Classification of clinical trials

Plan of lecture

- Superiority equivalence inferiority trials
- Multicenter trials
- Pharmacokinetics
- Dose finding trials

" … the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis."

R. A. Fisher

The Design of Experiments, 1935

Efficacy

- What do we want to demonstrate, Better, equivalent to or not worse than?

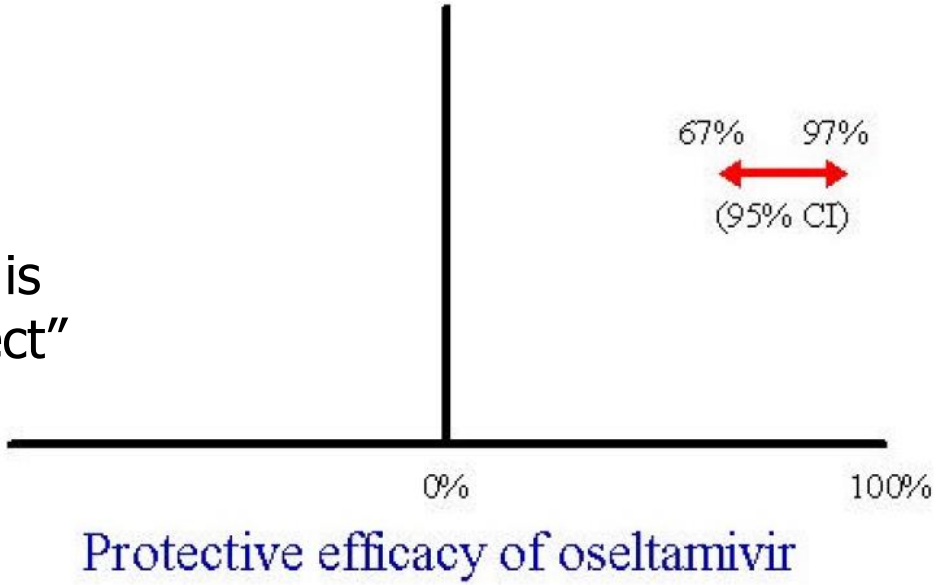
Better: Superiority

- A clinical trial to show that Oseltamivir is superior to placebo in prevention of influenza (Welliver et al)
- Of 206 placebo subjects exposed to influenza virus, 26 (12.6%) developed clinical influenza
- In the Oseltamivir group, of 209 subjects exposed to influenza virus, only 3 (1.4%) developed clinical influenza.
- The conventional way of expressing a treatment effect with this type of endpoint is to quote the "**protective efficacy**" (by analogy with vaccine trials).

$$PE = (UV-V)/UV$$
$$= (12.6-1.4)/12.6$$
$$= 89\%$$

Superiority

- So in this study the protective efficacy is 89%, with the 95% confidence interval ranging from [0.67; 0.97].
- If the drug did not work, we would expect a protective efficacy of 0%.
- The entire confidence interval is much higher than the "no-effect" value of 0%.
- Had the confidence interval contained 0%, we could not have been sure that the true effect was different from 0%.



Superiority

One-sided hypothesis

Efficacy of experimental treatment T is greater than that of current treatment C (placebo or active control)

$$H_0 : \mu_T \leq \mu_C \quad \text{vs.} \quad H_1 : \mu_T > \mu_C$$

We want to be able to reject H_0 to demonstrate superiority (efficacy)

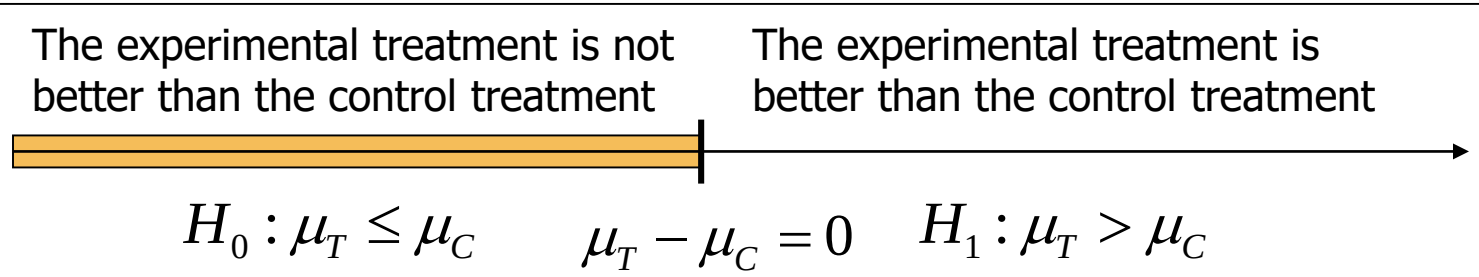
Superiority

Hypothesis: We want to demonstrate that the experimental treatment (T) is better than the control treatment (C).

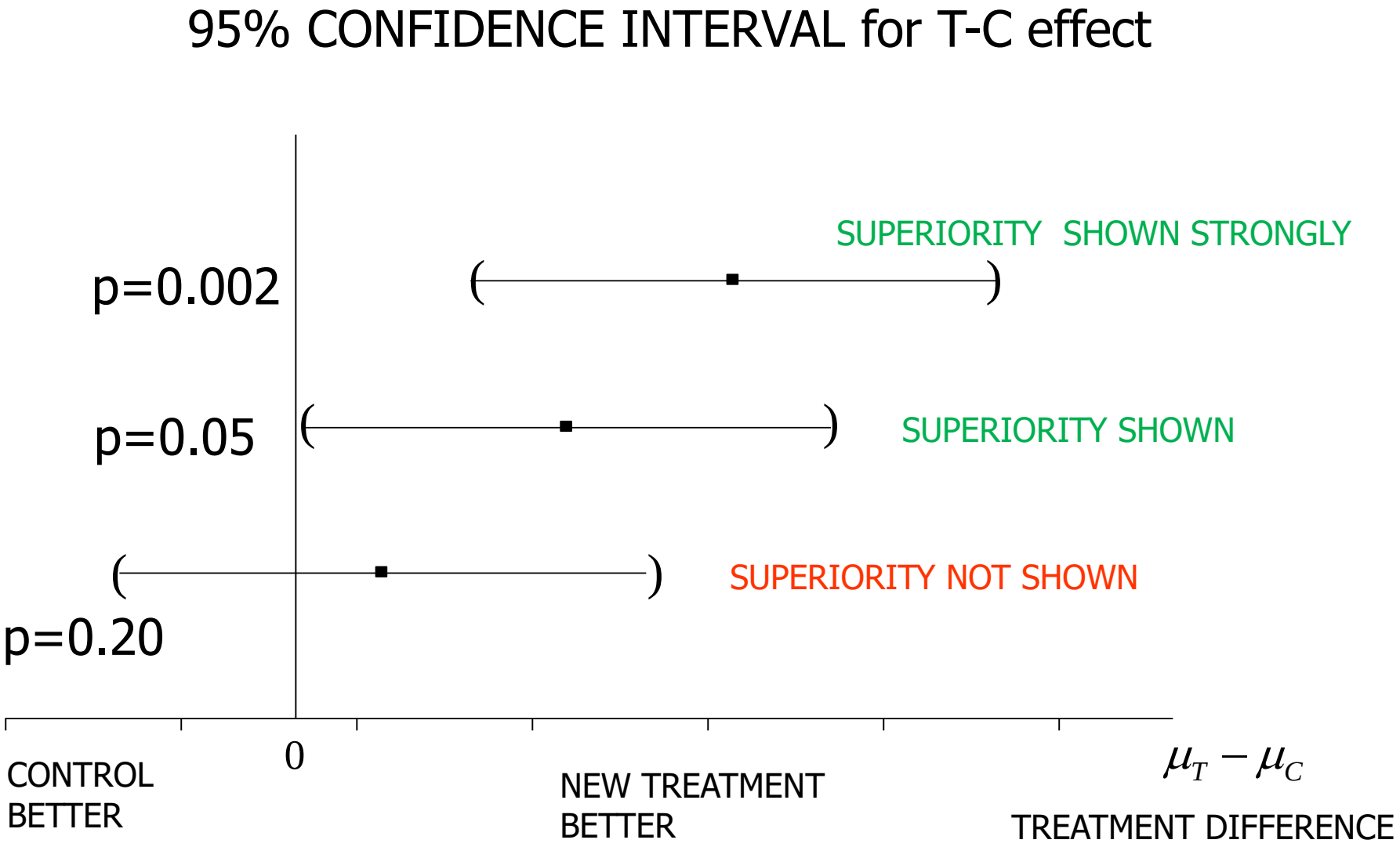
Test: $H_0 : \mu_T \leq \mu_C$ against $H_1 : \mu_T > \mu_C$ (One-sided)

Test based on: TREATMENT DIFFERENCE $\mu_T - \mu_C$

Test statistic based on: $\hat{\mu}_T - \hat{\mu}_C$ $(\text{e.g. } \hat{\mu}_T - \hat{\mu}_C = \overline{X}_T - \overline{X}_C)$



Superiority



Equivalence

- In some clinical trials, we do not expected that the new treatment will be superior to the existing standard. It may be realistic only to expect that a new treatment is "equivalent" to an existing established one, and it may be the objective of a clinical trial to provide adequate evidence of such equivalence.
- The standard method requires that a definition of equivalence should be stated in advance for the two treatments that are to be compared.
- This means setting numerical limits for the allowable difference between two treatments, such that any difference within these limits would be accepted by the clinical community as indicating that the two treatments produce essentially the same clinical effect.

What is similarity?

Due credit to Chris Miller AstraZeneca biostatistics USA

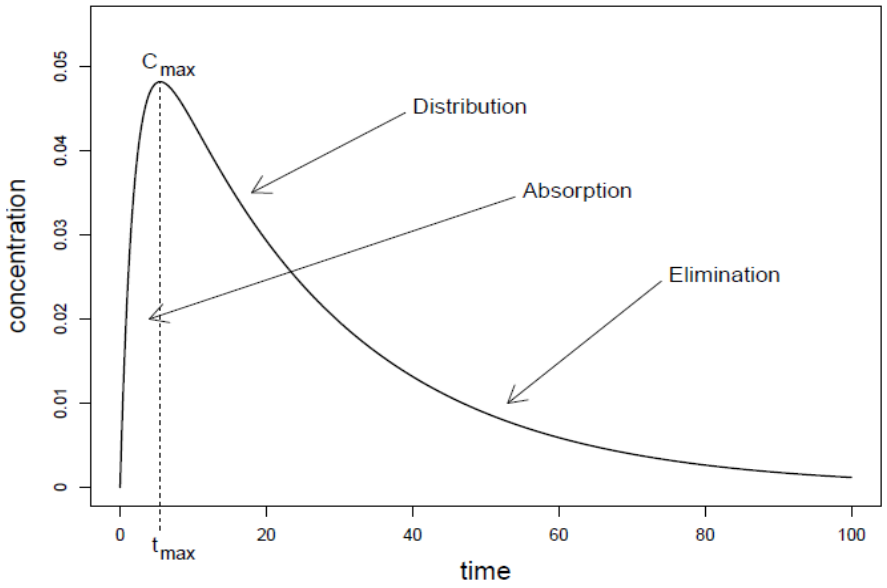
Example:

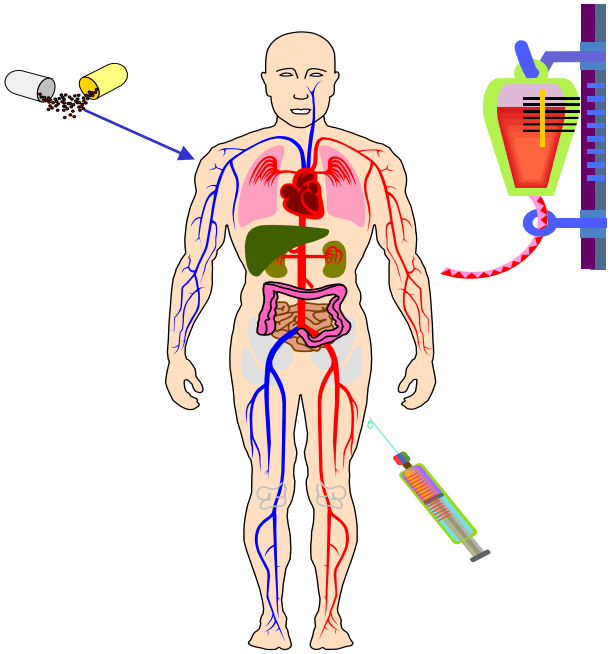
We have developed a new formulation (tablet) for our old best selling drug. How do we prove that the new formulation has the same effect as the old one without new big studies?

Bioequivalence

Pharmacokinetics

- **Pharmacokinetics** is the knowledge about **what happens to** a substance (e.g. **drug**) **after administration to the body**. It consists of absorption, distribution, metabolism and excretion, often abbreviated as **ADME**.
- To get an overview of this process, the concentration of the drug in blood plasma is observed over time.
- Compartment models are used to model such **concentration curves** and estimate the its parameters.





Bioequivalence

FDA definition:

> Pharmaceutical equivalents whose rate and extent of absorption are not statistically different when administered to patients or subjects at the same molar dose under similar experimental conditions

Operationalized as:

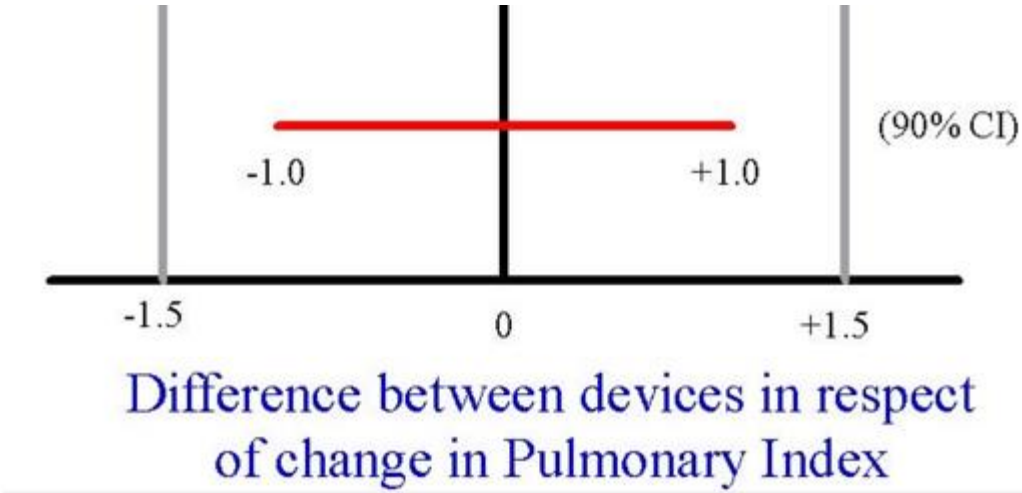
> Compare exposure in terms of AUC and Cmax of the plasma concentration vs time curve and conclude bioequivalence if the confidence interval for the ratio between the formulations lies between 0.8 and 1.25 for both AUC and Cmax.

Example

- Th Pulmonary Index (PI) is a simple score that is easily derived from clinical observation.
- The PI was derived from respiratory rate, wheezing, inspiratory-expiratory ratio, and use of accessory muscles.
- The PI usually correlates significantly with forced vital capacity ratio (FVC).
- The PI also correlates significantly with all common tests of pulmonary function.

Example

- A pediatric Asthma trial was planned to establish that an inhaler device (e.g. DPI: Dry Powder Inhaler) was clinically equivalent to a Nebulizer in treatment of children with recurrent wheezing.
- The primary endpoint in the study was a measure of lung function known as the Pulmonary Index (PI).
- Before the trial started, it was specified that if the difference between the two groups in in terms of PI was within ±1.5 units, then the two treatments would be considered clinically equivalent.
- In fact, in the final data analysis, the 90% confidence interval for the difference between the two groups in respect of PI change was [-1.0;+1].
- It was concluded that it is unlikely that the two devices really do differ by more than 1.5 units; hence, clinical equivalence has been adequately demonstrated.



- Note that when establishing equivalence, 90% confidence intervals are usually used, whereas in demonstrations of superiority, a 95% interval is more common.
- However, this convention is arbitrary and is subject of some debate amongst medical statisticians.

Equivalence

Assume we want to show that the efficacy of the new treatment T is similar to that of the active control treatment C. Due to the falsification principle, we need to formulate this hypothesis as our alternative hypothesis:

$$\mathbf{H_0}: \mu_T \neq \mu_C \qquad \mathbf{H_1}: \mu_T = \mu_C$$

T is similar to C

Opposite of how we usually formulate our hypotheses

- We need to specify "how close is close enough" to be considered the same? i.e. equivalence margin.
- Largest difference judged to be clinically acceptable, ignorable, or irrelevant

Equivalence

- Within a certain margin d, the efficacy of the new treatment T is similar to that of the active control treatment C.

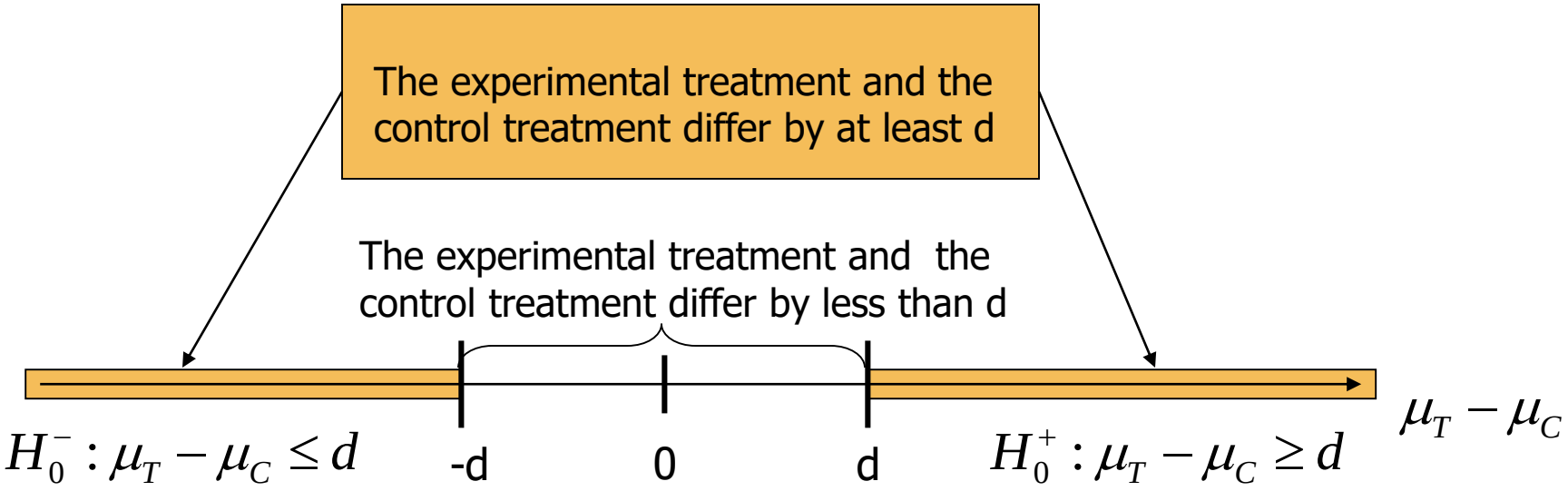
Ho: $|\mu_E - \mu_C| \geq d$ vs. **H1**: $|\mu_E - \mu_C| < d$

– We rewrite H_0 as:

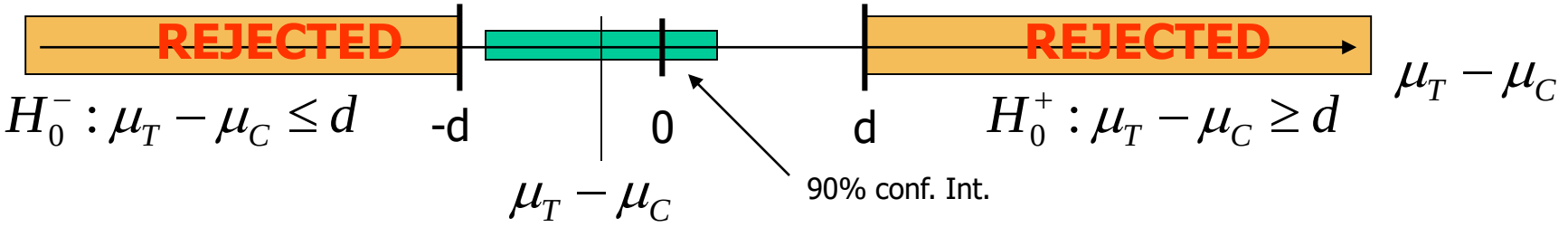
$$H_0^- : \mu_T - \mu_C \leq -d \quad \text{and} \quad H_0^+ : \mu_T - \mu_C \geq d$$

Two one-sided tests (TOST): need to reject both to reject H_0.

Equivalence

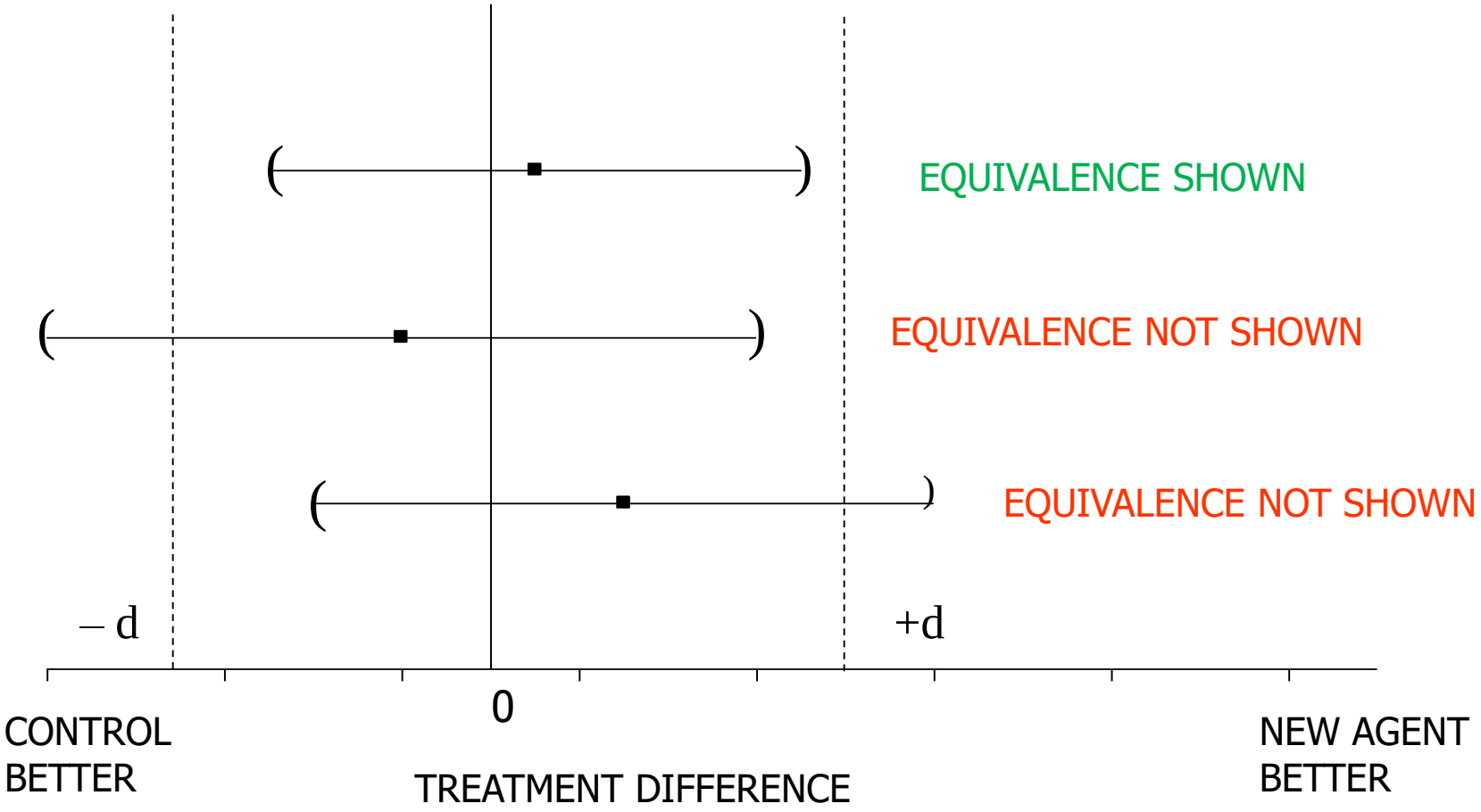


The combined null hypothesis H_0 can be tested at level α by testing each of the two <u>disjoint</u> components also at level α.



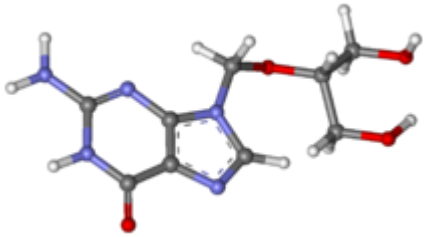
Equivalence

90% CONFIDENCE INTERVAL for E-C effect

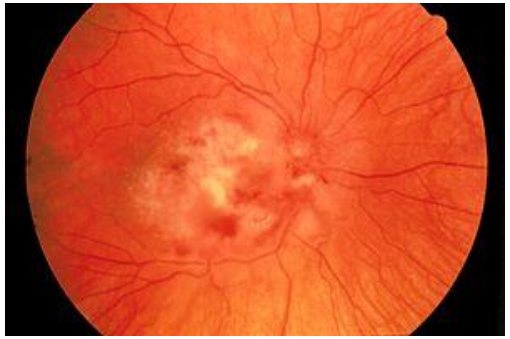


Non-inferiority: An example

- Cytomegalovirus Retinitis, also known as CMV Retinitis, is an inflammation of the retina of the eye that can lead to blindness.
- Caused by human Cytomegalovirus, it occurs predominantly in people whose immune system has been compromised, 15-40% of those with AIDS.
- There are different types of retinitis.



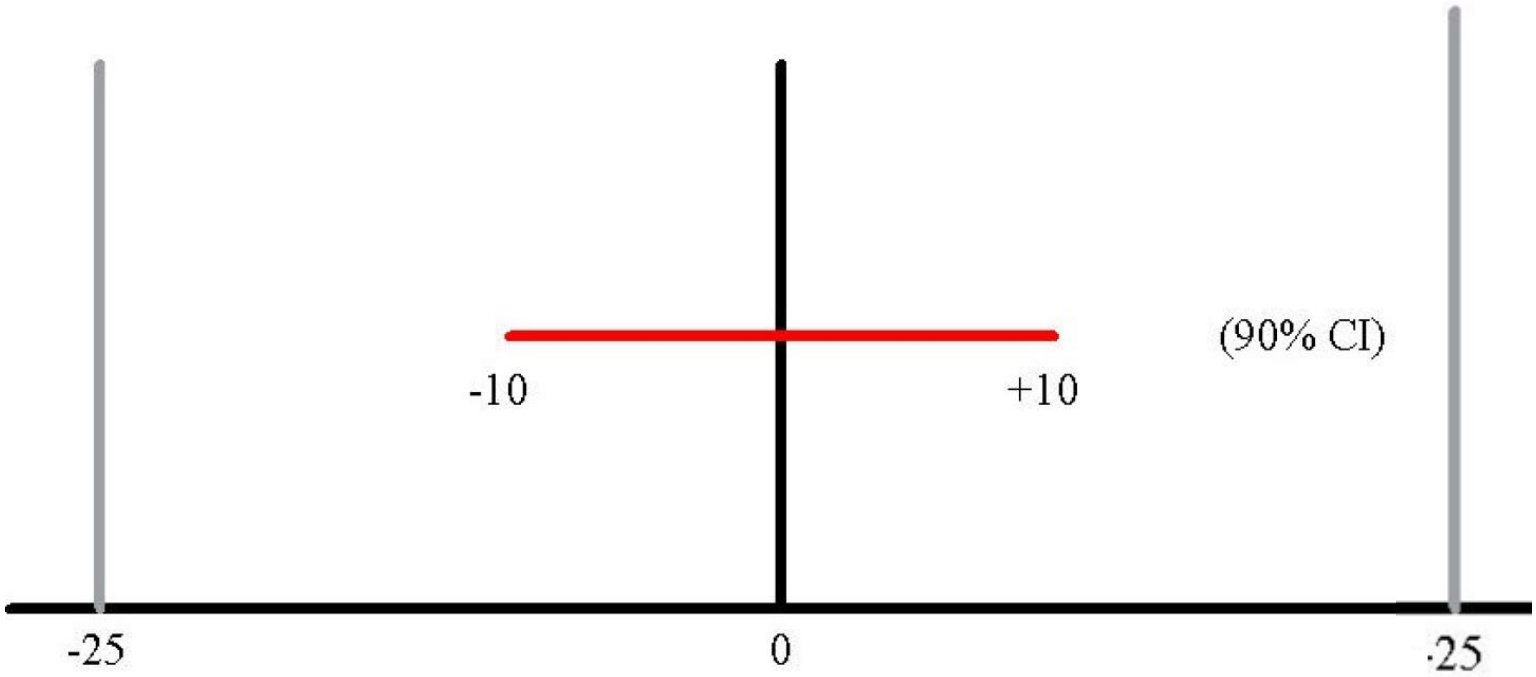
Ganciclovir



Non-inferiority: An example

- A new drug, Valganciclovir, was not expected to perform better than the existing standard Ganciclovir, but it was hoped to demonstrate that the new drug was not clinically inferior.
- Clinicians determined in advance that, as long as the proportion of patients who showed progression of CMV retinitis on Valganciclovir was not more than 25 percentage points worse than on Ganciclovir, then Valganciclovir could be regarded as no worse than (i.e. not inferior to) Ganciclovir in clinical practice in this indication.
- The 90% confidence interval for the difference between these two proportions is [-10%; +10%].

- Since 10 percentage points is not less than the pre-defined lower bound of -25 percentage points, it was concluded that Valganciclovir is not inferior and this conclusion was accepted by the FDA Antiviral Drugs Advisory Committee.

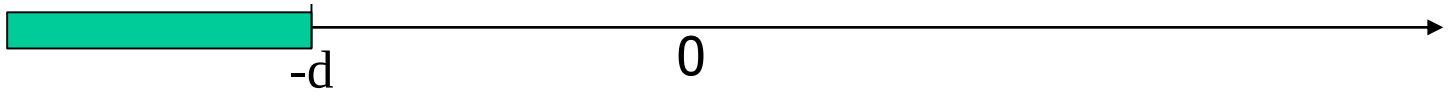


Non-inferiority

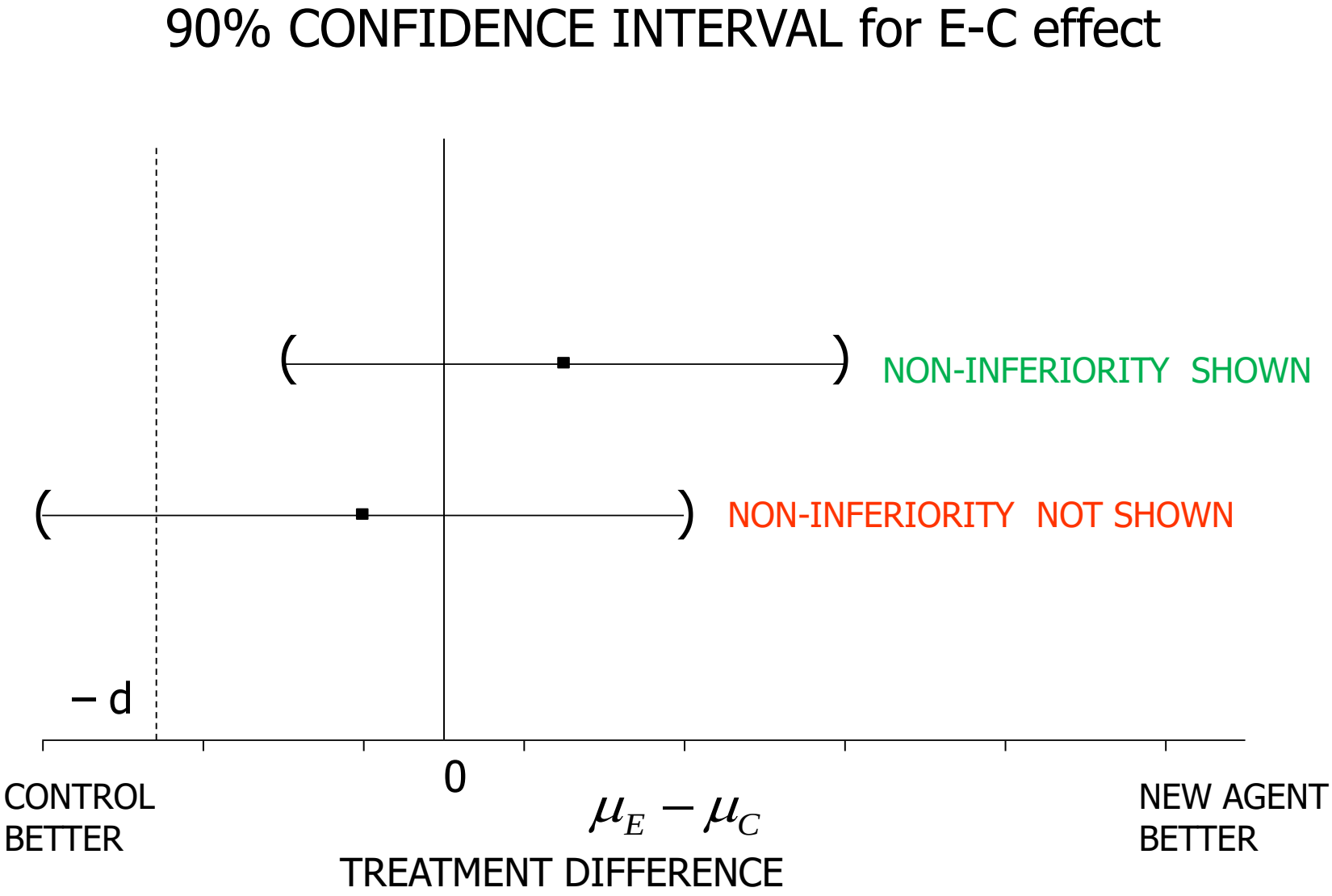
One-sided hypothesis

Within certain margin, efficacy of experimental treatment E is at least as good as that of active control treatment C.

Ho: $\mu_E - \mu_C \leq -d$ vs. H1: $\mu_E - \mu_C > -d$



Non-inferiority



Superiority, equivalence and non-inferiority

Experimental treatment with true mean effect: μ_T

Control treatment with true mean effect: μ_C

Superiority: The experimental treatment is better than the control treatment.

$$H_0 : \mu_T \leq \mu_C$$
$$H_1 : \mu_T > \mu_C$$

Equivalence: The experimental treatment and the control treatment are similar.

$$H_0 : \left|\mu_T - \mu_C\right| \geq d$$
$$H_1 : \left|\mu_T - \mu_C\right| < d$$

Non-inferiority: The experimental treatment is not that much worse than the control treatment.

$$H_0 : \mu_T \leq \mu_C - d$$
$$H_1 : \mu_T > \mu_C + d$$

Multicenter trials

All large studies are conducted at multiple centers.

Center=clinic=study site

Issues:

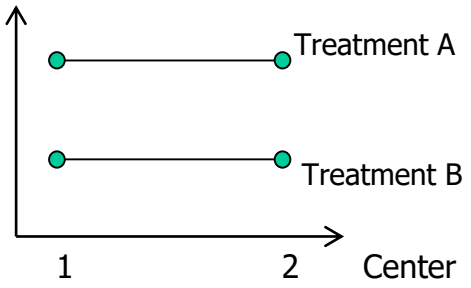
- Treatment by center interaction*
- Estimation of treatment effect.
- Randomization

*) partly Covered in the lecture on basic statistical concepts

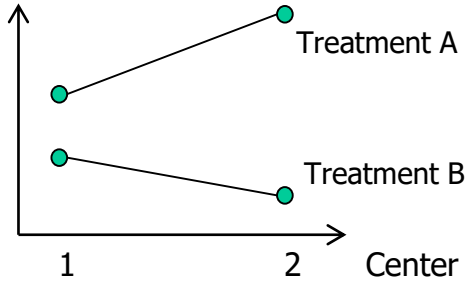
Multicenter trials: interaction

- Is to be expected but difficult to detect

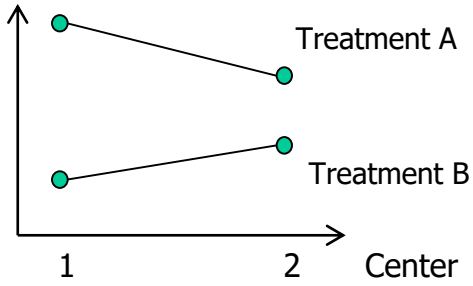
No interaction



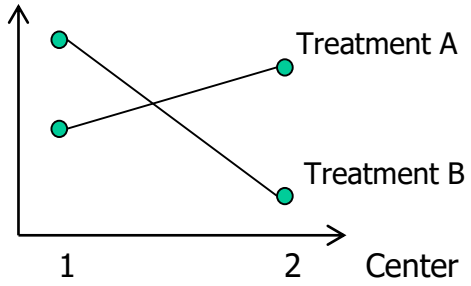
Quantitative interaction



Quantitative interaction



Qualitative interaction



- Quantitative (i.e., same direction only magnitude differs) very common
- Qualitative (i.e., treatment shows benefit in some centers , Placebo shows benefit in others) less common
- Qualitative interaction is of concern but not found very often and hard to establish

Multicanter trial: statistical model

$$y_{ijk} = \mu + \alpha_i + \tau_j + (\alpha\tau)_{ij} + \varepsilon_{ijk} \qquad \varepsilon_{ijk} \sim N(0, \sigma)$$

Obs_{ijk}=grand mean + $center_i$+$treatment_j$+$(center*treatment)_{ij}$+$error_{ijk}$

Alternatives:

- Fixed: Center and treatment*center interaction are fixed effects
- Random: Center and treatment*center interaction are random effects

Fixed:

- Gives a precise answer to a fairly well defined question, does the drug work for patients at **these** centers?
- The only option for a single center study
- Centers are carefully chosen, not randomly
- The definition of center is arbitrary

Random:

- We want to say something about patients in general and this is the best shot at this difficult question
- Wider confidence interval reflects the true uncertainty of centers are really different.
- Allows prediction of the effect in one specific center using information from all centers.

Multi center trials: interaction

ICH E9:

- Test main effect first , if significant:
- Test interaction as an exploratory analysis
- If there are a large number of centers, it is less important to consider interaction.

Sometimes poeple suggest to pool small centers.

Easy to see why (more patients/center) but what does it mean??

Multicenter trials

Assume k centers with

True treatment effect: τ_i

Estimated treatment effect: $\hat{\tau}_i$

Variance of estimated treatment effect: σ_i^2

Overall treatment effect estimated by $\hat{\tau} = \sum_{i=1}^{k} w_i \hat{\tau}_i$ where $\sum_{i=1}^{k} w_i = 1$

Type II estimator: $w_i = \dfrac{1/\sigma_i^2}{\sum_{i=1}^{k} 1/\sigma_i^2}$

Treatment effects averaged over center weighted according to precision (think n_i).

Type III estimator: $w_i = 1/k$

Treatment effects averaged over center with equal weight for all centers

Multicenter trials

Type II: $E[\hat{\tau}] = E\left[\dfrac{\sum_{i=1}^{k} \hat{\tau}_i / \sigma_i^2}{\sum_{i=1}^{k} 1/\sigma_i^2}\right] = \dfrac{\sum_{i=1}^{k} n_i \tau_i}{\sum_{i=1}^{k} n_i}$ $\qquad Var(\hat{\tau}) = \dfrac{4\sigma^2}{\sum_{i=1}^{k} n_i}$

Assuming (true) within center variance equal in all centers.

Type III: $E[\hat{\tau}] = \dfrac{1}{k}\sum_{i=1}^{k} \tau_i = \bar{\tau}$ $\qquad Var(\hat{\tau}) = \dfrac{1}{k^2}\sum_{i=1}^{k} \sigma_i^2$

PK/PD

- Pharmacokinetics: What happens to the drug?
- Pharmacodynamics
  - What happens to the body?
  - y=Response=f(dose)

PK: Compartment models

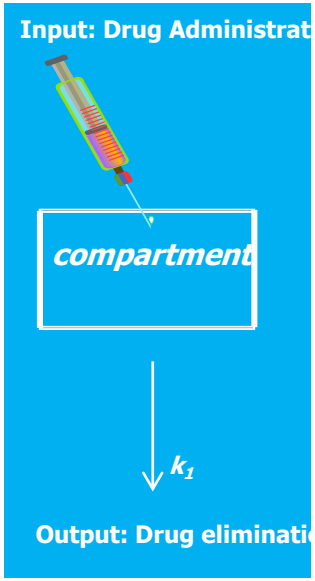
- A simplified way to understand what happens when a drug administered is to view the body as a system with a few compartments each of which representing is a different part of the body (bloodstream, liver, kidneys, tissues etc).
- Compartment models have many applications in physics, chemistry etc.
- In its simplest form the whole body is looked upon as one single compartment where only how the drug is absorbed into and eliminated from the body is considered.
- In more complex systems the rates for how to the drug is transported between different compartments are also taken into account.

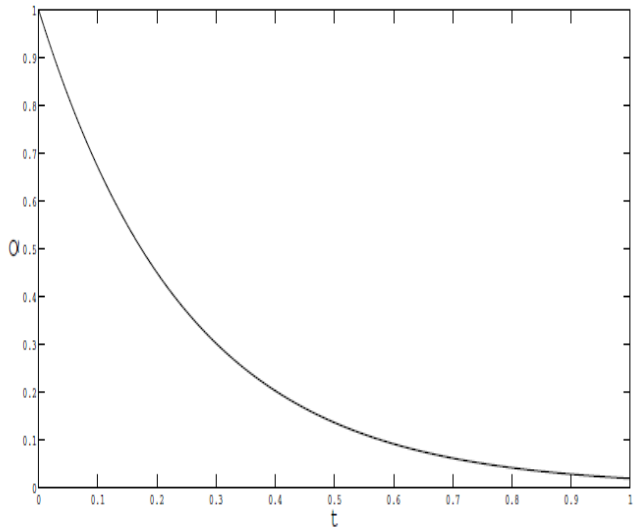
Compartment models: ODE

$$\frac{dA_1}{dt} = -k_1 A_1(t);\ A_1(0) = D$$

- $A_1(t)$: *amount of drug at the site of administration,*
- D *: dose of drug,*
- k_1 = *elimination rate.*

ODE Solution: $A_1(t) = De^{-k_1 t}$





Various types of mathematical models

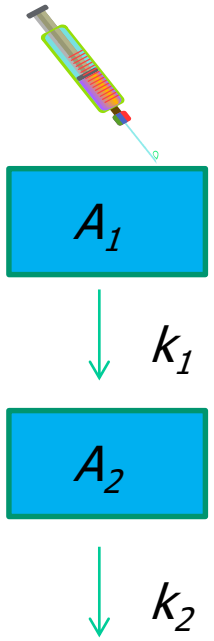
- Ordinary ODE
- Stochastic SDE
- Delay DDE
- Partial PDE
- Fredholm FIE
- Integro-diffetential IDE
- Markov processes MP

Compartment models: ODE

$$\frac{dA_1}{dt} = -k_1 A_1(t);\ A_1(0) = D$$

$$\frac{dA_2}{dt} = k_1 A_1(t) - k_2 A_2(t);\ A_2(0) = 0$$

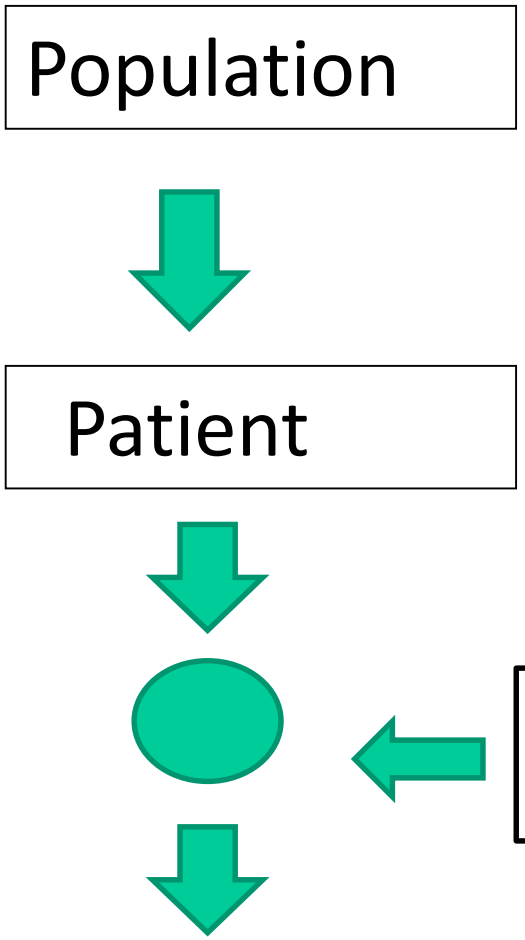
- $A_1(t)$: *amount of drug at the site of administration,*
- $A_2(t)$: *amount of drug in the compartment (blood),*
- D *: dose of drug,*
- k_1 = *absorption rate,*
- k_2 = *elimination rate.*



ODE Solution: $A_1(t) = De^{-k_1 t}$

$$A_2 = \frac{k_1 D}{k_1 - k_2}\left(e^{-k_2 t} - e^{-k_1 t}\right)$$

Types of variability



- Measurement errors
- Between-patient variability (population variability)
- Within-patient variability (intrinsic randomness, processes inside patient's body)

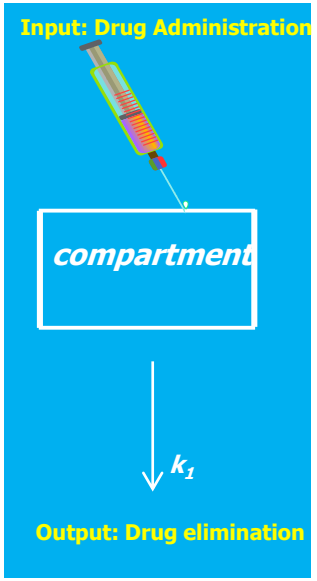
Instrumental error

Non-linear mixed models

Compartment models: SDE

$$dA_1 = -k_1 A_1(t)dt + \sigma_w dw(t);\ A_1(0) = D$$

- $A_1(t)$: *amount of drug at the site of administration,*
- D *: dose of drug,*
- k_1 = *elimination rate*
- A *state variable*
- *{w(t)} Wiener process*
- σ_w *diffusion term*

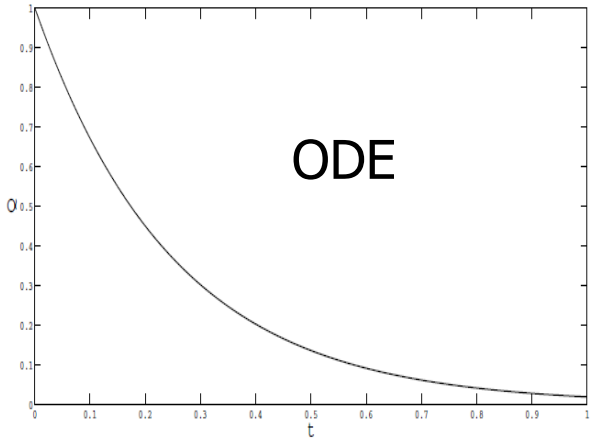


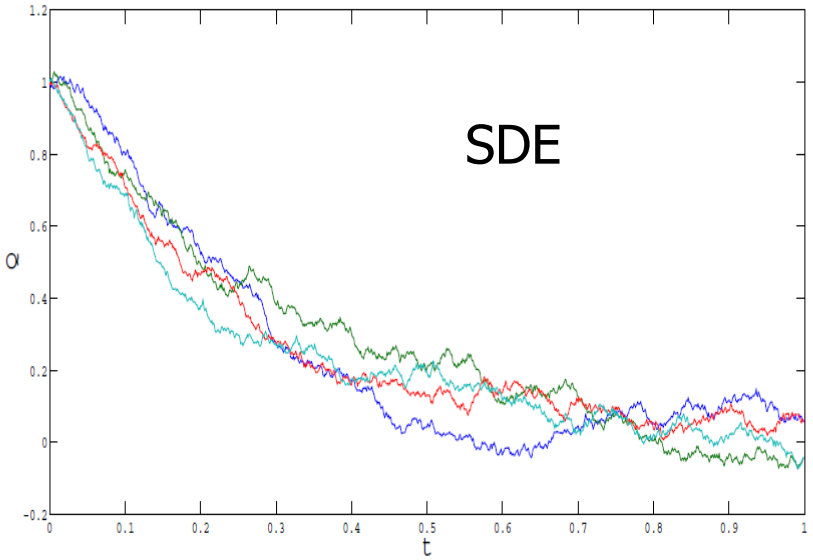
ODE Solution:

$$A(t) = A(0)e^{-kt} + \int_0^t \sigma_w e^{-k(t-s)} dw(t)$$

This equation should be interpreted as an informal way of expressing the corresponding integral equation (cf. Øksendal).

Compartment models: SDE



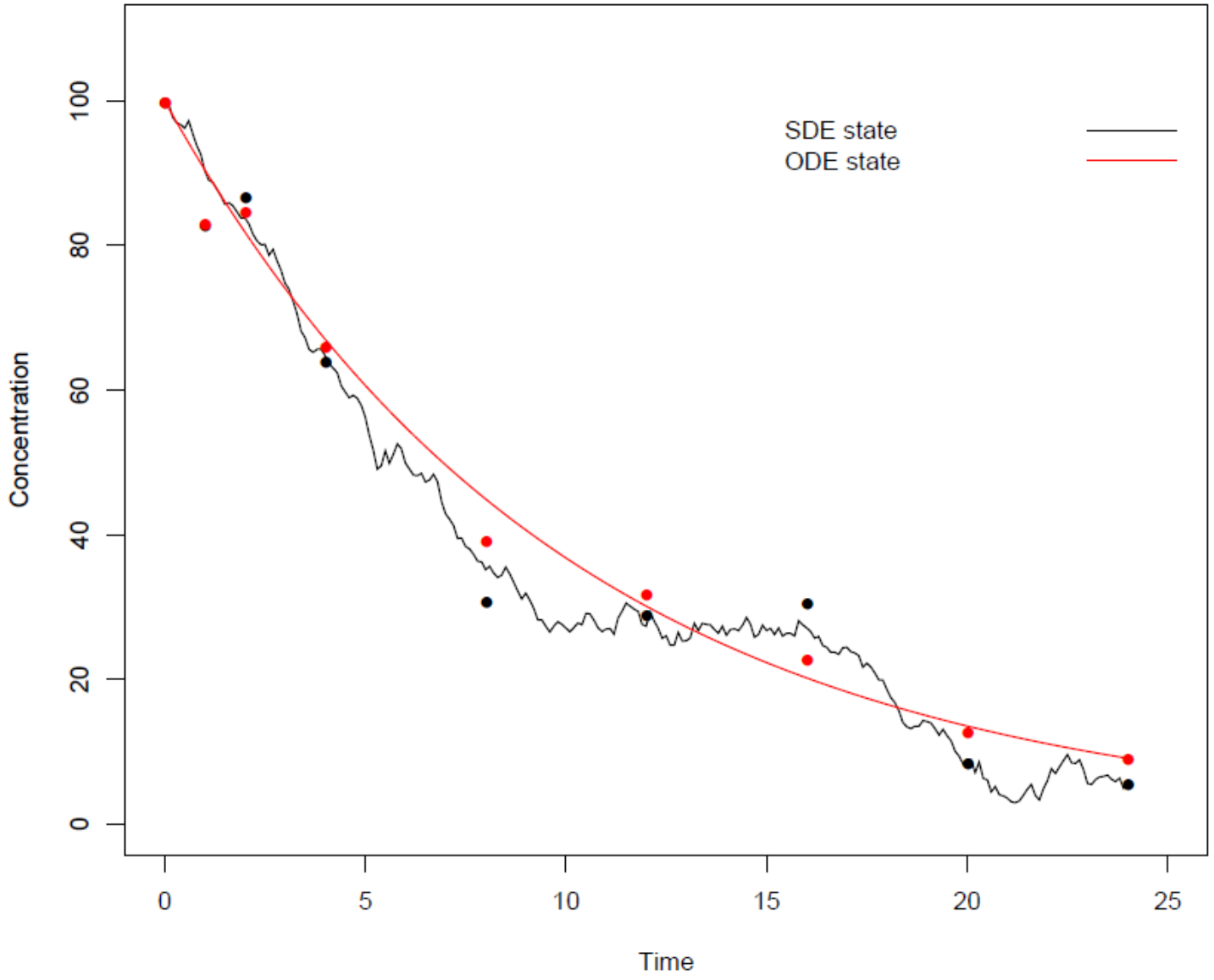


$$dA = -kAdt$$

$$A(t) = A(0)e^{-kt}$$

$$dA = -kAdt + \sigma_w dw(t)$$

$$A(t) = A(0)e^{-kt} + \int_0^t \sigma_w e^{-k(t-s)} dw(t)$$


SDE state
ODE state
Concentration
Time
0
20
40
60
80
100
0
5
10
15
20
25

PD: Dose response trial analysis options

- Typically Phase II are dose finding trials with many arms (doses). Most common approaches for analysis:

- **Pairwise comparisons:** The doses are compared using significance tests often adjusted for multiple comparisons and the aim is to show effect vs the comparator and to separate the doses.
    - Limited assumptions
    - Easy to compare doses
    - Need relatively many observations
    - No estimate of a dose reponse curve
- **Model based:** The effect is assumed to follow a parameteric model with parameters estimated from the data.
    - More assumptions
    - Tricky to compare doses
    - Need relatively few observations
    - Estimates a dose reponse curve

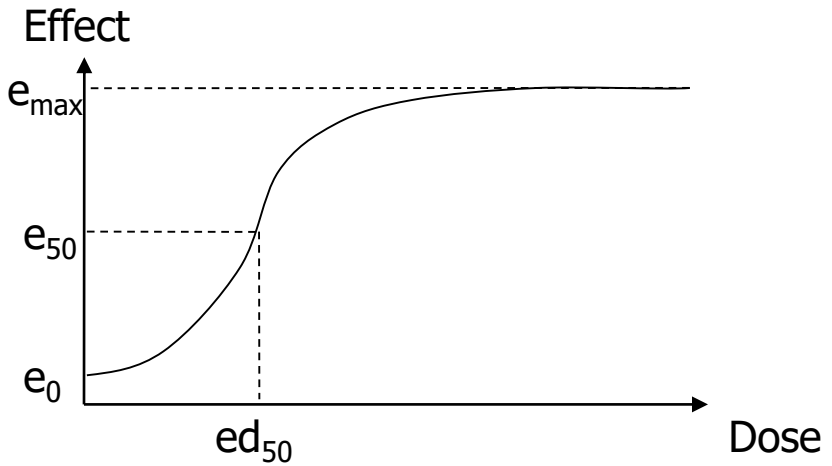
Dose Response trials models

Separate means with equal variance

$$y_{ij} = \mu_i + \varepsilon_{ij} \;\; ; \;\; \varepsilon_{ij} \text{ iid } N\left(0, \sigma^2\right)$$

Non-linear Regression model

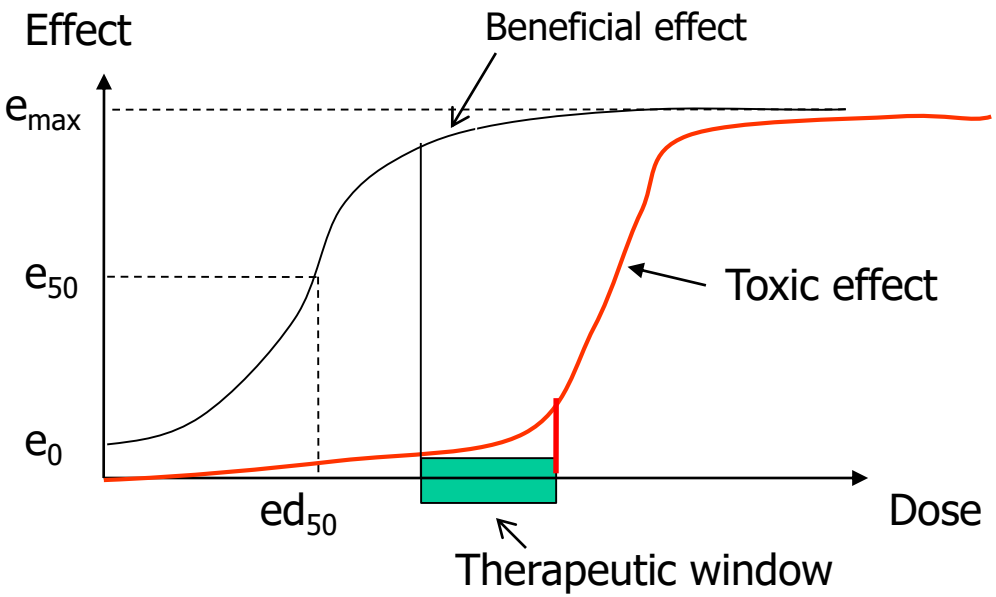
$$E = E_0 + \frac{E_{\max} \times d^{\gamma}}{ed_{50}^{\gamma} + d^{\gamma}}$$



Dose Response trials objectives

Objectives:
(ICH E9)

- Confirm efficacy
- Investigate shape of dose reponse curve
- Estimate an appropriate starting dose
- Determination of minimal effective dose
- Safety!



Chapter 7 Reading instructions

- 7.1 Introduction
- 7.2 Multicenter Trials: Read + extra material
- 7.3 Superiority Trials: Read
- 7.4 Equivalence/Non-inferiority Trials: Read
- 7.5 Dose Response Trials: Read
- 7.6 Combination Trials: Read
- 7.7 Bridging Trials: Skip
- 7.8 Vaccine Trails: Skip