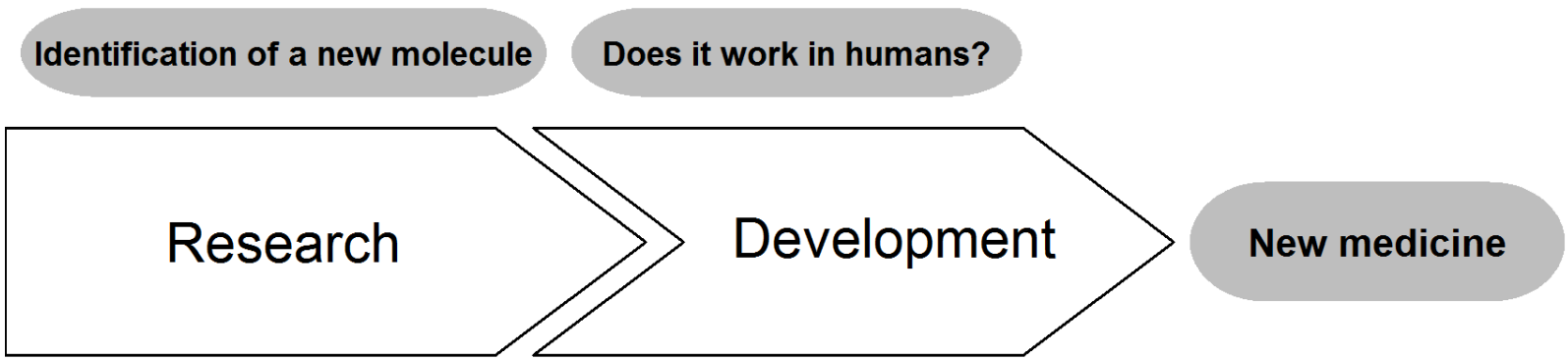
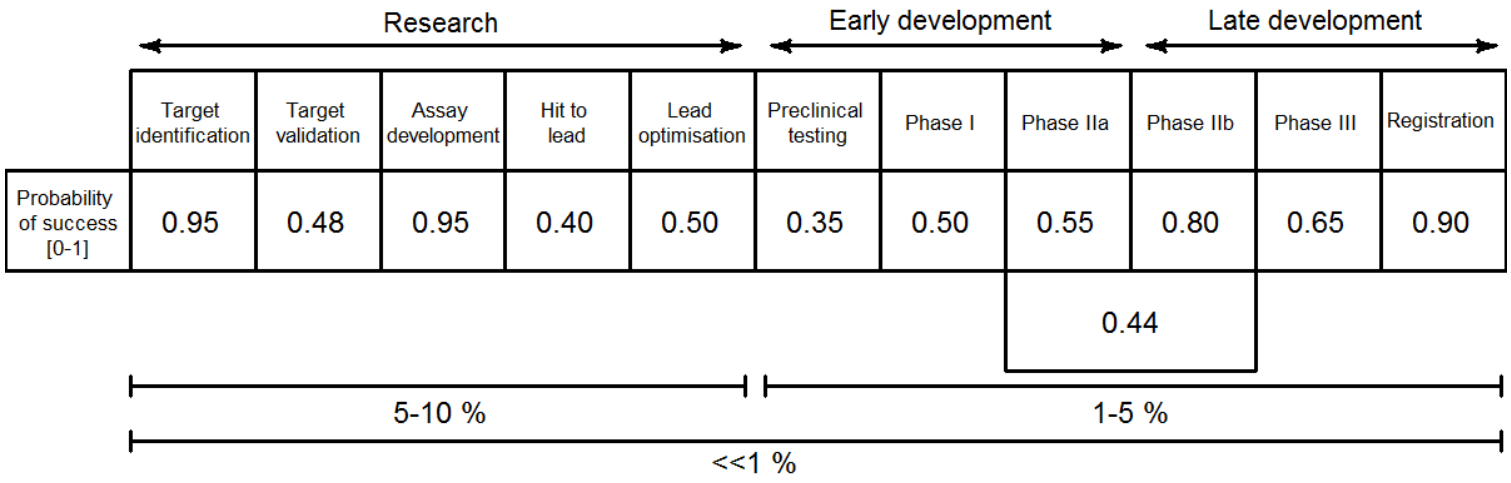
The Use of Biostatistics in Clinical Development

Søren Lophaven

Research and Development in the Pharmaceutical Industry



The Probability of Success



The Price of a Research and Developement Project

- For every 50000 molecules screened in research
  - 30 are entering preclinical testing
  - 10 are going to clinical testing
  - 1 is being approved by the regulatory authorities
- The total costs of a research and development project is 1.6-2.6 mia Dkr
- The aim is to close "bad projects"as soon as possible

Phase I Clinical Trial

- Investigation of the distribution of the drug in the human body
- Healthy volunteers
- Performed at specialised centres
- Single or multiple dosing
- What is the optimal dose level?
- 10 - 50 subjects

Phase II Clinical Trial

- Identification of an effective dose range
- Investigation performed in patients
- 50-400 patients
- Hospital centres in multiple countries

Phase III Clinical Trial

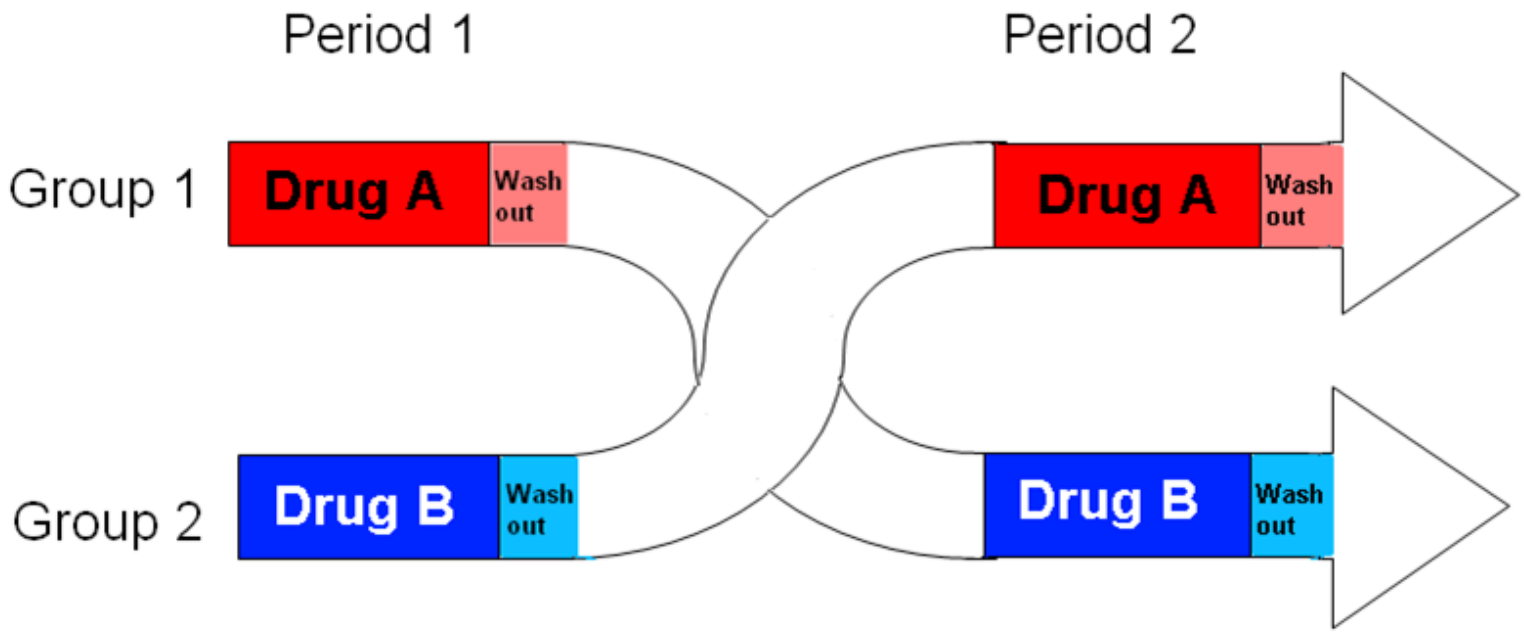
- Confirm the effect observed in phase II
- Further tolerability and safety data
- Comparison to existing treatment(s) and placebo
- 500 - 1.500 (+) patients
- Hospital centres in many countries involved

What do Biostatisticians Work With?

- Design of Trials
  - Design, blinding, randomisation, sample size, data collection (endpoints), comparator
- Conduct of Trials
  - Data issues
- Analysis and Reporting of Trials
  - Statistical Analysis Plan (SAP)
  - Data analysis
  - Reporting of results
- Analysis and Reporting Across Trials

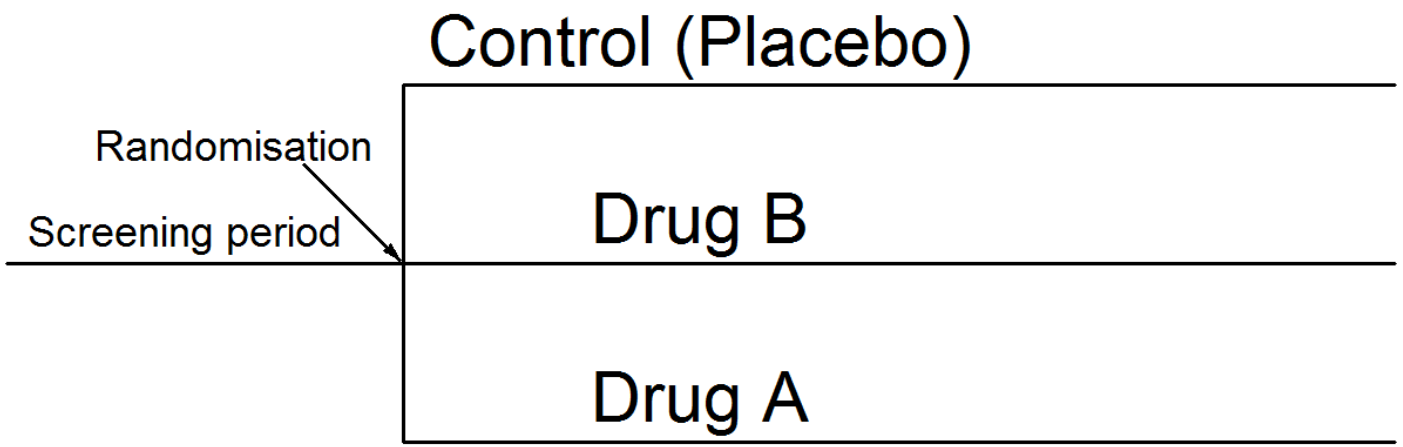
Cross-over or Parallel Group Design

- Phase I clinical trials
- Phase II/III clinical trials in chronic conditions



Parallel Group Design

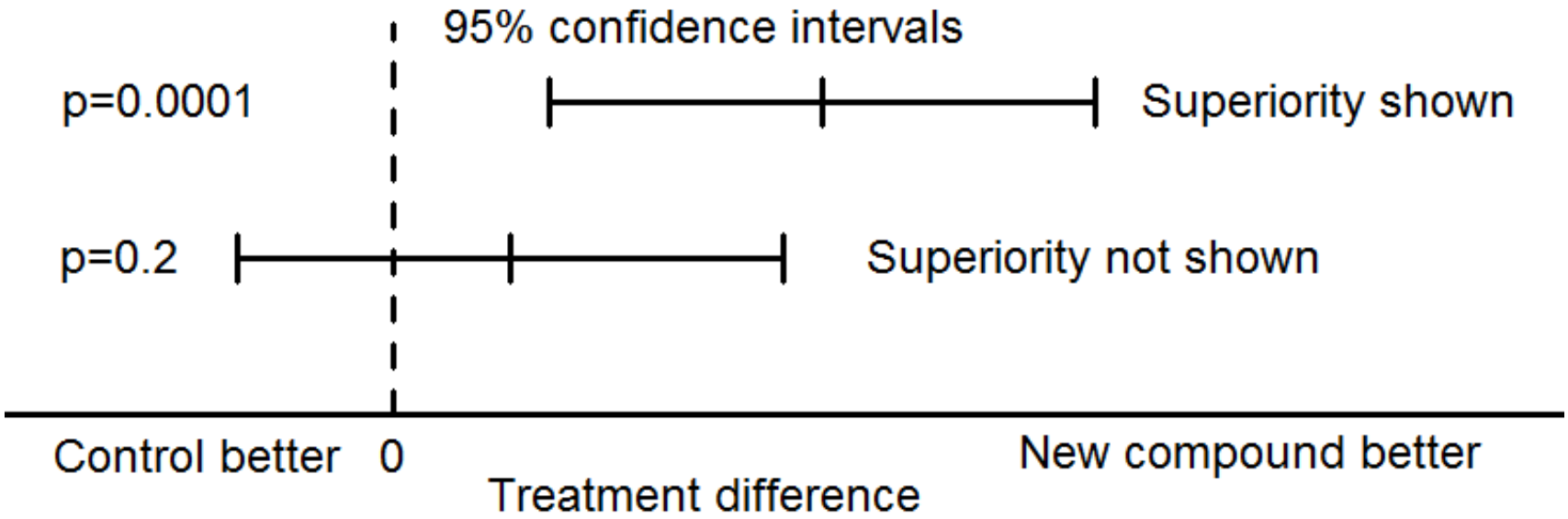
- Phase II/III clinical trials



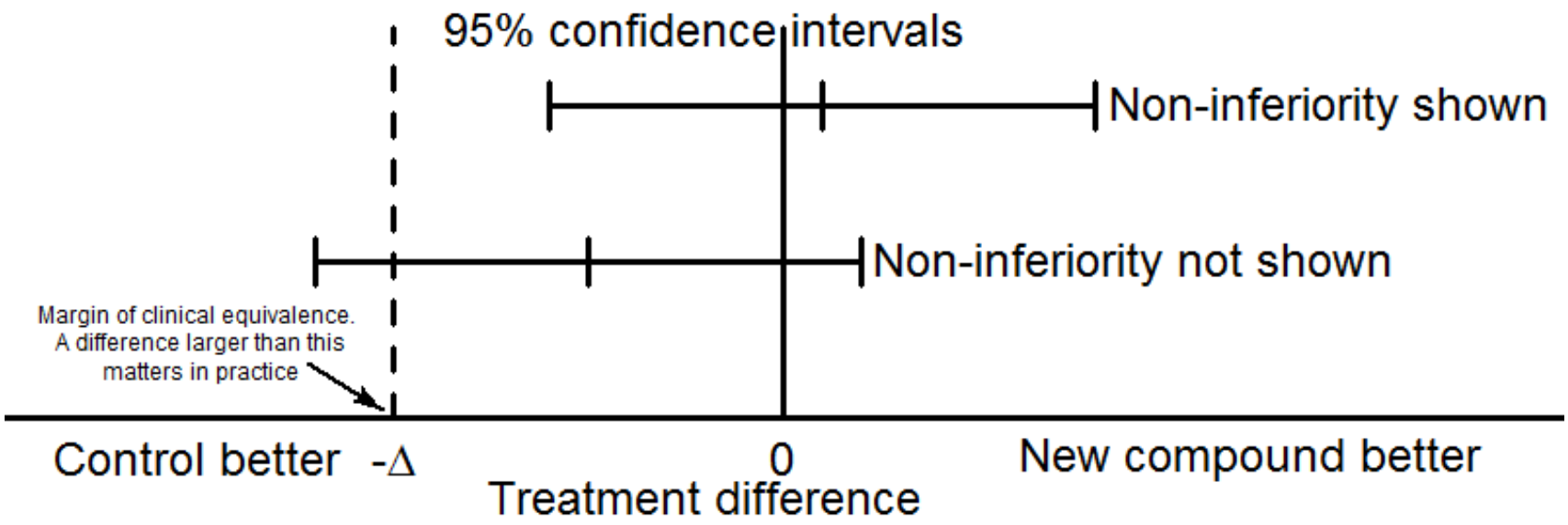
Trial Design - Comparators

- Placebo control
  - Estimates the true treatment effect
  - Can be unethical to use if well established treatment exists
- Active control
  - Active reference for validation of the trial design
  - Active comparator for comparing against an approved drug (superiority or non-inferiority)

Trial design - Showing Superiority



Trial design - Showing Non-inferiority



Drug Treatments of Asthma

- Chronic condition
- Abnormal degree of wheezing or breathlessness
- Severity of symptoms ebbs and flows over time
- Efficacy endpoint: Maximum rate at which the patient is able to exhale a fixed time after administration of the drug (PEF)
- Comparison of the two drugs Salbutamol and Formoterol for short-term relief of the symptoms

PEF Data (Parallel Group Design)

Formoterol	Salbutamol
310	270
310	260
370	300
410	390
250	210
380	350
330	365
385	370
400	310
410	380
320	290
340	260
220	90

PEF Data (Parallel Group Design) - Some Calculations

Mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Variance: $var(x) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, measures the total squared average "distance"from the mean to the observations

Standard deviation: $SD = \sqrt{var(x)}$, measures the total average "distance"from the mean to the observations

Standard error: $SE = \frac{SD}{\sqrt{n}}$, measures the variability of the mean

Formoterol: Mean=341.2, variance=3559.32, SD=59.66

Salbutamol: Mean=295.8, variance=6865.78, SD=82.86

PEF Data (Parallel Group Design) - Some Calculations for the Difference

Mean: $341.2 - 295.8 = 45.4$

Pooled variance: $(3559.32 + 6865.78)/2 = 5212.84$

Standard error:

$$SE = \sqrt{\frac{2 \cdot pooled\ \ variance}{n}} = \sqrt{\frac{2 \cdot 5212.84}{13}} = 28.3$$

95 % confidence interval for the difference:

$$mean \pm 2 \cdot SE = [-11.3; 102.0]$$

PEF Data (Cross-over Design)

Patient ID	Formoterol	Salbutamol	Difference
1	310	270	40
2	310	260	50
3	370	300	70
4	410	390	20
5	250	210	40
6	380	350	30
7	330	365	-35
8	385	370	15
9	400	310	90
10	410	380	30
11	320	290	30
12	340	260	80
13	220	90	130

PEF Data (Cross-over Design) - Some Calculations

Mean: 45.4

SD: 40.59

Standard error: $SE = \frac{SD}{\sqrt{n}} = \frac{40.59}{\sqrt{13}} = 11.26$

95 % confidence interval for the difference:

$$mean \pm 2 \cdot SE = [22.9; 67.9]$$

Endpoints

- Novo Nordisk: plasma glucose concentration
- Lundbeck: Montgomery and Åsberg Depression Rating Scale (MADRS)

- Objective versus subjective endpoints
- Continuous versus categorical

Montgomery and Åsberg Depression Rating Scale (MADRS)

- MADRS: rating scale designed to assess the severity of depressive symptoms
- Based on a clinical interview
- Administered by trained psychiatrists
- MADRS total score is the sum of the score of the 10 individual items
- Symptoms rated on 7-point scales from 0 (no symptom) to 6 (severe symptom) with detailed anchor points
- MADRS total score goes from 0 to 60

Two MADRS items

1. Apparent sadness

Representing despondency, gloom and despair, (more than just ordinary transient low spirits) reflected in speech, facial expression, and posture. Rate by depth and inability to brighten up.

- ☐ 0 No sadness.
- ☐ 1
- ☐ 2 Looks dispirited but does brighten up without difficulty.
- ☐ 3
- ☐ 4 Appears sad and unhappy most of the time
- ☐ 5
- ☐ 6 Looks miserable all the time. Extremely despondent.

4. Reduced sleep

Representing the experience of reduced duration or depth of sleep compared to the subject's own normal pattern when well.

- ☐ 0 Sleeps as usual.
- ☐ 1
- ☐ 2 Slight difficulty dropping off to sleep or slightly reduced, light or fitful sleep.
- ☐ 3
- ☐ 4 Sleep reduced or broken by at least two hours.
- ☐ 5
- ☐ 6 Less than two or three hours sleep.

Sample Size

- What things might affect the sample size needed for a trial?

?

Sample Size

- The number of patients needed in a clinical trial is determined by:

- Effect
  - How large an effect is expected on the primary endpoint? (from literature or previous trials, larger effect means fewer patients needed)
- Variation
  - How much variation is expected? (from literature or previous trials, larger effect means fewer patients needed)

Sample Size

- Power
  - The probability of finding a statistically significant effect, assuming that it exists (80%, 90%)
  - More patients provide higher power
- Significance level
  - The p-value has to be below 5% for comparing one drug to placebo, but below 2.5% if 2 are compared to placebo

Example of Sample Size Calculation

- Depression trial against placebo in a simple setting (two treatments, continuous endpoint, power=80%, significance level=5%)
- $n = \frac{2 \cdot (1.96 + 0.84)^2 \cdot SD^2}{effect^2}$
- effect=3, SD=9
- $n = \frac{2 \cdot (1.96 + 0.84)^2 \cdot 9^2}{3^2} = 142$ in each treatment group

Protocol Text

With 142 patients in each treatment group, a true difference between vortioxetine and placebo of 3 MADRS points, the standard deviation of 9 MADRS points and a significance level of 5% the probability of demonstrating a difference (in favor of vortioxetine) to Placebo is approximately 80%

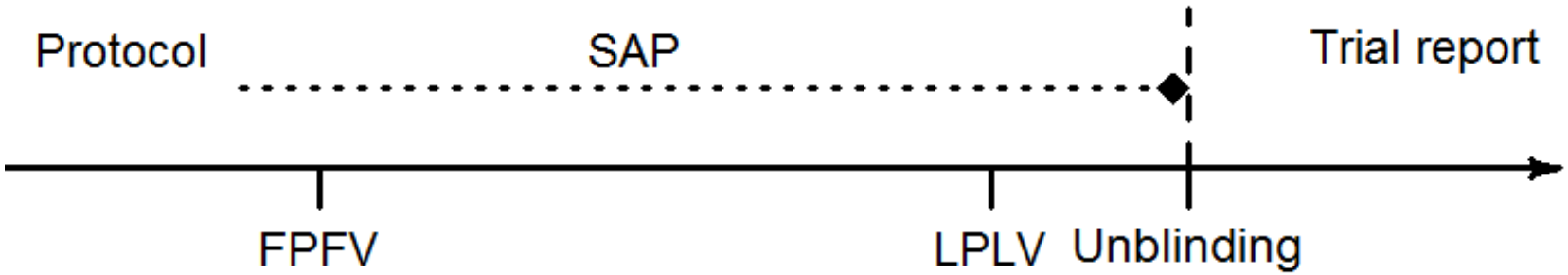
Randomisation and Blinding

- Blinding
  - Patients, investigators and trial personnel (including the biostatistician) are blinded to the treatment codes, until the trial is finalised
    - Prevents bias in judgement from investigator, patients and trial personnel
- Randomisation
  - Treatment has been allocated in a randomised fashion
    - Limits bias from underlying explanatory factors
- Stratification

Statistical Analysis Plan

- The trial protocol describes the activities that take place in connection with the trial and the key parts of the statistical analysis
- The Statistical Analysis Plan (SAP) specifies the details of the analysis of the trial data

Timing of the Statistical Analysis Plan



Statistical Analysis Plan

- Idea
  - To describe all parts of the statistical analyses before unblinding
- Content
  - Which patients are included in the analyses?
  - Statistical methods to be applied
  - Handling of missing data
  - Primary analysis and sensitivity analyses
  - Correcting for multiplicity

Patients to be Analysed

- In general data from all randomised patients should be analysed
- For reporting of safety data: all randomised patients who took at least one dose of investigational medicinal product (IMP)
- For reporting of efficacy data: Patients who took at least one dose of IMP and had a valid baseline assessment and at least one valid post-baseline assessment of the primary endpoint

Statistical Methods

- Statistical methods are chosen based on what is measured
- Binary endpoints: logistic regression, Fisher's exact test, chi-squared test
- Continuous endpoints: Analysis of covariance (ANCOVA), Mixed Model for Repeated Measures
- Time-to-event endpoints: Cox regression

Handling Missing Data

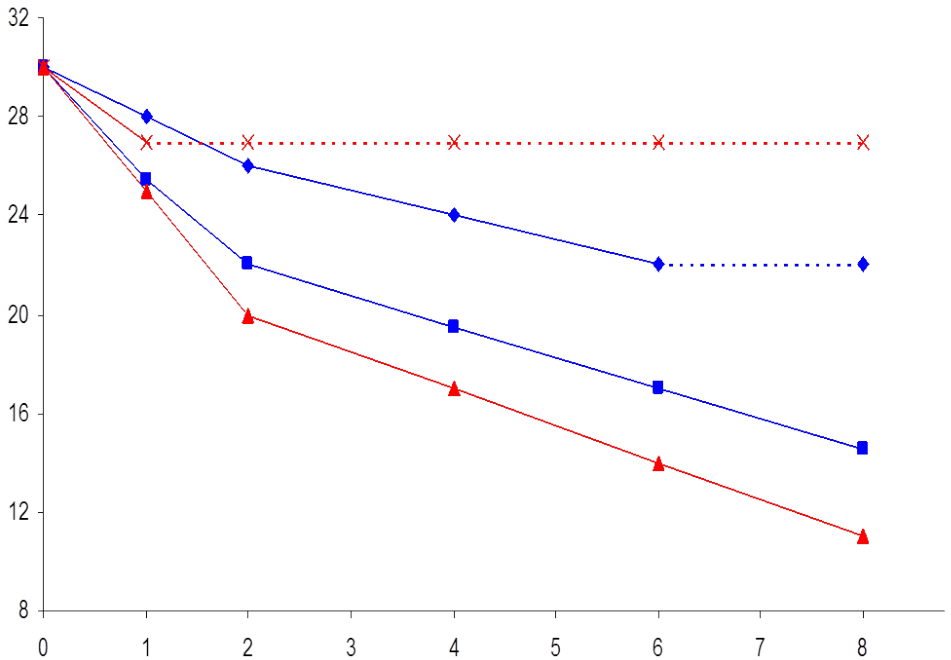
- Missing data due to e.g. patients withdrawing, missing visits, missing assessments at a visit
- Recommendations:
  - Avoid missing data
  - Pre-specified analyses for handling missing data
  - Analysis method should be conservative
  - Analyse the number, timing, and pattern of missing data
  - Use sensitivity analyses, to further investigate the robustness of the method chosen

Methods for Handling Missing Data

- Observed Cases (OC) - At each time point, only use the actual observations
- Last Observation Carried Forward (LOCF) - If an observation is missing, use the previous one
- Mixed Model Repeated Measures (MMRM)

Observed Cases and Last Observation Carried Forward analysis

An Observed Cases analysis would only have two observations at week 8. Last Observation Carried Forward has 4, but two of them are "punished"for dropping out earlier



Last Observation Carried Forward as a Conservative Approach?

- Alzheimer's disease and LOCF:
  - Patient's condition is expected to deteriorate over time
  - LOCF is likely to give overly optimistic estimates for both treatment (Active/Placebo)
  - Earlier withdrawals in the Active treatment group (e.g. due to adverse events)
  - Biased estimate of treatment difference in favour of Active

Last Observation Carried Forward as a Conservative Approach?

- Depression and LOCF:
  - Patient's condition improve over time
  - LOCF is likely to give conservative estimates for both treatment (Active/Placebo)
  - Earlier withdrawals in the Active treatment group (e.g. due to adverse events)
  - Biased estimate of treatment difference in favour of placebo
  - LOCF is considered as appropriate