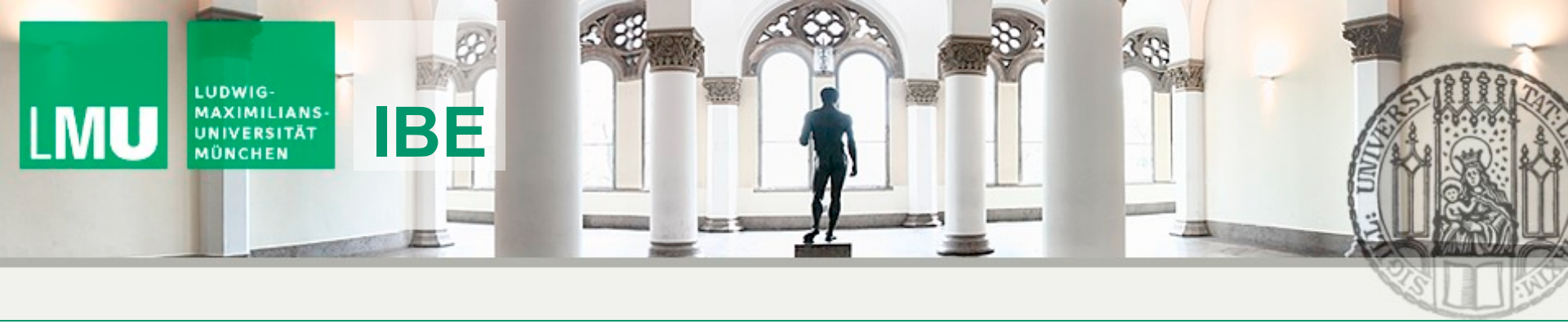


Research skills: therapy studies

Alexander Crispin, MD, MPH

cri@ibe.med.uni-muenchen.de

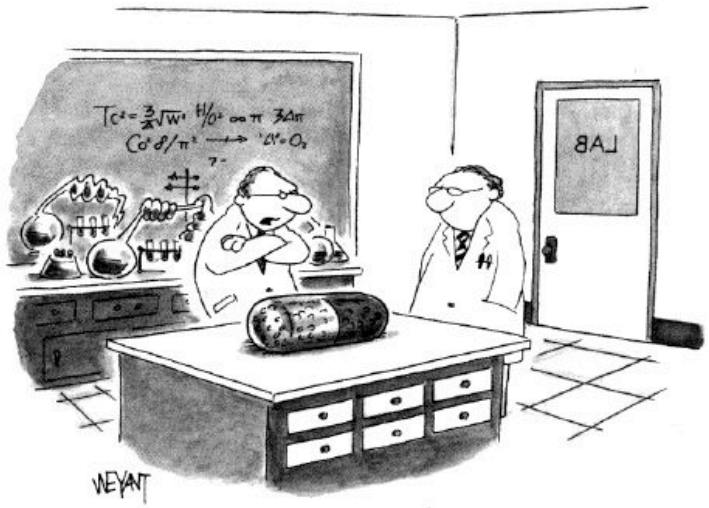
Department of Medical Informatics, Biometry, and Epidemiology

LMU – University of Munich

Course overview

Date and Time	Lecturer	Subject
Fri, 10/28/2011: 4-6 pm	Crispin	Introduction
Thu, 11/17/2011: 4-5 pm	Hasford	Developing a research question, ethics Developing a research question, data protection
Fri, 11/25/2011: 4-6 pm	Geldmacher	Literature research and management
Fri, 12/02/2011: 4-6 pm	Kalies, Grote	How to read a paper
Fri, 12/09/2011: 2-4 pm	Grill	How to write a protocol
Fri, 12/09/2011: 4-6 pm	Bodechtel	Questionnaire development, surveys
Thu, 12/15/2011: 2-4 pm	Rehfuess	How to write a paper
Thu, 12/15/2011: 4-6 pm	Strahwald	How to present scientific results
Fri, 12/16/2011: 4-6 pm	Wildner	Communicating scientific results to general public

Outline



"That's just great. I discover the cure for the common cold and all you can do is criticize."

- Evidence-Based Medicine
- Bench research vs. clinical research
- Clinical study designs
- Important methodological principles
- Typical flaws of clinical trials

Evidence-Based Medicine (David Sacket et al, BMJ, 1996)

- Conscientious, explicit, and judicious use of **current best evidence** in making decisions about the care of individual patients.
- The practice of EBM means integrating individual **clinical expertise** with the **best available external clinical evidence** from **systematic research**.

Levels of evidence (Oxford Centre for EBM)

Level	Evidence
1a	Systematic review (with homogeneity) of randomized controlled trials (RCT)
1b	One RCT with narrow confidence interval
1c	All or none
2a	Systematic review (with homogeneity) of cohort studies
2b	One cohort study or low quality RCT
2c	Outcomes research, ecological studies
3a	Systematic review (with homogeneity) of case-control studies
3b	Individual case-control study
4	Case-series (and poor quality cohort and case-control studies)
5	Expert opinion based on physiology, bench research, or "first principles"

"All or none": deterministic events

Smith GCS, Pell JP.

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials.

BMJ 2003; 327: 1459-1461

Grades of recommendation (Oxford Centre for EBM)

Grade	Basis
A	• Consistent level 1 studies
B	• Consistent level 2 or 3 studies • Extrapolations from level 1 studies
C	• Level 4 studies • Extrapolations from level 2 or 3 studies
D	• Level 5 evidence • Inconsistent or inconclusive studies of any level

EBM: what it ‘s not

Misunderstanding	Refutation
Everybody does it.	Difficulties of clinicians keeping abreast of medical advances reported in primary journals
EBM is medicine for ivory towers and armchairs.	Audits from the front lines of clinical care where inpatient clinical teams have provided evidence-based care to the vast majority of their patients
EBM is cookbook medicine.	EBM requires integration of best external evidence, individual expertise, and patients' choice. This cannot result in slavish cookbook approaches.
EBM will be hijacked by purchasers and managers to cut the costs of health care.	Evidence-based therapy may be more expensive than standard care.
EBM is restricted to RCT and meta-analyses.	If there is no level 1a evidence, one has to aim lower.

Bench research and clinical research

	Bench research	Clinical research
Methodology	**Natural science** and **animal experiments**	**Human experiments**
Strengths	**Rational, systematic, efficient** development of new therapies	**Direct** empirical assessment of **efficacy** and **safety**
Weaknesses	**Empirical** observations may contradict **theoretical** predictions.	**Ethical problem**

Surprising findings of clinical research (examples)

- Multiple organ failure after first dose of monoclonal antibody TGN1412
- Sodium channel blockers reduce risk of ventricular tachycardia after myocardial infarction, yet increase mortality.
- Higher risk of myocardial infarction in patients treated with selective COX-2 inhibitor Vioxx®
- Viagra®, designed as a therapy for angina pectoris, is effective in patients with erectile dysfunction.

Pre-clinical and clinical research (USA)

Years	Phase		
5	Drug discovery		10,000 compounds
1.5	Pre-clinical		250
	Investigational New Drug Application		5
6	Clinical trials	Phase I	
		Phase II	
		Phase III	1.2
	New Drug Application		
2	FDA review		1 FDA-approved drug
?	Post-marketing surveillance	Phase IV	

Pre-clinical and clinical research: summary

- **R&D time** until approval: **10–12 years**
- **Candidate compounds** needed to produce one approved drug: **5,000-10,000**
- **R&D cost** for one approved drug: up to **US$ 800 million**
- **Bet on the future**:
  - Current drug sales have to finance future R&D with its uncertain outcomes.
  - So the new drug better be a **"blockbuster"…**

Clinical trial: definition (A. Hofman)

- Therapeutic **experiment** with **human** participants to assess the **efficacy** of treatments in **quantitative** terms.
- **Interventional cohort study** (follow-up study)

Good Clinical Practice (GCP)

International
ethical and **scientific quality standard**
for **trials** that involve
the participation of **human subjects**

Study Design
- Health authority approval
- Ethics committee approval

Trial conduct

Recording data

Reporting results

Phase I trials

- **First** testing in **human subjects**
- Small group (20 to 50) of **healthy volunteers**
- Research interest: **pharmacokinetics**, tolerability, safety
- Typical setting:
  - Inpatient clinic
  - 24/7 observation for several half-lives of the compound

Phase II trials

- **First** testing in **patients**
- Smaller group (20 to 300) of patients with the **intended indication**
- Interest: **pharmacodynamics**, tolerability, safety
  - Phase IIA: dosing requirements
  - Phase IIB: efficacy at the prescribed dose
- Some trials **combine phases I and II**
  - E.g. cancer chemotherapy trials
- **Trial designs**
  - Case series
  - Randomised clinical trials

Phase III trials

- **Randomised controlled trials** on **large patient groups** (300–3,000 or more)
  - Typically multicentre trials
- **Pivotal trials** = decision basis for **approval** by **regulatory authorities** (FDA, EMEA, BfArM, PEI,...):
  - **Definitive efficacy assessment**
  - **Regulatory submission** after satisfactory phase III trials
  - Special case: "label expansion"

Clinical trials and safety

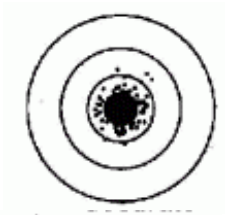
- Clinical trials are **excellent** for studying **intended effects** (reduction of morbidity or mortality).
- Clinical trials are **inefficient** for detecting **unintended effects:**
  - Every clinical trial collects data on:
    - Adverse events (AE)
    - Serious adverse events (SAE)
    - Suspected unexpected serious adverse reactions (SUSAR)
  - But even **major side-effects** may go **undetected**.
    - E.g. lethal complication with an incidence of 1 in 100.000 patients

Phase IV

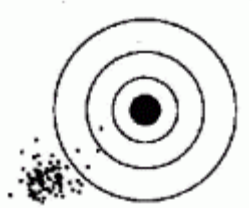
- **Post-marketing surveillance:**
  - **Safety** surveillance (pharmacovigilance)
    - Rare adverse effects
    - Long-term adverse effects
    - Interactions with other drugs
    - Adverse effects in populations not studied in phases I to III trials
  - Additional information on **benefits** and **optimal use**.
- Much **larger patient population** and **longer observation period** than in phases I to III

No empirical research without errors…

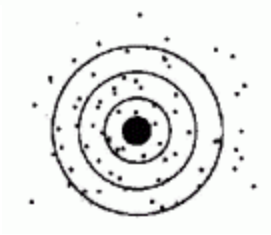
Truth



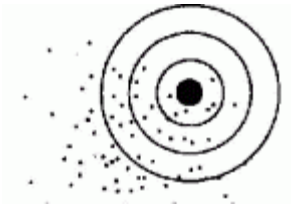
\+ Systematic error (bias)



\+ Random error



Study result



Systematic errors

- **Biased, invalid results**
- Only **limited control** during statistical **analysis**:
  - Multivariate methods
  - Sensitivity analysis
- Systematic errors must be **avoided**:
  - Adequate **study design!**
  - **Quality control!**
- Systematic errors are **always avoidable**.

Random error

- Results **unbiased**, yet **uncertain**
- **Unavoidable**
- **Inferential statistics: uncertainty** can be **quantified** using the empirical **sample data**
  - Standard errors
  - Confidence intervals
  - Hypothesis tests
- Uncertain knowledge + knowledge about the amount of uncertainty = **applicable knowledge**

Essential methods

- **Systematic error control**
  - Placebo or active control
  - Randomisation
  - Intention-to-treat (ITT) analysis
  - Follow-up of patients after protocol violations
  - Masking (blinding)
  - Relevant endpoints
- **Random error control**
  - Interval estimation of the therapy effect
  - Hypothesis tests
  - Sample-size estimation

Selected clinical study designs

- Uncontrolled therapy study
- Parallel group design
- Crossover design

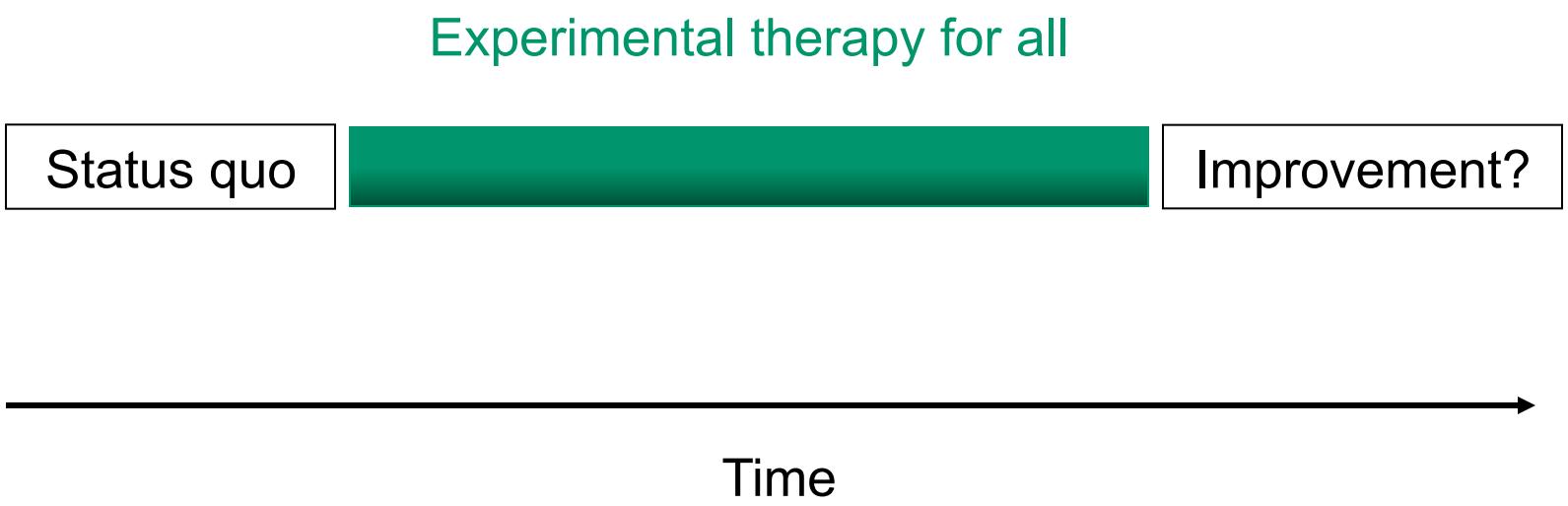
Components of the observed treatment effect

- Natural course of the disease
- Patient’s subjectivity
- Doctor’s subjectivity
- **Specific** (pharmacological) **effect**

Consequences

- In **phase III trials**, our one and only **scientific interest to prove** the **specific** (pharmacological) **effect**.
- We must make sure that:
  - Individual natural courses of disease are accounted for.
  - Patients' and doctors' attitudes do not influence the result.

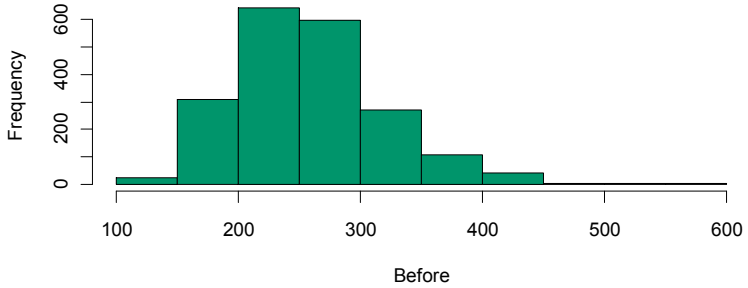
Flawed all-time classic: uncontrolled therapy study

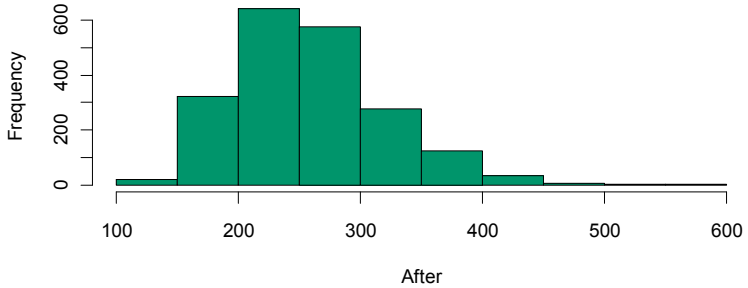


Uncontrolled studies: problem

- Clinical **improvement** is **no proof** of **efficacy**:
  - Spontaneous recovery?
  - Placebo effect?
  - Regression to the mean?
- Solution: comparison with **reference patients**
  - **Placebo control**
  - **Active control =** established standard therapy

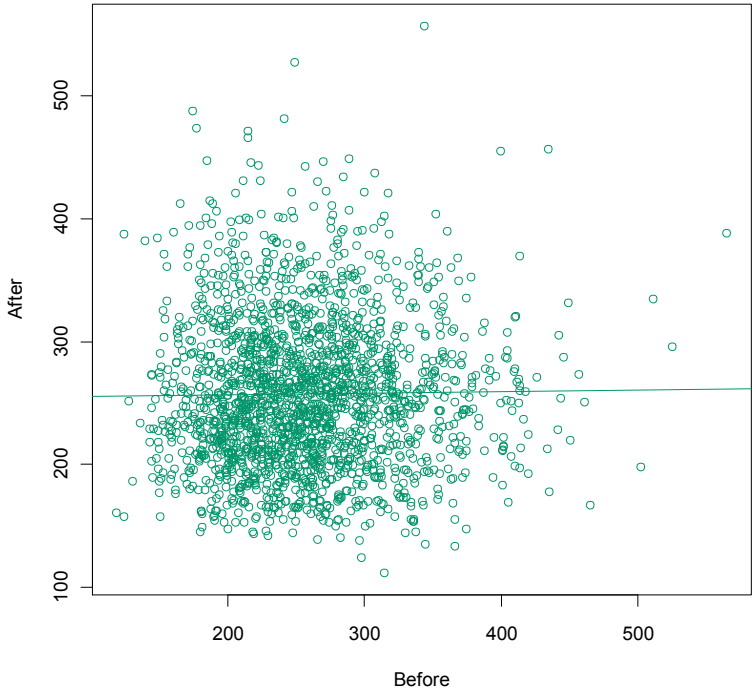
Regression to the mean: simulation study (I)





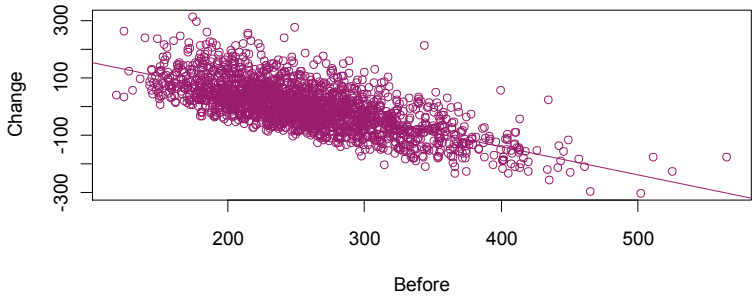
- The histograms depict total cholesterol values in mg/dl before and after therapy.
- These are simulated data:
  - Two vectors of $N(2.4,0.1)$-distributed random numbers
  - Exponentiation of the random numbers (base 10)
- Give or take some random variation, both distributions should be the same.

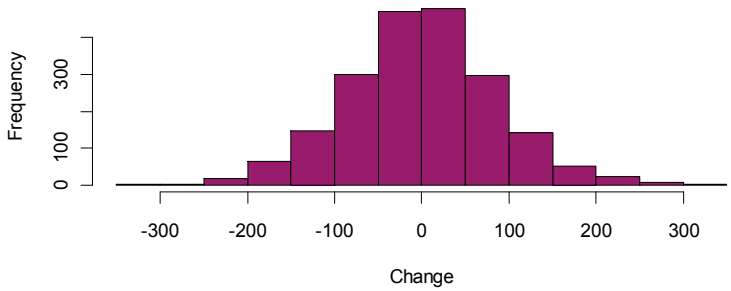
Simulation study (II): scatter-plot of values after vs. before



- Only random scatter.
- Slope of the regression line = 0 (give or take some random variation)
- (The slope is 0 because the simulated values are **uncorrelated**.)

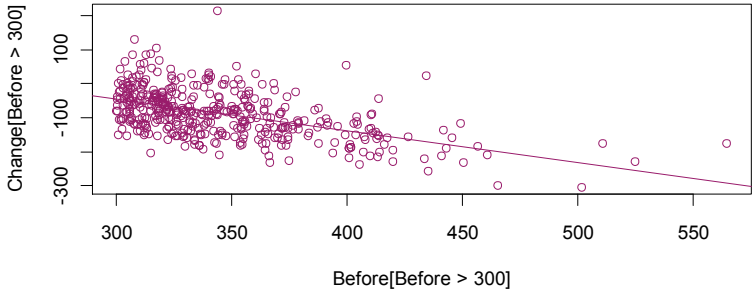
Simulation study (III): change after vs. before

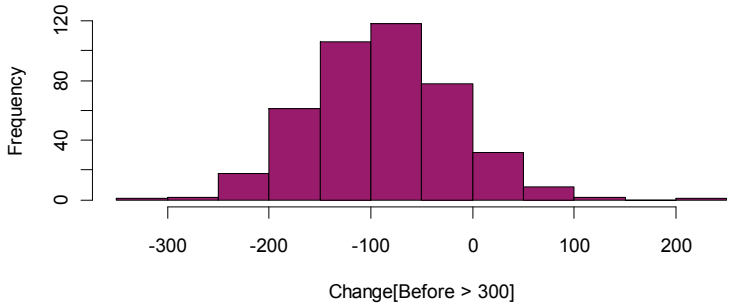




- Average change: 0
- Change as a function of initial values:
  - Low initial values → positive changes
  - High initial values → negative changes
- However, **nobody treats** patients with **low cholesterol levels**…

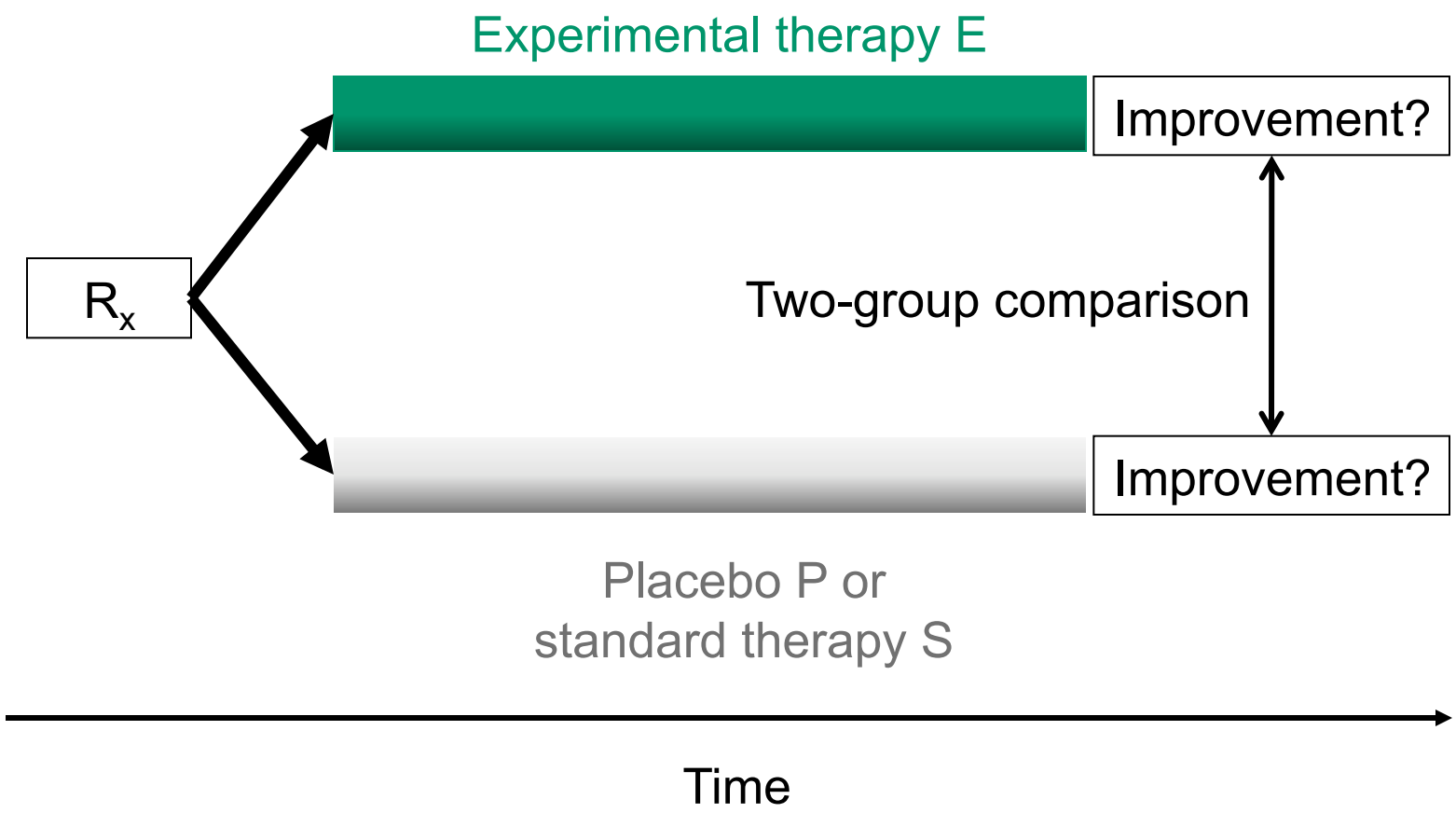
Simulation study (IV): restriction to initial values > 300 mg/dl



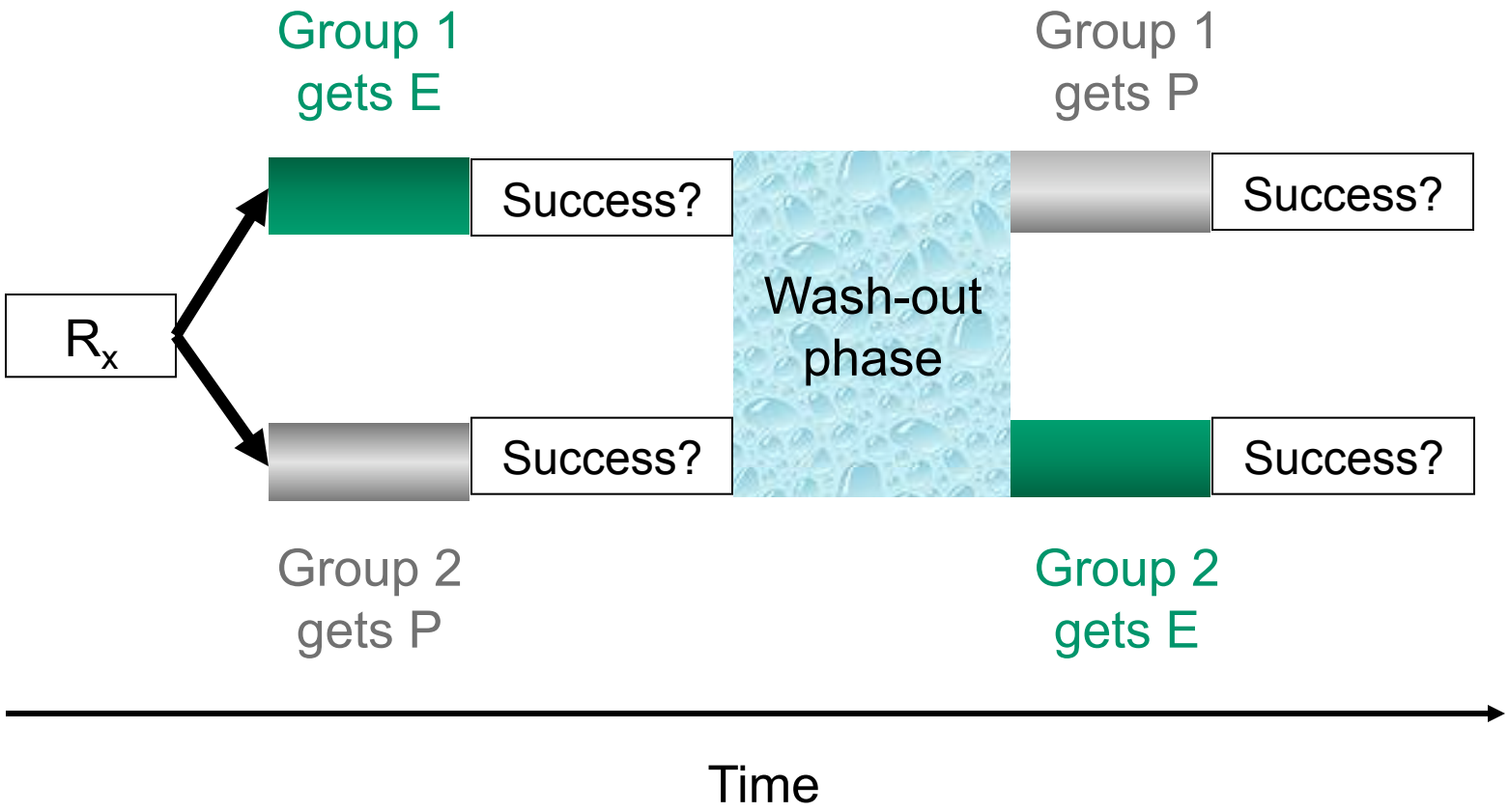


- The higher the initial value, the greater the decrease.
- Only few increases.
- Average change: –88.0

Parallel group design



Crossover design



Crossover design: pros and cons

Pro:

- Every patient receives both therapies
- Intra-individual comparison feasible
- No need to account for inter-individual variability as source of random error
- Higher statistical **power**
- Lower sample-size required

Con:

- Only feasible for **chronic recurrent conditions**
- Not feasible if therapy causes lasting improvement or cure
- **Carry-over effect**:
  - Is it the second therapy that works?
  - Or is it an after-effect of the first therapy?

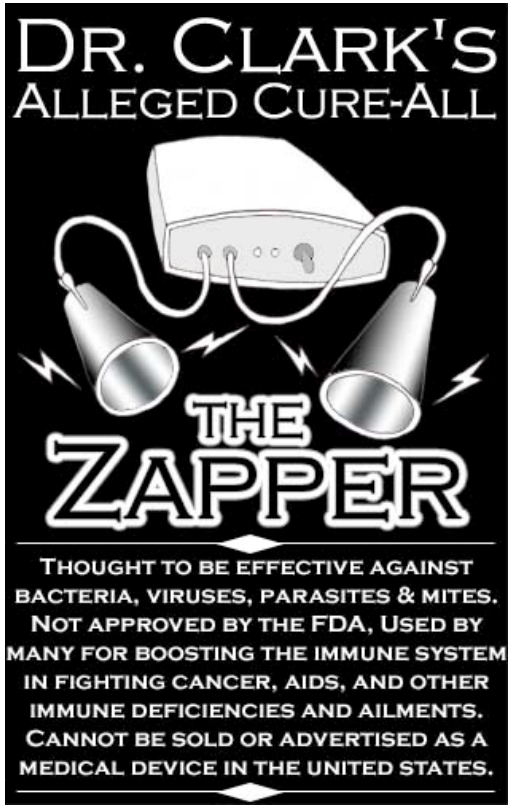
Randomisation

- **Random therapy allocation**
  = random allocation of study patients to therapy groups
- Not: drawing a representative population sample

Aim of randomisation: comparability of treatment groups at baseline

- **Balancing confounding factors**:
  - Most important confounder of all: **severity** of the disease (**confounding by indication**)
  - Confounders **need not be known**

Randomisation is no cure-all



- Try **stratified randomisation**, i.e. randomisation within strata with similar prognosis
- Use **multivariate models** to control for known confounders
- When reading a paper: check **table 1**

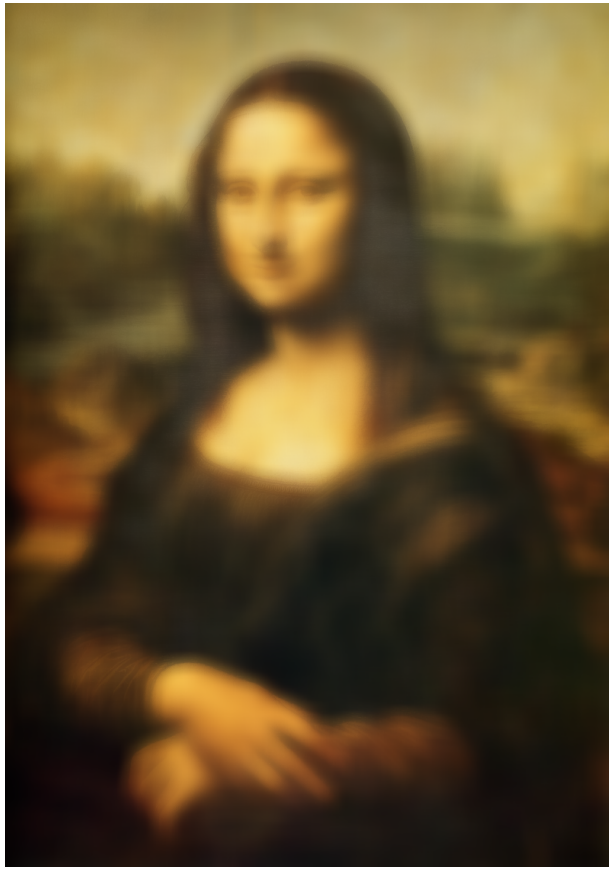
Randomisation: allocation concealment

- **Doctors** must **not** be able to **influence** the result of randomisation.
  - Best solution: central randomisation via telephone or internet
  - Suboptimal approaches:
    - Alternating allocation
    - Allocation based on day of birth, first letter of name etc.
- Definite **decision on participation** must be made **before** the result of **randomisation** becomes known.
- **No switch** of **therapy** after randomisation!

Intention-to-treat (ITT) analysis

- Some patients do not receive the treatment allocated by randomisation.
  - Placebo patients receive verum and vice versa.
  - Patients with contraindications for surgery receive conservative treatment.
- **ITT analysis:** patients are analysed as intended by randomisation, not as treated.
  - Despite receiving verum, placebo patients remain in the placebo group and vice versa.
  - Surgery patients without operation remain in the surgery group.
- ITT analysis requires **follow-up of all randomised patients**, regardless of possible protocol violations

ITT analysis: basic idea



- Preserving groups as randomised preserves **comparability** at baseline.
- Misclassification of patients **waters the difference** in **outcomes** between therapy groups.
- The observed difference **underestimates** the **true effect.**

Masking = blinding

- **Single blind:** researcher knows the treatment, patients do not
  - Outcome assessment unbiased by patients' attitudes
- **Double blind:** neither patients nor researchers know which patients receive which therapy
  - Outcome assessment unbiased by patients' and doctors' attitudes
  - Treatment intensity unbiased by doctors' attitudes

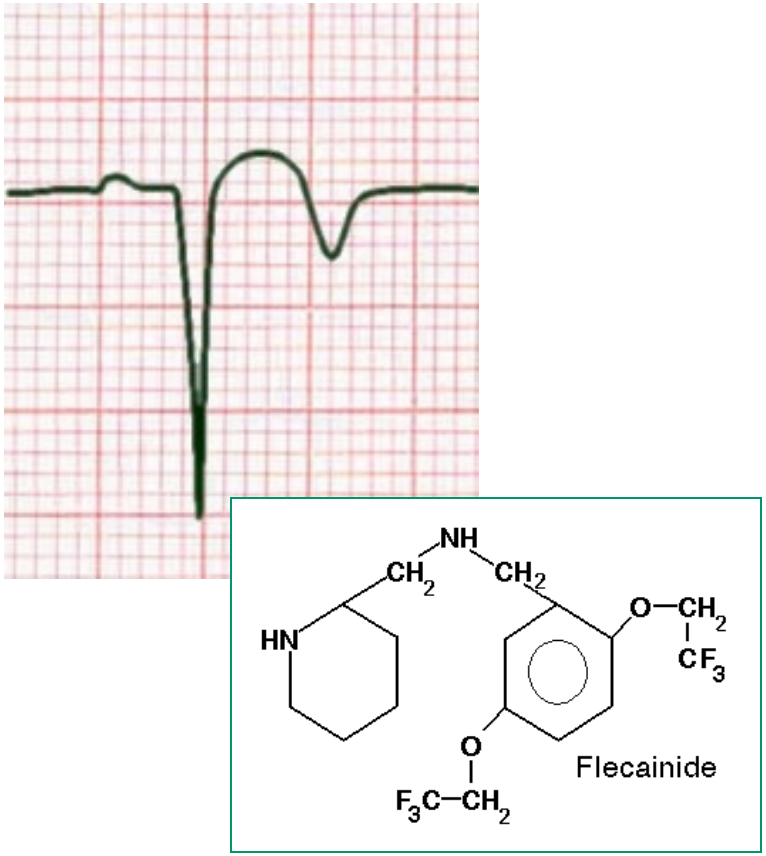
Typical study outcomes

- Cumulative incidence of events
- Quantitative changes in biomedical parameters
- Time-course of disease
  - Survival time, life-expectancy
  - Disease-free survival, time until disease progression
  - Survival probabilities after 1, 2, 5, or *x* years
- Quality of life (QoL)
- Survival and QoL (quality-adjusted life years, QALY)

Relevant outcomes vs. surrogate endpoints

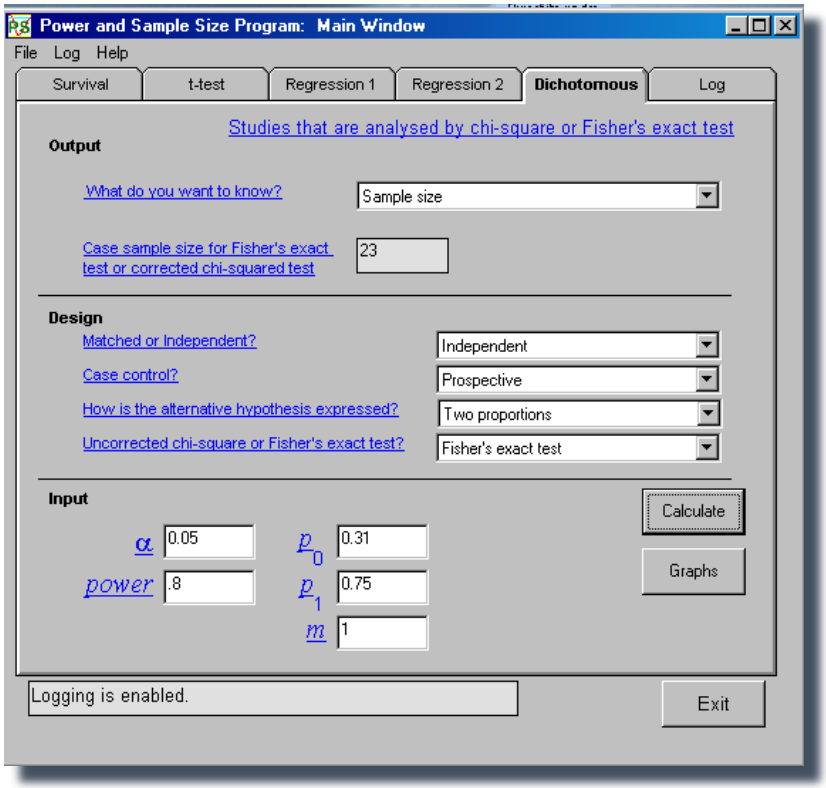
Indication	Surrogate endpoint	Relevant outcome
Type 2 diabetes mellitus	Hb_{A1c}	Complications, mortality
Heart failure	Haemodynamic parameters	Mortality, quality of life
Prevention of ventricular fibrillation after MI	VES suppression in the ECG	Mortality
Hypercholesterolaemia	Total cholesterol	Mortality
Oncological chemotherapy	Remission rate	Mortality, quality of life
Arterial hypertension	Blood pressure	Complications, mortality

Misleading surrogate endpoints: CAPS, CAST und SWORD



- *Cardiac Arrythmia Pilot Study*: sodium channel blockers *encainide* and *flecainide* reduce ventricular extrasystoles (VES) in MI patients
- However: doubled (!!) mortality among verum patient (*Cardiac Arrhythmia Suppression Trial*)
- *Survival With Oral d-Sotalol*: less VES, but increased mortality under verum

Sample size estimation



- The **statistical test** must be given a fair **chance** to become **significant**.
- **Underpowered** trials are **unethical**.

Multiple testing and Russian roulette

- Russian roulette:

  You only need to pull the trigger often enough, and you‘ll (almost) certainly be dead.

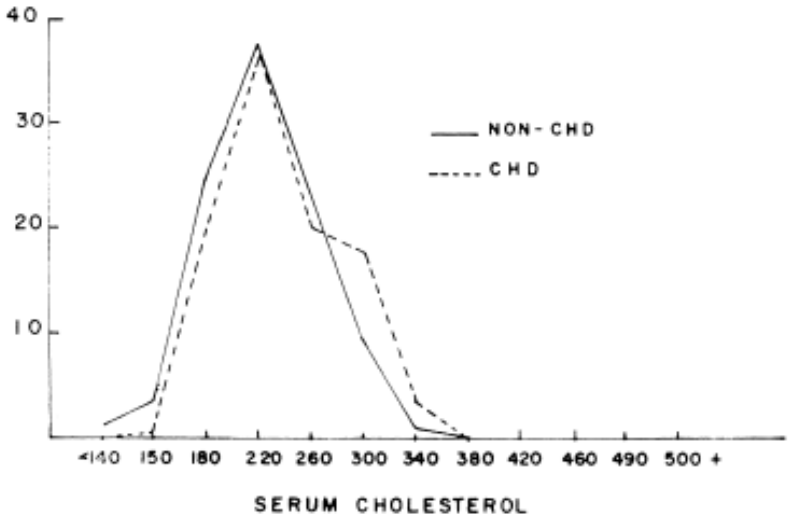
- Hypothesis tests:

  You only have to test a sufficient number of (true) null hypotheses to end up with a significant finding.

Multiple testing: solutions

- Make a clear distinction between **primary** and **secondary outcomes**
  - Test regarding the **primary outcome: confirmatory**
    - **Statistical "proof"** with defined type I error probability
  - Tests regarding **secondary outcomes: exploratory**
    - **Hypothesis generation** for further research
- Use **adequate tests** (e.g. ANOVA instead of pairwise comparisons)
- If there is **more than one primary outcome**, make sure to maintain the **global alpha level:**
  - Alpha adjustment
  - A priori ordering of hypotheses
  - Closed test procedures

Statistical significance vs. relevance



- The ***p* value** is a function of…
  - **Effect size** and
  - **Sample size**
- **Big trials** may yield statistically **significant**, but medically **irrelevant** findings.
- Consequence: always complement **test results** with the corresponding **confidence intervals.**

Instead of a conclusion: typical errors in clinical trials

- Irrelevant outcome
- Primary outcome not clear
- No control group
- Inadequate control group
- No or flawed randomisation
- No masking
- Separate analysis of therapy groups instead of comparison
- Confusing statistical significance and medical relevance
- Claims of efficacy based on secondary outcomes
- Wrong interpretation of non-significant results
- Inadequate power due to inadequate sample size