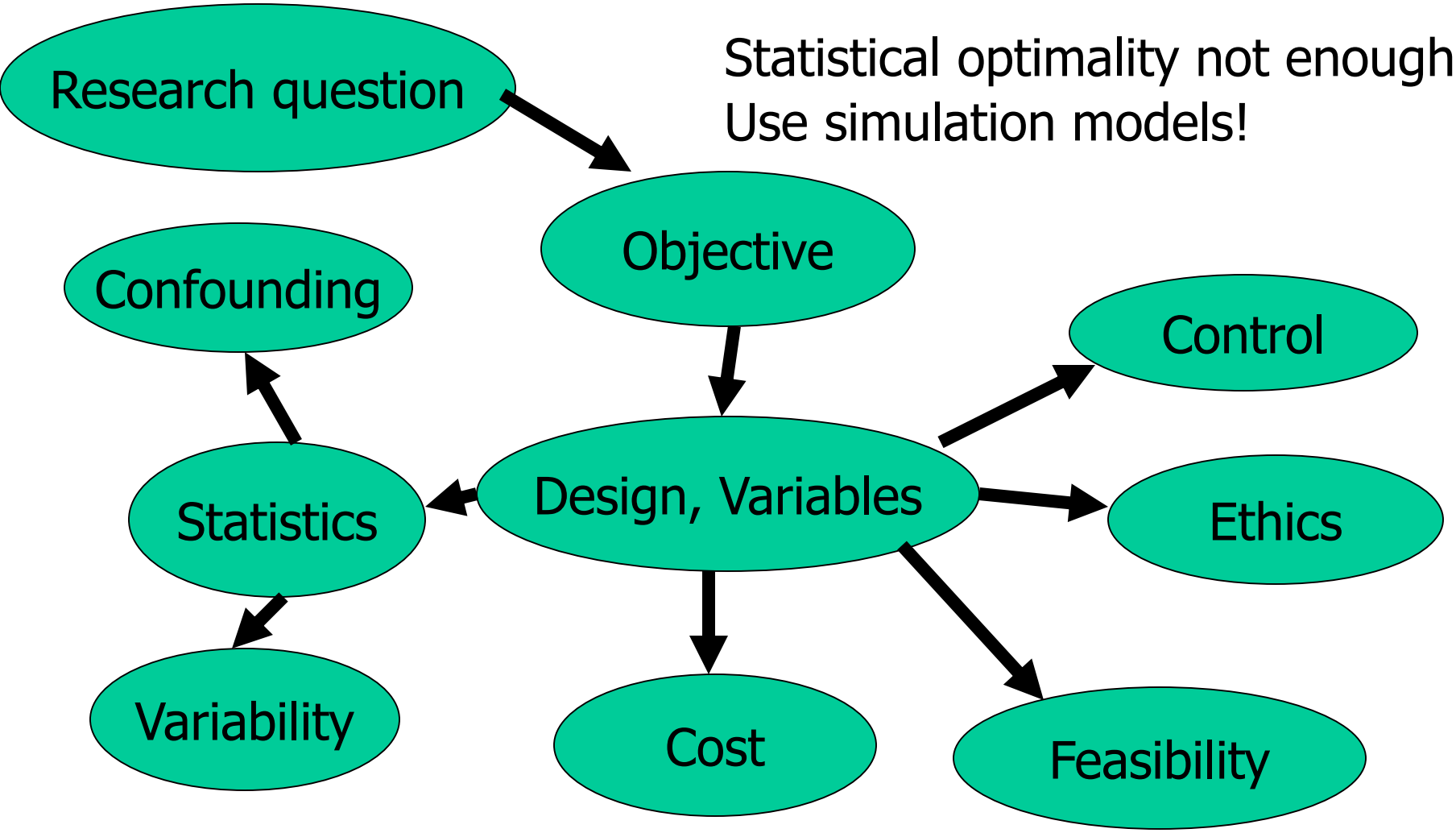
Designs for Clinical Trials

Lecture 4

31-01-2020

Design issues

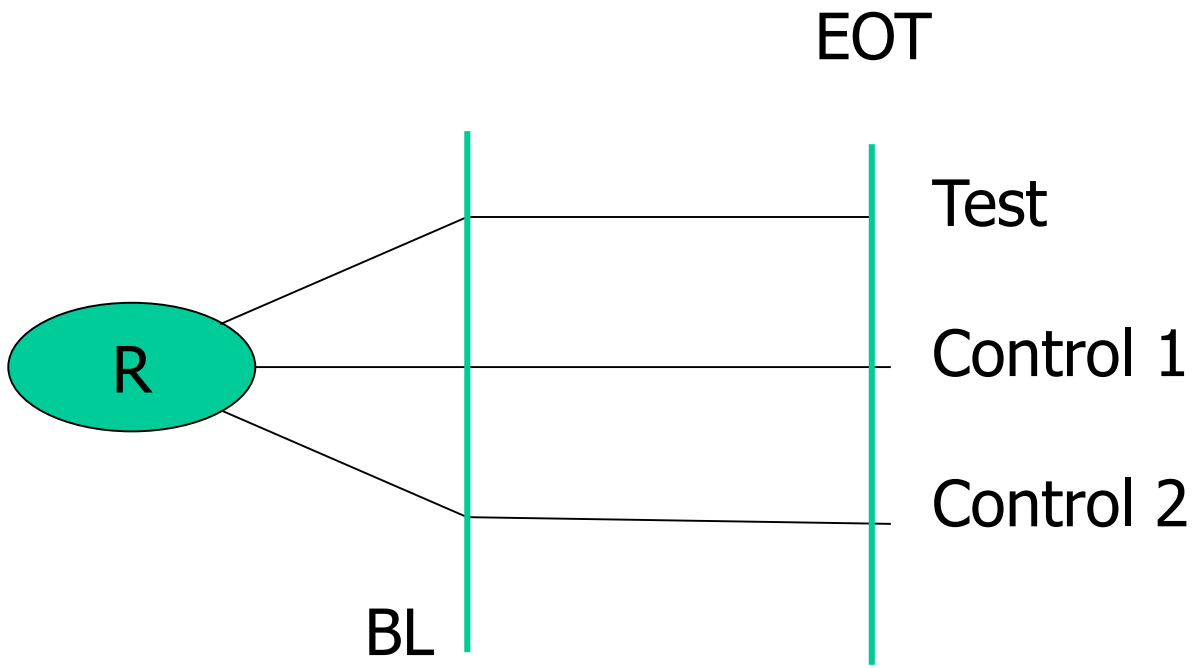


Selection of a design

The design of a clinical study is influenced by:

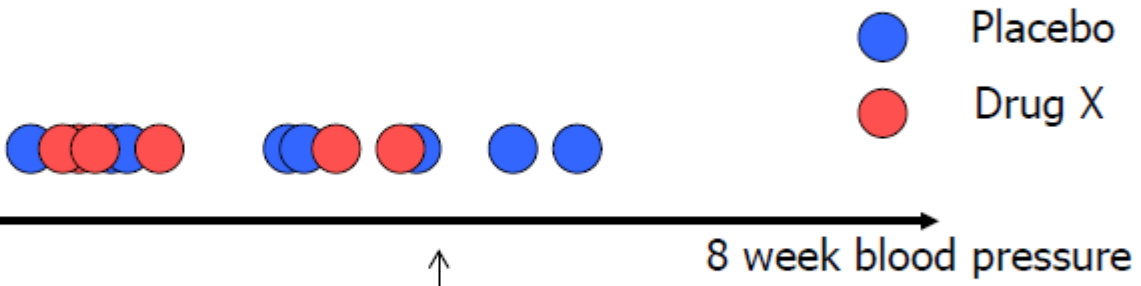
- Research question - Study objectives
- Number of treatments to be compared
- Characteristics of the treatment
- Characteristics of the disease/condition
- Inter and intra subject variability
- Possible confounding variables
- Duration of the study
- Accrual rate
- Drop out rates
- Cost
- Feasibility
- Ethics
- ...

Parallel Group Designs with Baseline measurements

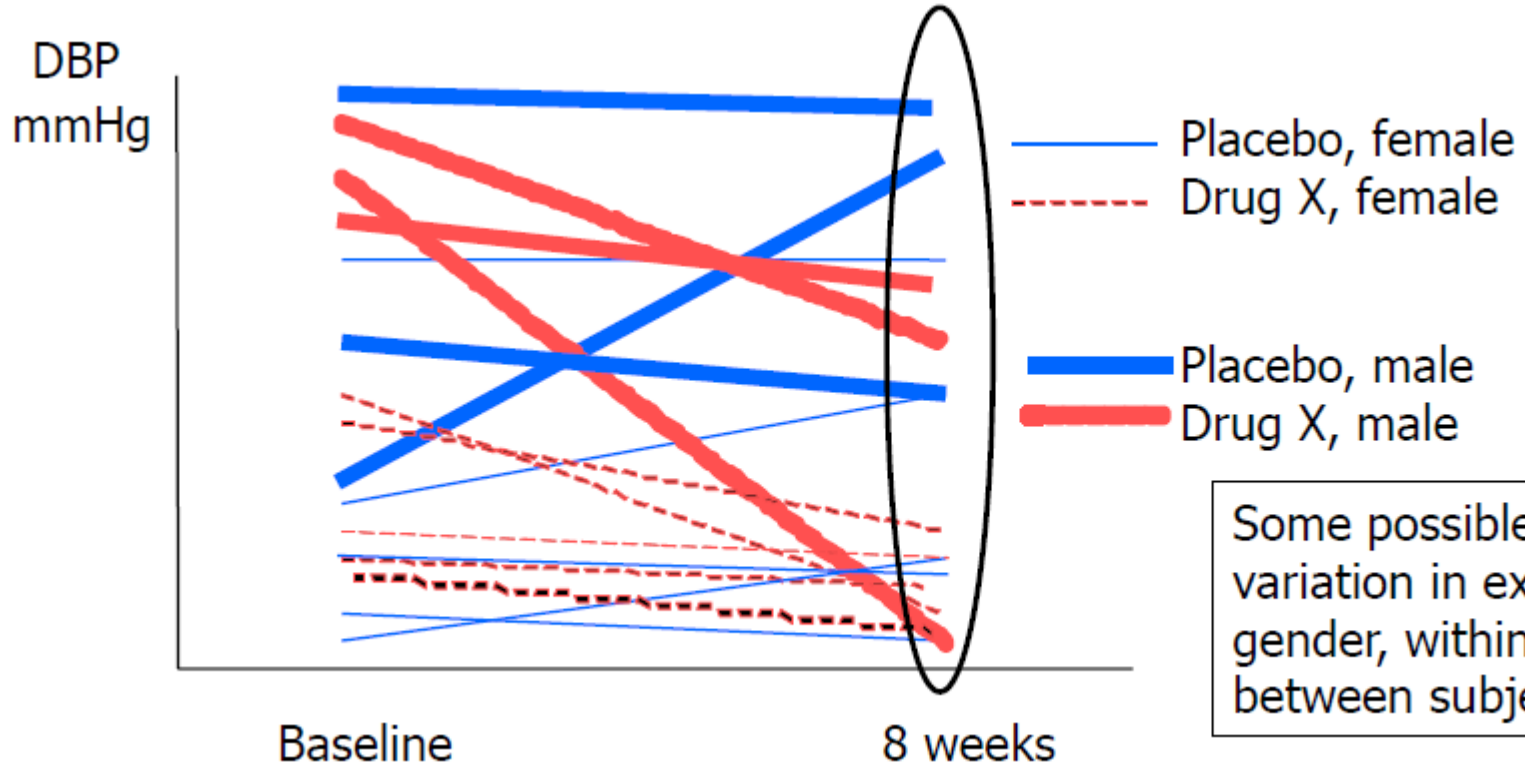


Example: Where does the variation come from?

Only 8 week:



Both baseline and 8 week:



Some possible sources of variation in example: gender, within- and between subject variation

Variance formulas for equations on prev slide

- Only end of study: $Var[Y_{end}] = \sigma^2$

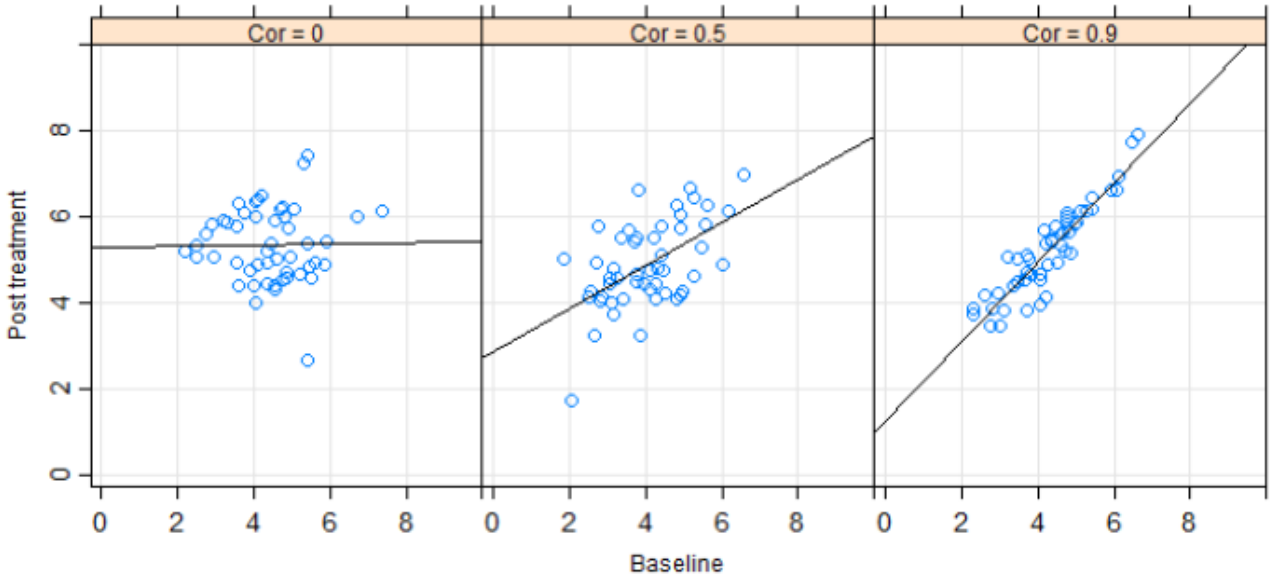
- Change from baseline

$$Var[Y_{end} - Y_{baseline}] = Var[Y_{end}] + Var[Y_{baseline}] - 2Cov[Y_{end}, Y_{baseline}]$$
$$= 2\sigma^2 - 2\sigma^2 Corr[Y_{end}, Y_{baseline}] = 2\sigma^2\left(1 - Corr[Y_{end}, Y_{baseline}]\right)$$
$$= 2\sigma^2(1-\rho)$$

- Baseline as covariate: Some other time ☺

Parallell group design with baseline measurements

Change from baseline: when is it an improvement over a single measurement?



cor < 0.5 : change increases variability

cor > 0.5 : change decreases variability

Cor = 0 → Worse	Cor = 0.5 → Same	Cor = 0.9 → Better
Sd_{change} = sqrt(2) * sd	Sd_{change} = sd	Sd_{change} = sqrt(0.2)* sd

Assume sd = $sd_{baseline}$ = sd_{end}

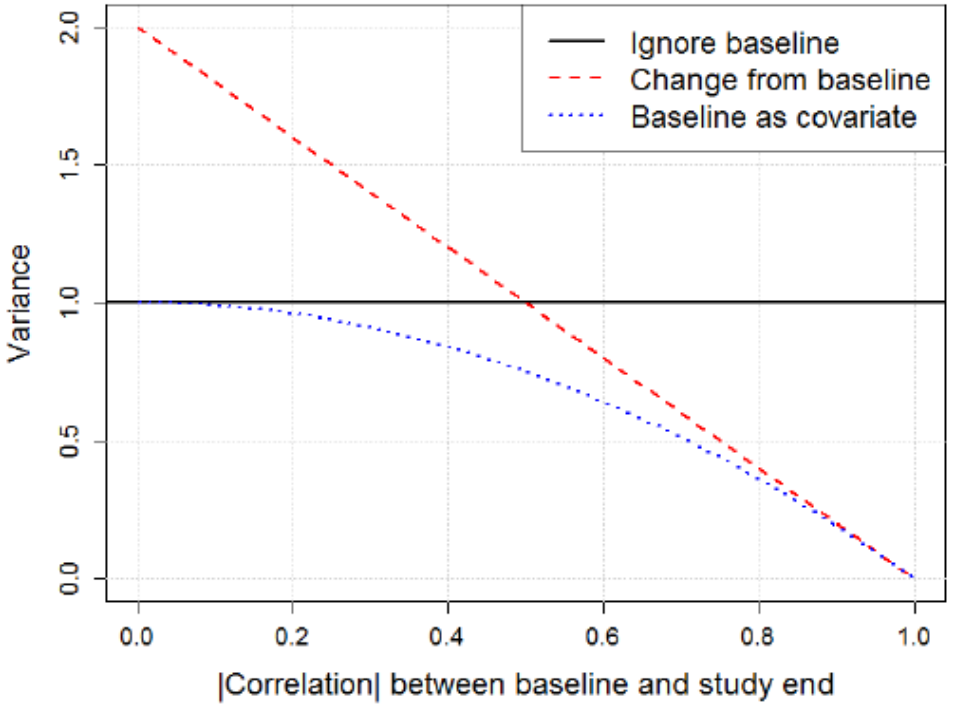
Sd_{change} = sd * sqrt(2*(1-corr))

Parallell group design with baseline measurements, contd

If corr < 0.5, should we just forget baseline?

NO !

- Use baseline as a covariate
  (eg ANCOVA rather than ANOVA)
- As soon as corr > 0, then baseline as a covariate decreases variability
- Baseline as covariate always better than "change from baseline"
- Change as endpoint + baseline as cov
  ↔
  Post-treat as endpoint + baseline as cov
  (same p-value, different estimates)

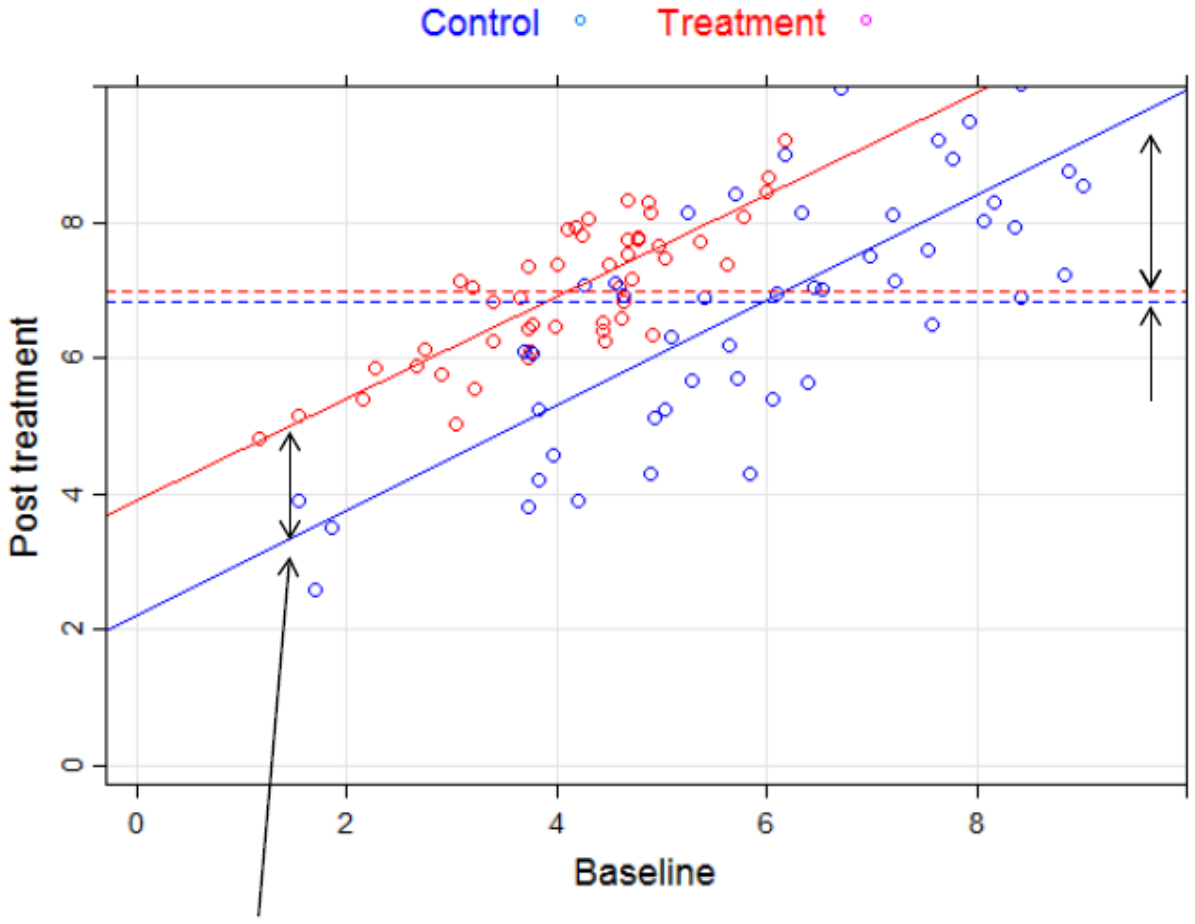


Variance of endpoint:
- Ignore baseline: σ^2
- Change from baseline: $2\ \sigma^2\ (1-\rho)$
- Baseline as covariate: $\sigma^2\ (1-\rho^2)$

Details next slide

Parallell group design with baseline measurements, contd

What do you compare when adjusting with baseline?



Treatment effect without covariate (difference in means)

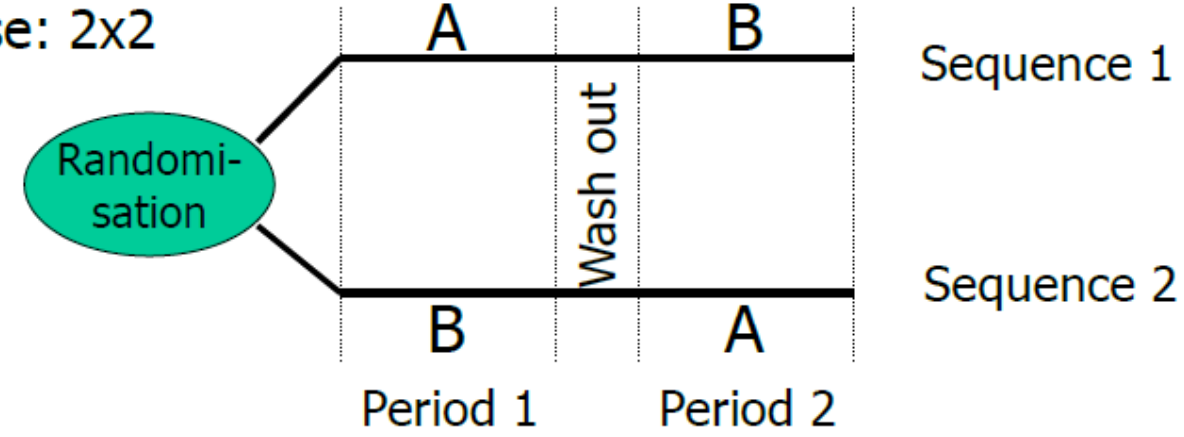
Same idea when using eg age, BMI as covariates

Treatment effect with covariate (adjusted for baseline)
(difference in ”lsmeans” = model adjusted means)

Crossover designs

- A repeated measurements design such that each patient receives different treatments during the different time periods, i.e., the patients cross over from one treatment to another

- Most simple case: 2x2
  - 2-sequence
  - 2-period
  - 2-treatment



- Pros:
  - Within subject comparison → Reduced sample size
  - Good for chronic conditions
- Cons:
  - Longer studies
  - Potential carry-over effects
  - Carryover effects may be confounded with treatment effects
  - Not suitable for acute conditions

Carryover effect: the effect of the treatment from the previous time period on the response at the current time period

Model for a cross over study

Obs=Period+sequence+subject+treament+carryover+error

$$Y_{ijk} = \mu + \gamma_i + \xi_{i(k)} + \pi_j + \tau_t + \quad \lambda_r + \varepsilon_{ijk}$$

γ_i = effect of sequence $i = 1,2$ $\gamma_1 + \gamma_2 = 0$

$\xi_{i(k)}$ = effect of subject $k = 1, \ldots, n_i$ within sequence i iid $N\left(0, \sigma_s^2\right)$

π_j = effect of period $j = 1,2$ $\pi_1 + \pi_2 = 0$

τ_t = effect of treatment $t \in \{A, B\}$ $\tau_A + \tau_B = 0$

ε_{ijk} = random error iid $N\left(0, \sigma^2\right)$

2 by 2 Crossover design

$$\mathrm{E}\left[Y_{ijk}\right]=\begin{array}{|c|c|}\hline \mu_{11} & \mu_{12} \\ \hline \mu_{21} & \mu_{22} \\ \hline \end{array}$$

$\gamma_1+\pi_1+\tau_A$	$\gamma_1+\pi_2+\tau_B+\lambda_A$
$\gamma_2+\pi_1+\tau_B$	$\gamma_2+\pi_2+\tau_A+\lambda_B$

$$\frac{1}{2}\left(\left(\mu_{11}-\mu_{12}\right)+\left(\mu_{22}-\mu_{21}\right)\right)=$$

Obs! μ everywhere

$$\frac{1}{2}\left(\left(\pi_1-\pi_2+\tau_A-\tau_B-\lambda_A\right)+\left(\pi_2-\pi_1+\tau_A-\tau_B-\lambda_B\right)\right)=$$

$$\tau_A-\tau_B-\frac{1}{2}\left(\lambda_A+\lambda_B\right)$$

Effect of treatment and carry over can not be separated!

Matrix formulation

Model: $$Y_{ijk} = \mu + \gamma_i + \xi_{i(k)} + \pi_j + \tau_t + \varepsilon_{ijk}$$

Sum to zero:
$$\gamma_1 + \gamma_2 = 0$$
$$\pi_1 + \pi_2 = 0$$
$$\tau_1 + \tau_2 = 0$$

(Restrictions made so these parameters become uniquely defined and not only up to some additive constants)

Matrix formulation $\mu = \mathbf{X}\beta$

$$\beta = \begin{pmatrix} \mu & \gamma_1 & \pi_1 & \tau_A \end{pmatrix}^T$$
$$\mu = \begin{pmatrix} \mu_{11}, \mu_{12}, \mu_{21}, \mu_{22} \end{pmatrix}^T$$

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Matrix formulation

Parameter estimate: $\hat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T Y$

$$\hat{\sigma}^2 = \left(\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{Y}\right) - \hat{\beta}\mathbf{X}^{\boldsymbol{T}}\boldsymbol{Y}$$

$$Cov\left(\hat{\beta}\right) = \left(\mathbf{X}^T\mathbf{X}\right)\hat{\sigma}^2$$

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0.25 \end{bmatrix}$$

Estimates independent and

$$Var\left(\hat{\tau}_A\right) = 0.25\hat{\sigma}^2$$

Alternatives to 2*2

Compare

A	B
B	A

to

A	B	B
B	A	A

Same model but with 3 periods and a carry over effect

$$Y_{ijk} = \mu + \gamma_i + \xi_{i(k)} + \pi_j + \tau_t + \quad \lambda_r + \varepsilon_{ijk}$$

$\xi_{i(k)} = \text{effect of subject } k = 1,\ldots,n_i \text{ within sequence } i \text{ iid } N\left(0,\sigma_s^2\right)$

$\gamma_i = \text{effect of sequence } i = 1,2 \qquad \gamma_1 + \gamma_2 = 0$

$\pi_j = \text{effect of period } j = 1,2,3 \qquad \pi_1 + \pi_2 = 0$

$\tau_t = \text{effect of treatment } t \in \{A, B\}$

$\varepsilon_{ijk} = \text{random error iid } N\left(0,\sigma^2\right) \qquad \tau_A + \tau_B = 0$

Parameters of the AAB, BBA design

$$Y_{ijk} = \mu + \gamma_i + \xi_{i(k)} + \pi_j + \tau_t + \quad \lambda_r + \varepsilon_{ijk} \quad E\left[Y_{ijk}\right] = \mu_{ijk}$$

μ_{11}	μ_{12}	μ_{13}
μ_{21}	μ_{22}	μ_{23}

Effect of treatment and carry over can be estimated independently!

$$\mu_{11} = \mu + \gamma_1 + \pi_1 + \tau_A$$
$$\mu_{12} = \mu + \gamma_1 + \pi_2 + \tau_B + \lambda_A$$
$$\mu_{13} = \mu + \gamma_1 + \pi_3 + \tau_B + \lambda_B$$
$$\mu_{21} = \mu + \gamma_2 + \pi_1 + \tau_B$$
$$\mu_{22} = \mu + \gamma_2 + \pi_2 + \tau_A + \lambda_B$$
$$\mu_{23} = \mu + \gamma_2 + \pi_3 + \tau_A + \lambda_B$$

$$\lambda_1 + \lambda_2 = 0 \quad \pi_1 + \pi_2 + \pi_3 = 0 \quad \tau_1 + \tau_2 = 0 \quad \lambda_A + \lambda_B = 0$$

Matrix again

$$\mathbf{X}\beta = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \gamma_1 \\ \pi_1 \\ \pi_2 \\ \tau_A \\ \lambda_A \end{bmatrix}$$

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{bmatrix} 0.17 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.19 & 0 & 0 & 0.06 & 0 \\ 0 & 0 & 0.33 & -0.17 & 0 & 0 \\ 0 & 0 & -0.17 & 0.33 & 0 & 0 \\ 0 & 0.06 & 0 & 0 & 0.19 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.25 \end{bmatrix}$$

Other 2 sequence 3 period designs

1

A	B	B
B	A	A

2

A	B	A
B	A	B

3

A	A	B
B	B	A

	1	2	3
$Var(\hat{\tau}_A)$	0.19	0.75	0.25
$Var(\hat{\lambda}_A)$	0.25	1.00	1.00
$Corr(\hat{\tau}_A, \hat{\lambda}_A)$	0.0	0.87	0.50

Comparing the AB, BA and the ABB, BBA designs

A	B
B	A

A	B	B
B	A	A

$$Var(\hat{\tau}_A) = 0.25\hat{\sigma}^2$$

Can't include carry over

2 treatments per subject

Shorter duration

$$Var(\hat{\tau}_A) = 0.19\hat{\sigma}^2$$

Carry over estimable

3 treatments per subject

Longer duration

More than 2 treatments

Tool of the trade: Define the model

Investigate $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$

A	B
B	C
C	A
A	C
B	C
C	B

A	B	C
B	C	A
C	A	B
A	C	B
B	A	C
C	B	A

A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

Watch out for drop outs!

Titration Designs

Increasing dose panels (Phase I):

- SAD (Single Ascending Dose)
- MAD (Multiple Ascending Dose)

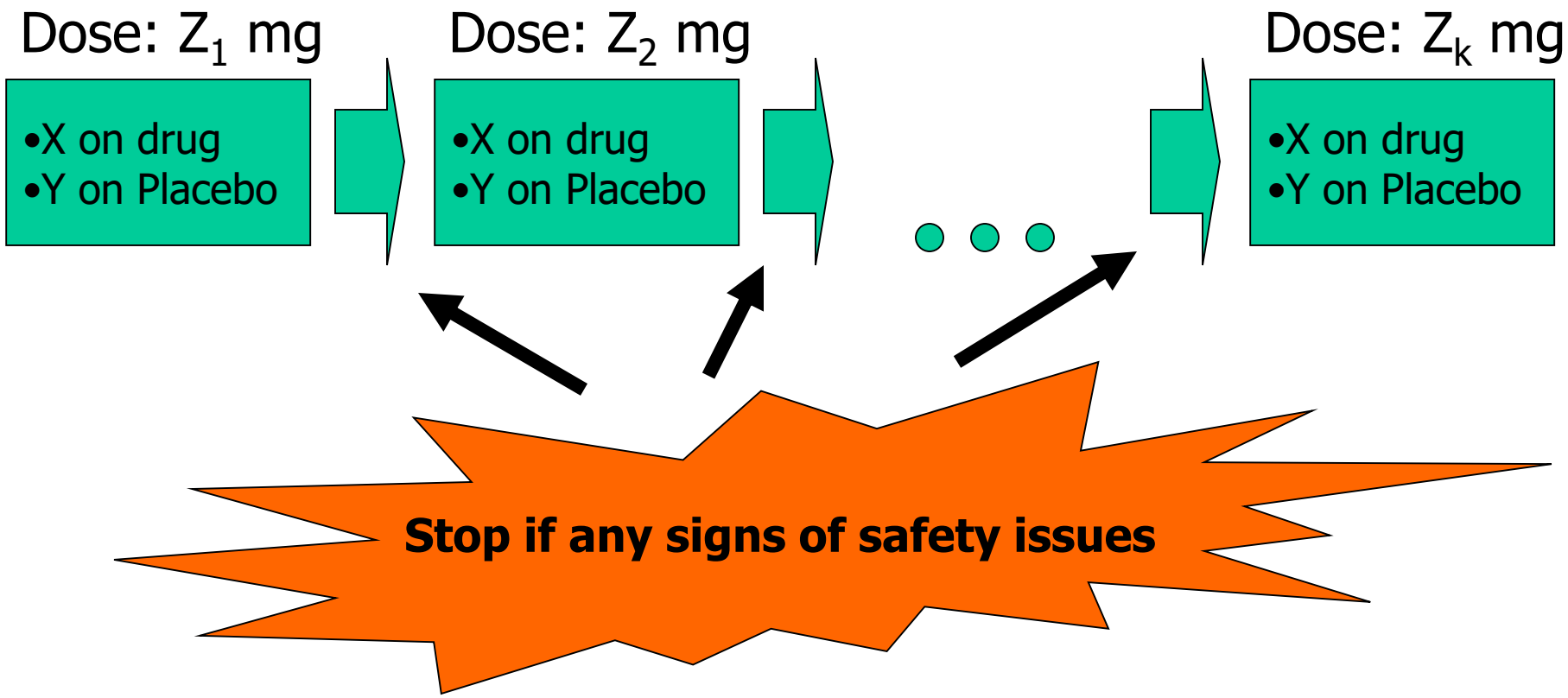
Primary Objective:

- Establish Safety and Tolerability
- Estimate Pharmaco Kinetic (PK) profile

Increasing dose panelse (Phase II):

Dose - response

Titration Designs (SAD, MAD)



VERY careful with first group!

Titration Designs

Which dose levels?

- Start dose based on exposure in animal models.
- Stop dose based on toxdata from animal models.
- Doses often equidistant on log scale.

Which subject?

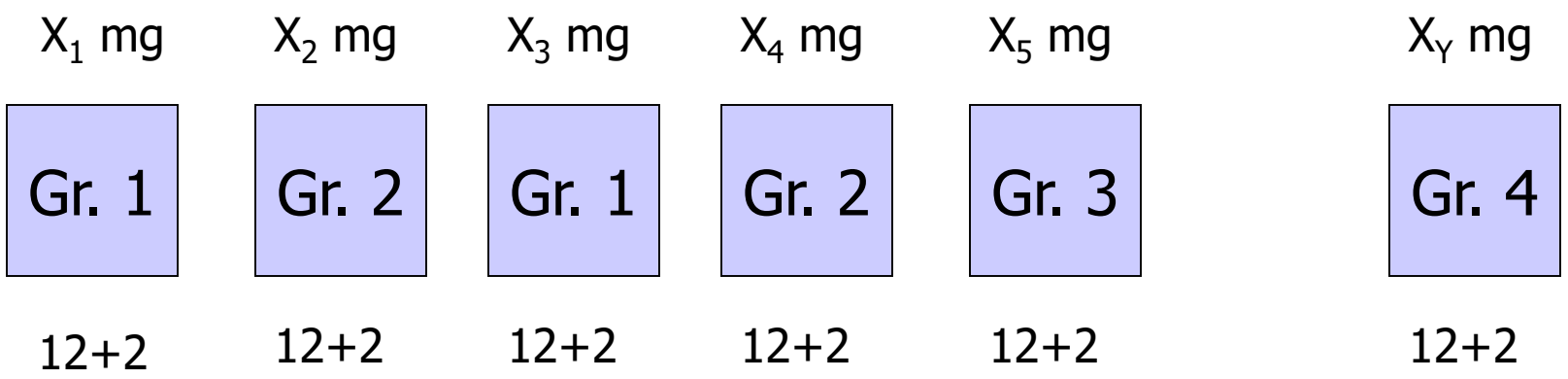
- Healty volunteers
- Young
- Male

How many subjects?

- Rarely any formal power calculation.
- Often 2 on placebo and 6-8 on drug.

Titration Designs

Not mandatory to have new subject for each group.



- Slighty larger groups to have sufficiently many exposed.
- Dose in fourth group depends on results so far.
- Possible to estimate within subject variation.

Adaptive Designs

Dragalin V. Adaptive Designs: Terminology and Classification. *Drug Information Journal*. 2006; 40(4): 425-436.

Definition

Adaptive Design

- uses accumulating data to decide on how to modify aspects of the study
- without undermining the *validity* and *integrity* of the trial

Validity means

- providing correct statistical inference (such as adjusted p-values, unbiased estimates and adjusted confidence intervals, etc)
- assuring consistency between different stages of the study
- minimizing operational bias

Integrity means

- providing convincing results to a broader scientific community
- preplanning, as much as possible, based on intended adaptations
- maintaining confidentiality of data

- An adaptive design should be adaptive by "design" not an *ad hoc* change of the trial conduct and analysis
- Adaptation is a prospective design feature, not a remedy for poor planning

General Structure

- An adaptive design requires the trial to be conducted in several stages with access to the accumulated data
- An adaptive design may have one or more rules:
  - **Allocation Rule**: how subjects will be allocated to available arms
  - **Sampling Rule**: how many subjects will be sampled at next stage
  - **Stopping Rule**: when to stop the trial (for efficacy, harm, futility)
  - **Decision Rule**: the terminal decision rule and interim decisions pertaining to design change not covered by the previous three rules
- At any stage, the data may be analyzed and next stages redesigned taking into account all available data

Examples

- Group Sequential Designs: only **Stopping Rule**
- Response Adaptive Allocation: only **Allocation Rule**
- Sample Size Re-assessment: only **Sampling Rule**
- Flexible Designs:
  - Changing the randomization ratio
  - Changing the timing of the next interim analysis
  - Changing the stopping Rule
  - Changing the target treatment difference; changing the primary endpoint; varying the form of the primary analysis; modifying the patient population; etc

Anything goes?

Achieving the goals

- There are plenty of available designs on statistician's shelf
- The greatest challenge is their implementation

- Adaptive designs have much more to offer than the rigid conventional parallel group designs in clinical trials

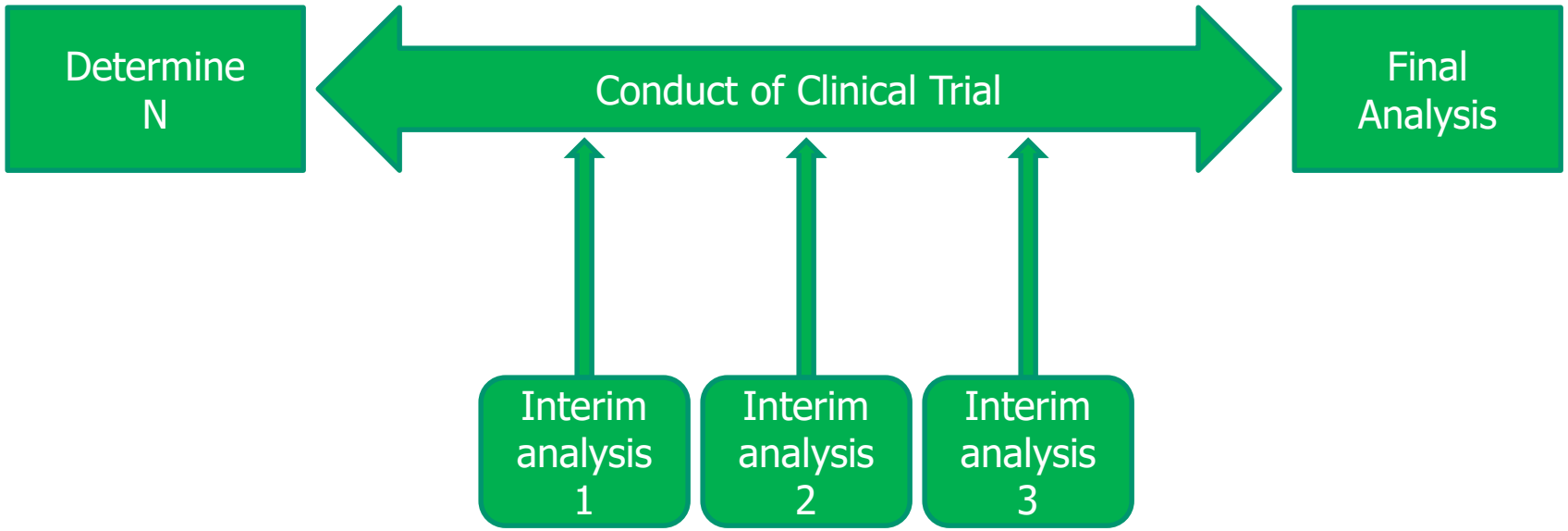
Group sequential designs

A large study is a huge investment, $, ethics

- What if the drug doesn't work or is much better than expected?
- Could we take an early look at data and stop the study is it look good (or too bad)?

Group Sequential Methods with Applications to Clinical Trials by Christopher Jennison and Bruce W. Turnbull. http://people.bath.ac.uk/mascj/

Group Sequential Design



Scope of Early termination of trial

- *Overwhelming efficacy*
- *futility of the drug*

Group Sequential Designs

- Group sequential designs are used to facilitate the conduct of interim analysis (see section 4.5 and Glossary).

- While group sequential designs are not the only acceptable types of designs permitting interim analysis, they are the most commonly applied because it is more practicable to assess grouped subject outcomes at periodic intervals during the trial than on a continuous basis as data from each subject become available.

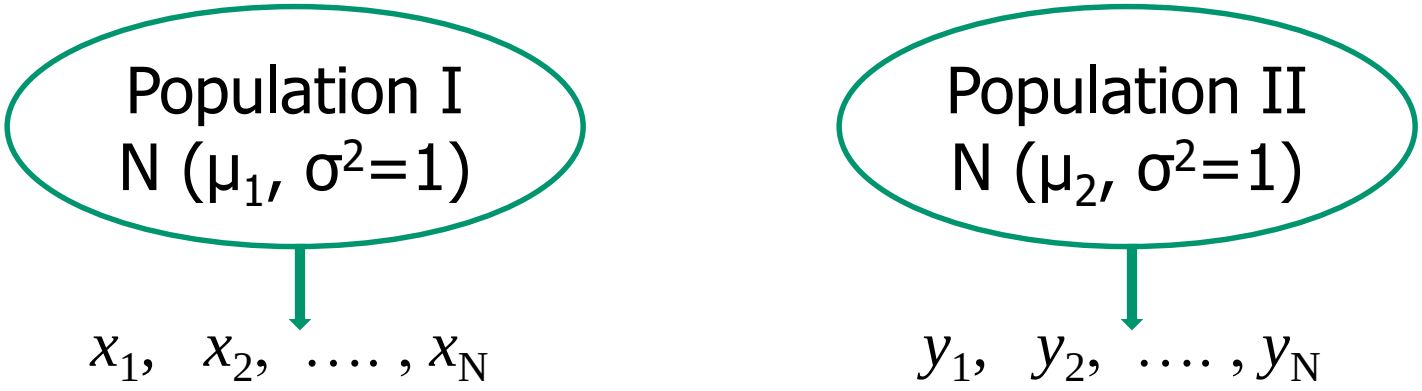
- The statistical methods should be fully specified in advance of the availability of information on treatment outcomes and subject treatment assignments (i.e. blind breaking, see Section 4.5).

- An Independent Data Monitoring Committee (see Glossary) may be used to review or to conduct the interim analysis of data arising from a group sequential design (see Section 4.6).

- While the design has been most widely and successfully used in large, long-term trials of mortality or major nonfatal endpoints, its use is growing in other circumstances.

- In particular, it is recognized that safety must be monitored in all trials and therefore the need for formal procedures to cover early stopping for safety reasons should always be considered.

Theoretical set-up



Effect size, $\Delta = \mu_1 - \mu_2$

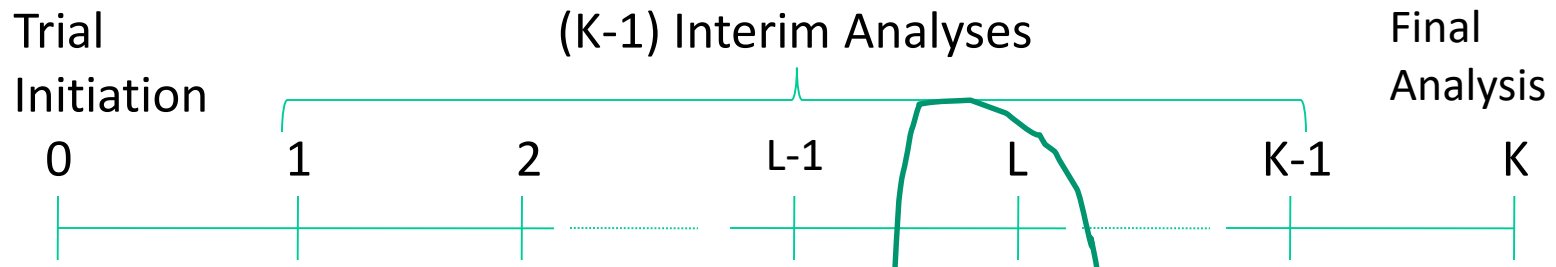
Our interest is to test (using two sample Z-test)

$H_o : \Delta = 0$ vs $H_a : \Delta > 0$

Assuming $\Delta = \delta$, total sample size (N) per population

$$N = 2\left(\frac{z_\alpha + z_\beta}{\delta}\right)^2$$

Group-sequential structure



Additional Subjects	n_1	n_2	n_{L-1}	n_L	n_{K-1}	n_K
Cumulative Subjects	N_1	N_2	N_{L-1}	N_L	N_{K-1}	$N_K=N$
Information Time	$t_1 = \frac{N_1}{N}$	$t_2 = \frac{N_2}{N}$	$t_{L-1} = \frac{N_{L-1}}{N}$	$t_L = \frac{N_L}{N}$	$t_{K-1} = \frac{N_{K-1}}{N}$	$t_K = \frac{N_K}{N}$
Observed Effect size	Δ_1	Δ_2	Δ_{L-1}	Δ_L	Δ_{K-1}	Δ_K
2-sample Z Test Statistic	T_1	T_2	T_{L-1}	T_L	T_{K-1}	T_K
Critical values	C_1	C_2	C_{L-1}	C_L	C_{K-1}	C_K
Reject H_o & Stop trial if:	$T_1>C_1$	$T_2>C_2$	$T_{L-1}>C_{L-1}$	$T_L>C_L$	$T_{K-1}>C_{K-1}$	$T_K>C_K$

Conditional Power

- The conditional power evaluated at the L^{th} interim analysis

$$CP_L(\Delta) = \Pr\left[T_K > C_K \middle| T_L, \Delta\right]$$

$$= \emptyset\left(-\frac{C_K - T_L\sqrt{t_L} - (1-t_L)\Delta\sqrt{\frac{N}{2}}}{\sqrt{1-t_L}}\right)$$

- $\{ T_K > C_K \}$ is the Rejection Region.

Repeated significance test

can we use same rejection limit for all interims?

Let: $Y_{ij} \sim N\left(\mu_j, \sigma^2\right)$ Test: $H_0 : \mu_1 = \mu_2$

Test statistic: $Z_{mk} = \dfrac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{2\sigma^2 / mk}}$ $\hat{\mu}_j = \dfrac{1}{mk}\sum_{i=1}^{mk} Y_{ij}$

For $k = 1, \ldots K-1$ If $\left|Z_k\right| \geq C_\alpha$ Stop, reject H_0

otherwise Continue to group $k+1$

For $k = K$ If $\left|Z_k\right| \geq C_\alpha$ Stop, reject H_0

otherwise stop accept H_0

True type I error rate

Repeat testing until H_0 rejected

Tests	Critical value	P(type I error)
1	1.96	0.05
2	1.96	0.08
3	1.96	0.11
4	1.96	0.13
5	1.96	0.24

Too high! Want to control error rate so that overall error rate = 5%

Rejection limits need to be adjusted somehow. Many methods available.

Pocock's test

Suppose we want to test the null hypothesis 5 times using the same critical value each time and keep the overall significance level at 5%

For $k = 1, \ldots K-1$ If $|Z_k| \geq C_p(\alpha, K)$ Stop, reject H_0

otherwise Continue to group $k+1$

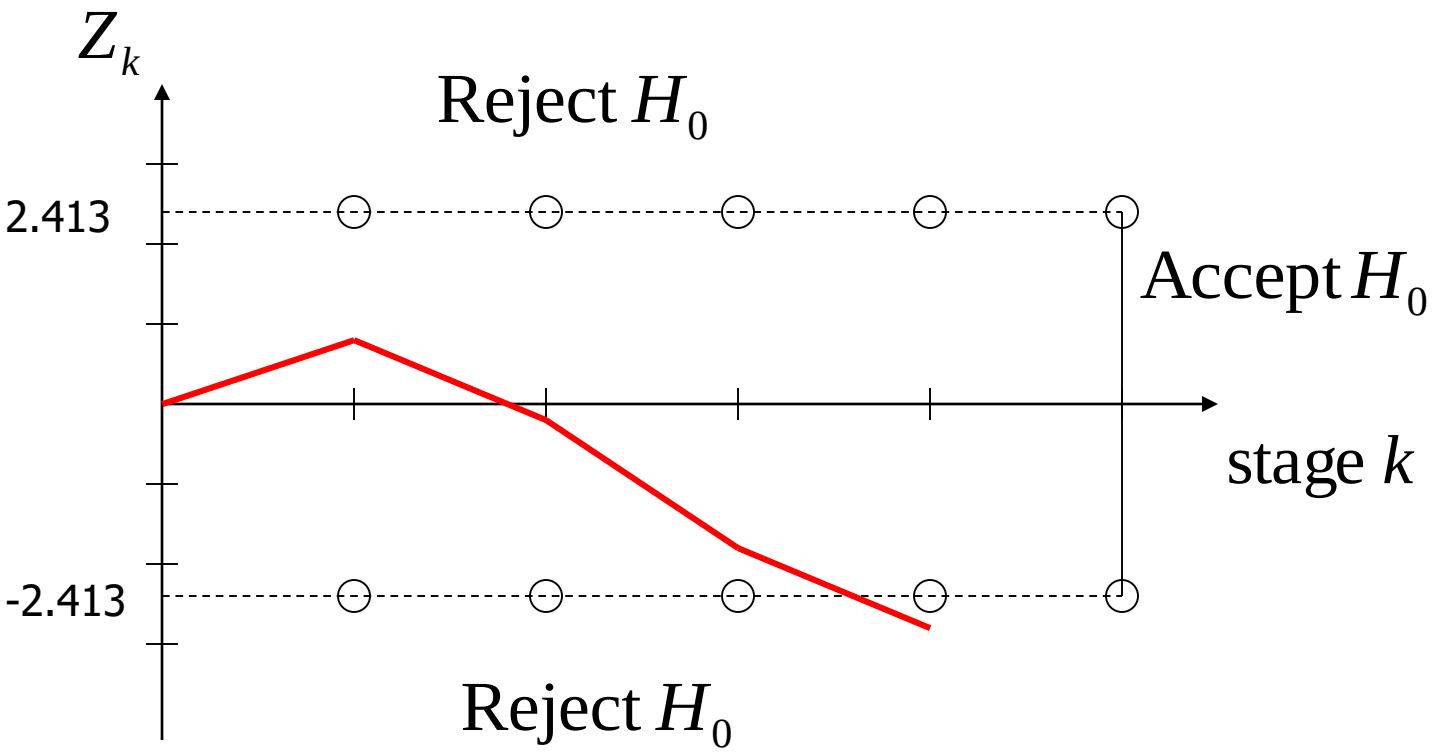
After group K If $|Z_k| \geq C_p(\alpha, K)$ Stop, reject H_0

otherwise stop accept H_0

Choose $C_p(\alpha, K)$ Such that

$$P(\text{Reject } H_0 \text{ at any analysis } k = 1 \ldots K) = \alpha$$

Pocock's test



2.413 ↔ Nominal sign level ≈ 1.6%

Group sequential tests

K	Interim Analysis Number	O'Brien-Fleming		Haybittle-Peto*		Pocock	
		B	α	C	α	C	α
2	1	2.782	0.0054	3.0	0.002	2.178	0.0294
	2	1.967	0.0492	1.960	0.0500	2.178	0.0294
3	1	3.438	0.0006	3.291	0.0010	2.289	0.0221
	2	2.431	0.0151	3.291	0.0010	2.289	0.0221
	3	1.985	0.0471	1.960	0.0500	2.289	0.0221
4	1	4.084	0.00005	3.291	0.00100	2.361	0.0182
	2	2.888	0.0039	3.291	0.00100	2.361	0.0182
	3	2.358	0.0184	3.291	0.00100	2.361	0.0182
	4	2.042	0.0412	1.960	0.0500	2.361	0.0182
5	1	4.555	0.000005	3.291	0.00100	2.413	0.0158
	2	3.221	0.0013	3.291	0.00100	2.413	0.0158
	3	2.630	0.0085	3.291	0.00100	2.413	0.0158
	4	2.277	0.0228	3.291	0.00100	2.413	0.0158
	5	2.037	0.0417	1.960	0.0500	2.413	0.0158

O'Brian & Flemmings test

Increasing nominal significance levels – decreasing rejection limits

For $k = 1, \ldots K-1$: If $|Z_k| \geq C_p(\alpha, K)\sqrt{K/k}$ Stop, reject H_0

otherwise Continue to group $k+1$

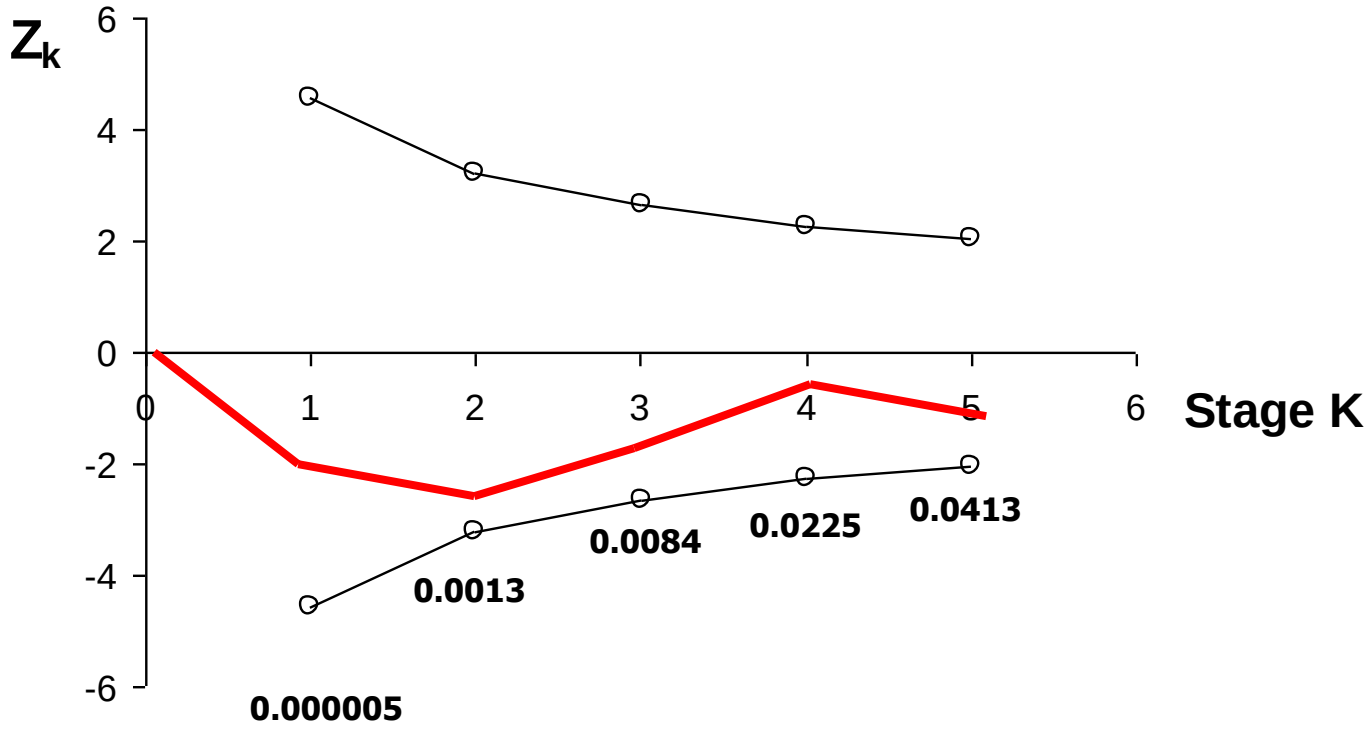
After group K : If $|Z_k| \geq C_p(\alpha, K)$ Stop, reject H_0

otherwise stop accept H_0

Choose $C_p(\alpha, K)$ Such that

$$P(\text{Reject } H_0 \text{ at any analysis } k = 1 \ldots K) = \alpha$$

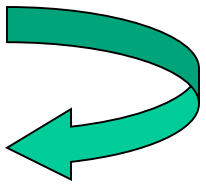
O'Brian & Flemmings test



O'Brian & Flemmings test

Critical values and nominal significance levels for a O'Brian Flemming test with 5 interrim tests.

Test (k)	$C_B(K,\alpha)$	$C_B(K,\alpha)*Sqrt(K/k)$	α'
1	2.04	4.56	0.000005
2	2.04	3.23	0.0013
3	2.04	2.63	0.0084
4	2.04	2.28	0.0225
5	2.04	2.04	0.0413



Rather close to 5%

Group sequential tests

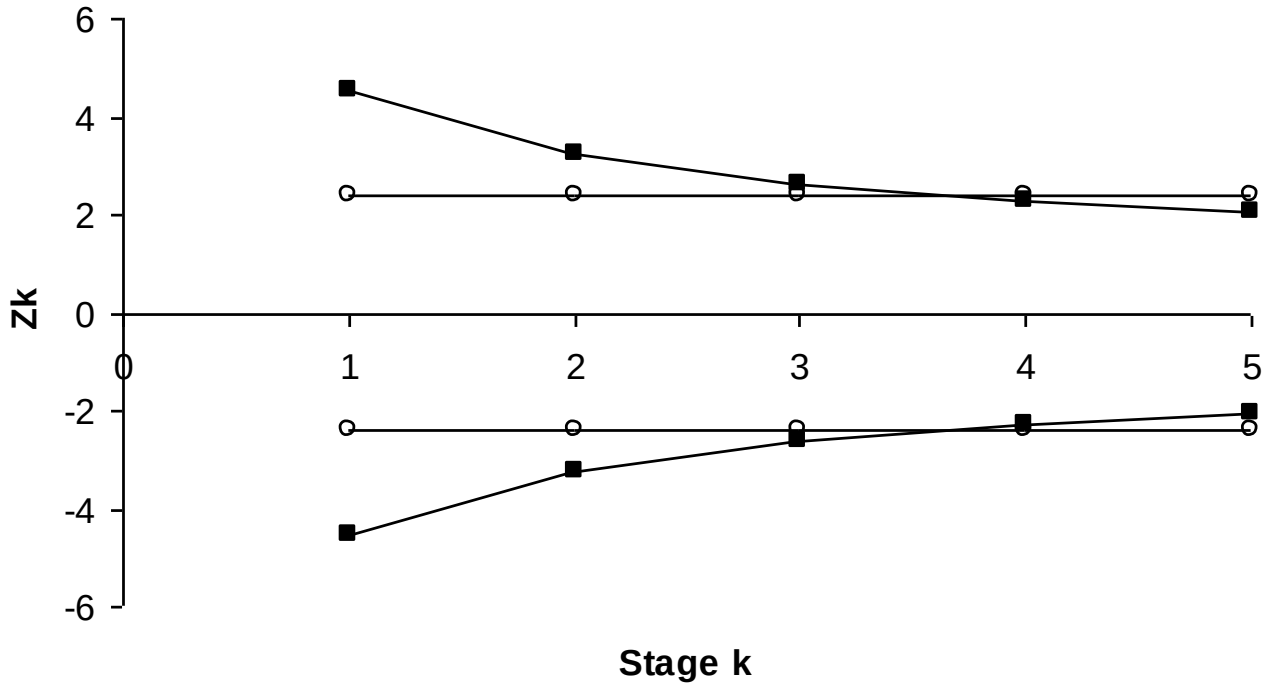
K	Interim Analysis Number	O'Brien-Fleming		Haybittle-Peto*		Pocock	
		B	α	C	α	C	α
2	1	2.782	0.0054	3.0	0.002	2.178	0.0294
	2	1.967	0.0492	1.960	0.0500	2.178	0.0294
3	1	3.438	0.0006	3.291	0.0010	2.289	0.0221
	2	2.431	0.0151	3.291	0.0010	2.289	0.0221
	3	1.985	0.0471	1.960	0.0500	2.289	0.0221
4	1	4.084	0.00005	3.291	0.00100	2.361	0.0182
	2	2.888	0.0039	3.291	0.00100	2.361	0.0182
	3	2.358	0.0184	3.291	0.00100	2.361	0.0182
	4	2.042	0.0412	1.960	0.0500	2.361	0.0182
5	1	4.555	0.000005	3.291	0.00100	2.413	0.0158
	2	3.221	0.0013	3.291	0.00100	2.413	0.0158
	3	2.630	0.0085	3.291	0.00100	2.413	0.0158
	4	2.277	0.0228	3.291	0.00100	2.413	0.0158
	5	2.037	0.0417	1.960	0.0500	2.413	0.0158

Rather close to 5%

Comparing Pocock and O'Brian Flemming

	O'Brian Flemming		Pocock	
Test (k)	**$C_B(K,a)*Sqrt(K/k)$**	**α'**	**$C_P(K,a)$**	**α'**
1	4.56	0.00001	2.413	0.0158
2	3.23	0.0013	2.413	0.0158
3	2.63	0.0084	2.413	0.0158
4	2.28	0.0225	2.413	0.0158
5	2.04	0.0413	2.413	0.0158

Comparing Pocock and O'Brian Flemming



Group Sequential Designs

Pros:

- Efficiency Gain (Decreasing marginal benefit)
- Establish efficacy earlier
- Detect safety problems earlier

Cons:

- Smaller safety data base
- Complex to run
- Need to live up to stopping rules!

Chapter 5 Reading Instructions

- 5.1: Introduction
- 5.2: Parallel Group Designs (read)
- 5.3: Cluster Randomized Designs (less important)
- 5.4: Crossover Designs (read+copies)
- 5.5: Titration Designs (read)
- 5.6: Enrichment Designs (less important)
- 5.7: Group Sequential Designs (read include 10.6)
- 5.8: Placebo-Challenging Designs (less important)
- 5.9: Blinded Reader Designs (less important)
- 5.10: Discussion