Statistical Issues in the Design of Clinical Trials

Søren Lophaven

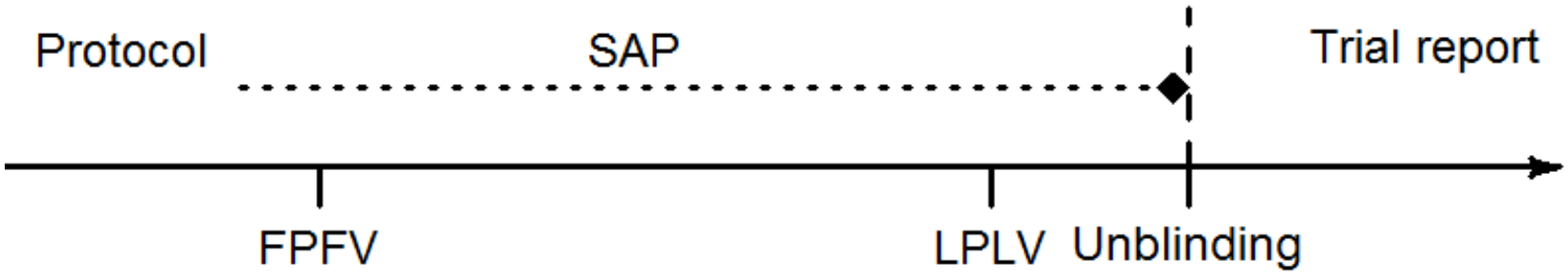
What do biostatisticians in the pharmaceutical industry do?

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

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

Randomisation and blinding

- Blinding
  - Ideally 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

Randomisation

- Netiher the selection of investigators nor the recruitment of patients is random
- However, patients entering a clinical trial are assigned to treatment groups at random
- Limits bias from underlying explanatory factors

ICH-E9:

- Randomisation provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects
- In combination with blinding, randomisation helps to avoid possible bias in the selection and allocation of subjects arising from the predictability of treatment assignments
- Details of the randomisation that facilitate predictability (e.g. block length) should not be contained in the trial protocol

Stratification

- Stratification can be used to ensure equal allocation of subgroups of patients to each treatment group.
- Examples of stratification variables: disease severity, age, gender
- Stratification does not aim at ensuring that e.g. at least a certain proportion of patients are women. Instead it aims at ensuring that the proportion of women is the same in different treatment groups
- Most clinical trials use block randomisation, i.e. clinical site is a stratification variable

Types of blinding

Open-label:

- No blinding employed
- Investigators and patients know which treatment the patient receives
- Not adequate as well-controlled clinical trial
- Used for post-marketing surveillance, long-term studies etc.

Single-blind:

- Treatment assignment not known by the patient
- Investigator's clinical evaluation may be biased

Double-blind:

- Investigators and patients not aware of treatment

Triple-blind:

- In addition to the patients and investigators all members of the clinical project associated with the study are blinded

The integrity of blinding

Table 1 Results of Patients' Guess on Treatment for the Prophylactic Use

	Actual assignment	
Patient's guess	Ascorbic acid	Placebo
Ascorbic acid	40	11
Placebo	12	39
Do not know	49	39
Total	101	89

Table 2 Results of Patients' Guess on Treatment for Weight Loss

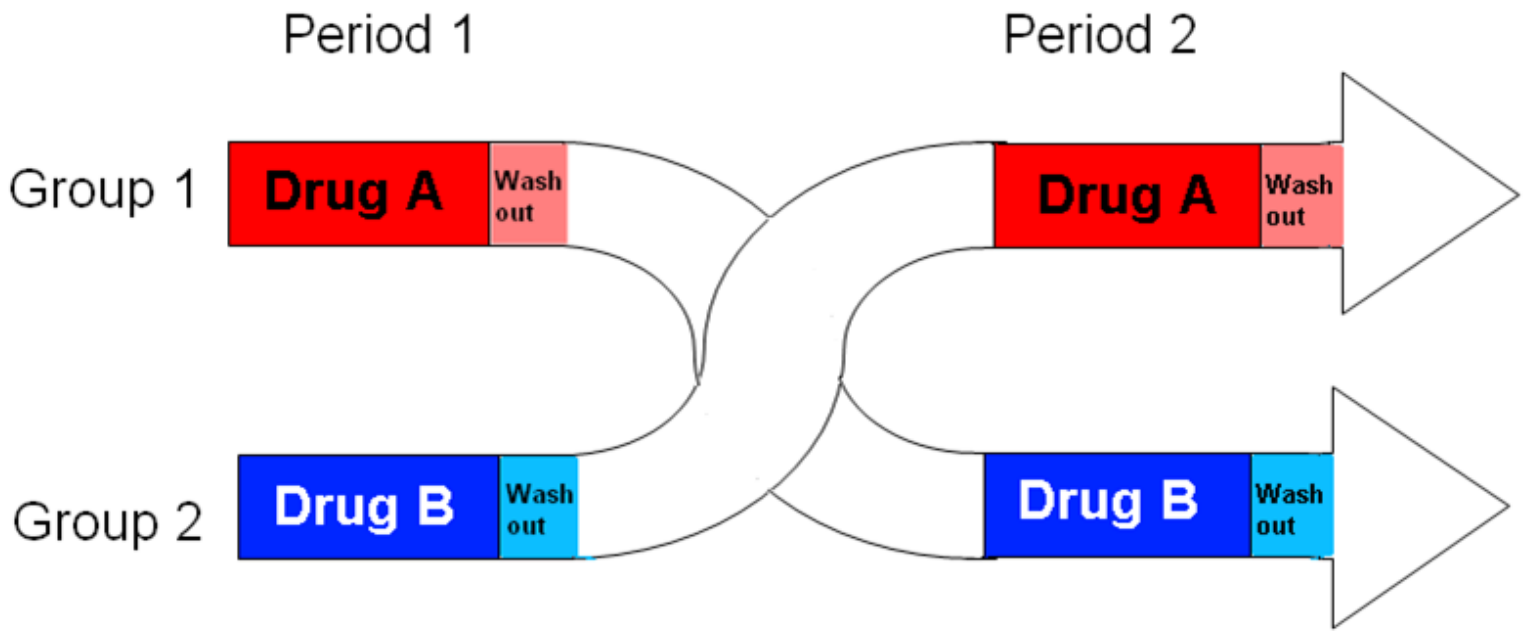
	Actual assignment	
Patient's guess	Active drug	Placebo
Active drug	19	3
Placebo	3	16
Do not know	2	6
Total	24	25

Blinding in ICH-E9

- Blinding is done to avoid potential influence on the trial results caused by subjective behaviour, reporting, evaluation, data processing, and analysis due to personal preference of treatment
- Also the placebo effect of a trial may differ depending on what treatment the patient think he/she is receiving
- If a double-blind trial is not feasible, then the single-blind option should be considered
- In single-blind or open-label trials every effort should be made to minimise the various known sources of bias and primary variables should be as objective as possible
- The sponsor should have adequate standard operating procedures to ensure that access to the treatment code is appropriately restricted during the process of cleaning the database prior to its release for analysis

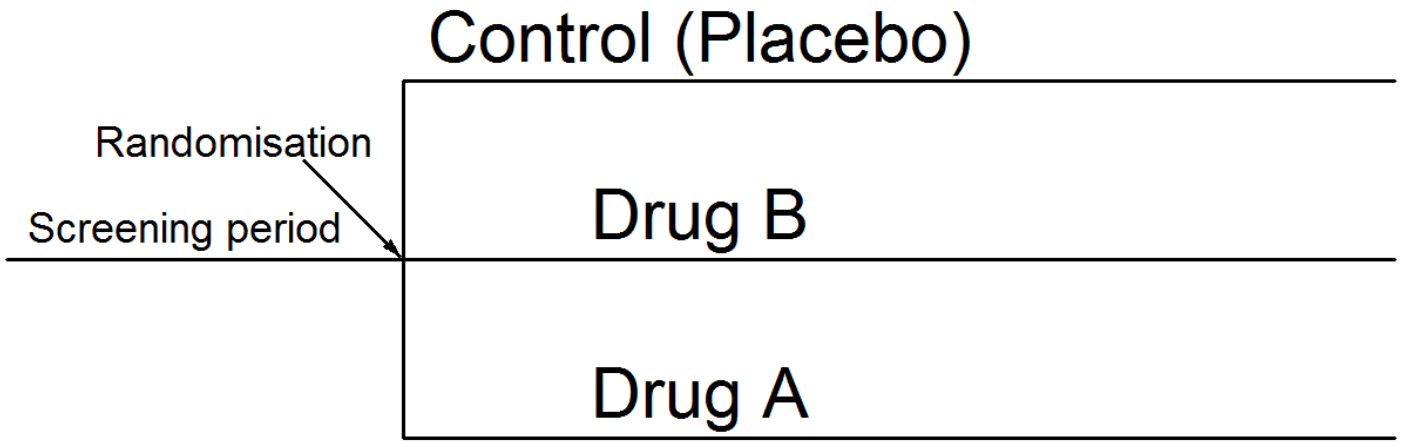
Cross-over or parallel group design

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



Parallel group design

- Phase II/III clinical trials



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

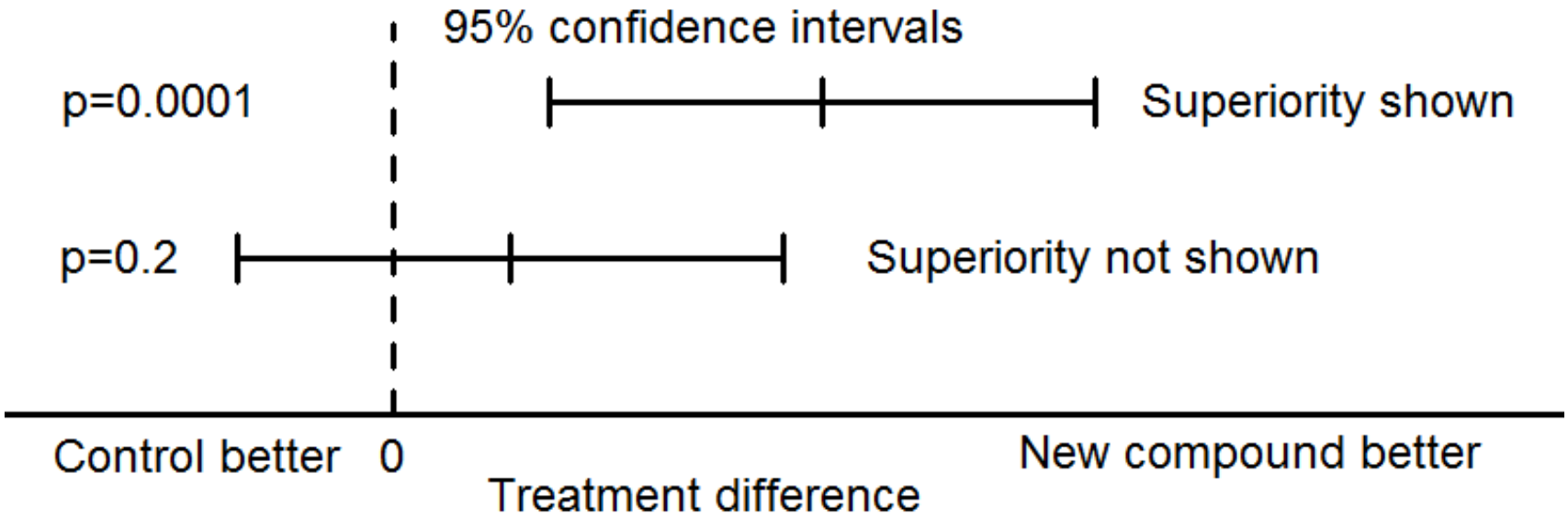
95 % confidence interval for the difference:

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

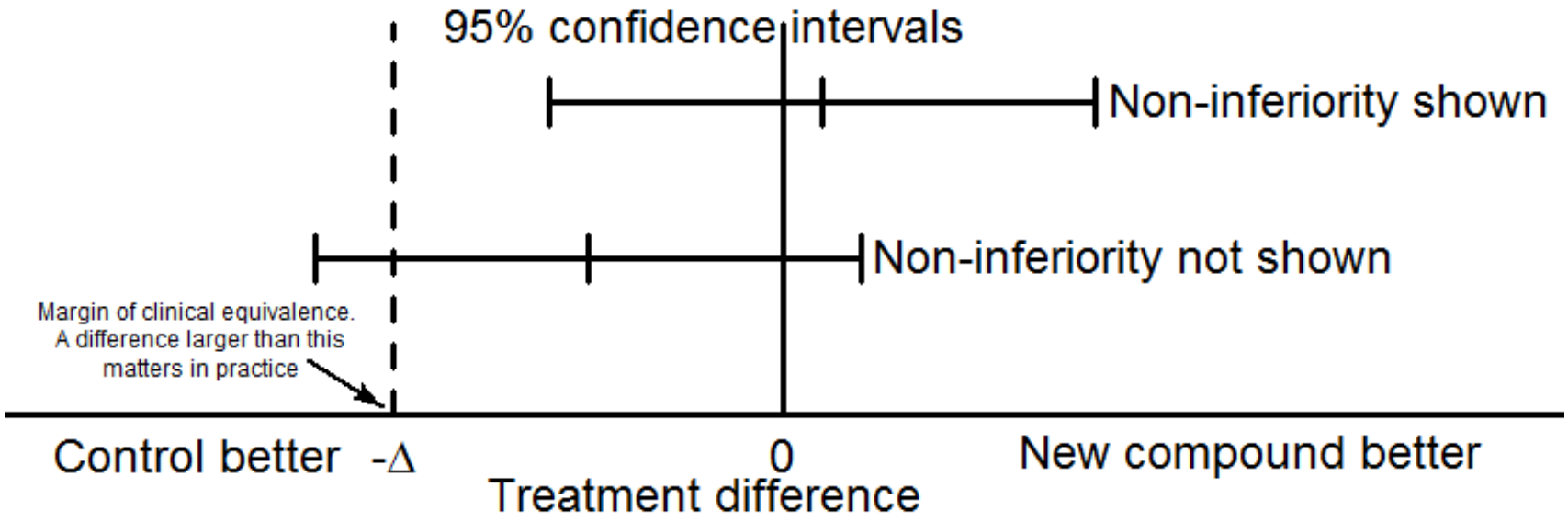
Trial design - comparators

- 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

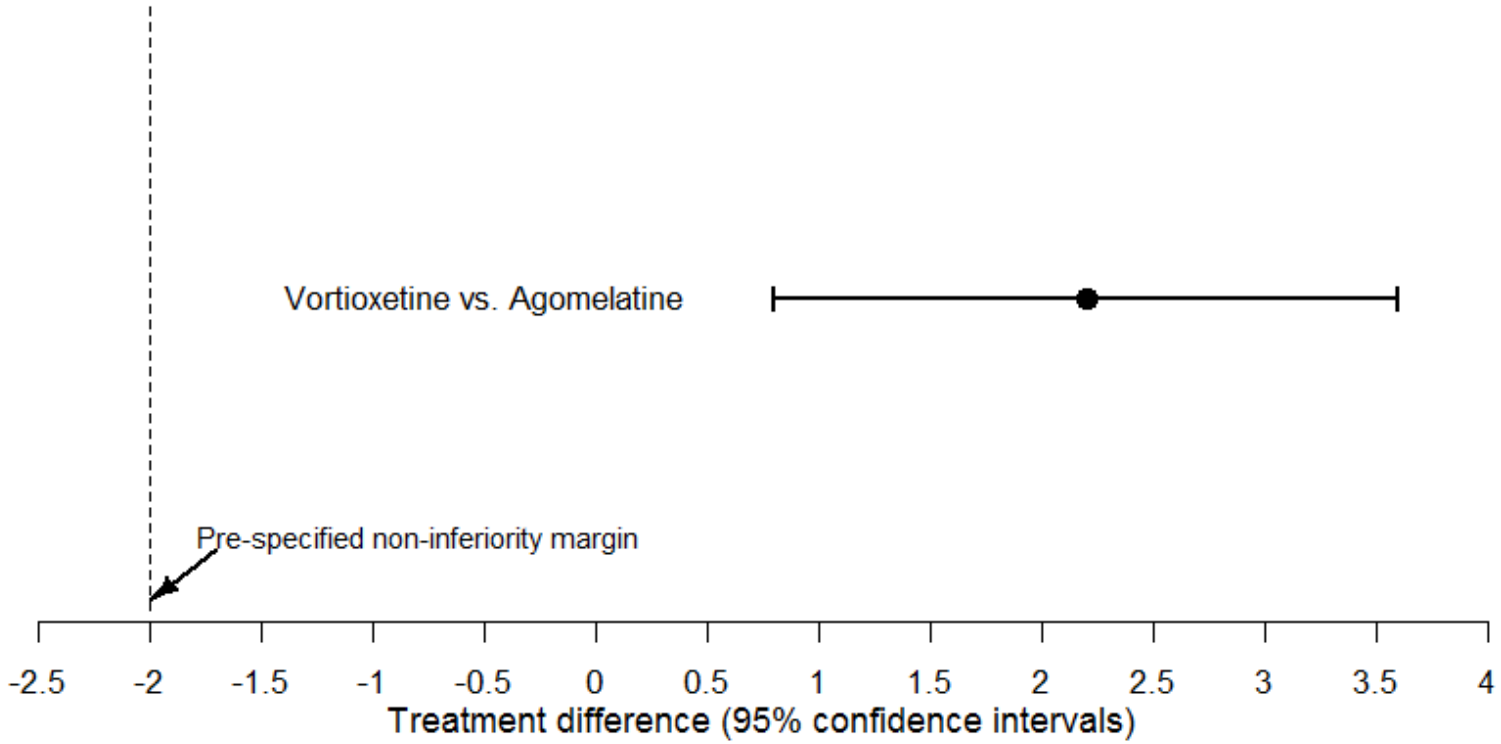


Example of non-inferiority trial

Objective: This randomised double-blind, 12-week study compared efficacy and tolerability of flexible-dose treatment with vortioxetine (10-20 mg/day) versus agomelatine (25-50 mg/day) in major depressive disorder patients with inadequate response to selective serotonin reuptake inhibitor (SSRI)/serotonin-noradrenaline reuptake inhibitor (SNRI) monotherapy

Methods: Patients were switched directly from SSRI/SNRI to vortioxetine or agomelatine. Primary endpoint was change from baseline to week 8 in the Montgomery-Åsberg Depression Rating Scale (MADRS) total score analysed by mixed model for repeated measurements, using a noninferiority test followed by a superiority test.

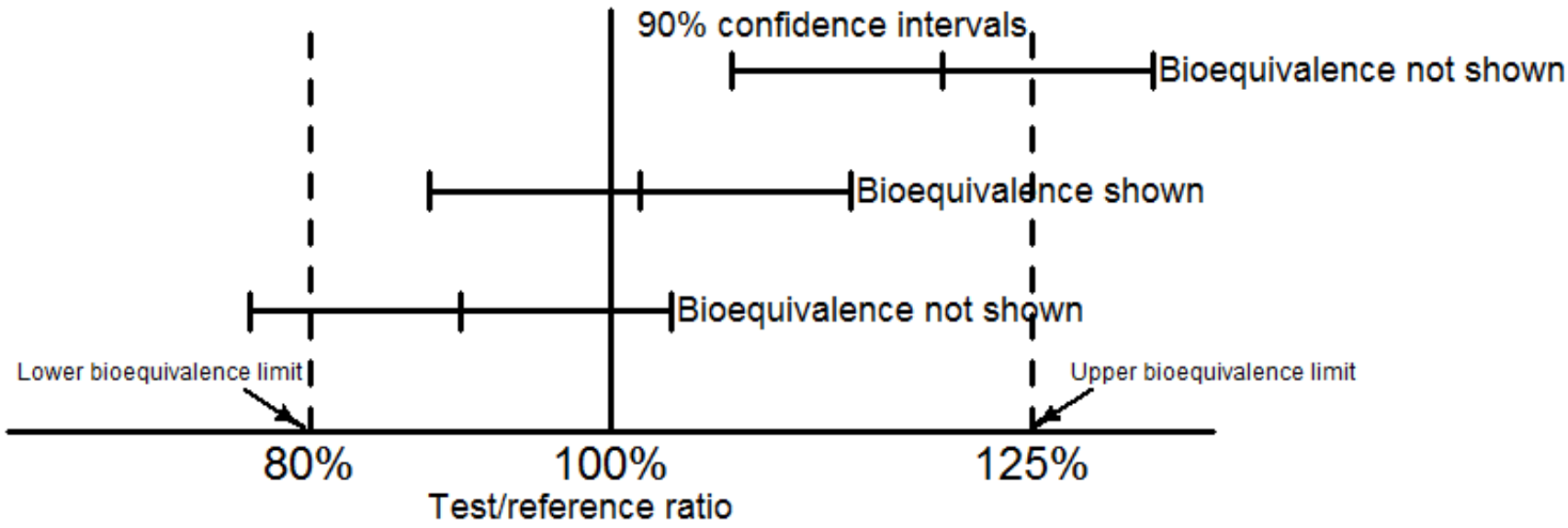
Example of non-inferiority trial



Example of non-inferiority trial

Vortioxetine was noninferior and significantly superior to agomelatine in major depressive disorder patients with previous inadequate response to a single course of SSRI/SNRI monotherapy

Trial design - showing equivalence

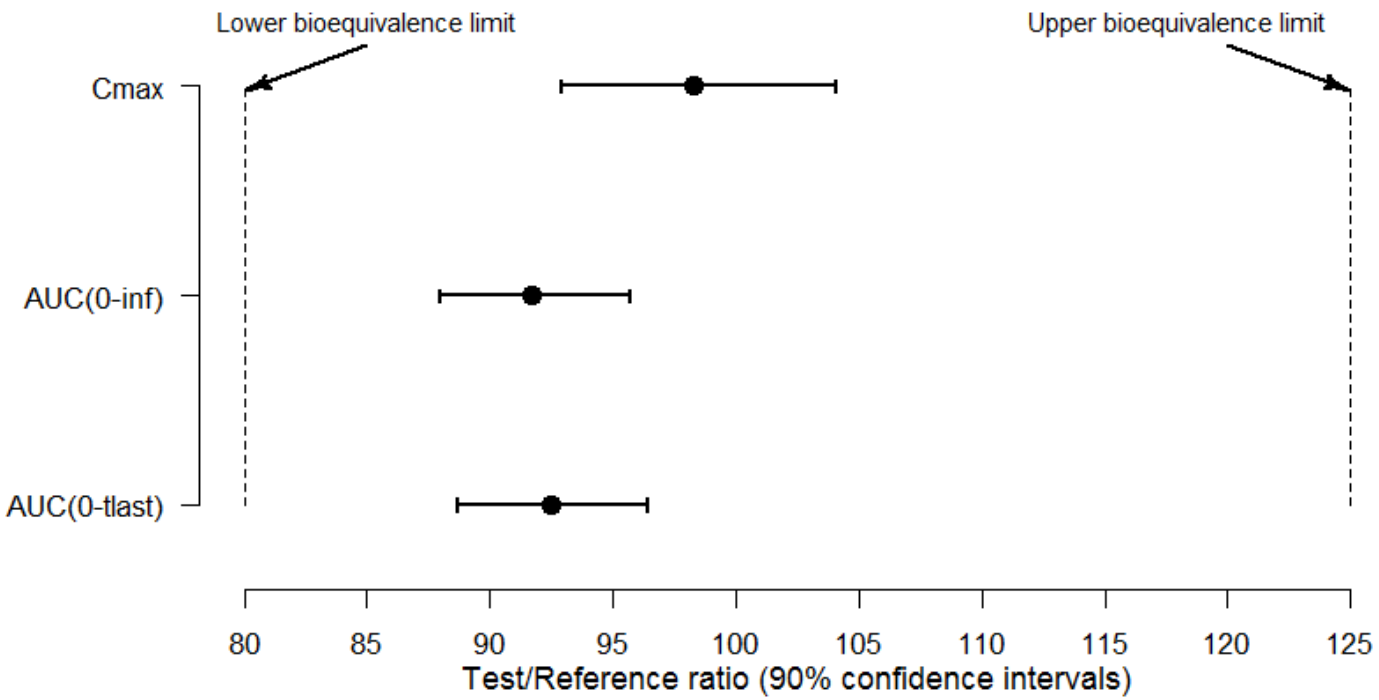


Example of a bioequivalence trial

Objective: Compare the pharmacokinetic profiles of the two transdermal fentanyl patches Matrifen (100 microg/h) and Durogesic (100 microg/h), used to manage severe chronic pain, and show that Matrifen and Durogesic are bioequivalent despite different technologies

Example of a bioequivalence trial

Methods: Transdermal fentanyl patches [Matrifen or Durogesic (100 microg/h)] were applied for 72 h to 30 healthy subjects in a randomized, four-period (two replicated treatment sequences), crossover study. The pharmacokinetic parameters of fentanyl were determined for 144 h after application using plasma samples.



Patients to be analysed

- In general data from all randomised patients should be analysed
- For reporting of safety data (as treated): all randomised patients who took at least one dose of investigational medicinal product (IMP)
- For reporting of efficacy data (ITT, as randomised): 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

Patients to be analysed

- For non-inferiority and equivalence testing the per-protocol and ITT analyses are of equal importance and should lead to a consistent conclusion
- For superiority testing the ITT analysis is the primary analysis

Analysis sets in depression studies

all-patients-randomised set (APRS)

- All randomised patients

all-patients-treated set (APTS)

- All patients in the APRS who took at least one dose of IMP

full-analysis set (FAS)

- All patients in the APTS who had a valid baseline assessment and at least one valid post-baseline assessment of the primary variable

per-protocol set (PPS) - all patients in FAS who:

- had more than 70% IMP compliance while in the study
- had more than 14 days of double-blind IMP exposure in total while in the study
- did not have >6 consecutive days of non-compliance with IMP during treatment
- did not have any other major protocol violation that could interfere with the efficacy outcomes, such as violation of eligibility criteria, prohibited concomitant medication or medical history

Interim analysis

Any analysis of data prior to formal completion of the trial

- Terminate the trial early for safety reasons
- Terminate the trial early for effectiveness
- Terminate the trial early for futility
- Modify design based on data (adaptive trial design)

Possibility for effective new treatments to reach the market earlier

Interim analysis

ICH-E9:

The execution of an interim analysis should be a completely confidential process because unblinded data and results are potentially involved. All staff involved in the conduct of the trial should remain blind to the results of such analyses, because of the possibility that their attitudes to the trial will be modified and cause changes in the characteristics of patients to be recruited or biases in treatment comparisons. This principle may be applied to all investigator staff and to staff employed by the sponsor except for those who are directly involved in the execution of the interim analysis.

Involvement of DMC to make recommendations and decisions

Sample size

- Why is it of interest?

?

Sample size

- Ethical - to bother unneccessarily many patients versus to bother patients unneccesary
- Economical - The cost of collecting information versus the cost of getting useless information
- Requirement - It is a requirement in clinical trials to state how many and why in the protocol

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)

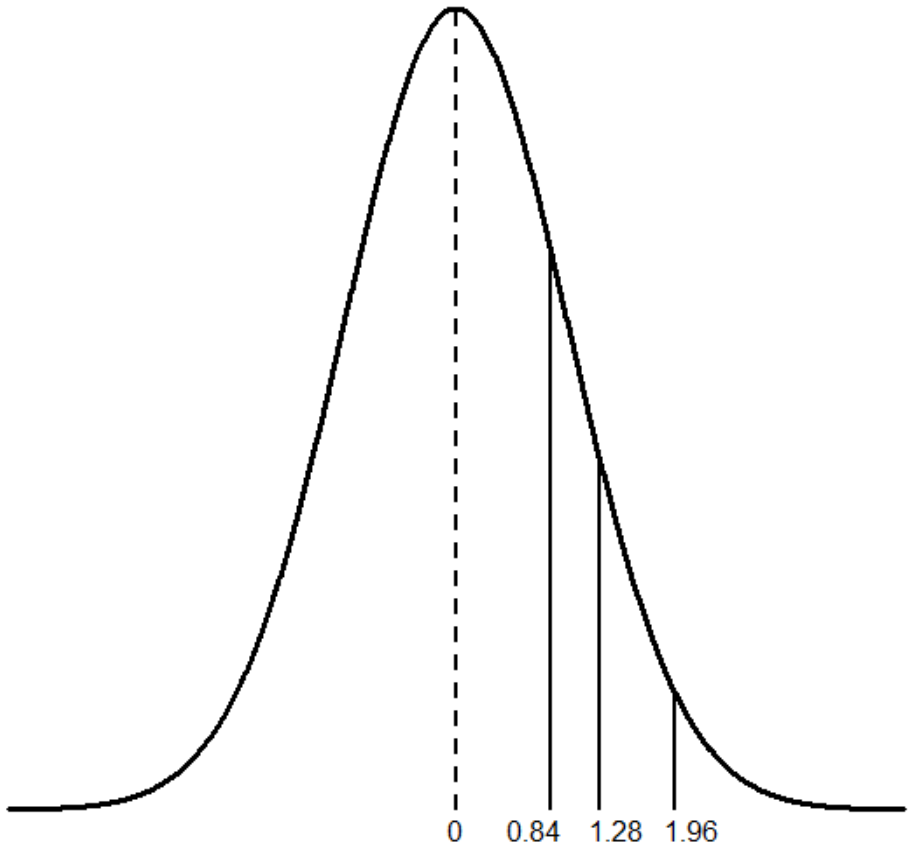
Sample size

- 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 - showing superiority

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

Quantiles from the normal distribution to determine Z



Protocol text

With 142 patients in each treatment group, a true difference between vortioxetine and placebo of 3 MADRS points, a 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%

Example of sample size calculation - showing non-inferiority

- Depression trial against agomelatine (two treatments, continuous endpoint, power=80%, significance level=5%, non-inferiority)
- $n = \frac{2 \cdot (Z_\alpha + Z_\beta)^2 \cdot SD^2}{effect + \delta}$
- effect=0, SD=9, δ=2
- $n = \frac{2 \cdot (1.645 + 0.84)^2 \cdot 9^2}{(0+2)^2} = 251$ in each treatment group

Protocol text

With 251 patients in each treatment group, a true difference between vortioxetine and agomelatine of 0 MADRS points, a non-inferiority margin of 2 MADRS points, a standard deviation of 9 MADRS points and a significance level of 5% the probability of demonstrating non-inferiority against agomelatine is approximately 80%