Multi-arm Group Sequential Designs with a Simultaneous Stopping Rule

Susanne Urach, Martin Posch

ICODOE 2016 Memphis, Tennessee, USA

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013-603160.
ASTERIX Project - http://www.asterix-fp7.eu/

Objectives of multi-arm multi-stage trials

Aim: Comparison of several treatments to a common control

Compared to separate, fixed sample two-armed trials

- less patients needed than for separate controlled clinical trials
- larger number of patients are randomised to experimental treatments
- possibility to stop early for efficacy or futility

Objective: Identify **all** treatments that are superior to control

Objective: Identify **at least one** treatment that is superior to control

Which stopping rule?

Design setup: group sequential Dunnett test

- Comparison of two treatments to a control
- Normal endpoints, variance known
- One sided tests: $H_A : \mu_A - \mu_C \leq 0$ and $H_B : \mu_B - \mu_C \leq 0$
- Control of the FamilyWise Error Rate (FWER) $= 0.025$
- Two stage group sequential trial: one interim analysis at $\frac{N_{max}}{2}$
- $Z_{A,i}$, $Z_{B,i}$ are the cumulative z-statistics at stage i=1,2

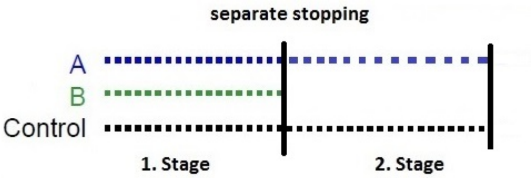
Classical group sequential Dunnett tests with “separate stopping”

Classical group sequential Dunnett tests

Objective: Identify all treatments that are superior to control

"separate stopping rule":
Treatment arms, for which a stopping boundary is crossed, stop.

E.g.:

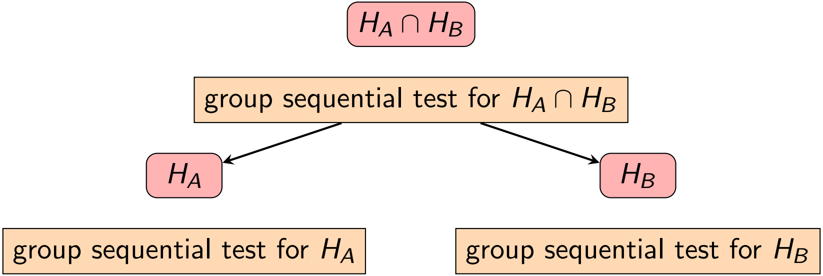


→ H_B is rejected at interim
→ A can go on and is tested again at the end

Magirr, Jaki, Whitehead (2012)

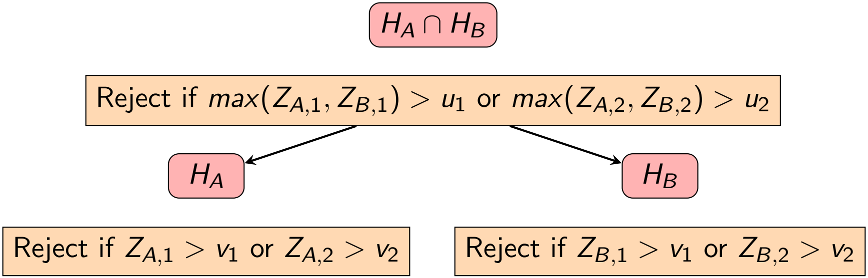
Closed group sequential tests

Local group sequential tests for $H_A \cap H_B$ and H_A, H_B are needed!!!



A hypothesis is rejected with FWER α if the intersection hypothesis and the corresponding elementary hypothesis are rejected locally at level α.

Closed group sequential tests



u_1, u_2...global boundaries

v_1, v_2...elementary boundaries

Koenig, Brannath, Bretz and Posch (2008)
Xi, Tamhane (2015)
Maurer, Bretz (2013)

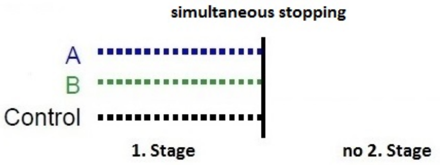
Group sequential Dunnett tests with “simultaneous stopping”

Group sequential simultaneous stopping designs

"simultaneous stopping rule":
If at least one rejection boundary is crossed, the whole trial stops.

Objective: Identify at least one treatment that is superior to control

If, e.g., H_B is rejected at interim then the trial is stopped:



Simultaneous versus Separate Stopping

- The **FWER** is controlled when using the boundaries of the separate stopping design.
- The **expected sample size (ESS)** is lower compared to separate stopping designs.
- The **power to reject**
  - **any** null hypothesis is the **same** as for separate stopping designs.
  - **both** null hypotheses is **lower** than for separate stopping designs.

$\rightarrow$ **Trade-off between ESS and conjunctive power**

Construction of efficient simultaneous stopping designs

1. Can one **relax the interim boundaries** when stopping simultaneously?
2. How large is the impact on **ESS and power** when stopping simultaneously or separately?
3. How to **optimize** the critical boundaries for either stopping rule?

Question 1: Relaxation of interim boundaries?

For simultaneous stopping:

- The boundaries u_1, u_2 for the local test of $H_A \cap H_B$ cannot be relaxed.
- The boundaries v_1, v_2 for the local test of H_j can be relaxed.

Intuitive explanation

If, e.g., H_B is rejected at interim, but H_A not, H_A is no longer tested at the final analysis and not all α is spent.

It's possible to choose improved boundaries for the elementary tests.

(similar as for group sequential multiple endpoint tests in Tamhane, Metha, Liu 2010).

What changes when stopping simultaneously?

Example: O'Brien Fleming boundaries

$H_A \cap H_B$

Reject if $max(Z_{A,1}, Z_{B,1}) > u_1$ or $max(Z_{A,2}, Z_{B,2}) > u_2$

$u_1 = 3.14, \; u_2 = 2.22$

H_A

Reject if $Z_{A,1} > v_1$ or $Z_{A,2} > v_2$

$v_1 = 2.80, \; v_2 = 1.98$

H_B

Reject if $Z_{B,1} > v_1$ or $Z_{B,2} > v_2$

$v_1 = 2.80, \; v_2 = 1.98$

What changes when stopping simultaneously?

Example: O'Brien Fleming boundaries

$H_A \cap H_B$

Reject if $max(Z_{A,1}, Z_{B,1}) > u_1$ or $max(Z_{A,2}, Z_{B,2}) > u_2$

$u_1 = 3.14, \ u_2 = 2.22$

H_A

Reject if $Z_{A,1} > v_1$ or $Z_{A,2} > v_2$

$v_1 = 2.80, \ v_2 = 1.98$

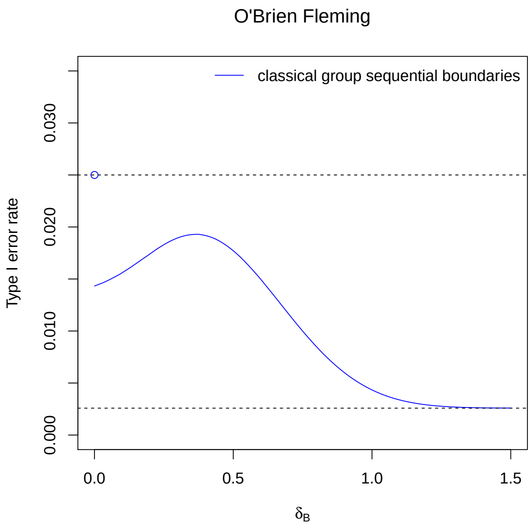
H_B

Reject if $Z_{B,1} > v_1$ or $Z_{B,2} > v_2$

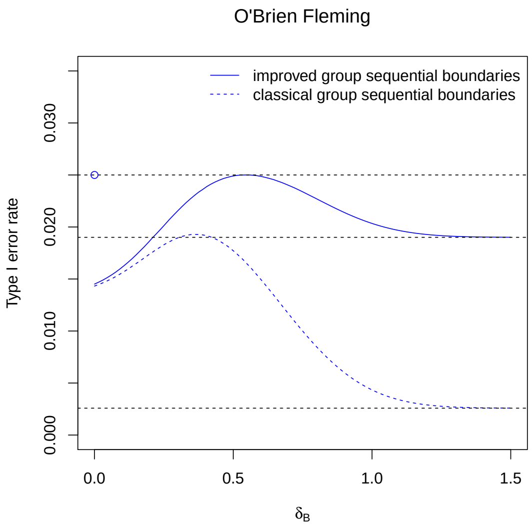
$v_1 = 2.80, \ v_2 = 1.98$

For simultaneous stopping there is no second stage test if one of the null hypotheses can already be rejected at interim.

FWER for simultaneous stopping if only H_A holds ($\delta_A = 0$)



FWER for simultaneous stopping if only H_A holds ($\delta_A = 0$)



Question 2: Impact on ESS and power?

For $\alpha = 0.025$ and $\delta_A = \delta_B = 0.5$

Conjunctive Power = Power to reject both false hypotheses
Disjunctive Power = Power to reject at least one false hypothesis

	separate stopping rule	simultaneous stopping rule	improved simultan.
Boundaries u_i for $H_1 \cap H_2$	$u_1 = 3.14,\ u_2 = 2.22$		
Interim boundary v_1	2.80	2.80	2.08
Final boundary v_2	1.98	1.98	1.98
Maximum α for test of H_j	0.025	0.019	0.025
Disj. power	0.97	0.97	0.97
N	324	324	324
ESS	230	205	205
Conj. power	0.89	0.69	0.76

Question 2: Impact on ESS and power?

For $\alpha = 0.025$ and $\delta_A = \delta_B = 0.5$

Conjunctive Power = Power to reject both false hypotheses
Disjunctive Power = Power to reject at least one false hypothesis

	separate stopping rule	simultaneous stopping rule	improved simultan.
Boundaries u_i for $H_1 \cap H_2$	$u_1 = 3.14,\ u_2 = 2.22$		
Interim boundary v_1	2.80	2.80	2.08
Final boundary v_2	1.98	1.98	1.98
Maximum α for test of H_j	0.025	0.019	0.025
Disj. power	0.97	0.97	0.97
N	324	324	324
ESS	230	205	205
Conj. power	0.89	0.69	0.76

Optimized multi-arm multi-stage designs

Optimized designs

For $\alpha = 0.025$ and $\delta_A = \delta_B = 0.5$.

Design	**"Separate stopping"**	**"Simultaneous stopping"**	**"Improved simult. stopping"**
Boundaries	group sequential	group sequential	improved group sequential
Stopping rule	separate stopping rule	simultaneous stopping rule	simultaneous stopping rule

Optimized designs

For $\alpha = 0.025$ and $\delta_A = \delta_B = 0.5$.

Design	**"Separate stopping"**	**"Simultaneous stopping"**	**"Improved simult. stopping"**
Boundaries	group sequential	group sequential	improved group sequential
Stopping rule	separate stopping rule	simultaneous stopping rule	simultaneous stopping rule
N_{max}	chosen to achieve disjunctive power of 0.9		
Obj. function to optimize u_1, u_2	expected sample size		

Optimized designs

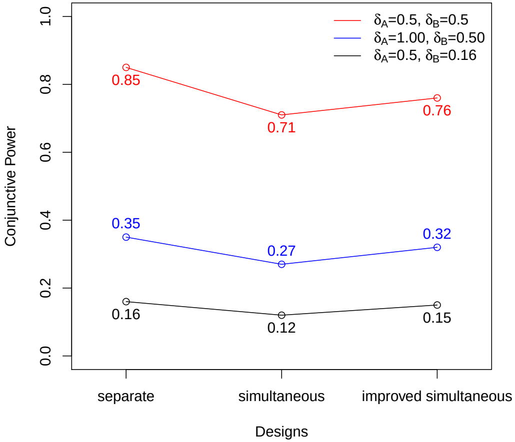
For $\alpha = 0.025$ and $\delta_A = \delta_B = 0.5$.

Design	**"Separate stopping"**	**"Simultaneous stopping"**	**"Improved simult. stopping"**
Boundaries	group sequential	group sequential	improved group sequential
Stopping rule	separate stopping rule	simultaneous stopping rule	simultaneous stopping rule
N_{max}	chosen to achieve disjunctive power of 0.9		
Obj. function to optimize u_1, u_2	expected sample size		
Obj. function to optimize v_1, v_2	expected sample size	conjunctive power	

Optimized boundaries $\delta_A = 0.5$, $\delta_B = 0.5$

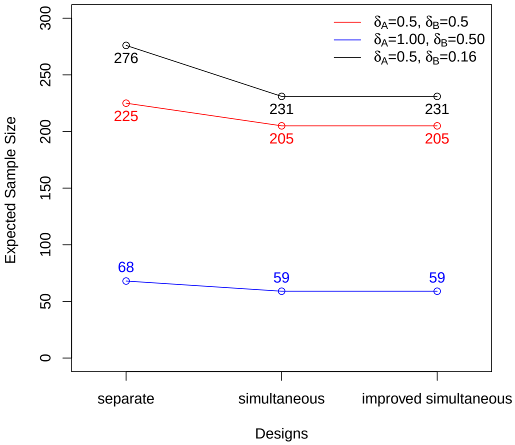
	separate	simultaneous	improved simult.
u_1	2.47	2.41	2.41
u_2	2.38	2.43	2.43
v_1	2.05	2.06	2.00
v_2	2.38	2.37	2.06
conj. power	0.85	0.71	0.76
ESS	225	205	205
N_{max}	318	324	324

Power to reject both null hypotheses



Power to reject at least one false hypothesis = 90% for all designs.

Optimal expected sample size (ESS)



ESS reduction between 8% and 16%.

Unknown variance: Extension to the t test

- P-value approach: z-score boundaries are converted to p-value boundaries and then applied to t-test p-values
- Simulation of t-statistics for p-value approach (optimized for $\delta_A = \delta_B = 1$) for $\sigma = 1$.

Design	N	α
separate	8	0.0259
	12	0.0257
	100	0.0251
improved	8	0.0261
	12	0.0258
	100	0.0250

Summary

- The **optimal design** depends on the type of objective:
  - Reject **all** hypotheses
  - Reject **at least one** hypothesis

- **Simultaneous stopping compared to separate stopping** leads to
  - lower expected sample size
  - the same power to reject any hypothesis
  - lower power to reject both hypotheses

  Improved boundaries can be used to regain some of the power to reject both null hypotheses.

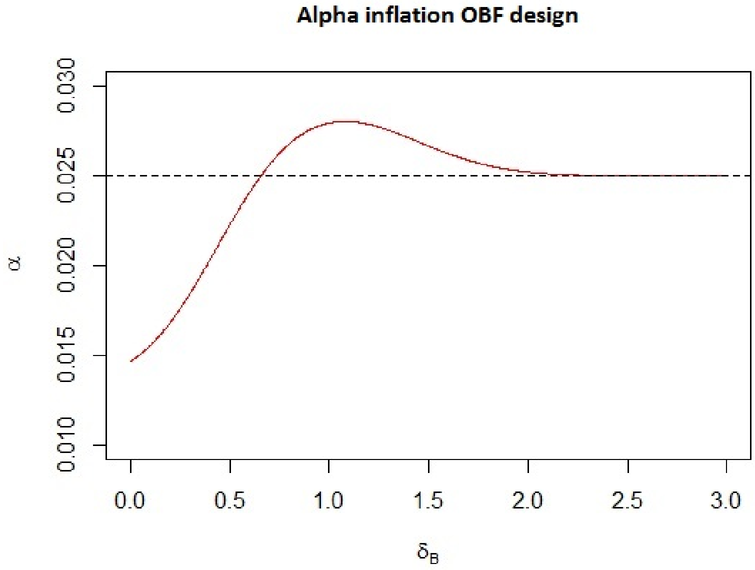
- **Limitation:** If improved boundaries are used, the simultaneous stopping rule must be adhered to!

- **Extensions:**
  - more treatment arms, stopping for futility
  - optimal choice of first stage sample size/allocation ratio

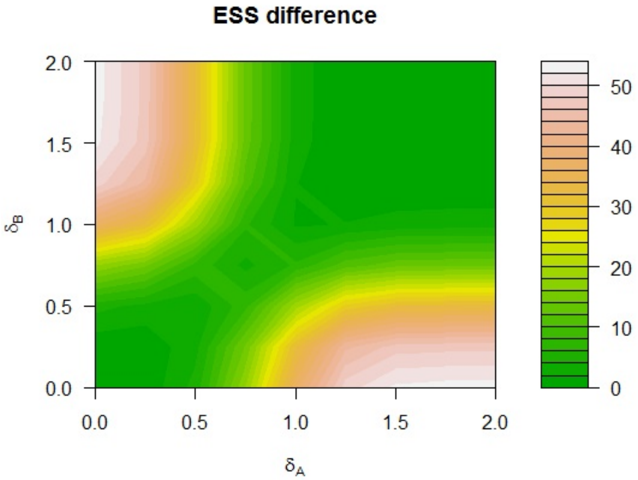
References

- Thall et al. (1989): one treatment continues, futility stopping, two stages, power comparisons under LFC
- Follmann et al. (1994): Pocock and OBF MAMS designs, Dunnett and Tukey generalisations, several stages
- Stallard & Todd (2003): only one treatment is taken forward, several stages, power comparisons
- Stallard & Friede (2008): stagewise prespecified number of treatments
- **Magirr, Jaki, Whitehead (2012)**: FWER of generalised Dunnett
- Koenig, Brannath, Bretz (2008): closure principle for Dunnett test, adaptive Dunnett test
- **Magirr, Stallard, Jaki (2014)**: Flexible sequential designs
- Di Scala & Glimm (2011): Time to event endpoints
- **Wason & Jaki (2012)**: Optimal MAMS designs
- **Tamhane & Xi (2013)**: multiple hypotheses and closure principle
- Maurer & Bretz (2013): Multiple testing using graphical approaches

FWER inflation when $u_1^* = z_{1-\alpha}$=1.96



Difference in expected sample size: OBF design



Difference in conjunctive power: OBF design