B2. Adaptive clinical trials and response-adaptive randomisation

David Robertson

david.robertson@mrc-bsu.cam.ac.uk

29th October 2020

Overview

Introduction to adaptive clinical trials

Response-adaptive randomisation

Flexible adaptive designs

Introduction to adaptive clinical trials

Traditional clinical trials

Traditional clinical trials:

1. The trial is designed, satisfying type I error and power constraints
2. The trial is conducted (as prescribed by the design) until planned sample size has been recruited
3. After all patients are assessed for their response to treatment, the data are analysed according to a pre-specified analysis plan

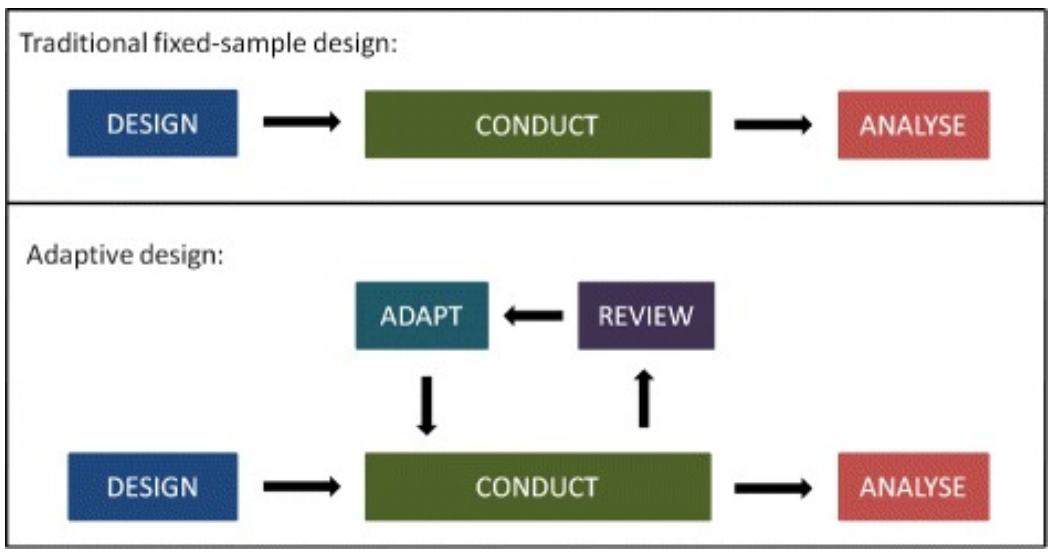
This is straightforward, but inflexible: no option to make any changes that may become desirable or necessary during the course of the trial

Q: Can we improve on this fixed-sample trial design?

A: Adaptive clinical trials provide an alternative

What is an adaptive clinical trial?

A clinical trial that allows for changes in the design during the course of the study, based on data generated during the study, while preserving statistical integrity



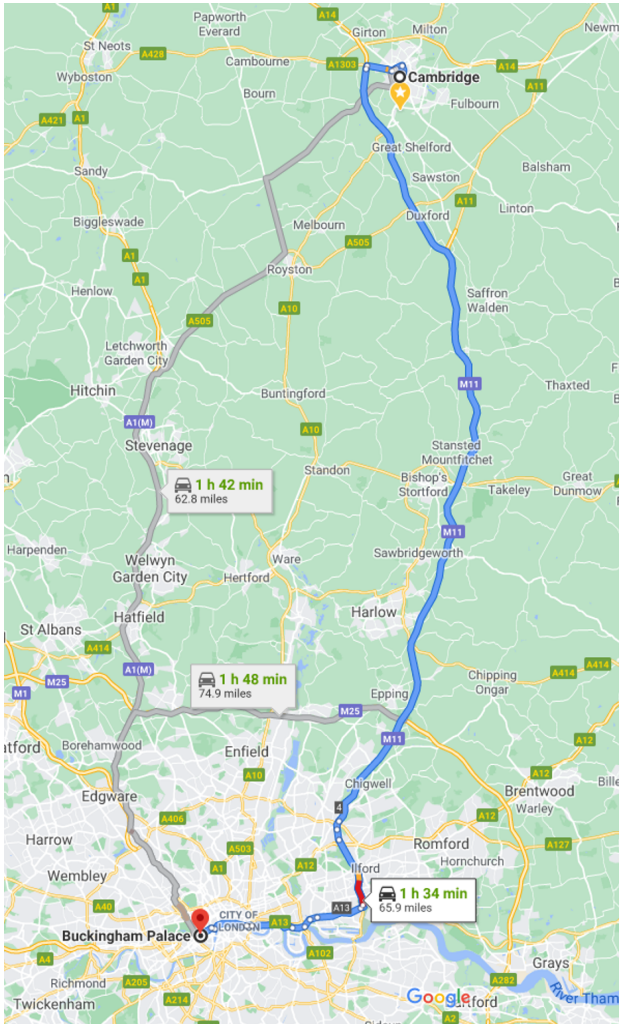
Pallman et al. (2018). Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Medicine, 16:29

Examples of adaptations

- Sample size
- Stopping recruitment to a treatment (or the whole trial) at an earlier stage for success or lack of efficacy
- Allocation probabilities of patients to trial arms
- Identifying patients most likely to benefit from treatment and focusing recruitment efforts on them
- Treatment regimens of the different groups
- ...

Types of adaptive designs

- Traditional fixed-sample (non-adaptive) designs
- Pre-specified adaptive designs
- Flexible adaptive designs



Why adapt?

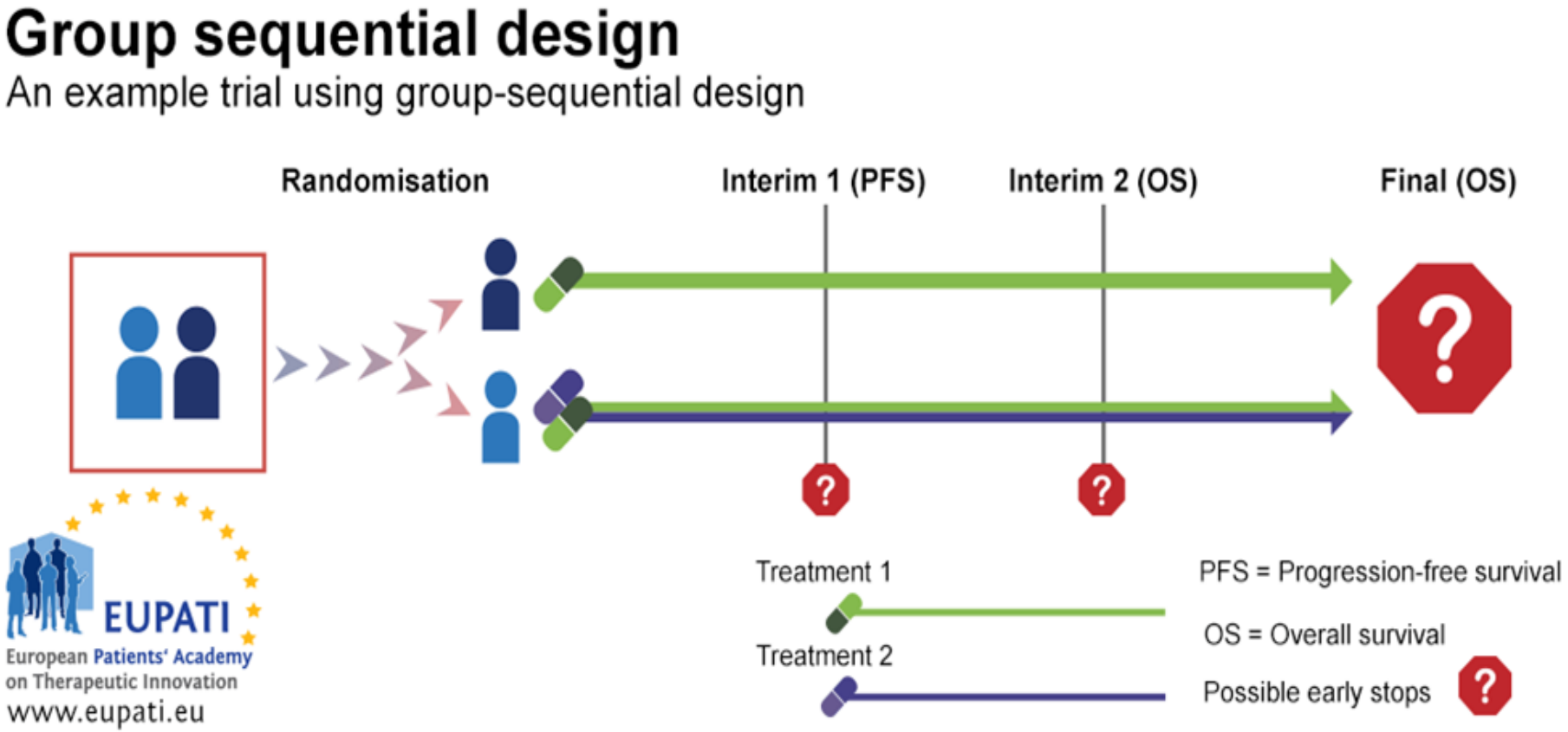
- *Ethical rationale*: lower the number of patients potentially exposed to unsafe/ineffective drugs.
- *Economic rationale*: lower the sample sizes, allow early stopping, shorten development timelines.
- *Administrative rationale*: check on accrual, eligibility and compliance, and generally ensure trial is being carried out as per protocol.
- *Scientific rationale*: can provide more information, and can increase the probability of success.

Group sequential trials

Lecture B3

? How to set the time for an interim analyses

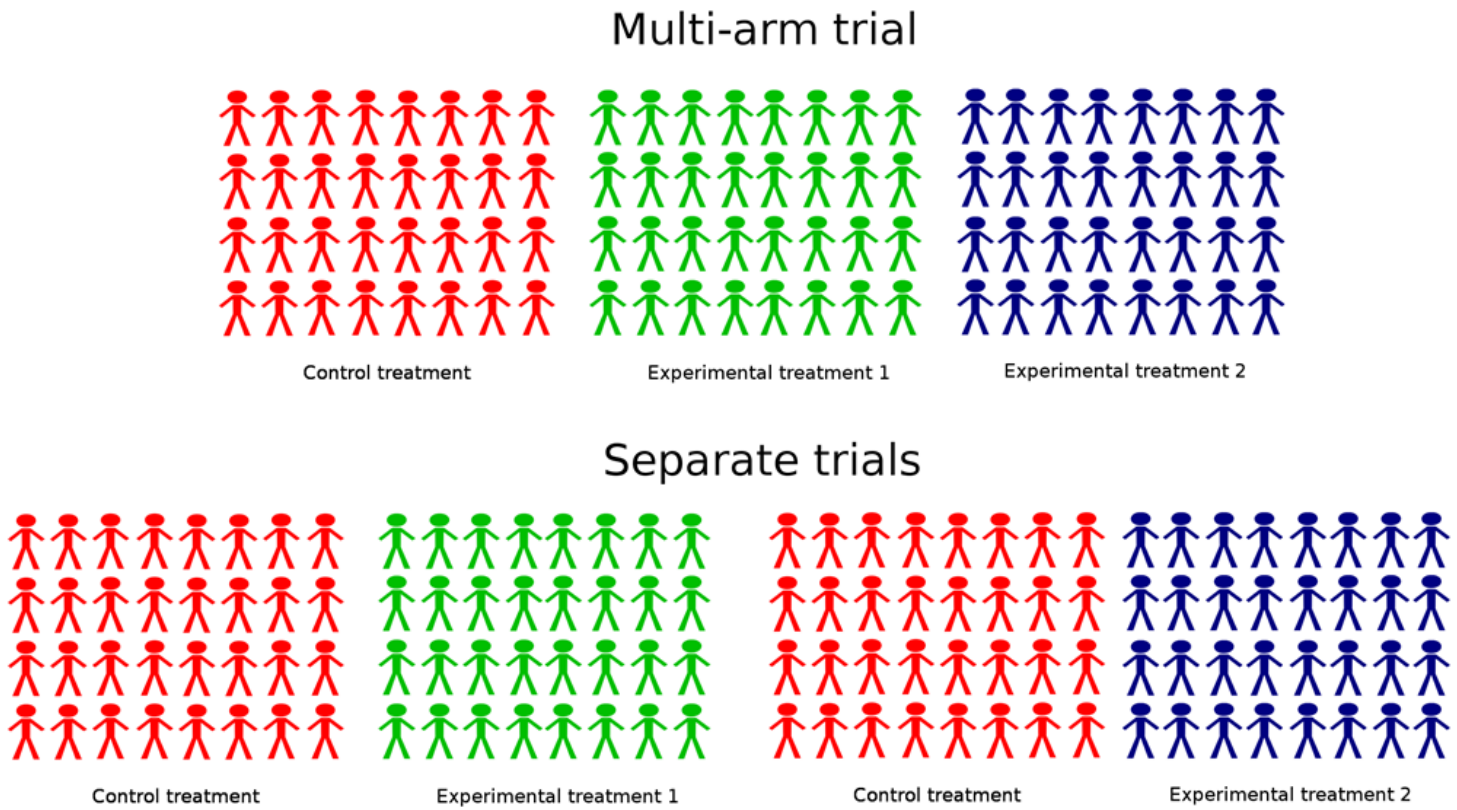
- Group-sequential designs allow us to use pre-specified *interim analyses* to test H_0 before the planned sample size is recruited.
- Have rules for *early stopping* for futility and efficacy.
- These designs are more efficient and more ethical than traditional designs:
  - More efficient as we have the chance to stop the trial early – fewer patients on average are needed
  - More ethical because the trial may stop early if the new treatment is considerably worse than the control treatment


Group sequential design
An example trial using group-sequential design
Randomisation
Interim 1 (PFS)
Interim 2 (OS)
Final (OS)
EUPATI
European Patients' Academy
on Therapeutic Innovation
www.eupati.eu
Treatment 1
Treatment 2
PFS = Progression-free survival
OS = Overall survival
Possible early stops

Multi-arm trials

- In some therapeutic areas, there may be several novel treatments awaiting evaluation in controlled clinical trials
- In a multi-arm trial, multiple treatments are compared in parallel
- Advantages:
  - Efficient and cheaper, since a shared control group is used
  - More treatments can be tested with a limited set of patients
  - More popular with patients as a greater chance of being allocated to a new treatment

Multi-arm trials



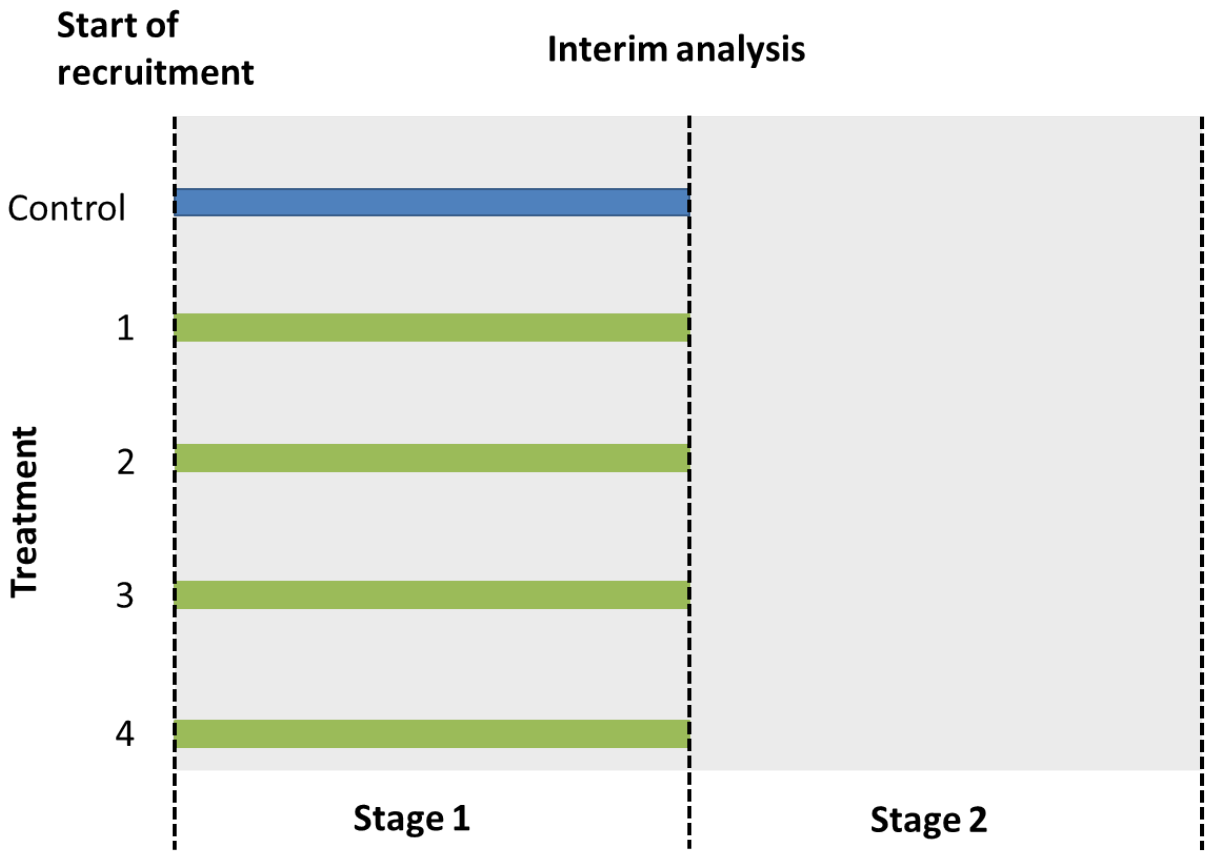
MAMS designs

Multi-arm multi-stage (MAMS) designs

- Can also add interim analyses to a multi-arm trials
- One approach is to set stopping boundaries, which the test statistics are compared to at each analysis
- This is a generalisation of group-sequential designs to multi-arm trials
- Increasingly popular in practice

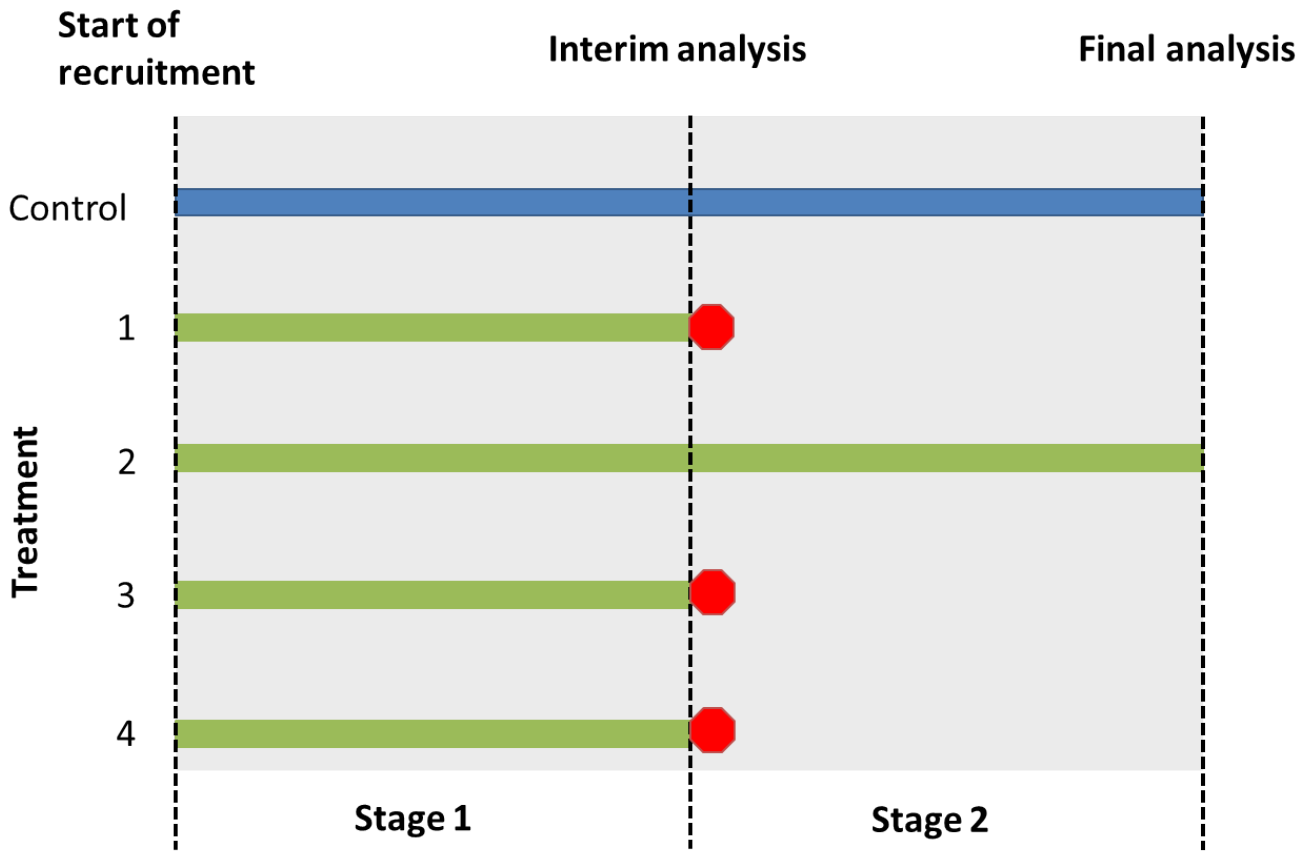
Seamless phase II/III trials

- Stage 1 is used to select the most promising treatment



Seamless phase II/III trials

- Stage 1 is used to select the most promising treatment
- Stage 2 is for confirmatory analysis



Response-adaptive randomisation

Beyond equal randomisation

- Ethical dilemma inherent in every clinical trial: earning vs. learning
- Consider a clinical trial comparing a new treatment against a control, using equal randomisation and with a power of 80%
- Patients in the trial: Probability of receiving a (potentially) better treatment is fixed, by design, and is equal to $1/2$. **Learning**
- Patients outside the trial: Probability of receiving a better treatment is (potentially) equal to $(1-\beta) = 80\%$. **Earning**
- All things being equal, the higher $(1-\beta)$ is, the larger the size of the trial n, and the larger the number of patients in the worse arm $n/2$ will be. There is a conflict between the individual and collective ethics
- **Response-adaptive randomisation** and **bandit models** (optimal decision-theoretic models) are two responses to this dilemma

Randomisation in clinical trials

Why are patients randomised in clinical trials?

There are some classical reasons in the medical literature:

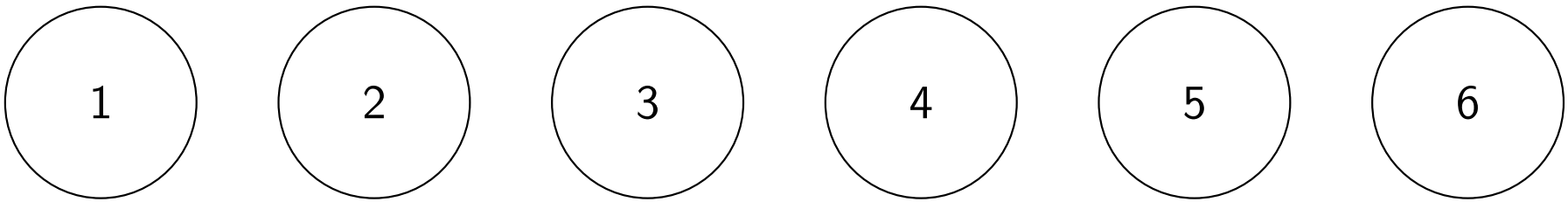
(1) **Mitigate accidental bias:** it promotes comparability among the study groups (specially with respect to unknown important covariates)

(2) **Mitigate intentional bias:** by adding an element of unpredictability to the treatment assignment process

(3) **Basis of inference:** It makes possible, at the end of the trial, to answer the question "In how many experiments could a difference of this magnitude have arisen by chance alone if the treatment truly has no effect?"

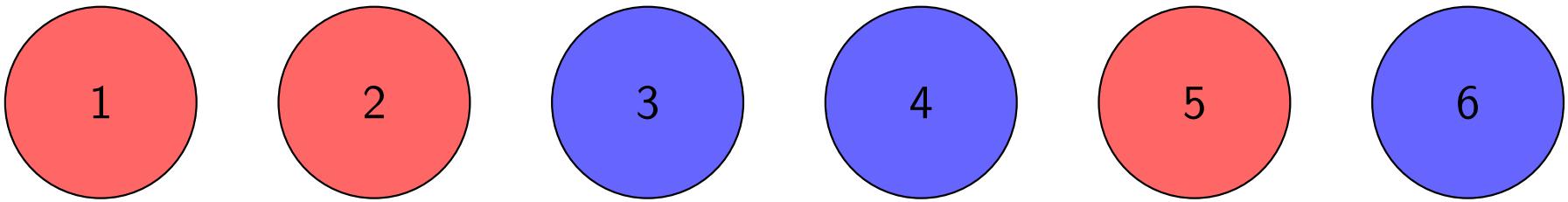
How do we implement randomisation in clinical trials

Randomisation sequences



How do we implement randomisation in clinical trials

Randomisation sequences

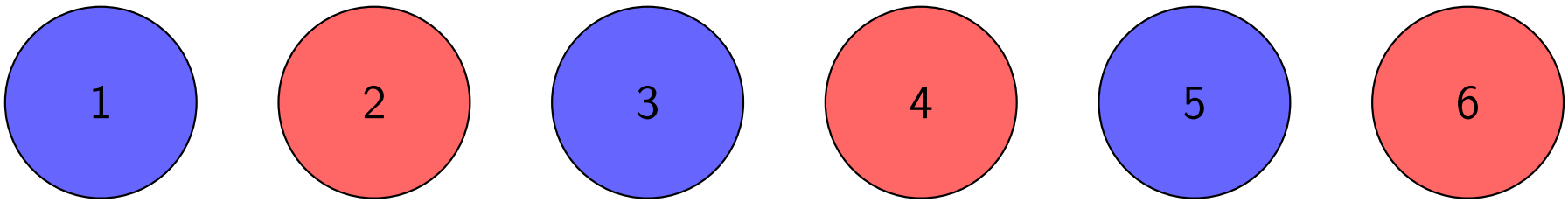


Realisation: 1

A	0	0	1	1	0	1
B	1	1	0	0	1	0

How do we implement randomisation in clinical trials

Randomisation sequences

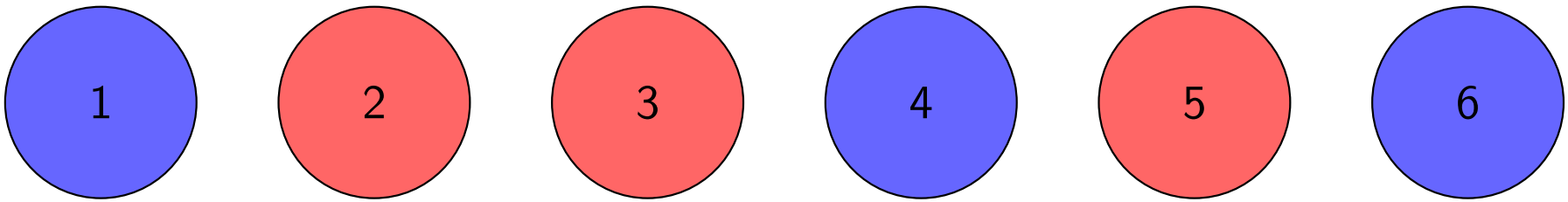


Realisation: 2

A	1	0	1	0	1	0
B	0	1	0	1	0	1

How do we implement randomisation in clinical trials

Randomisation sequences



Realisation: 3

A	1		0	0	1	0	1
B	0		1	1	0	1	0

We can represent the randomisation sequences through a **random** matrix T

$$\boldsymbol{T} = (\boldsymbol{T}_1, \boldsymbol{T}_2, \ldots, \boldsymbol{T}_n) \text{ where } \boldsymbol{T}_i \in \{\boldsymbol{e}_1, \boldsymbol{e}_2\} \text{ for } i = 1, \ldots, n$$

$\boldsymbol{e}_j$ is a column vector with a 1 in the j^{th}position if the j^{th} arm is allocated

Type of randomisation procedures

Complete, restricted and response-adaptive randomisation

A randomisation procedure is defined by the conditional probability $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{iK})$ of allocating treatments $1, \ldots, K$ to the i-th patient, which may be conditional on the history of allocations $(\boldsymbol{T}_{i-1}, \ldots, \boldsymbol{T}_1)$ and observed patient outcomes $(Y_{i-1}, \ldots, Y_1)$:

$$\boldsymbol{\phi}_i = \mathrm{E}(\boldsymbol{T}_i | \boldsymbol{T}_{i-1}, \ldots, \boldsymbol{T}_1, Y_{i-1}, \ldots, Y_1)$$

- **Complete randomisation** (CR):

$$\boldsymbol{\phi}_i = \mathrm{E}(\boldsymbol{T}_i) = \boldsymbol{\phi} \quad \text{e.g. if } K = 2,\ \boldsymbol{\phi} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

- **Restricted randomisation** (RR):

$$\boldsymbol{\phi}_i = \mathrm{E}(\boldsymbol{T}_i | \boldsymbol{T}_{i-1}, \ldots, \boldsymbol{T}_1)$$

- **Response-adaptive randomisation** (RAR):

$$\boldsymbol{\phi}_i = \mathrm{E}(\boldsymbol{T}_i | \boldsymbol{T}_{i-1}, \ldots, \boldsymbol{T}_1, Y_{i-1}, \ldots, Y_1)$$

Type of randomisation procedures

Response-adaptive randomisation

- For **RAR** procedures allocation probabilities are changed based on previous allocations *and* *outcomes* in order to meet a certain objective
- Usual objectives in clinical trials are: maximising **power**; maximising **patient benefit**; maximising **patient benefit subject to a power constraint**
- Two broad approaches to define RAR rules are:
  - **Non-parametric** (e.g. based on urn models)
  - **Parametric** (those that sequentially substitute updated estimates of those parameters)
- In what follows, we focus on RAR procedures where the patient outcomes are **binary**, and there are two treatments (A and B) with true probabilities of success p_A and p_B

Non-Parametric RAR procedures

Randomised Play the winner (RPTW)

- An urn that contains blue (treatment A) and red (treatment B) balls. We draw a ball, allocate a patient to that treatment and replace the ball. If the outcome was a success ($Y_i = 1$) we add 1 ball to the urn of the colour drawn, otherwise we add 1 of the other colour.

  Let $\mathbf{C}_i$ be the urn composition after treating patient i.

- Initial urn composition: $\mathbf{C}_0 = (\delta, \delta)$ (Equipoise)

- Design Matrix: $D(Y_i = 1) = \begin{array}{c} \\ b \\ r \end{array} \begin{array}{c} \begin{array}{cc} b & r \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$ $D(Y_i = 0) = \begin{array}{c} \\ b \\ r \end{array} \begin{array}{c} \begin{array}{cc} b & r \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$

  (Rows: colour drawn; Columns: balls to add to the colour)

- Recursive formula: $\mathbf{C}_i = \mathbf{C}_{i-1} + \mathbf{T}_i D(Y_i = y_i)$. Let $\|\mathbf{C}_i\| = \sum_{j=1}^{2} C_{ij}$

- Allocation prob.: $\phi_i = \mathrm{E}(\boldsymbol{T}_i | \boldsymbol{T}_{i-1}, \ldots, \boldsymbol{T}_1, Y_{i-1}, \ldots, Y_1) = \left(\frac{\mathbf{C}_{i-1}}{\|\mathbf{C}_{i-1}\|} \right)'$

Non-Parametric RAR procedures

Randomised Play the winner (RPTW)

- **Objective**: maximise exposure to best treatment (not explicitly). Not concerned with power to reject a null hypothesis of the form $H_0 : p_A - p_B = 0$
- Asymptotic behaviour – e.g., in terms of $\mathrm{E}(N_j(i))$: the expected sample size per arm – depends on $\mathrm{E}(D)$
- For RPTW it is intuitive that: $\mathrm{E}(D) = \begin{array}{c} \\ b \\ r \end{array} \begin{array}{c} \begin{array}{cc} b & r \end{array} \\ \begin{pmatrix} p_A & (1-p_A) \\ (1-p_B) & p_B \end{pmatrix} \end{array}$

Parametric RAR procedures

Optimal allocation ratios: minimise sample size

- Wald Test: $Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{s^2_{\Delta\hat{p}}(n_A, n_B)}}$, $\quad s^2_{\Delta\hat{p}}(n) = \frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}$.

  Maximising power = minimising $s^2_{\Delta\hat{p}}(n_A, n_B)$

Q: What is the minimum sample size $n = n_A + n_B$ given a power constraint (or a fixed variance level)?

Let $\rho = \frac{n_A}{n_A+n_B}$ (and $1 - \rho = \frac{n_B}{n_A+n_B}$),

$\min_\rho n_A + n_B \quad$ s.t. $s^2_{\Delta\hat{p}}(n_A, n_B) = C$

Solution (a.k.a., Neyman allocation):

$$\rho^*(p_A, p_B) = \frac{\sqrt{p_A(1-p_A)}}{\sqrt{p_A(1-p_A)} + \sqrt{p_B(1-p_B)}} \tag{1}$$

- Randomise patients by letting: $\phi_{i1} = \rho^*(\hat{p}_A, \hat{p}_B)$, where $\hat{p}_A, \hat{p}_B$ are the MLEs of p_A, p_B

Neyman allocation

Sketch of solution

(1) $$\min_\rho n_A + n_B \quad \text{s. t. } s^2_{\Delta\hat{p}}(n_A, n_B) = C$$

$$\begin{aligned} s^2_{\Delta\hat{p}}(n_A, n_B) &= C \\ \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B} &= C \end{aligned} \tag{2}$$

(2) Letting $n_A = \rho n$ and $n_B = (1-\rho)n$ and solving (2) for n:

$$n = \frac{p_A(1-p_A)}{\rho C} + \frac{p_B(1-p_B)}{(1-\rho)C} \tag{3}$$

- To minimise (3), set to 0 its first derivative with respect to ρ:

$$-\frac{p_A(1-p_A)}{\rho^2} + \frac{p_B(1-p_B)}{(1-\rho)^2} = 0 \tag{4}$$

(3) Now express ρ^* as a function of p_A and p_B to get (1).

Parametric RAR procedures

Optimal allocation ratios to minimise failures

- Wald Test: $Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{s^2_{\Delta\hat{p}}(n_A, n_B)}}$ where $s^2_{\Delta\hat{p}}(n) = \frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}$.

  Power constraint = constraint on $s^2_{\Delta\hat{p}}(n_A, n_B)$ level

Q: How would patients be allocated to minimise expected failures (or maximise expected successes) given a power constraint?

$\min_\rho = [(1 - p_A)n_A + (1 - p_B)n_B]$

s.t. $s^2_{\Delta\hat{p}}(n_A, n_B) = C$

Solution (Rosenberger et al., 2001):

$$\rho^*(p_A, p_B) = \frac{\sqrt{p_A}}{\sqrt{p_A} + \sqrt{p_B}} \tag{5}$$

- Randomise patients by letting: $\phi_{i1} = \rho^*(\hat{p}_A, \hat{p}_B)$, where $\hat{p}_A, \hat{p}_B$ are the MLEs of p_A, p_B

Response-Adaptive randomisation

Simulation Results

$n = 148$	$H_0 : p_A = p_B = 0.3$			
	α	n_B/n (s.e.)	$(1 - \rho^*)$	ENS (s.e.)
CR	0.049	0.500 (0.04)	**0.5**	44.33 (5.57)
RPTW	0.048	0.503 (0.28)	**0.5**	44.43 (5.48)
Neyman	0.058	0.501 (0.05)	**0.5**	44.29 (5.49)
Min. Failures	0.055	0.499 (0.05)	**0.5**	44.29 (5.66)
$n * p_B$				44.40 (0.00)
$n = 148$	$H_1 : p_A = 0.3\ p_B = 0.5$			
	$(1 - \beta)$	n_B/n (s.e.)	$(1 - \rho^*)$	ENS (s.e.)
CR	0.805	0.500 (0.04)	**0.500**	59.25 (5.94)
RPTW	0.659	0.592 (0.25)	**0.583**	62.10 (9.40)
Neyman	0.817	0.519 (0.04)	**0.522**	59.75 (5.77)
Min. Failures	0.809	0.557 (0.05)	**0.564**	60.83 (5.99)
$n * p_B$				74.00 (0.00)

(5000 trial replications)

Advantages:

- It is a procedure that can be designed to expose fewer patients (on average) to a (potentially) worse treatment, compared to fixed sample-size designs.
- No need to define stopping rules at interim analyses – but some RAR rules implicitly make a treatment choice before the end of the trial.
- For multi-armed trials, one can define RAR rules that achieve both higher power and higher patient benefit compared to fixed sample-size designs.

Disadvantages:

- For two-armed trials, trials may need to be larger to attain the power level of a fixed sample-size design.
- Requires frequent interim analyses - more statistical resources and may introduce hidden biases if not cautious in the analysis.

Flexible adaptive designs

Flexible designs

- Often the analyses and any adaptations for a clinical trial are planned in advance
  - e.g. number of analyses and the rules for adaptation are pre-specified
- All very well in theory, but in practice it may not be possible (or desirable) to have fully pre-specified rules
  - In practice, at an interim analysis the original assumptions on which the design was based may not hold (e.g. unplanned early stopping of a treatment arm due to safety concerns)
  - Decision rules can often be complex and involve new, possibly external information that cannot be known in advance (e.g. in a combined phase II/III trial, at the end of phase II, new information from external trials may be available)

Flexible designs

- Hence there is a need for flexible designs: guarantee type I error control after design adaptations that are not completely specified in advance
  - Note distinction between *flexible* and *adaptive* designs. Many adaptive designs incorporate adaptivity that is not flexible, i.e. pre-specified adaptivity
- In flexible designs, only need to pre-specify one stage of the study at a time
- Leaves the study open for adaptation of almost any type after each interim analysis, while still controlling the type I error
  - The design can be changed based on all (unblinded) interim data as well as any external information

Flexible designs

How to construct statistical tests for flexible adaptive designs that control the type I error?

- *Conditional Invariance Principle*
  1. Tests based on Combination Tests
  2. Tests based on Conditional Error Rate

- We'll concentrate on the two-stage framework

Conditional invariance principle

Adaptive designs follow a common *conditional invariance principle* in order to control the type I error rate (Brannath et al., 2007)

- Consider a generic adaptive trial with two stages, separated by an interim analysis
  - The design of stage 2 is chosen based on the interim data from stage 1, as well as any external information
- Suppose we are testing a specific null hypothesis H_0

Conditional invariance principle

Adaptive designs follow a common *conditional invariance principle* in order to control the type I error rate (Brannath et al., 2007)

- Consider a generic adaptive trial with two stages, separated by an interim analysis
  - The design of stage 2 is chosen based on the interim data from stage 1, as well as any external information
- Suppose we are testing a specific null hypothesis H_0
- Let T_2 denote the test statistic for H_0 from the stage 2 data
  - Due to the data dependent choice of the stage 2 design, T_2 will depend on the interim data
- Conditional invariance principle: construct a modified statistic $\tilde{T}_2$ from the stage 2 data, so that the conditional distribution under H_0 (given the interim data and the stage 2 design) equals a fixed pre-specified distribution and hence is *invariant* to the interim data and design adaptations

Conditional invariance principle

- An invariant conditional distribution is typically found by transforming T_2 into a p-value q which is uniformly distributed under H_0 (conditional on the interim data and stage 2 design)

Conditional invariance principle

- An invariant conditional distribution is typically found by transforming T_2 into a p-value q which is uniformly distributed under H_0 (conditional on the interim data and stage 2 design)
- We know the joint distribution of the interim data and q is known and invariant with respect to the (potentially unknown) adaptation rules
- Hence can specify a level α *rejection region* in terms of the interim data and q
- This gives a test with type I error rate α that is independent from the adaptation rules
- The combination test and conditional error approaches are two (equivalent) ways of defining an invariant rejection region

Combination tests

Suppose we are testing a one-sided null hypothesis H_0 using a general two-stage design with early stopping rules

Stage 1

- Fix design (sample size, test statistic etc.)
- Also fix a *combination function* $C(\cdot, \cdot)$ that is used when the trial proceeds to stage 2, which combines the p-values from stage 1 and stage 2
- Calculate p-value p from stage 1 data
- Early stopping for futility and efficacy:
  - If $p \leq \alpha_1$, we reject H_0
  - If $p > \alpha_0$, we accept H_0
  - If $\alpha_1 < p \leq \alpha_0$, we proceed to stage 2

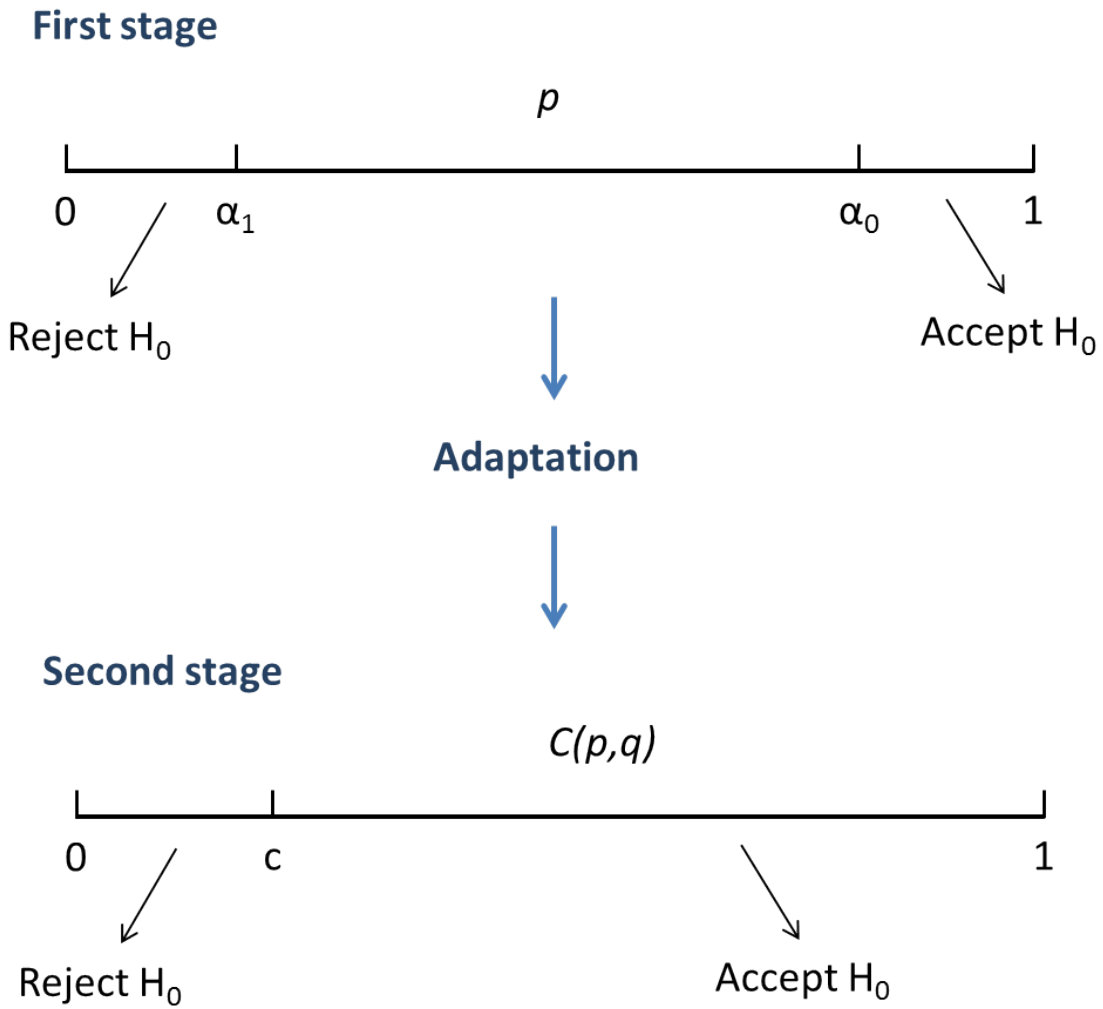
Combination tests

- All the data from stage 1 (as well as external information) can be used to design stage 2, e.g.
  - Reassessing the sample size
  - Redefining the test procedure

Stage 2

- Calculate p-value q from stage 2 data
- Reject H_0 if and only if $C(p, q) \leq c$
  - c is determined by $\alpha, \alpha_1, \alpha_0$ and $C(\cdot, \cdot)$ to control the level α

Combination tests



Combination tests

Examples of combination functions:

- Fisher's product test: $C(p, q) = pq$
- Weighted inverse normal test:

$$C(p, q) = 1 - \Phi\left[\sqrt{w}\Phi^{-1}(1-p) + \sqrt{1-w}\Phi^{-1}(1-q)\right]$$

(where Φ is the cdf of the standard normal distribution)

Computation of Stopping Boundaries

The critical value c is chosen so that the type I error rate of the combination test is at most α:

$$P_{H_0}(p \leq \alpha_1) + P_{H_0}\{\alpha_1 < p \leq \alpha_0, C(p, q) \leq c\} \leq \alpha \tag{6}$$

Computation of Stopping Boundaries

The critical value c is chosen so that the type I error rate of the combination test is at most α:

$$P_{H_0}(p \leq \alpha_1) + P_{H_0}\{\alpha_1 < p \leq \alpha_0, C(p,q) \leq c\} \leq \alpha \tag{6}$$

Assume that the distribution of the p-values p and q under H_0 satisfies the 'p-clud' condition: for all $\alpha \in [0,1]$

$$P_{H_0}(p \leq \alpha) \leq \alpha \quad \text{and} \quad P_{H_0}(q \leq \alpha | p) \leq \alpha$$

This applies if e.g. independent patients are recruited at each stage, and conservative tests are used

Given the p-clud condition, equation (6) becomes:

$$\alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^1 1_{[C(p,q) \leq c]} \, dp \, dq = \alpha$$

Type I error control

- Do not pool the data of the stages – combine the stage-wise p-values
- The distribution of the combination function under the null does not depend on design modifications
- *Hence the adaptive test is still a valid test for the modified design*

- Recursive application of the combination test allows for a flexible number of looks (Brannath et al. 2002)

Design considerations

Choice of α_1

- What is the probability of early stopping for efficacy?
- Is it likely that the results will be 'convincing enough' after early stopping for efficacy?
- Is it ethical to continue the trial if a large treatment effect is observed?

Choice of α_0

- Decreasing α_0 means a larger critical value c
- However, deviating from a binding futility stopping rule may inflate the type I error rate
- Suggestion in the literature is a *non-binding* futility rule
- Set $\alpha_0 = 1$, then stopping early for futility (for any reason) will lead to a strictly conservative procedure

Summary: Type I error control in adaptive designs

Sources of Type I error inflation

- Early rejection
- (Flexible) adaptive trial design
- Multiple hypothesis testing

Approaches for error rate control

- Group sequential trials
- Combination of p-values / Conditional error functions
- Multiple testing methodology

Maurer et al. (2009)

References


W. Brannath et al. (2002).
Recursive Combination Tests
Journal of the American Statistical Association, 97:236–244.


W. Brannath et al. (2007).
Multiplicity and flexibility in clinical trials
Pharmaceutical Statistics, 6:205–216.


W. Maurer et al. (2009).
Adaptive designs and confirmatory hypothesis testing
Multiple testing problems in pharmaceutical statistics, Chapman & Hall Biostatistics Series, Baton Raton.


Pallman et al. (2018).
Adaptive designs in clinical trials: why use them, and how to run and report them
BMC Medicine, 16:29


Rosenberger, W.F., Stallard, N, Ivanova, A, Harper, C. And Ricks, M. (2001).
Optimal adaptive designs for binary response trials.
Biometrics, 57:909–913.


Hu, F., and Rosenberger, W. F. (2006).
The theory of response-adaptive randomization in clinical trials. John Wiley & Sons.