Group sequential designs for Clinical Trials with multiple treatment arms

Susanne Urach, Martin Posch

Cologne, June 26, 2015

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

Advantages:

- less patients needed than for separate controlled clinical trials
- especially important for limited set of patients (rare diseases, children)
- larger number of patients are randomised to experimental treatments
- allows changes to be made during the trial using the trial data so far, e.g. stopping for efficacy or futility

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

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

$\rightarrow$ **different kind of stopping rules!!**

Design setup: group sequential Dunnett test

- control of the FamilyWise Error Rate (FWER) $= 0.025$
- comparison of two treatments to a control
- normal endpoints, variance known
- one sided tests: $H_A : \mu_A - \mu_C \leq 0$ and $H_B : \mu_B - \mu_C \leq 0$
- two stage group sequential trial: one interim analysis at $\frac{N_{max}}{2}$
- power to reject at least one hypothesis $= 0.8$
- $Z_{A,i}$, $Z_{B,i}$ are the cumulative z-statistics at stage i=1,2

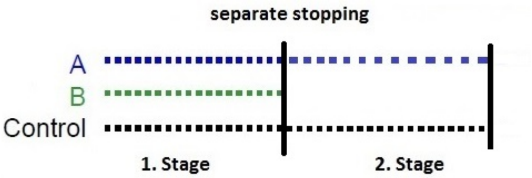
Classical group sequential Dunnett tests with “separate stopping”

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.:

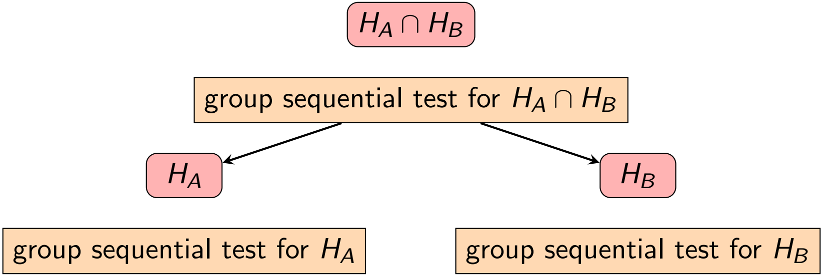


→ H_B is rejected at interim
→ A can go on and is tested again at the end

Magirr, Jaki, Whitehead (2012)

Closed testing - sequentially rejective tests

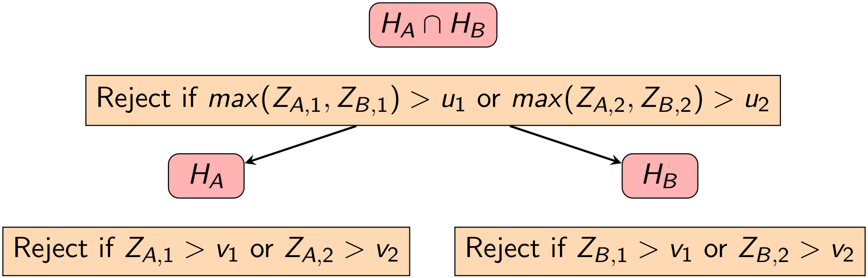
Local group sequential tests for $H_A \cap H_B$ and H_A, H_B are needed!!!



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

Xi, Tamhane (2015)
Maurer, Bretz (2013)

Closed testing - sequentially rejective tests



u_1, u_2...global boundaries

v_1, v_2...elementary boundaries

Koenig, Brannath, Bretz and Posch (2008)

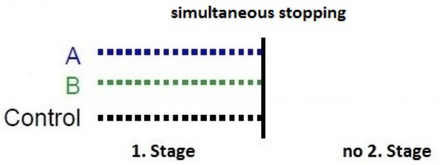
Group sequential Dunnett tests with “simultaneous stopping”

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

E.g.: H_B is rejected at interim



$\rightarrow$ There is no second stage!

Simultaneous versus Separate stopping

- **FWER** is controlled using the separate stopping design boundaries.
- **Lower expected sample size** 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.

Example: O'Brien Flemming boundaries

What changes when stopping simultaneously?

$H_A \cap H_B$

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 $Z_{A,2} > v_2$

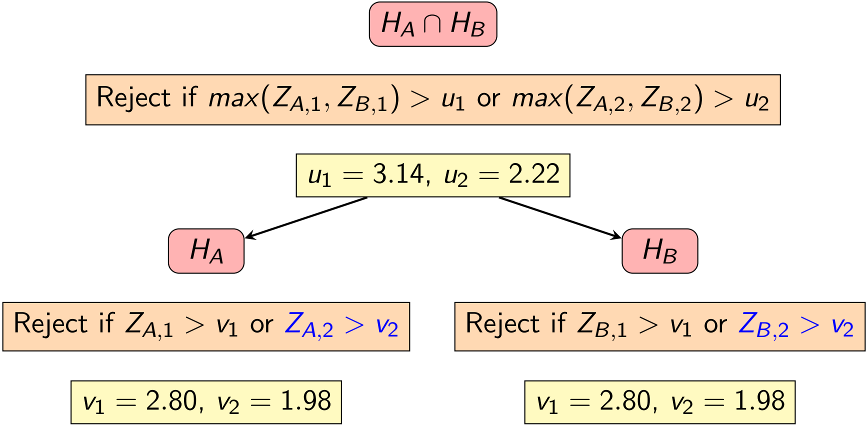
$v_1 = 2.80,\ v_2 = 1.98$

H_B

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

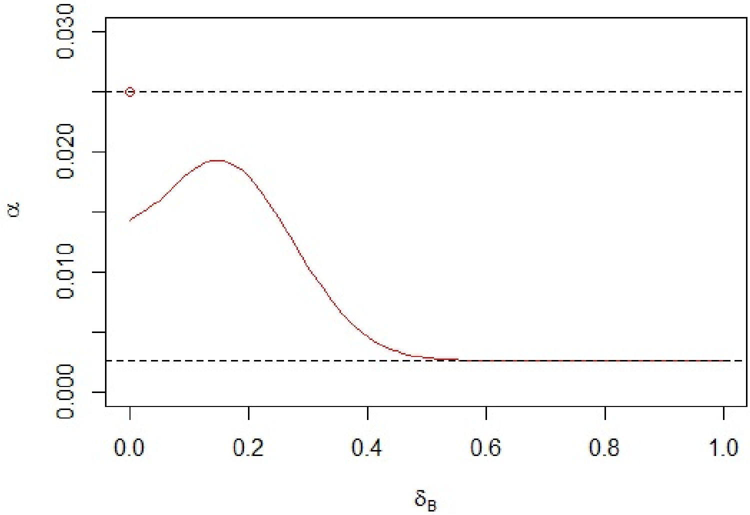
$v_1 = 2.80,\ v_2 = 1.98$

Example: O'Brien Flemming boundaries

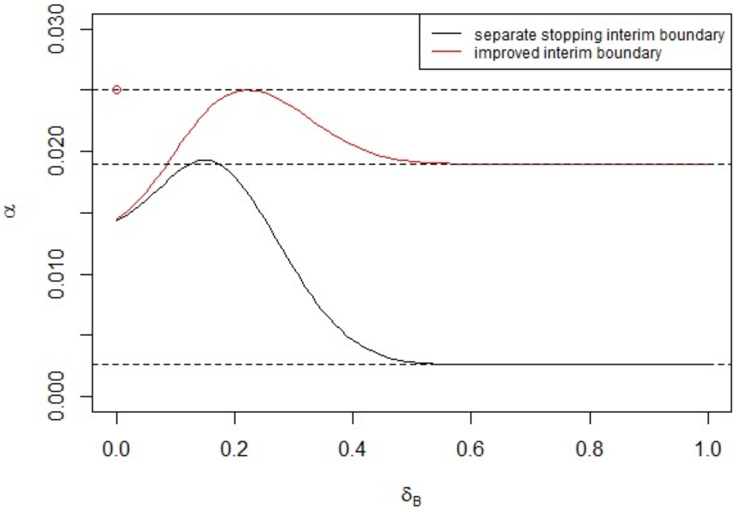


For simultaneous stopping there is no second stage test if one of the null hypotheses can already be rejected at interim.

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



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



Example: O'Brien Flemming form of rejection boundaries

Improved boundary at interim for simultaneous stopping:

$H_A \cap H_B$

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 $Z_{A,2} > v_2$

$v_1' = 2.08,\ v_2 = 1.98$

H_B

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

$v_1' = 2.08,\ v_2 = 1.98$

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

Question 2: Impact on ESS and power?

global boundaries	$u_1 = 3.14,\ u_2 = 2.22$		
local α for test of $H_A \cap H_B$	$\alpha = 0.025$		
	separate stopping rule	simultaneous stopping rule	improved simultan.
local α for test of H_j	0.025	0.019	0.025
interim boundary v_1	2.80	2.80	2.08
final boundary v_2	1.98	1.98	1.98

Question 2: Impact on ESS and power?

global boundaries	$u_1 = 3.14,\ u_2 = 2.22$		
local α for test of $H_A \cap H_B$	$\alpha = 0.025$		
	separate stopping rule	simultaneous stopping rule	improved simultan.
local α for test of H_j	0.025	0.019	0.025
interim boundary v_1	2.80	2.80	2.08
final boundary v_2	1.98	1.98	1.98
disj. power	0.8	0.8	0.8
N	162	162	162

Question 2: Impact on ESS and power?

global boundaries	$u_1 = 3.14,\ u_2 = 2.22$		
local α for test of $H_A \cap H_B$	$\alpha = 0.025$		
	separate stopping rule	simultaneous stopping rule	improved simultan.
local α for test of H_j	0.025	0.019	0.025
interim boundary v_1	2.80	2.80	2.08
final boundary v_2	1.98	1.98	1.98
disj. power	0.8	0.8	0.8
N for $\delta_A = \delta_B = 0.5$	162	162	162
ESS for $\delta_A = \delta_B = 0.5$	154	149	149
conj. power for $\delta_A = \delta_B = 0.5$	0.59	0.50	0.56

Optimized multi-arm multi-stage designs

Optimal designs

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

Optimal designs

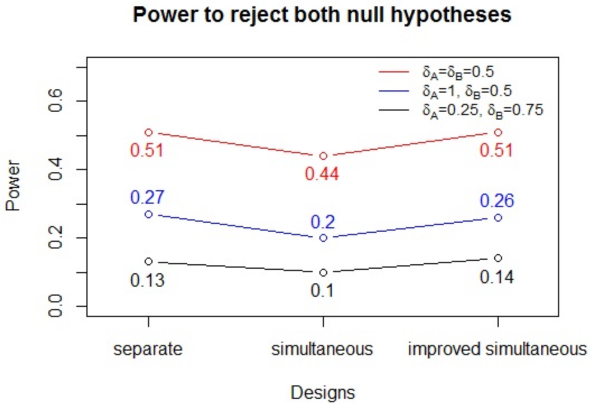
Scenario	"Separate stopping"	"Simultaneous stopping"	"Improved simult. stopping"
Boundaries	classical group sequential	classical group sequential	improved group sequential
Stopping rule	separate stopping rule	simultaneous stopping rule	simultaneous stopping rule
N_{max}	chosen to achieve disjunctive power of 0.8		
Obj. function optimize u_1, u_2	minimize ESS under certain parameter configuration		

Optimal designs

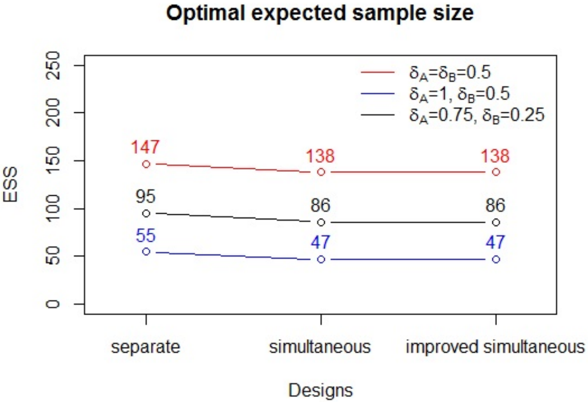
Scenario	"Separate stopping"	"Simultaneous stopping"	"Improved simult. stopping"
Boundaries	classical group sequential	classical group sequential	improved group sequential
Stopping rule	separate stopping rule	simultaneous stopping rule	simultaneous stopping rule
N_{max}	chosen to achieve disjunctive power of 0.8		
Obj. function optimize u_1, u_2	minimize ESS under certain parameter configuration		
Obj. function optimize v_1, v_2	minimize ESS	maximize conjunctive power	

Power to reject both null hypotheses

Power to reject at least one hypothesis = 0.8



Optimal expected sample size (ESS)



Remarks: Percentual reduction gets bigger, but stays between 5 and 12%

Summary

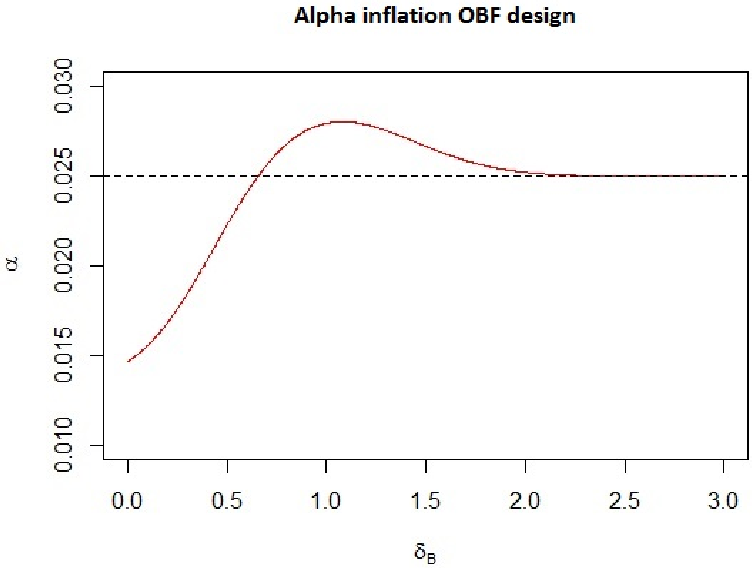
- The **optimal design** depends on the objective:
  - Reject **all** hypotheses
  - Reject **at least one** hypothesis
- **Simultaneous stopping compared to separate stopping** leads to
  - lower expected sample size
  - the same power to reject any hypothesis
  - lower power to reject both hypotheses

  Improved boundaries can be used to rescue some of the power to reject both null hypotheses.
- **Limitation:** If improved boundaries are used, the simultaneous stopping rule must be adhered to!
- **Extensions:**
  - unknown variance: t-test: p-value approach
  - more treatment arms, stopping for futility
  - optimal choice of first stage sample size/allocation ratio

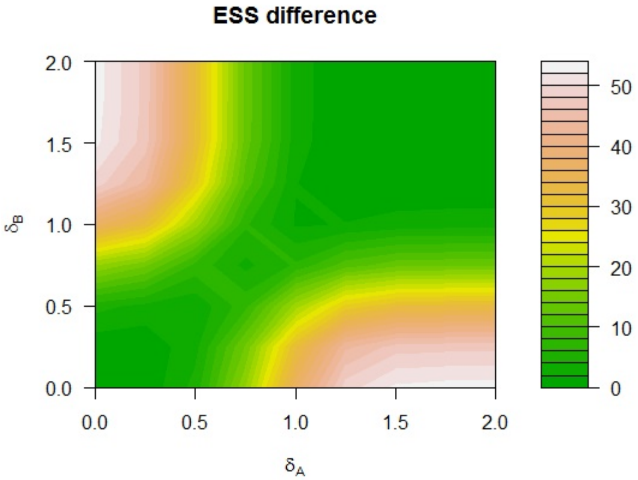
References

- 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

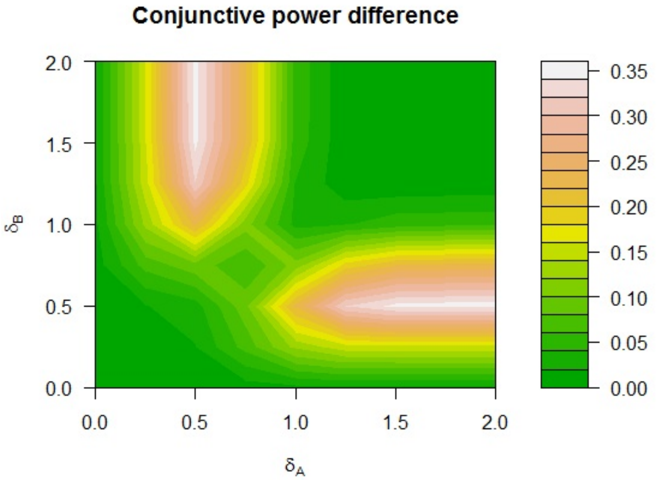
FWER inflation when $u_1^* = z_{1-\alpha}$=1.96



Difference in expected sample size: OBF design



Difference in conjunctive power: OBF design



Unknown variance: Extension to the t test

- **p-value approach = quantile substitution (Pocock (1977)):** z-score boundaries are converted to p-value boundaries and then converted to t-score boundaries:

$$u_i' = T_{2n_i-2}(\Phi^{-1}(u_i))$$

- **for known variance:** sample size per arm per stage n of **8** for a power to reject at least one of 0.8 at $\delta_A = \delta_B = 1$

(separate: ESS=32/power=0.61; improved simultaneous: ESS=30/power=0.51)

Simulation of t-statistics for p-value approach ($\delta_A = \delta_B = 1$)					
Design	n	α	power at least one	power both	ESS
separate	8	0.0260	0.80	0.56	34
separate	10	0.0258	0.89	0.70	43
imp. sim.	8	0.0260	0.79	0.49	32
imp. sim.	10	0.0258	0.88	0.61	39

Optimal boundaries

$\delta_A = 0.5,\ \delta_B = 0.5$			
Design	separate	simultaneous	improved simult.
u_1	2.64	2.48	2.48
u_2	2.29	2.37	2.37
v_1	2.09	2.16	2.05
v_2	2.29	2.20	1.97
conj. power	0.51	0.44	0.51
ESS	147	138	138