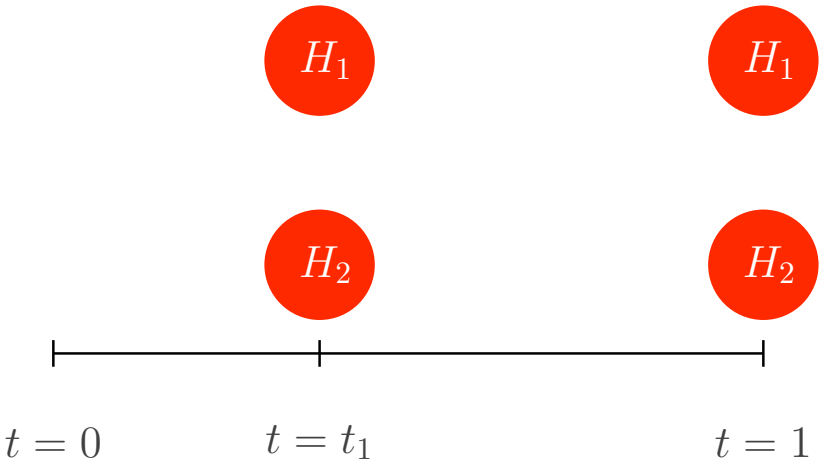
H_1, Null hypothesis of no effect on PFS

H_2, Null hypothesis of no effect on OS

Multiplicity adjustment

Several multple testing procedures will be considered to control the overall Type I error rate at a one-sided $\alpha = 0.025$ at each decision point (fixed-sequence and Holm procedures)

Decision rules



p_{11} and p_{12}, P-values for testing H_1 at the interim and final

p_{21} and p_{22}, P-values for testing H_2 at the interim and final

Stopping boundary

Design with a single interim analysis

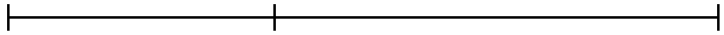
Analysis	Stopping boundary	
Interim	$U_1 = 2.80$	$u_1 = 0.0026$
Final	$U_2 = 1.98$	$u_2 = 0.0240$

Interim analysis with a 50% information fraction ($t_1 = 0.5$)

O'Brien-Fleming spending function with $\alpha = 0.025$ (one-sided)

Decision rules for the primary endpoint (PFS)

$p_{11} \leq u_1$

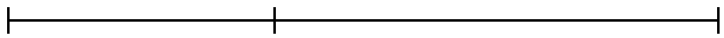


$t = 0$ $t = t_1$ $t = 1$

Reject H_1 (stop early due to superior efficacy) if $p_{11} \leq u_1 = 0.0026$; continue to the final analysis if $p_{11} > u_1$

Decision rule for the primary endpoint (PFS)

$p_{12} \leq u_2$



$t = 0$ $t = t_1$ $t = 1$

Reject H_1 (significant treatment effect on PFS) if $p_{12} \leq u_2 = 0.0240$; negative trial outcome if $p_{12} > u_2$

Scenario 1

Testing strategy

Both PFS and OS will be tested at the interim and final analyses

Multiplicity adjustment

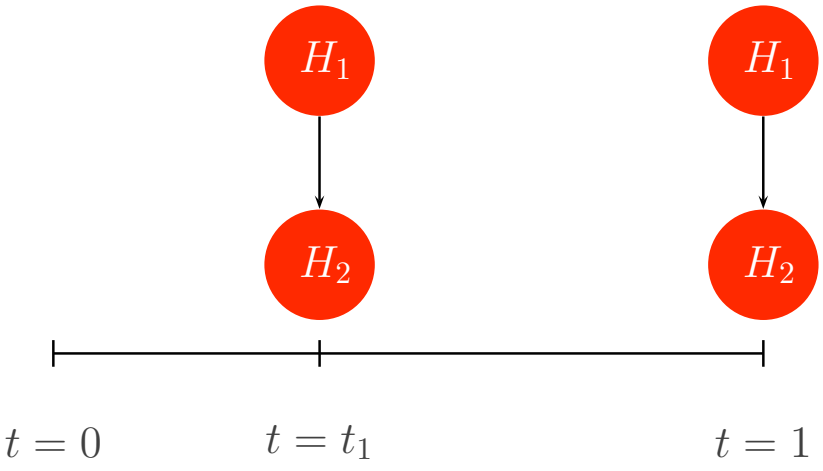
Overall Type I error rate will be controlled using the fixed-sequence procedure

Fixed-sequence procedure

Step 1: H_1 will be tested first

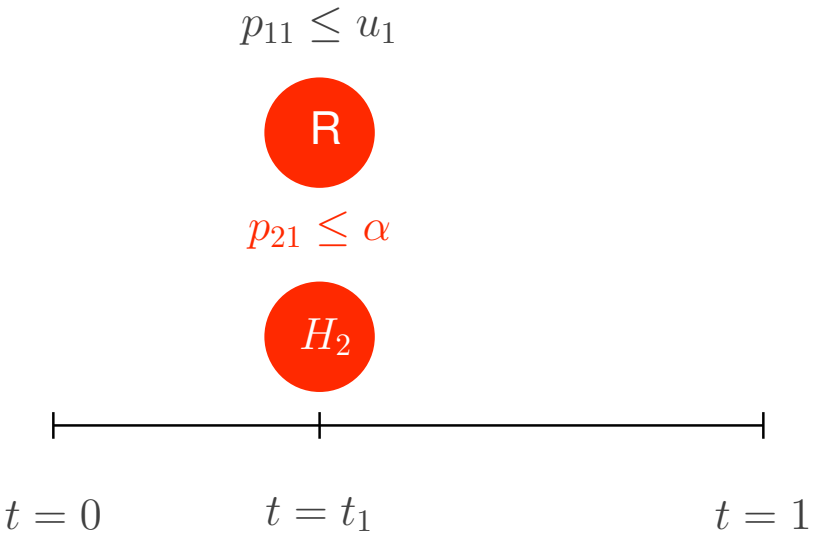
Step 2: H_2 will be tested if H_1 is rejected

Testing strategy



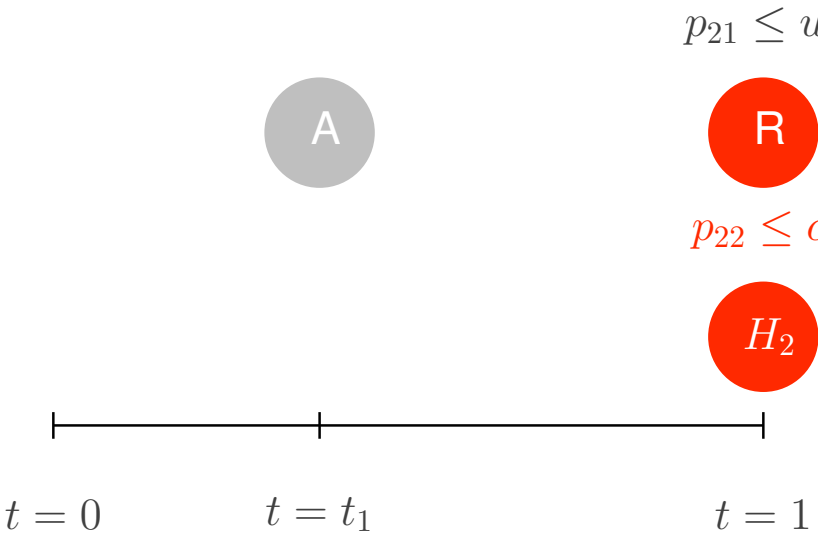
Fixed-sequence procedure will be applied at each decision point

Decision rules for the secondary endpoint (OS)



Since OS can be examined only once, test H_2 at the full $\alpha = 0.025$, i.e., reject H_2 (significant treatment effect on OS) if $p_{21} \leq \alpha$

Decision rules for the secondary endpoint (OS)



Again test H_2 at the full $\alpha = 0.025$, i.e., reject H_2 (significant treatment effect on OS) if $p_{22} \leq \alpha$

Decision rules for the secondary endpoint

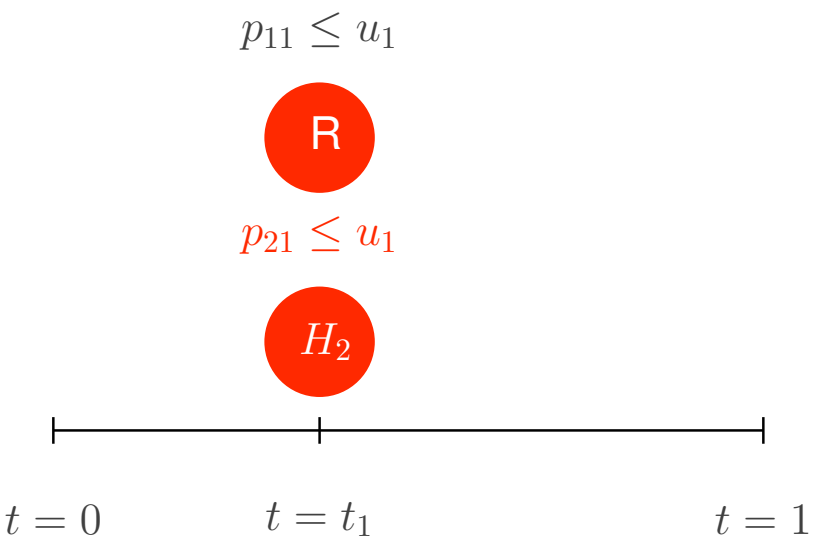
Type I error rate inflation

Overall Type I error rate will be inflated if H_2 is tested at the full $\alpha = 0.025$ at the interim and final analyses (Hung, Wang and O'Neill, 2007)

Type I error rate control

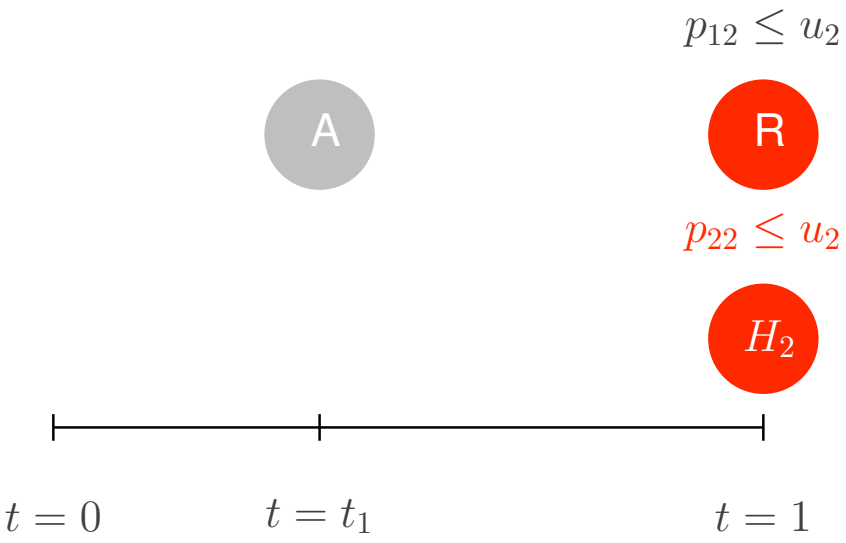
Common or endpoint-specific α spending functions must be applied to protect the overall Type I error rate (Glimm, Maurer and Bretz, 2010)

Decision rules for the secondary endpoint (OS)



Assuming a common α spending function, reject H_2 (significant treatment effect on OS) if $p_{21} \leq u_1 = 0.0026$

Decision rules for the secondary endpoint (OS)



Assuming a common α spending function, reject H_2 (significant treatment effect on OS) if $p_{22} \leq u_2 = 0.0240$

Scenario 2

Testing strategy

PFS will be tested at the interim and final analyses

OS will be tested only at the final analysis

Multiplicity adjustment

Overall Type I error rate will be controlled using the fixed-sequence procedure

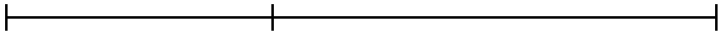
Fixed-sequence procedure

Step 1: H_1 will be tested first

Step 2: H_2 will be tested if H_1 is rejected

Decision rules for the primary endpoint

$p_{11} \leq u_1$

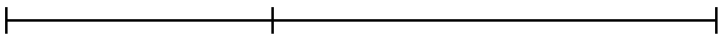


$t = 0$ $t = t_1$ $t = 1$

Reject H_1 (stop early due to superior efficacy) if $p_{11} \leq u_1 = 0.0026$; continue to the final analysis if $p_{11} > u_1$

Decision rule for the primary endpoint

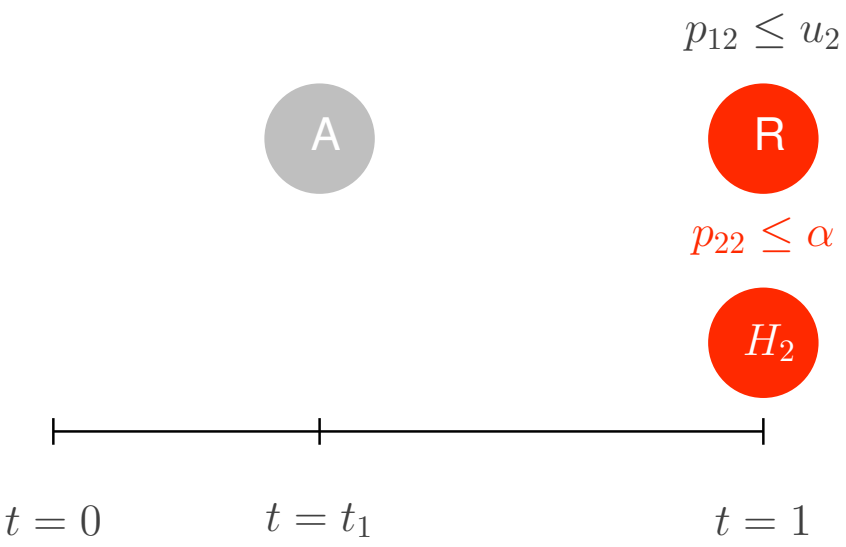
$p_{12} \leq u_2$



$t = 0$ $t = t_1$ $t = 1$

Reject H_1 (significant treatment effect on PFS) if $p_{12} \leq u_2 = 0.0240$; negative trial outcome if $p_{12} > u_2$

Decision rules for the secondary endpoint



Reject H_2 (significant treatment effect on OS) if $p_{22} \leq \alpha = 0.025$

Scenario 3

Testing strategy

Both PFS and OS will be tested at the interim and final analyses

Positive trial outcome at the interim or final if a significant treatment effect is established on OS or PFS

Multiplicity adjustment

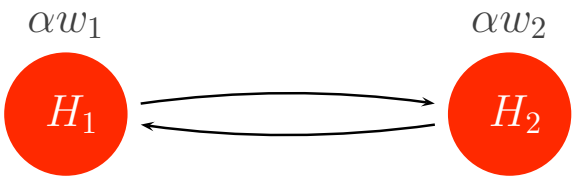
Overall Type I error rate will be controlled using the Holm procedure

$w_1 = 0.8$, PSF weight

$w_2 = 0.2$, OS weight

Scenario 3

Holm procedure



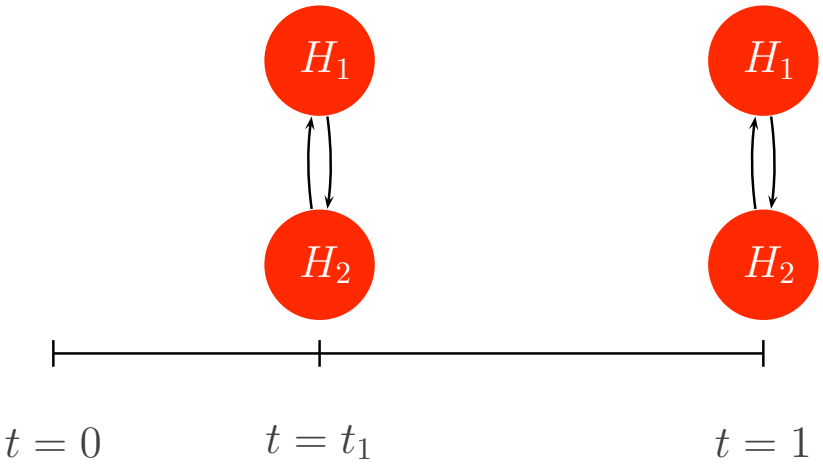
Stepwise procedure

Step 1 (PFS test): Reject H_1 if $p_1 \leq \alpha w_1$

Step 2 (OS test): Reject H_2 if (1) $p_2 \leq \alpha$ and H_1 is rejected or (2) $p_2 \leq \alpha w_2$ and H_1 is not rejected

Step 3 (PFS test): If H_1 is not rejected in Step 1, reject H_1 if $p_1 \leq \alpha$ and H_2 is rejected

Testing strategy



Holm procedure will be applied at each decision point

Stopping boundary

Stopping boundary for PFS test (H_1)

Analysis	Stopping boundary	
	Type I error rate $= \alpha$	Type I error rate $= \alpha w_1$
Interim	$u_1(\alpha) = 0.0026$	$u_1(\alpha w_1) = 0.0017$
Final	$u_2(\alpha) = 0.0240$	$u_2(\alpha w_1) = 0.0193$

Interim analysis with a 50% information fraction ($t_1 = 0.5$)

O'Brien-Fleming spending function with $\alpha = 0.025$ (one-sided)

Stopping boundary

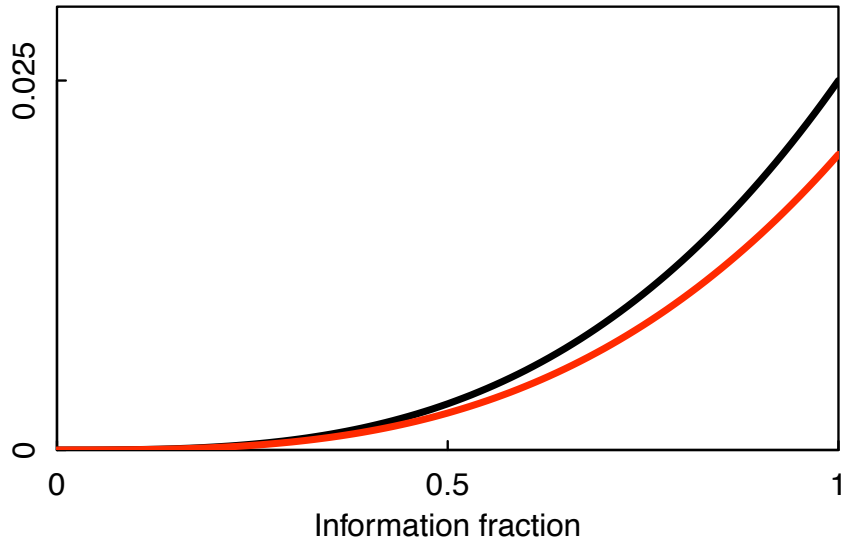
Stopping boundary for OS test (H_2)

Analysis	Stopping boundary	
	Type I error rate $= \alpha$	Type I error rate $= \alpha w_2$
Interim	$u_1(\alpha) = 0.0026$	$u_1(\alpha w_2) = 0.0001$
Final	$u_2(\alpha) = 0.0240$	$u_2(\alpha w_2) = 0.0049$

Interim analysis with a 50% information fraction ($t_1 = 0.5$)

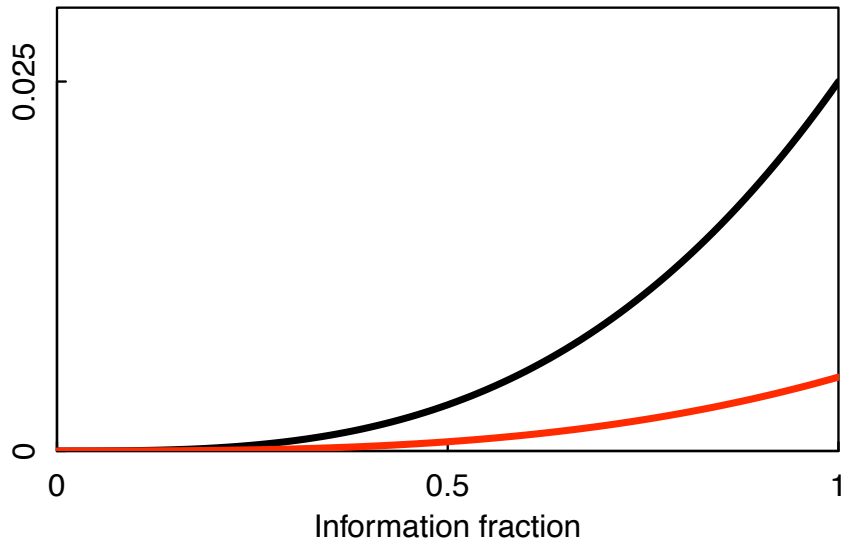
O'Brien-Fleming spending function with $\alpha = 0.025$ (one-sided)

α spending functions for PFS test (H_1)



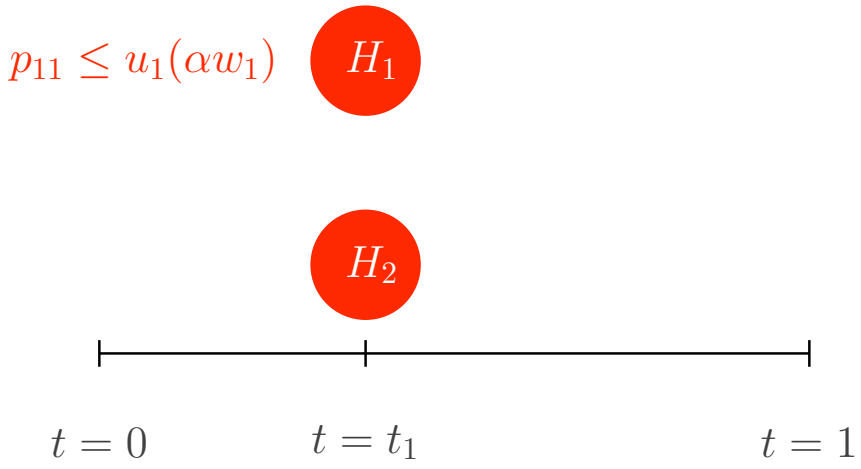
Black curve: α-spending function (Type I error rate $= \alpha$)
Red curve: α-spending function (Type I error rate $= \alpha w_1$)

α spending functions for OS test (H_2)



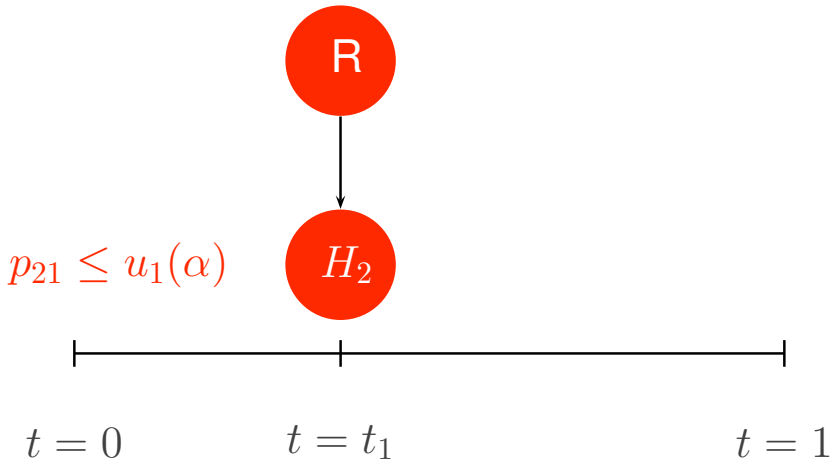
Black curve: α-spending function (Type I error rate $= \alpha$)
Red curve: α-spending function (Type I error rate $= \alpha w_2$)

Decision rules at the interim analysis



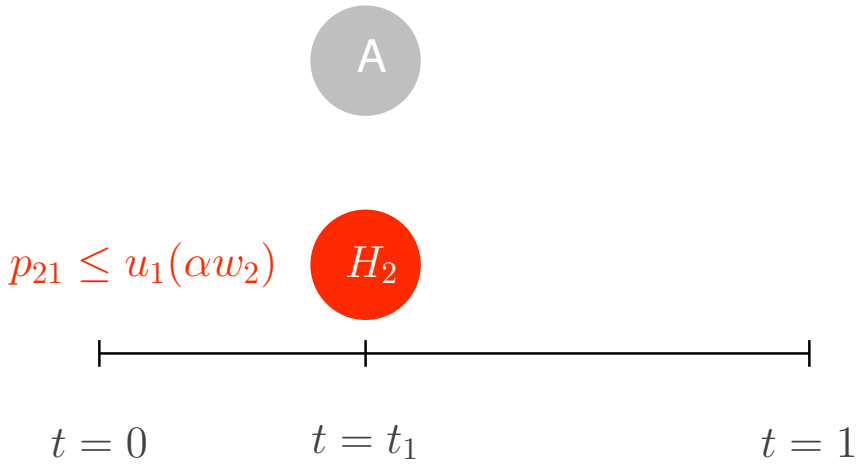
Step 1 (significant effect on PFS): Reject H_1 if $p_{11} \leq u_1(\alpha w_1) = 0.0017$

Decision rules at the interim analysis



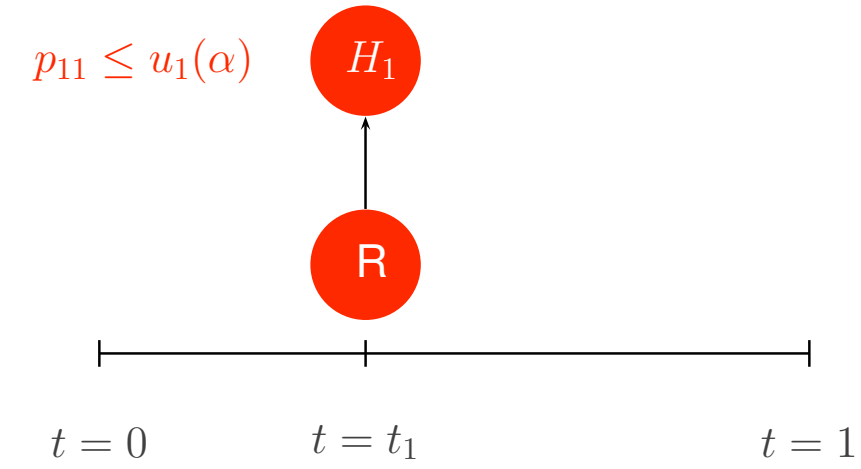
Step 2A (significant effect on OS): Reject H_2 if $p_{21} \leq u_1(\alpha w_1 + \alpha w_2) = u_1(\alpha) = 0.0026$ and H_1 is rejected in Step 1

Decision rules at the interim analysis



Step 2B (significant effect on OS): Reject H_2 if $p_{21} \le u_1(\alpha w_2) = 0.0001$ and H_1 is not rejected in Step 1

Decision rules at the interim analysis



Step 3 (significant effect on PFS): Reject H_1 if $p_{11} \le u_1(\alpha w_1 + \alpha w_2) = u_1(\alpha) = 0.0026$ if H_2 is rejected in Step 2

Group-sequential trials with multiple objectives

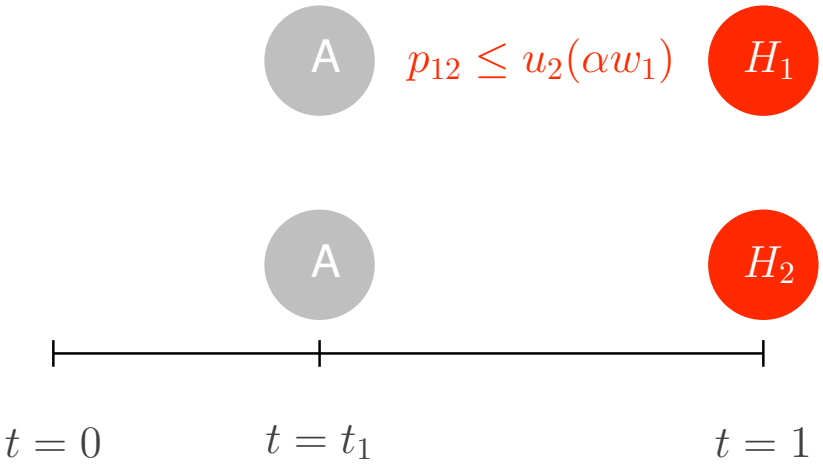
Decision rules

A trial may continue to the next decision point (e.g., final analysis) even if superior efficacy is established at the current decision point (e.g., interim analysis)

Consistency restrictions

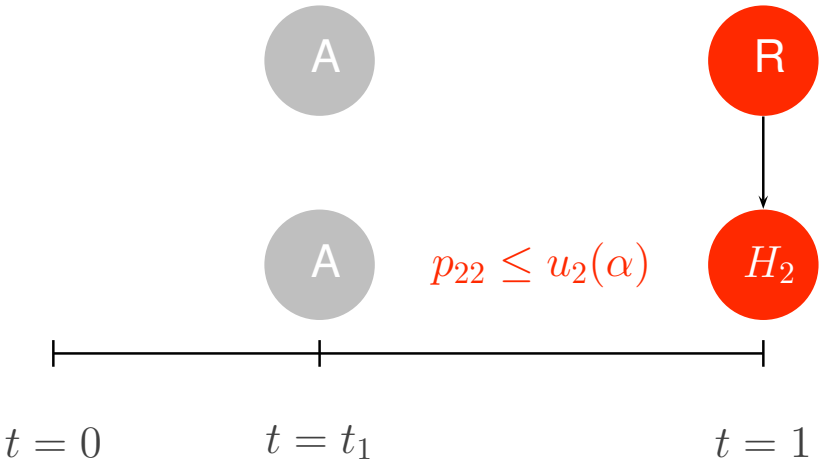
A hypothesis rejected at a decision point is automatically rejected at all subsequent decision points without testing

Decision rules at the final analysis



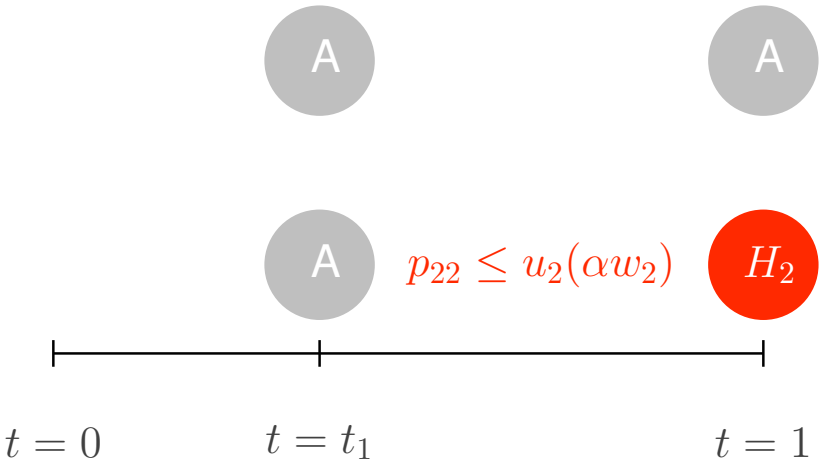
Step 1 (significant effect on PFS): Reject H_1 if $p_{12} \leq u_2(\alpha w_1) = 0.0193$

Decision rules at the final analysis



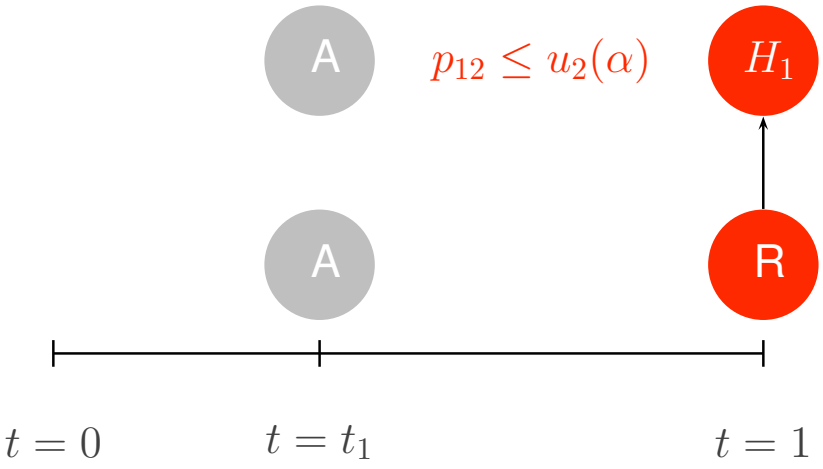
Step 2A (significant effect on OS): Reject H_2 if $p_{22} \leq u_2(\alpha w_1 + \alpha w_2) = u_2(\alpha) = 0.0240$ and H_1 is rejected in Step 1

Decision rules at the final analysis



Step 2B (significant effect on OS): Reject H_2 if $p_{22} \leq u_2(\alpha w_2) = 0.0049$ and H_1 is not rejected in Step 1

Decision rules at the final analysis



Step 3 (significant effect on PFS): Reject H_1 if $p_{12} \leq u_2(\alpha w_1 + \alpha w_2) = u_2(\alpha) = 0.0240$ if H_2 is rejected in Step 2

Scenario 4

Testing strategy

Both PFS and OS will be tested at the interim and final analyses

O'Brien-Fleming spending function for PFS test

Pocock spending function for OS test

Multiplicity adjustment

Overall Type I error rate will be controlled using the Holm procedure

$w_1 = 0.8$, PSF weight

$w_2 = 0.2$, OS weight

Stopping boundary

Stopping boundary for OS test (H_2)

Analysis	Stopping boundary	
	Type I error rate $= \alpha$	Type I error rate $= \alpha w_2$
Interim	$u_1(\alpha) = 0.0147$	$u_1(\alpha w_2) = 0.0028$
Final	$u_2(\alpha) = 0.0147$	$u_2(\alpha w_2) = 0.0028$

Interim analysis with a 50% information fraction ($t_1 = 0.5$)

Pocock spending function with $\alpha = 0.025$ (one-sided)

Software implementation

Software implementation in SAS

SAS/STAT module

Group-sequential designs with popular α spending functions: SEQDESIGN procedure

Software implementation in R

gsDesign package

Group-sequential designs with popular α spending functions

http://cran.r-project.org/web/packages/gsDesign

Group-sequential designs in R

O'Brien-Fleming spending function

Single interim analysis (50% information fraction)

Type I error rate $\alpha = 0.025$

```
gsDesign(k=2, timing=c(0.5, 1), test.type=1,
             sfu="OF", alpha=0.025, beta=0.1)
```

Output

```
           Sample
            Size
  Analysis Ratio*   Z   Nominal p  Spend
         1  0.504 2.80    0.0026 0.0026
         2  1.007 1.98    0.0240 0.0224
     Total                       0.0250
```


Group-sequential designs in R

Pocock spending function

Single interim analysis (50% information fraction)

Type I error rate $\alpha = 0.025$

```
gsDesign(k=2, timing=c(0.5, 1), test.type=1,
             sfu="Pocock", alpha=0.025, beta=0.1)
```

Output

```
           Sample
            Size
  Analysis Ratio*   Z   Nominal p  Spend
         1   0.55 2.18    0.0147 0.0147
         2   1.10 2.18    0.0147 0.0103
     Total                       0.0250
```

Group-sequential designs in R

O'Brien-Fleming spending function

Single interim analysis (50% information fraction)

Type I error rate $\alpha = 0.025 w_1$ with $w_1 = 0.8$

```
gsDesign(k=2, timing=c(0.5, 1), test.type=1,
             sfu="OF", alpha=0.025 * 0.8, beta=0.1)
```

Output

```
           Sample
            Size
  Analysis Ratio*   Z   Nominal p  Spend
         1  0.503 2.92     0.0017 0.0017
         2  1.006 2.07     0.0193 0.0183
     Total                        0.0200
```


Module F
General Adaptive Trials with Multiple Objectives

Module F outline

F1. Decision rules in adaptive designs

Review of commonly used decision rules in adaptive designs

F2. Adaptive population selection trials

Multiplicity issues in adaptive trials aimed at selecting the best patient population

Section F1
Decision Rules in Adaptive Designs

Adaptive trials with multiple objectives

Examples of data-dependent adaptations

Sample size or event count adjustment

Treatment selection

Population selection

Other sources of multiplicity

Several endpoints (primary and secondary endpoints)

Several dose-control comparisons

Several patient populations

Example 5: Breast cancer trial

Trial arms

Experimental treatment versus control

Primary endpoint

Progression-free survival

Two sources of multiplicity

Two pre-defined patient populations

Two decision points

Example 5: Breast cancer trial

Biomarker

Potential predictive biomarker (PIK3CA mutation status)

Two pre-defined patient populations

Overall population (OP)

Biomarker-positive subpopulation (M+): Patients in M+ are expected to experience a strong beneficial treatment effect

Treatment effect will not be evaluated in the biomarker-negative subpopulation (M–)

Example 5: Breast cancer trial

Two decision points

Interim analysis and final analysis

Decision rules at the interim analysis

Early stopping due to futility or superior efficacy (based on a strong or weak treatment effect in the overall population or biomarker-positive subpopulation)

Adjustment of the target event count

Population selection

Adaptive design methodology

Adaptive trials

Traditional setting

Statistical methods used in an adaptive design setting with a single source of multiplicity

Combination function approach

Broadly used in clinical trials with general data-driven decision rules (Bauer and Köhne, 1994)

Other approaches

Conditional error function approach is more commonly used in clinical trials with sample size adjustment rules (will not be discussed)

Two-stage adaptive clinical trial

Single source of multiplicity

Interim analysis and final analysis

Treatment effect test statistics and p-values

Z_1 and Z_2, Test statistics at interim and final analyses

p_1 and p_2, P-values at interim and final analyses

Two-stage adaptive clinical trial

Patient-wise separation

Stage 1 (before interim analysis) with n_1 patients

Stage 2 (after interim analysis) with n_2 patients

Stagewise test statistics and p-values

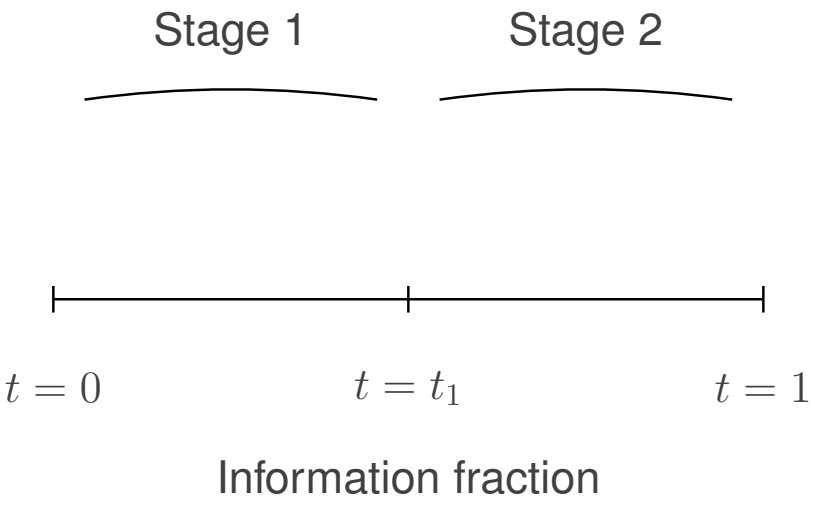
Z_1^* and Z_2^*, Test statistics in Stages 1 and 2

p_1^* and p_2^*, P-values in Stages 1 and 2

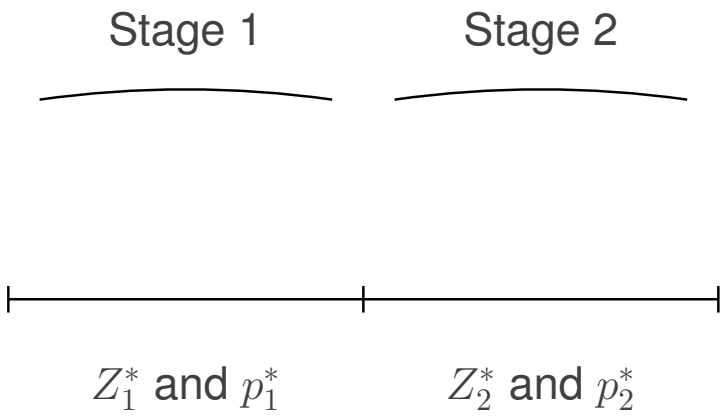
Independence of test statistics

Test statistics in Stages 1 and 2 are independent

Two-stage adaptive clinical trial



Stagewise test statistics and p-values



Two-stage adaptive clinical trial

Pre-defined stage weights

w_1, Stage 1 weight

w_2, Stage 2 weight

$w_1 > 0$, $w_2 > 0$ and $w_1 + w_2 = 1$

Example

Stage weights are typically proportional to the information fractions, e.g.,

$$w_1 = \frac{n_1}{n_1 + n_2}, \; w_2 = \frac{n_2}{n_1 + n_2}.$$

Two-stage adaptive clinical trial

Test statistics

Interim analysis: $Z_1 = Z_1^*$

Final analysis: $Z_2 = \sqrt{w_1} Z_1^* + \sqrt{w_2} Z_2^*$

P-values

Interim analysis: $p_1 = p_1^*$

Final analysis: $p_2 = C(p_1^*, p_2^*)$, where $C(x, y)$ is the pre-defined combination function

Two-stage clinical trial

Combination function

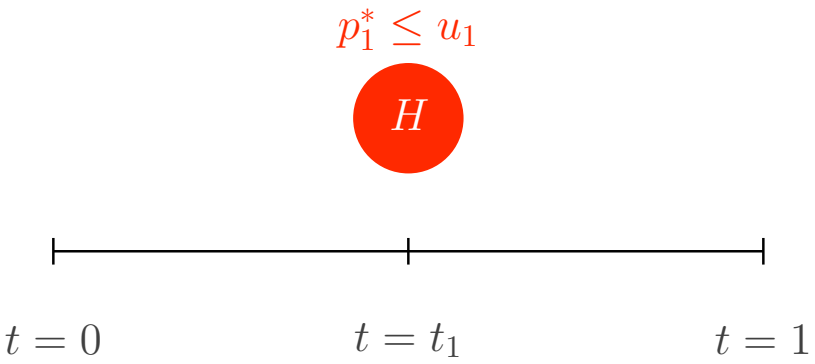
Multiple combination functions have been proposed and the inverse normal combination function is very commonly used

$$C(p_1^*, p_2^*) = 1 - \Phi(\sqrt{w_1}\Phi^{-1}(1-p_1^*) + \sqrt{w_2}\Phi^{-1}(1-p_2^*))$$

Note that

$$\Phi^{-1}(1-p_1^*) = Z_1^*, \ \Phi^{-1}(1-p_2^*) = Z_2^*$$

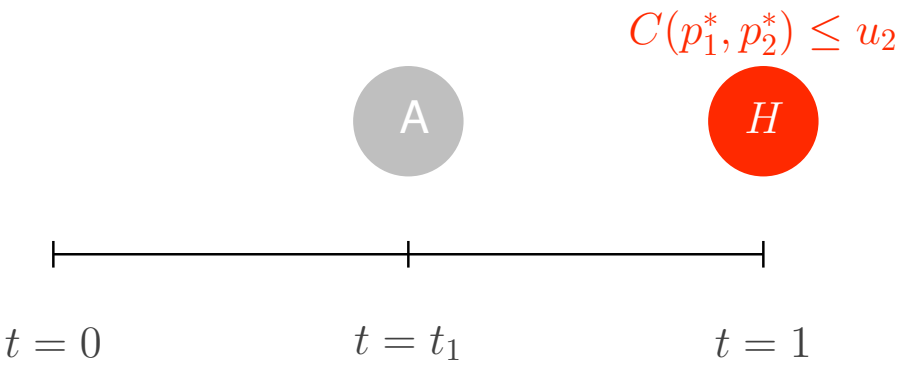
Decision rule at the interim analysis



Reject H (stop early due to superior efficacy) if $p_1^* \leq u_1$

Continue to the final analysis if $p_1^* > u_1$

Decision rule at the final analysis



Reject H (significant treatment effect) if $C(p_1^*, p_2^*) \leq u_2$

Negative trial outcome if $C(p_1^*, p_2^*) > u_2$

Combination function principle

Group sequential and multi-stage designs

Decision rules based on a combination function are equivalent to decision rules used in group sequential designs if the trial design is not modified at the interim analysis

Combination function principle

General adaptive designs

Combination function principle protects overall Type I error rate for any design adaptations provided the stage weights are pre-defined

Examples

Sample size or event count may be increased

Patient population may be modified in Stage 2

Combination function principle

Type I error rate inflation

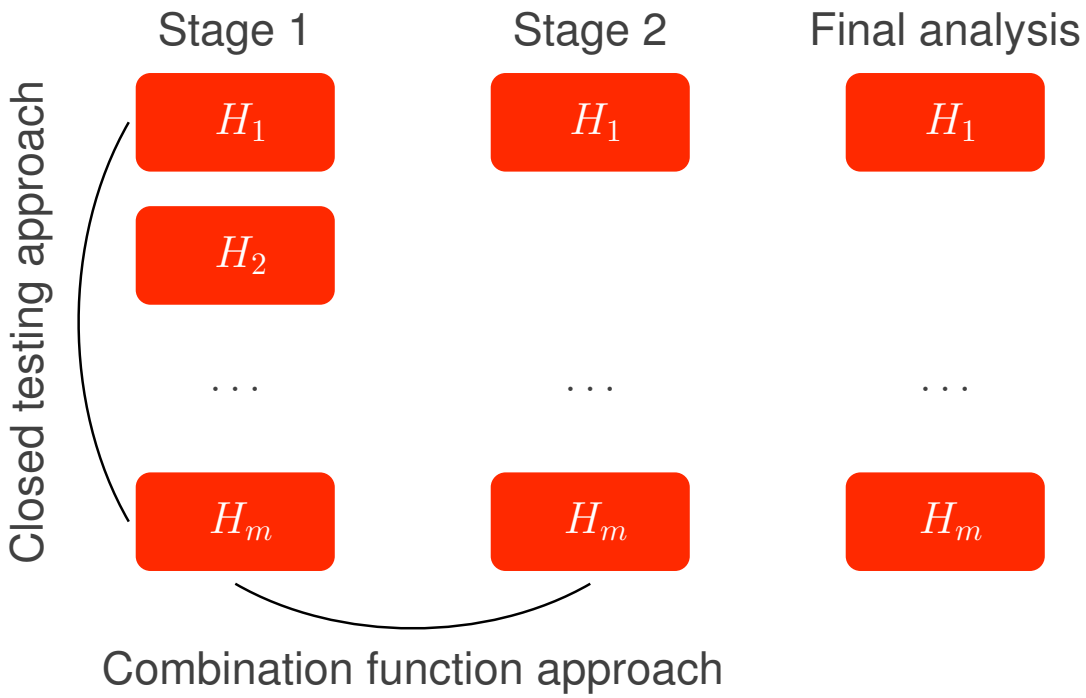
Virtually all adaptive (data-driven) decisions at an interim analysis change the distribution of Stage 2 statistic (Z_2^*)

Type I error rate is inflated if no adjustment is introduced

Type I error rate control

Combination function approach provides an adjustment to control overall Type I error rate for a very broad class of adaptive designs

Adaptive trials with several objectives



Adaptive trials with several objectives

Recent publications

Chain procedures/graphical procedures in group-sequential and general adaptive trials (Maurer and Bretz, 2013; Sugitani, Bretz and Maurer, 2016)

More powerful gatekeeping procedures, e.g., Hommel-based gatekeeping procedures, in adaptive trials (Kordzakhia, Dmitrienko and Ishida, 2018)

Section F2
Adaptive Population Selection Trials

Example 5: Breast cancer trial

Two-stage trial design

Stage 1: All patients enrolled prior to the interim analysis

Stage 2: All patients enrolled after the interim analysis

Stagewise test statistics and p-values

Z_1^* and Z_2^*, Test statistics in Stages 1 and 2

p_1^* and p_2^*, P-values in Stages 1 and 2

Test statistics and p-values in Stages 1 and 2 are independent

Two-stage adaptive clinical trial

Stage 1 data

PFS outcomes will be censored at the end of the trial

Stage 2 data

PFS outcomes will be censored at the end of the trial

Data-driven adjustment of the target event count

Suppose a decision is made at the interim analysis to increase the original target event count

Two-stage adaptive clinical trial

Event count adjustment

If the original target event count is increased at the interim analysis, the length of event follow-up in Stage 1 will depend on the data collected in Stage 2

Stage 1 and Stage 2 test statistics will no longer be independent and Type I error rate may no longer be protected

To control Type I error rate, the length of event follow-up in Stage 1 must be fixed (Magirr et al., 2016)

General decision rules

Decision rules at the interim analysis

Task 1. Develop the futility stopping rules

Task 2. Develop the population selection rules

Task 3. Develop the event count adjustment rules

Multiplicity adjustment

Task 4. Develop the multiplicity adjustment strategy based on the closure principle and combination function approach to control the alpha at a one-sided 0.025

Inferences at the interim and final analyses

Example 5: Breast cancer trial

Null hypotheses

H_1, Null hypothesis of no effect in the overall trial population (OP)

H_2, Null hypothesis of no effect in the biomarker-positive subpopulation (M+)

Multiplicity adjustment

Hochberg procedure to control the overall Type I error rate at $\alpha = 0.025$ (one-sided) at each decision point

Hochberg procedure

Ordered p-values and null hypotheses

p_1 and p_2, P-values for testing H_1 and H_2

$p_{(1)} < p_{(2)}$, Ordered p-values

$H_{(1)}$ and $H_{(2)}$, Ordered hypotheses

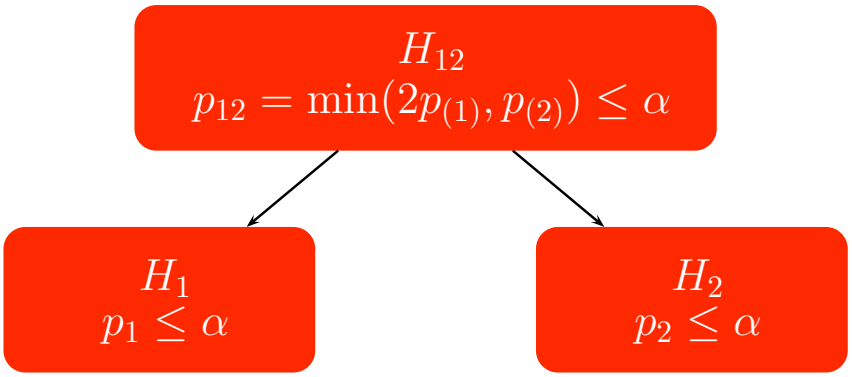
Testing algorithm

If $p_{(2)} \leq \alpha$, reject both $H_{(1)}$ and $H_{(2)}$

If $p_{(2)} > \alpha$ and $p_{(1)} \leq \alpha/2$, reject $H_{(1)}$ only

Multiplicity adjustment

Hochberg procedure as a closed testing procedure



Significant treatment effect in OP if H_{12} and H_1 are rejected

Significant treatment effect in M+ if H_{12} and H_2 are rejected

Two-stage design

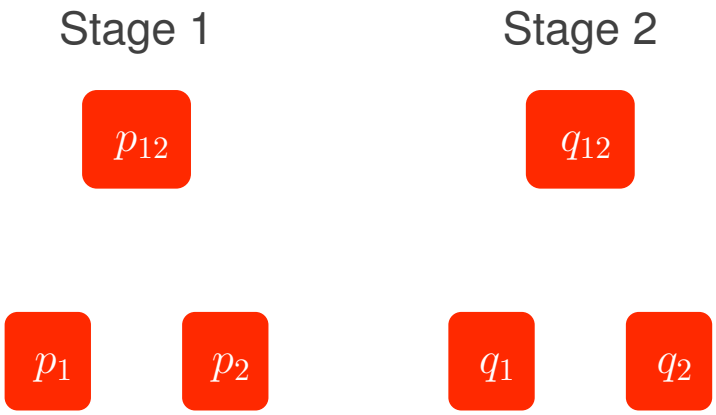
Notation

Hypothesis	P-value	
	Stage 1	Stage 2
H_{12}	p_{12}	q_{12}
H_1	p_1	q_1
H_2	p_2	q_2

Note: $p_{12} = \min(2p_{(1)}, p_{(2)})$ and $q_{12} = \min(2q_{(1)}, q_{(2)})$

Two-stage design

Notation



Efficacy stopping rule

Interim analysis

t, Information fraction at the interim analysis, e.g., $t = 0.5$ if the interim look is taken after 50% of the total number of PFS events

α spending function

Choose an α spending function, e.g., O'Brien-Fleming spending function can be used

Efficacy stopping rule

Stopping boundaries

u_k, $k = 1, 2$, Stopping boundary on a p-value scale

Example

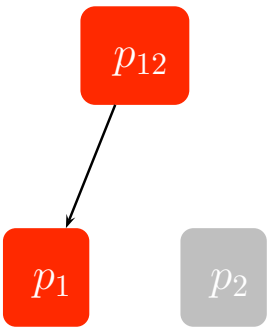
Apply the O'Brien-Fleming spending function with $t = 0.5$

$u_1 = 0.0026$, One-sided significance level at the interim analysis

$u_2 = 0.0240$, One-sided significance level at the final analysis

Inferences at the interim analysis

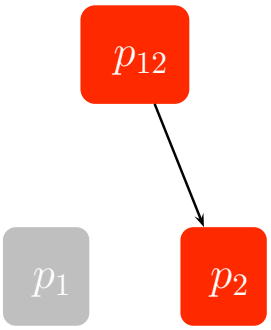
Stage 1



Stop early due to superior efficacy in OP if $p_{12} \leq u_1 = 0.0026$ and $p_1 \leq u_1 = 0.0026$

Inferences at the interim analysis

Stage 1



Stop early due to superior efficacy in M+ if $p_{12} \leq u_1 = 0.0026$ and $p_2 \leq u_1 = 0.0026$

Inferences at the final analysis

Combination function

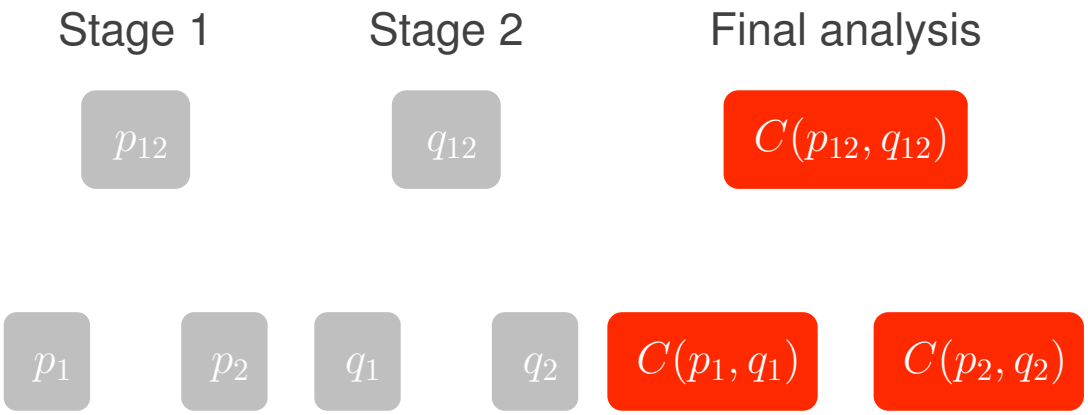
Inverse normal combination function will be used to perform inferences at the final analysis

$$C(x, y) = 1 - \Phi[\sqrt{w_1}\Phi^{-1}(1 - x) + \sqrt{w_2}\Phi^{-1}(1 - y)]$$

Stage weights

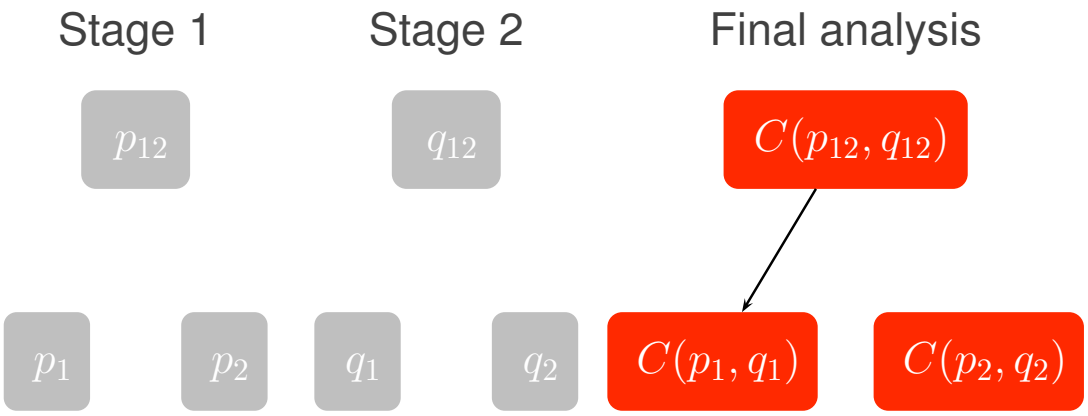
$w_1 = t = 0.5$, $w_2 = 1 - t = 0.5$, Pre-specified stage weights

Inferences at the final analysis



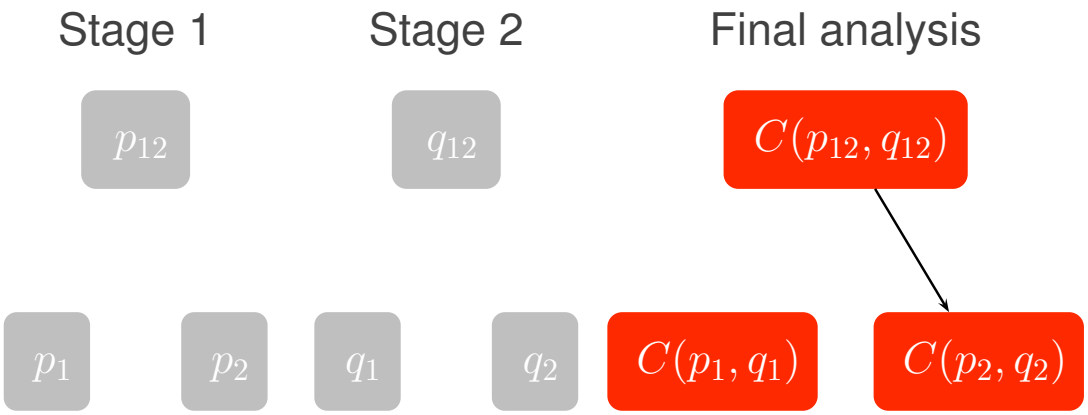
Scenario 1: Both OP and M+ are selected at the interim analysis

Inferences at the final analysis



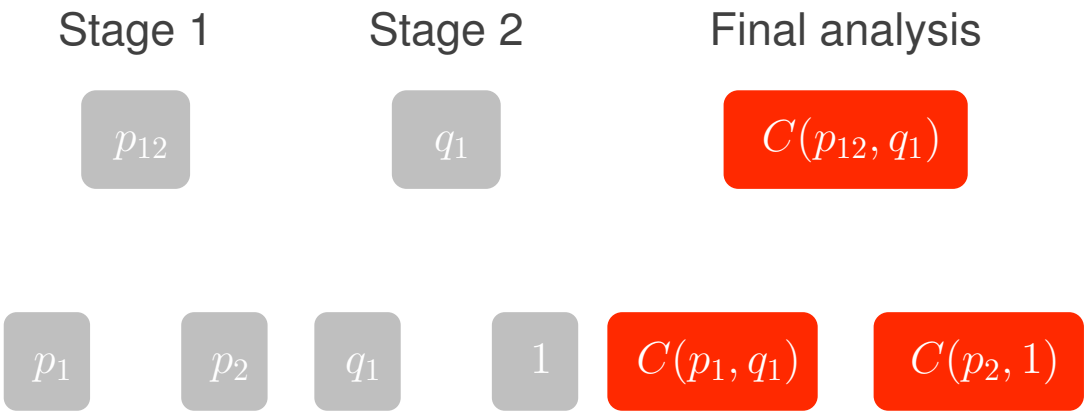
Treatment effect in OP is significant if
$C(p_{12}, q_{12}) \leq u_2 = 0.0240$ and $C(p_1, q_1) \leq u_2 = 0.0240$

Inferences at the final analysis



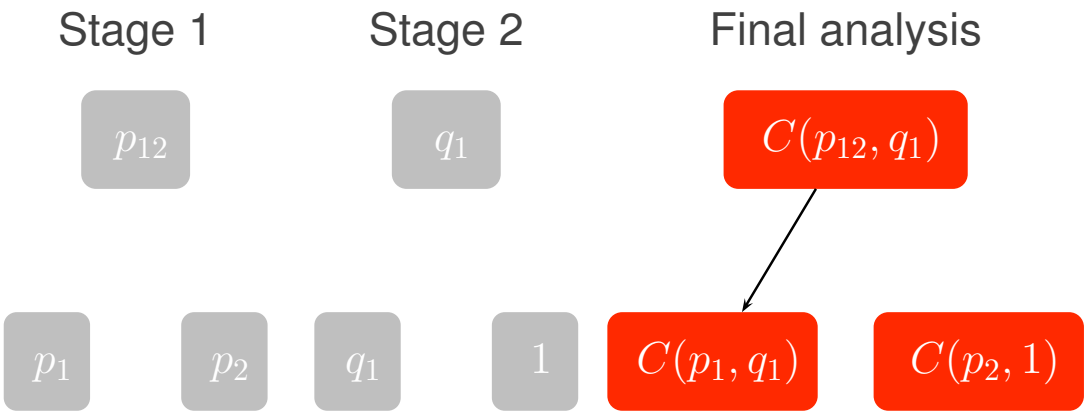
Treatment effect in M+ is significant if
$C(p_{12}, q_{12}) \leq u_2 = 0.0240$ and $C(p_2, q_2) \leq u_2 = 0.0240$

Inferences at the final analysis



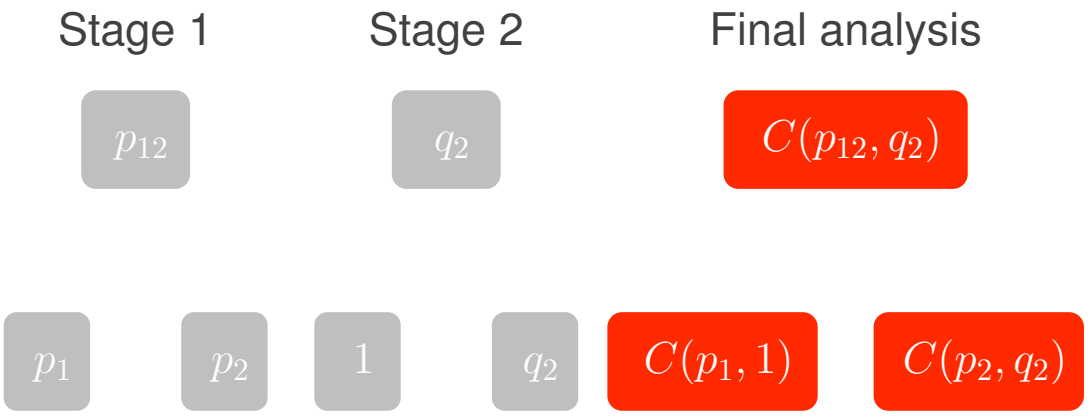
Scenario 2: Only OP is selected at the interim analysis
Since there are no M+ data in Stage 2, $q_2 = 1$ and $q_{12} = q_1$

Inferences at the final analysis



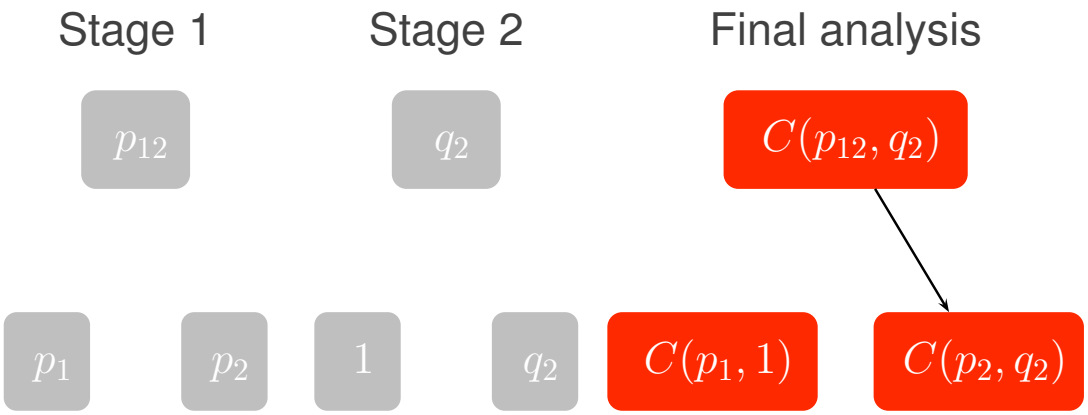
Treatment effect in OP is significant if
$C(p_{12}, q_1) \leq u_2 = 0.0240$ and $C(p_1, q_1) \leq u_2 = 0.0240$

Inferences at the final analysis



Scenario 3: Only M+ is selected at the interim analysis
Since there are no OP data in Stage 2, $q_1 = 1$ and $q_{12} = q_2$

Inferences at the final analysis



Treatment effect in M+ is significant if
$C(p_{12}, q_2) \leq u_2 = 0.0240$ and $C(p_2, q_2) \leq u_2 = 0.0240$

Summary

Multiplicity issues in adaptive trials

Adaptive approaches in multi-stage clinical trials

Flexible alternative to traditional trial designs

Support multiple types of adaptations, e.g., treatment selection, population selection, sample size adjustment, etc

Example: Adaptive population selection designs facilitate the development of tailored therapies that are likely to benefit a subset of the general patient population

Module G
Power Calculations

Module G outline

G1. Power calculations in clinical trials with multiple objectives

Analytical and simulation-based power calculation approaches

Introduction to Clinical Scenario Evaluation

Mediana package: New R package for clinical trial simulations

Section G1
Power calculations in clinical trials with multiple objectives

Power and sample size calculations

Analytical approach

Closed-form expressions used in traditional sample size calculations often rely on simplifying assumptions

Are we cutting corners and looking for shortcuts in multi-million dollar clinical trials?

Simulation-based approach

Much more reliable approach to power and sample size calculations in trials with complex design and analysis strategies

Simulation-based approaches

FDA guidance (FDA, 2017)

"Determination of an appropriate study sample size to ensure that the study is appropriately powered can be difficult in these cases, and often will be dependent upon computer simulations rather than an analytic formula, which can be used for simpler situations"

Introduction to Clinical Scenario Evaluation

Clinical scenario evaluation

General framework

Clinical scenario evaluation (CSE) approach was developed in Benda et al. (2010), Friede et al. (2010) and other publications

Motivation

Clinical trial researchers have recognized the importance of employing quantitative, comprehensive and disciplined approaches to evaluating the design and analysis of clinical trials to enable better decision making

Key components

Data models (Assumptions)

Describe the data generation mechanism in a clinical trial

Analysis models (Options)

Define statistical tests, descriptive statistics and other analysis tools, e.g., multiplicity adjustments, computed from the trial data

Evaluation models (Metrics)

Specify measures for evaluating performance of the analysis strategies

Clinical trial optimization

Clinical scenario evaluation

An important application of general CSE approach is clinical trial optimization

General theme

Utilize CSE to transition from traditionally used approaches to optimal approaches to selecting trial designs and analysis strategies

Inform decision making at different stages of a drug development program to maximize the overall probability of success

Examples of clinical trial optimization

Phase II and Phase III trials

Optimal multiplicity adjustment in clinical trials with several objectives

Optimal decision rules at interim looks in clinical trials with adaptive designs

Optimal decision rules in clinical trials with several patient populations

Clinical trial optimization

Publications

Review of general approaches to clinical trial optimization (*Clinical Trial Optimization Using R* edited by Dmitrienko and Pulkstenis, 2017)

Optimal selection of multiplicity adjustments in Phase III trials (Dmitrienko, Paux and Brechenmacher, 2015)

Optimal selection of multiplicity adjustments and adaptive trial designs in Phase II and III trials (Dmitrienko, Paux, Pulkstenis and Zhang, 2016)

Mediana package

Mediana: New R package

Goals

Implement the Clinical Scenario Evaluation approach

Provide general framework for simulation-based power and sample size calculations typically performed in late-phase trials (Phase II and III trials)

Support clinical trial optimization aimed at identifying optimal trial designs and analysis strategies

Mediana package

Release

First version (Version 1.0.1) was released in July 2015

CRAN web site

https://cran.r-project.org/web/packages/Mediana

Online manual

http://gpaux.github.io/Mediana/

Example 3
Schizophrenia trial

Example 3: Schizophrenia trial

Objective

Evaluate the short-term efficacy and safety of an experimental treatment for acute schizophrenia

Two dose-placebo comparisons

Two doses (Dose L and Dose H) versus placebo

Two endpoints

Primary endpoint: Change from baseline to Week 6 in PANSS

Secondary endpoint: Change from baseline to Week 6 in CGI-S

Sample size calculations

Goal

Compute the number of patients to guarantee a sufficiently high level of overall success probability

Recommended approach

Simulation-based approach to compute appropriately defined power for a range of sample sizes

Simulation-based approach will be illustrated using the Mediana package

Data model

Key components

Outcome distribution

Samples (independent sets of patients, e.g., treatment arms in a trial)

Parameters of individual samples (outcome distribution parameters and sample sizes)

Trial design

Data model

Three samples

Sample 1: Placebo arm

Sample 2: Dose L arm

Sample 3: Dose H arm

Outcome distribution

Primary and secondary endpoints are normally distributed (*NormalDist*)

Trial design

Fixed design

Data model

Outcome distribution parameters

Sample	ID	Endpoint	Mean	SD
Sample 1	Placebo (P)	Endpoint P	-12	20
	Placebo (S)	Endpoint S	-0.8	1
Sample 2	Dose L (P)	Endpoint P	-18	20
	Dose L (S)	Endpoint S	-1.2	1
Sample 3	Dose H (P)	Endpoint P	-20	20
	Dose H (S)	Endpoint S	-1.3	1

Endpoint P: Primary endpoint

Endpoint S: Secondary endpoint

Mediana code

Data model: Outcome distribution parameters

```
placebo.par = parameters(
              parameters(mean = -12, sd = 20),
              parameters(mean = -0.8, sd = 1))

dosel.par = parameters(
              parameters(mean = -18, sd = 20),
              parameters(mean = -1.2, sd = 1))

doseh.par = parameters(
              parameters(mean = -20, sd = 20),
              parameters(mean = -1.3, sd = 1))
```

Mediana code

Data model: Correlation and variable types

```
corr.matrix = matrix(c(1.0, 0.3,
                       0.3, 1.0), 2, 2)

var.type = list("NormalDist", "NormalDist")

outcome.placebo = parameters(type = var.type,
                             par = placebo.par,
                             corr = corr.matrix)
outcome.dosel = parameters(type = var.type,
                           par = dosel.par,
                           corr = corr.matrix)
outcome.doseh = parameters(type = var.type,
                           par = doseh.par,
                           corr = corr.matrix)
```


Mediana code

Data model

```
ex3.data.model = DataModel() +
  OutcomeDist(outcome.dist = "MVNormalDist") +
  SampleSize(c(110, 120, 130, 140)) +
  Sample(id = list("Placebo (P)", "Placebo (S)"),
      outcome.par = parameters(outcome.placebo)) +
  Sample(id = list("Dose L (P)", "Dose L (S)"),
      outcome.par = parameters(outcome.dosel)) +
  Sample(id = list("Dose H (P)", "Dose H (S)"),
      outcome.par = parameters(outcome.doseh))
```

Analysis and evaluation models

Key components of analysis model

Dose-placebo tests and their parameters, including a multiplicity adjustment

Key components of evaluation model

Success criteria and their parameters

Analysis and evaluation models

Analysis model

Treatment effect for each endpoint test is assessed using the two-sample t test (*TTest*)

Multiplicity adjustment based on a general gatekeeping procedure with several sequences of hypotheses (*MultipleSequenceGatekeepingAdj*)

Evaluation model

Disjunctive power for the primary and secondary tests (*DisjunctivePower*)

Mediana code

Analysis model: Multiplicity adjustment

```
families = families(family1 = c(1, 2),
                    family2 = c(3, 4))

component.procedures =
        families(family1 ="HochbergAdj",
                 family2 = "HochbergAdj")

gamma = families(family1 = 0.7,
                 family2 = 1)
```


Mediana code

Analysis model

```
ex3.analysis.model = AnalysisModel() +
  MultAdjProc(proc =
         "MultipleSequenceGatekeepingAdj",
          par = parameters(family = families,
                     proc = component.procedures,
                     gamma = gamma),
          tests = tests("Placebo vs Dose L (P)",
              "Placebo vs Dose H (P)",
              "Placebo vs Dose L (S)",
              "Placebo vs Dose H (S)")) +
```

Mediana code

Analysis model (continued)

```
Test(id = "Placebo vs Dose L (P)",
   method = "TTest",
   samples = samples("Dose L (P)", "Placebo (P)")) +
Test(id = "Placebo vs Dose H (P)",
   method = "TTest",
   samples = samples("Dose H (P)", "Placebo (P)")) +
Test(id = "Placebo vs Dose L (S)",
   method = "TTest",
   samples = samples("Dose L (S)", "Placebo (S)")) +
Test(id = "Placebo vs Dose H (S)",
   method = "TTest",
   samples = samples("Dose L (S)", "Placebo (S)"))
```


Mediana code

Evaluation model

```
ex3.evaluation.model = EvaluationModel() +
  Criterion(id = "Disjunctive power (P)",
            method = "DisjunctivePower",
            tests = tests("Placebo vs Dose L (P)",
                          "Placebo vs Dose H (P)"),
            labels = tests("Disjunctive power"),
            par = parameters(alpha = 0.025)) +
  Criterion(id = "Disjunctive power (S)",
            method = "DisjunctivePower",
            tests = tests("Placebo vs Dose L (S)",
                          "Placebo vs Dose H (S)"),
            labels = tests("Disjunctive power"),
            par = parameters(alpha = 0.025))
```

Mediana code

Run simulations

```
ex3.sim.parameters =
    SimParameters(n.sims = 10000,
                  proc.load = "full",
                  seed = 42938001)

ex3.results = CSE(ex3.data.model,
                  ex3.analysis.model,
                  ex3.evaluation.model,
                  ex3.sim.parameters)

summary(ex3.results)
```


Example 3: Schizophrenia trial

Simulation results

Sample size per arm	Disjunctive power	Value
110	Endpoint P	82.0%
	Endpoint S	62.5%
140	Endpoint P	90.4%
	Endpoint S	78.0%

Mediana package

Simulation report

Microsoft Word document that provides a detailed summary of assumptions and simulation results

Presentation model

User needs to define a presentation model to customize the report's structure (create sections and subsections, specify how the rows will be sorted within tables)

Mediana package

More information

Online manual at http://gpaux.github.io/Mediana

Case studies

Multiple case studies to illustrate CSE and clinical trial simulations in numerous settings at http://gpaux.github.io/Mediana/CaseStudies.html

Other Mediana code

Mediana code from *Clinical Trial Optimization Using R* (edited by Dmitrienko and Pulkstenis, 2017)

Key Multiplicity Issues in Clinical Trials (Part II)

Alex Dmitrienko (Mediana Inc)
admitrienko@medianainc.com