MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Adaptive Methods in Clinical Research

Lecture 3: Multi-Arm Multi-Stage (MAMS) Designs

Alessandra Serra

MRC Biostatistics Unit

September 2024

Multiple treatments

Treatments $T_1, \ldots, T_K$ are available

One of them is to be selected for further study

- All have passed Phase I
- They are distinct therapies or different doses of the same drug were no dose-response relationship is assumed
- In Phase III the selected treatment will be compared to the control

Naive approach

Conduct separate Phase II trials for each T_k

- sample size will get large
- especially if α is reduced to allow for **multiple comparisons**

$\rightarrow$ special methods have been developed

The fixed sample size (single-stage) design

Single stage selection screens

Suppose that responses to T_k are

$$Y_{ki} \sim N(\mu_k, \sigma^2),\ k = 0, 1, \ldots, K, i = 1, \ldots, n_k$$

T_0 is the control treatment, there are n_k patients treated on T_k, and the mean of their responses is $\bar{y}_k$

Put

$$Z_k = \frac{1}{\sigma}\sqrt{\frac{n_0 n_k}{n_0 + n_k}}(\bar{y}_k - \bar{y}_0)$$

then

$$\boldsymbol{Z} \sim N(\boldsymbol{m}, \boldsymbol{V})$$

with

$$m_k = \frac{1}{\sigma}\sqrt{\frac{n_0 n_k}{n_0 + n_k}}(\mu_k - \mu_0); \quad v_{kk'} = \sqrt{\frac{n_k n'_k}{(n_0 + n_k)(n_0 + n'_k)}},\ k \neq k'; \quad v_{kk} = 1$$

Single stage selection screens

Select treatment $T_{k^\star}$ where

$$Z_{k^\star} = \max(Z_1, \ldots, Z_K)$$

provided that

$$Z_{k^\star} \geq u$$

Otherwise, make no selection

Finding a design

- Put $n_k = r_k n_0,\ k = 1, \ldots, K$, for chosen values of $r_1, \ldots, r_K$
- Often $r_1 = \cdots = r_K$;
  although the common value may be < 1
- Have to find n_0 and $u \Rightarrow$ need 2 equations

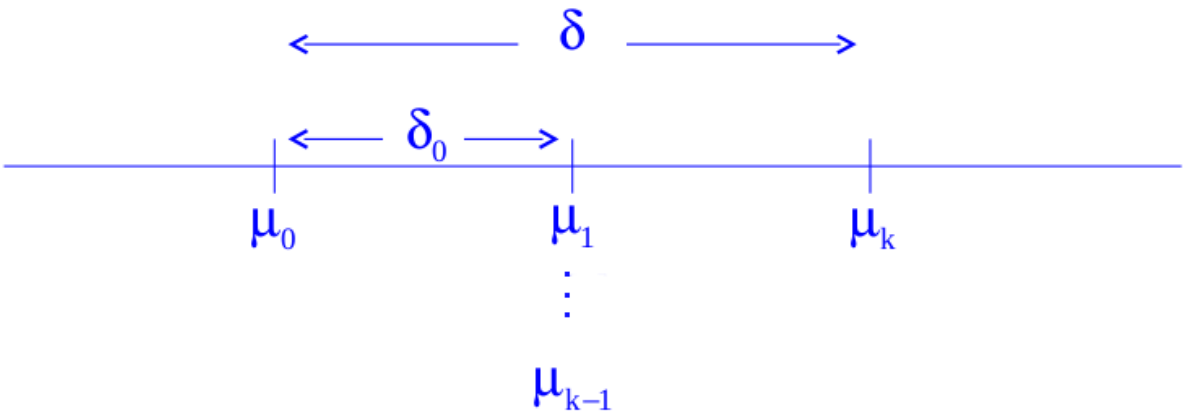
Specifying the accuracy

$$
\begin{aligned}
\nu &= P(\text{select a treatment} \mid \mu_0 = \mu_1 = \cdots = \mu_K) \\
1-\eta &= P(\text{select } T_k \mid \mu_k = \mu_0 + \delta \text{ and} \\
&\qquad \mu_k = \mu_0 + \delta_0,\ k = 1, \ldots, K-1)
\end{aligned}
$$

were $0 \leq \delta_0 < \delta$.

Set ν to be small and $1 - \eta$ to be large

Least favorable configuration LFC



$$1 - \eta = P(\text{select } T_k \mid \text{LFC})$$

if $\mu_{r+1}, \ldots, \mu_k \geq \mu_0 + \delta$ and $\mu_1, \ldots, \mu_r \leq \mu_0 + \delta_0$ then $T_{r+1}, \ldots, T_k$ are good treatments, while the rest is not good enough. If follows that

$$P(\text{select one of } T_{r+1}, \ldots, T_k) \geq 1 - \eta$$

Example

$\sigma = 4.4$
$\nu = 0.05,\ 1 - \eta = 0.9$
$K = 4,\ r_1 = \cdots = r_4 = 1$

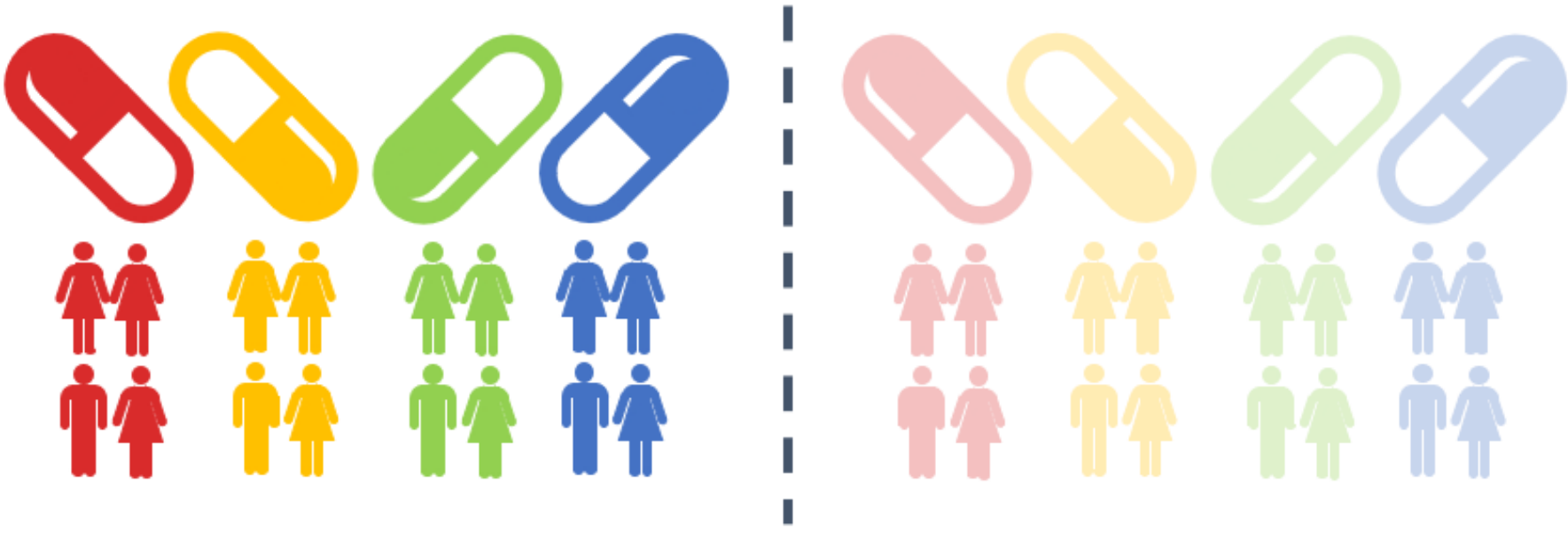
δ	δ_0	n	n_0
2.5	0.625	375	75
2.0	0.500	585	117
1.5	0.375	1040	208

$n = (K + 1)n_0$

Comparison

- 4 separate comparative trials;
- $\alpha = 0.05,\ 1 - \beta = 0.9$ and $\delta = 2$
- Sample size per trial 204
- Total sample size 816
- Reduce sample size by $\approx 30\%$
- Without adjusting for multiplicity

Two-stage selection process



Two-stage selection screens

Suppose that responses to T_k are
$Y_{ki} \sim N(\mu_k, \sigma^2),\ k = 0, 1, \ldots, K, i = 1, \ldots, n_k.$

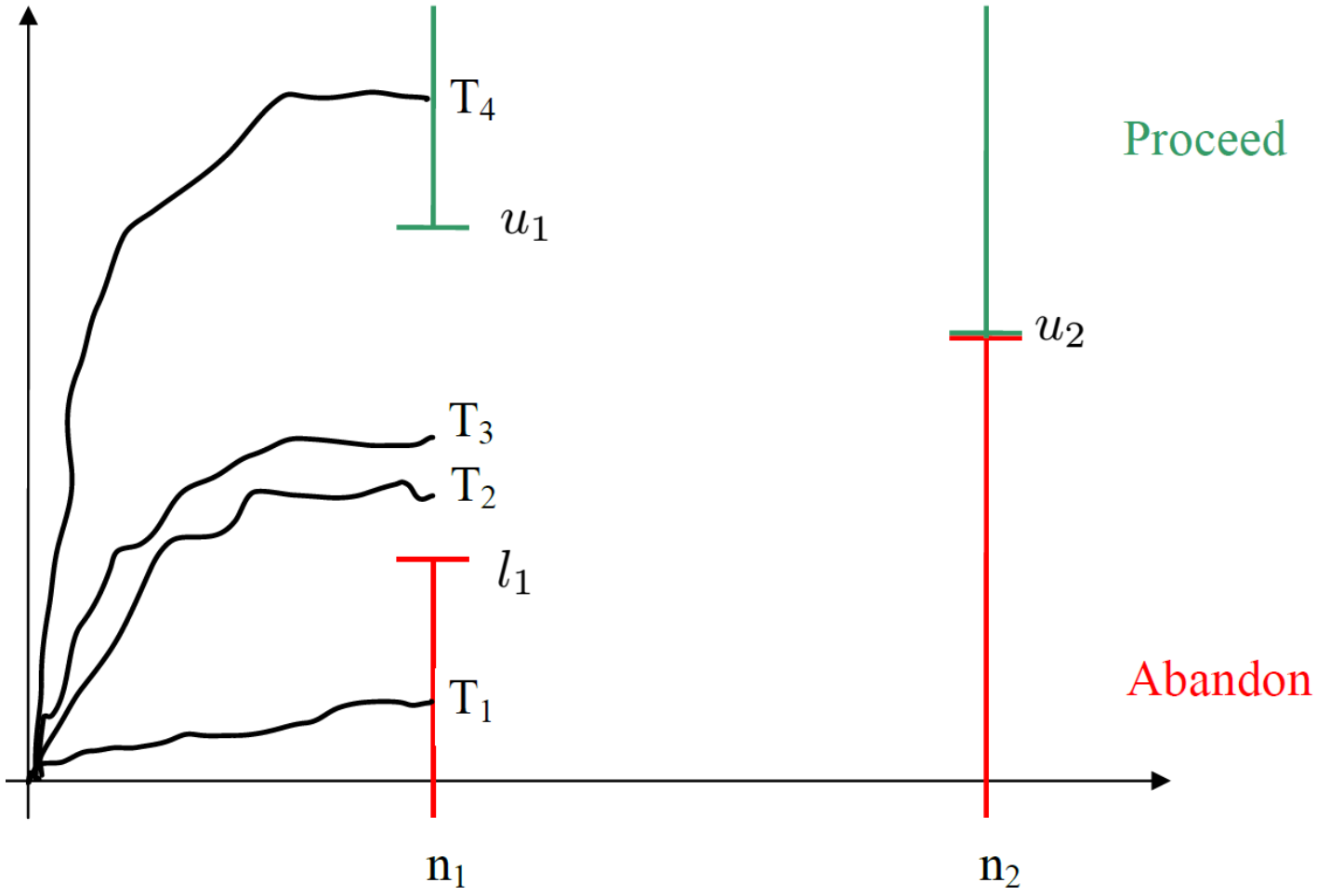
In Stage 1, n_k patients are treated on T_k, with mean response $\bar{y}_{1k}$

$$Z_{1k} = \frac{1}{\sigma}\sqrt{\frac{n_0 n_k}{n_0 + n_k}}(\bar{y}_{1k} - \bar{y}_{10})$$

After Stage 1, an interim analysis is conducted

- If $Z_{1k} < l_1$, then T_k will be dropped
- If $Z_{1k'} = \max(Z_{11}, \ldots, Z_{1K}) > u_1$, then $T_{k'}$ will be selected

Stopping early



T_1 eliminated and T_4 selected at interim

Best treatment only

(e.g. Stallard & Todd, 2003; Whitehead & Jaki 2009)

- If no treatment is selected after Stage 1
- and not all are dropped

then n_k further patients are treated on the control and **the best remaining active** treatment, $T_{k'}$.

At the end of Stage 2, the statistic

$$Z_{2k'} = \frac{1}{\sigma}\sqrt{\frac{2n_0 n_{k'}}{n_0 + n_{k'}}}(\bar{y}_{2k'} - \bar{y}_{20})$$

is calculated, where $\bar{y}_{2k'}$ denotes the mean response on $T_{k'}$ over all $n_{k'}$ patients.

If $Z_{2k'} > u_2$, then $T_{k'}$ will be selected.

All promising treatments

(Magirr et al, 2012)

- If no treatment is selected after Stage 1
- and not all are dropped

then n_k further patients are treated on the control and and **on each of the remaining active** treatments.

At the end of Stage 2, the statistics

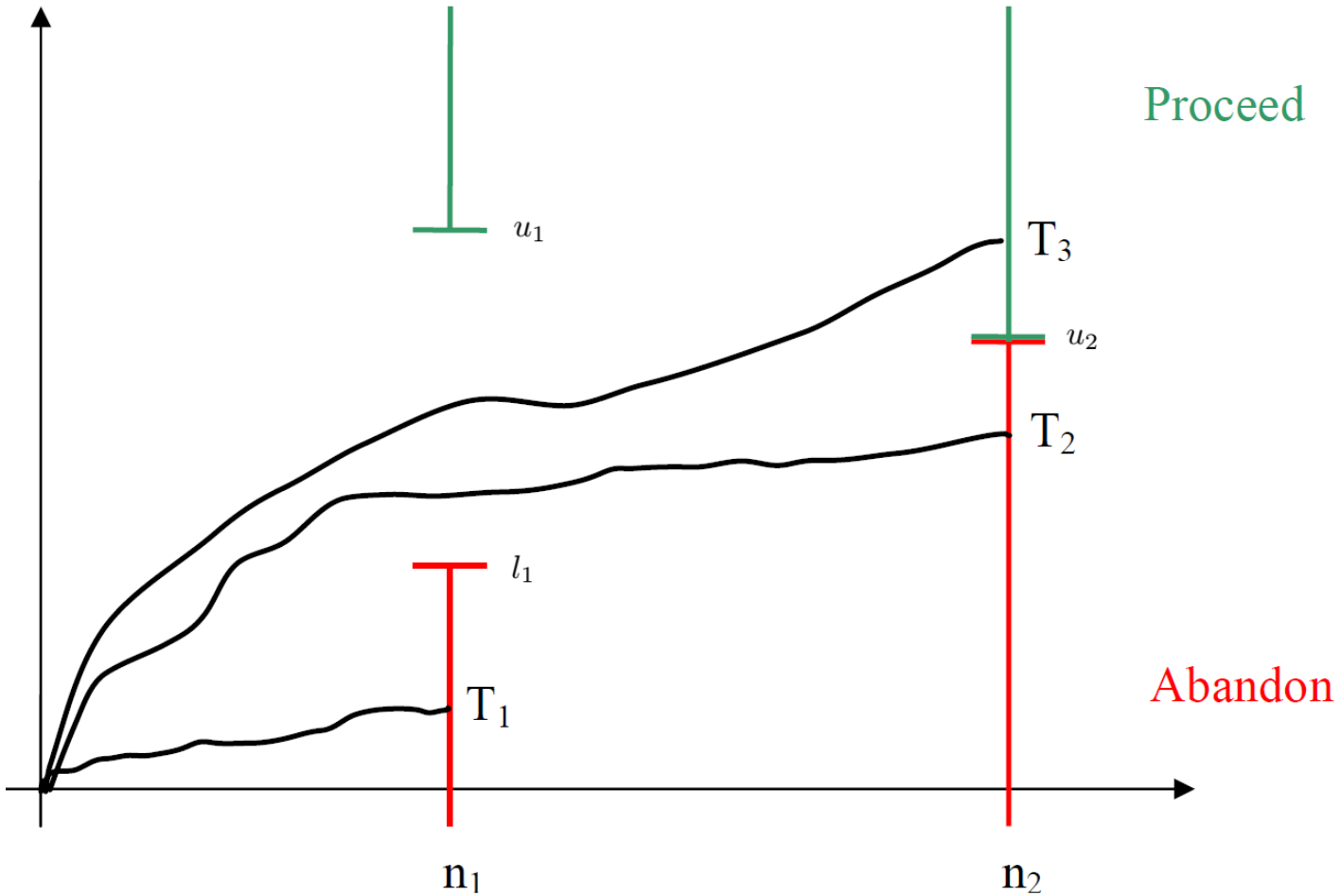
$$Z_{2k} = \frac{1}{\sigma}\sqrt{\frac{2n_0 n_k}{n_0 + n_k}}(\bar{y}_{2k} - \bar{y}_{20})$$

are calculated, where $\bar{y}_{2k}$ denotes the mean response on T_k over all n_k patients.

If $Z_{2k'} = \max(Z_{21}, \ldots, Z_{2k}) > u_2$, then $T_{k'}$ will be selected.

All promising treatments

(Magirr et al, 2012)



T_1 eliminated at interim and T_3 selected at final analysis

Drop the loser

- An alternative is to fix the number of treatments that will be dropped during each stage;
- This is called as the “drop-the-loser” design
- The design controls the FWER and the sample size can be pre-specified in advance;
- However, one has to stick with the pre-specified rules for dropping (even if two treatments perform very similar).

Finding a design

$$\begin{aligned} \nu &= P(\text{select a treatment} \mid \mu_0 = \mu_1 = \cdots = \mu_K) \\ 1-\eta &= P(\text{select } T_k \mid \mu_k = \mu_0 + \delta \text{ and} \\ &\qquad \mu_{k'} = \mu_0 + \delta_0,\ k' = 1, \ldots, K-1) \end{aligned}$$

as before.

- Fix the sample size ratios $r_k = m_k/m_0 \Rightarrow$ find $l_1,\ u_1,\ u_2,\ m_0$
- Further constraints can be set, such as

  (i) $l_1 = 0,\ u_1 = \infty$
  (ii) $l_1 = -\frac{u_2}{3}, u_1 = \infty$
  (iii) $l_1 = 0, u_1 = \frac{4}{3}u_2$

  and a solution can be computed and evaluated

Example: all promising

$\sigma = 4.4,\ \delta = 2,\ \delta_0 = 0.5$
$\nu = 0.05,\ 1 - \eta = 0.9$
$K = 4,\ r_1 = \cdots = r_4 = 1$

l_1	u_1	u_2	n_0	n_{max}	$ESS_0(N)$	$ESS_{LFC}(N)$
0	∞	2.15	59	590	460	542
$-\frac{u_2}{3}$	∞	2.17	59	590	532	574
0	$\frac{4u_2}{3}$	2.18	61	610	474	460

n_{max} is the maximum sample size

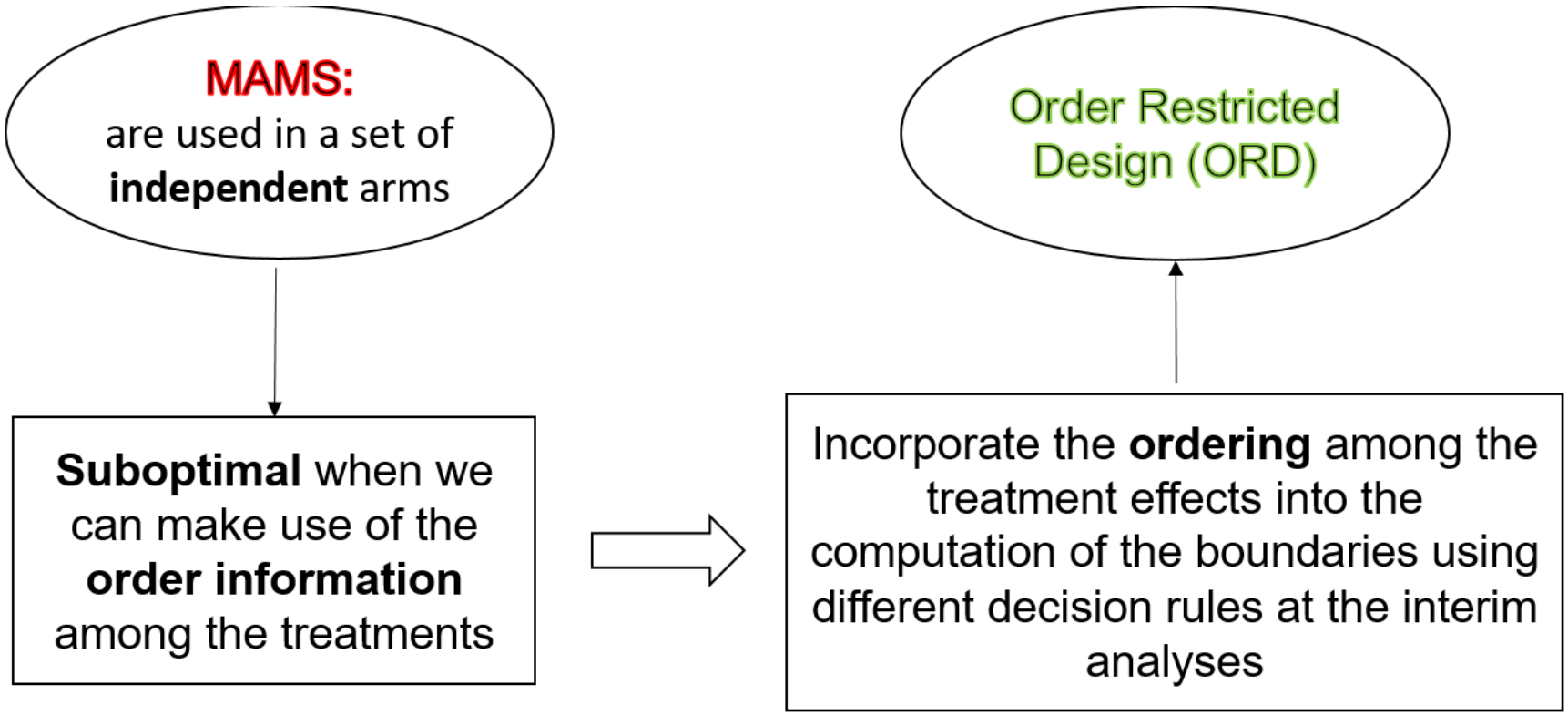
$ESS_0(N)$ and $ESS_{LFC}(N)$ are the expected total sample sizes under the null and least favourable configuration, respectively

Extensions

- Covariates can be included under orthogonality
  - Orthogonality holds if same number of subjects on each treatment and stage
  - approximately holds for large n
- Non-normal endpoints based on asymptotics

Beyond Independent Treatment Arms

Aim of the study: select all the promising treatment arms.



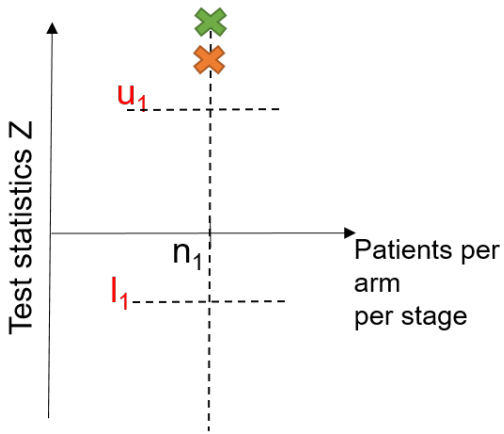
Notations

Let consider a MAMS with:

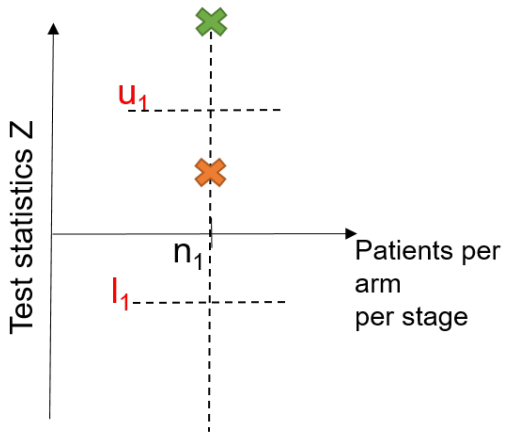
- **three treatment arms:** control, treatment L and treatment S
- **one interim analysis**
- **pre-specified allocation** of patients
- Z_{1k} the test statistic relative to the arm k at the first stage
- $\theta^{(L)} \geq \theta^{(S)}$
- $\theta^{(L)}$ and $\theta^{(S)}$ are the effects at the longest and shortest treatment durations, respectively

Decision rules when $\theta^{(L)} \geq \theta^{(S)}$, $Z_{1L} \geq u_1$

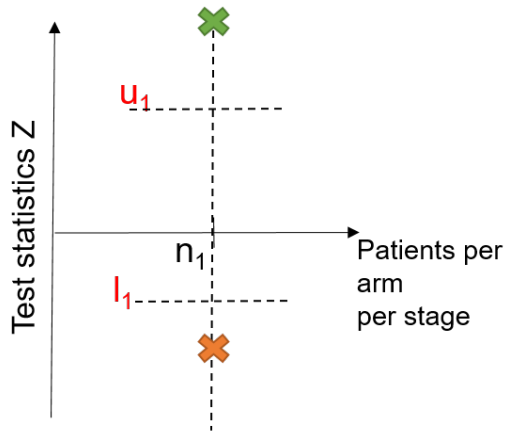
Treatments: L and S



Stop the trial: select L and S



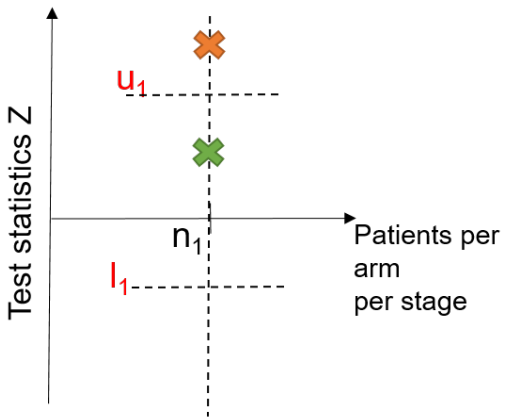
Continue with S



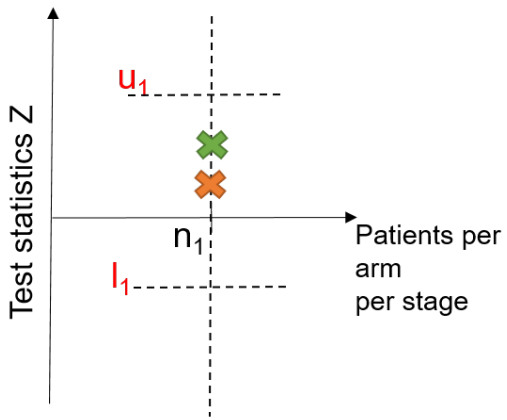
Stop the trial: select L

Decision rules when $\theta^{(L)} \geq \theta^{(S)},\ l_1 < Z_{1L} < u_1$

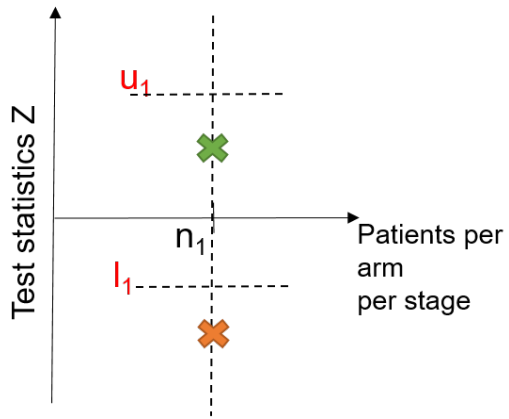
Treatments: L and S



Continue with L and S



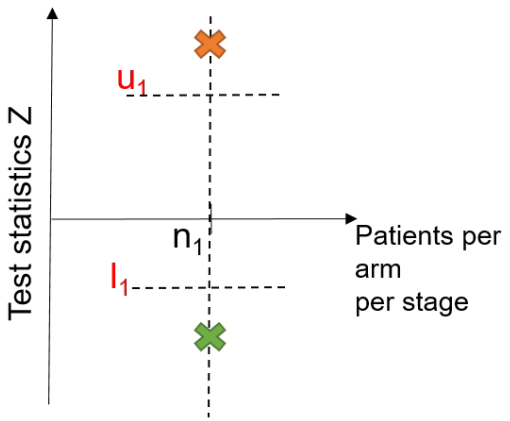
Continue with L and S



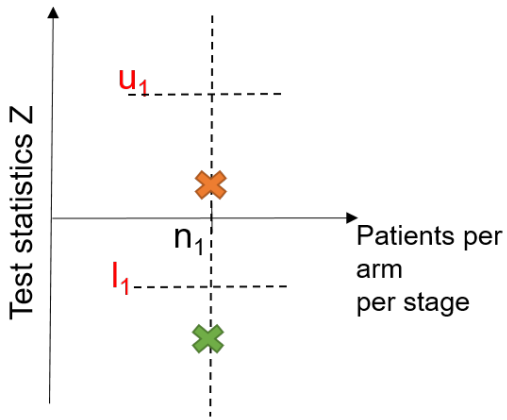
Continue with L

Decision rules when $\theta^{(L)} \geq \theta^{(S)}$, $Z_{1L} \leq l_1$

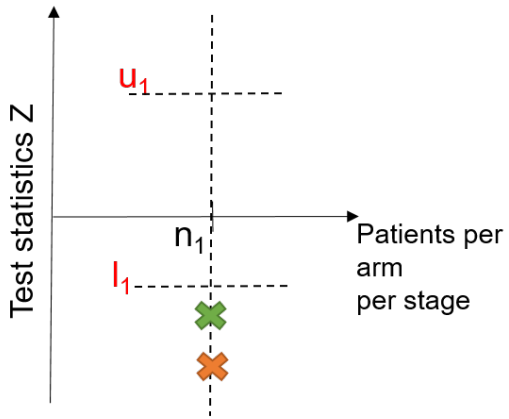
Treatments: L and S



Continue with L and S



Stop the trial



Stop the trial

Family-wise Error Rate and Power requirement

- **Family-wise Error Rate (FWER)**

  Control of the FWER:

  $P(\text{reject at least one true } H_{0k}, k \in \{L, S\} | \text{null treatment effect}) \leq \nu$

- **Power requirement**

  Power the study at $(1 - \eta)$ to reject both hypotheses under $\boldsymbol{\theta} = (\theta^{(L)}, \theta^{(S)})$, where $\theta^{(L)} \geq \theta^{(S)} \geq \delta_0 > 0$ and δ_0 the minimum clinically relevant difference.

Simulations

Simulations were run to compare the 3-arm 2-stage Order Restricted Design (ORD) with:

a) Fixed Sample Design (FSD)

b) Fixed Sample Design with hierarchical test (FSD(h)): that is the 3-arm 1-stage ORD

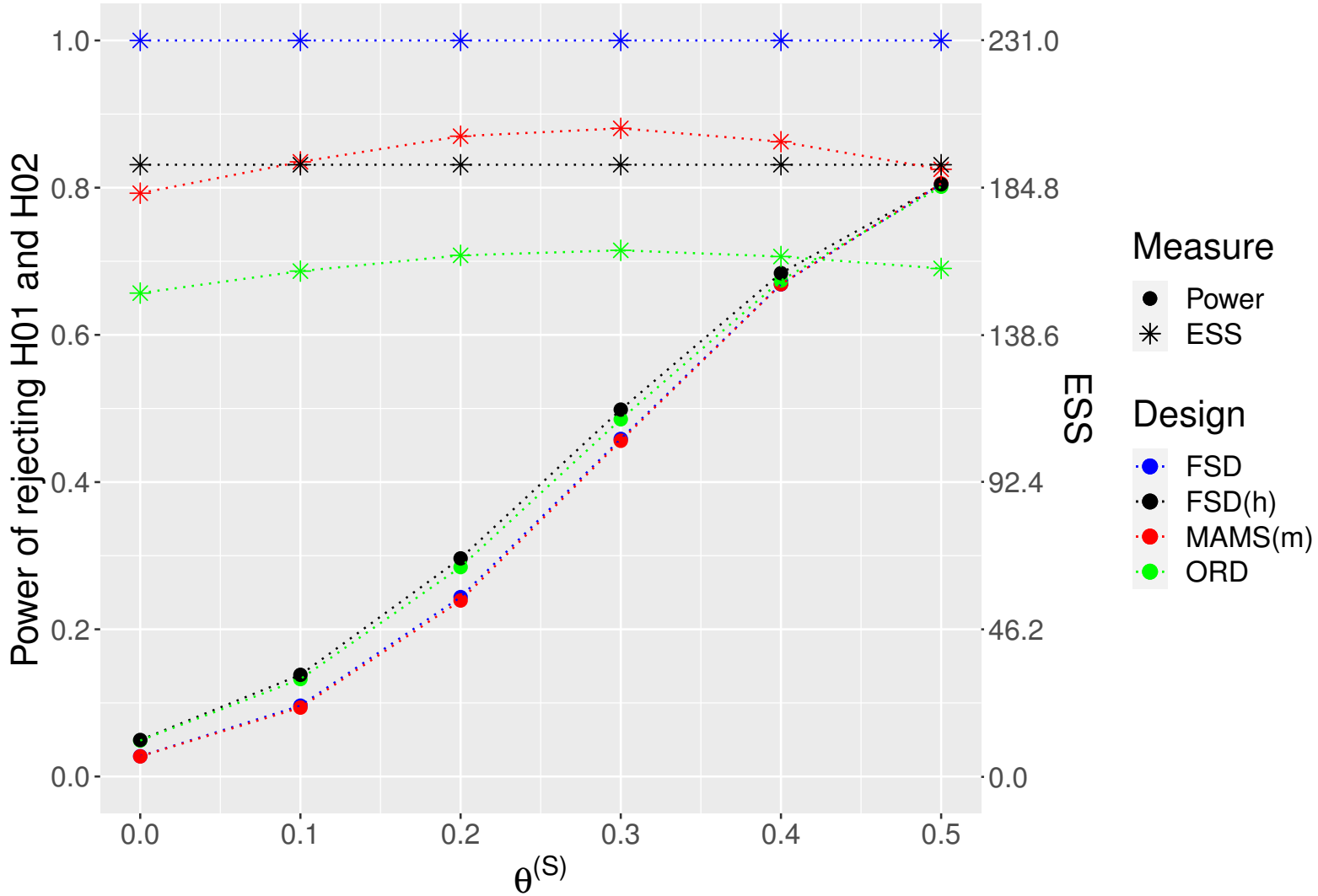
c) modified MAMS design (MAMS(m)): the trial is continued until the decision on each arm has been made

The **measure of performance** is the probability of rejecting both hypotheses.
The **efficiency** of the proposed design is measured by its expected sample size (ESS).

Settings

- The FWER at level $\nu = 0.05$
- all designs are powered at 80% to reject both hypotheses under $\boldsymbol{\theta} = (0.5, 0.5)$
- different scenarios of $\boldsymbol{\theta} = (0.5, \theta^{(S)}),\ \theta^{(L)} \geq \theta^{(S)}$
- use of triangular bounds for the 3-arm 2-stage ORD and MAMS(m) designs

Probability to reject both hypotheses



Discussion

- MAMS trials can provide huge efficacy gains (compared to the several two parallel arm trials) while maintaining type I and type II errors;
- Methods for MAMS are developed and recognised by both funders and HAs;
- MAMS designs are implemented in many `R`-packages including `MAMS`, `rpact`, and `gsDesign`.

References (1)

Dunnett CW (1984) Selection of the best treatment in comparison to a control with an application to a medical trial. In: Santner TJ, Tamhane AC (Eds). Design of experiments: Ranking and selection. New York: Marcel Dekker.

Jaki T, Magirr D. (2013) Considerations on covariates and endpoints in multi-arm multi-stage clinical trials selecting all promising treatments. Statistics in Medicine. 32, 1150-1163.

Jennison, C. and Turnbull, B.W. (2000). Group Sequential methods with Applications to Clinical Trials. Boca Raton: CRC.

Magirr D, Jaki T, Whitehead J (2012). A generalized Dunnett test for multi-arm multi-stage clinical studies with treatment selection. Biometrika, 99, 494-501.

Stallard N, Todd S (2003). Sequential designs for phase III clinical trials incorporating treatment selection. Statistics in Medicine, 22, 689-703.

Wason J, Stallard N, Bowden J and Jennison C (2017). A multi-stage drop-the-losers design for multi-arm clinical trials. Statistical methods in medical research, 26(1), pp.508-524.

Whitehead J, Jaki T (2009). One- and two-stage design proposals for a phase II trial comparing three active treatments with control using an ordered categorical endpoint. Statistics in Medicine, 28, 828-847.