Applied Analysis of Variance and Experimental Design

Lukas Meier, Seminar für Statistik

About Me

- Studied mathematics at ETH.
- Worked at the statistical consulting service and did a PhD in statistics (at ETH).
- Excursion to the insurance industry.
- Since 2011: Senior scientist, Seminar für Statistik, ETH.

About You

- About 50 people of CAS / DAS in applied statistics (“WBL”)
- About 190 students
  - Food science
  - Statistics / (applied) mathematics
  - Environmental science
  - Biology
  - PhD students from various fields
  - …
- You (should) all have in common that you have attended (at least) an **introductory course** to probability and statistics.
- We use this knowledge as a **basis**.

Lecture Style

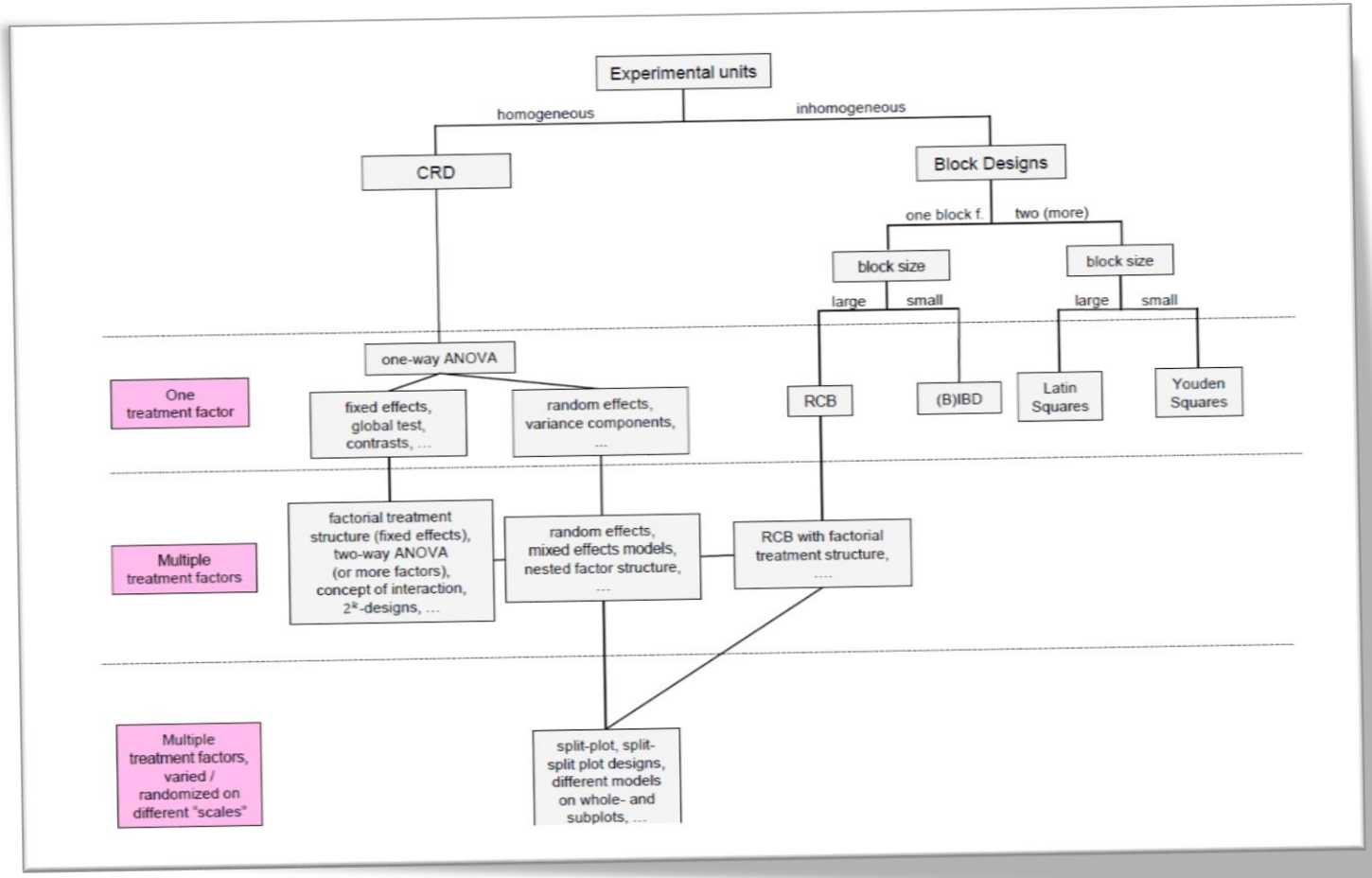
- This is an **applied** lecture.
- We will skip many of the mathematical details.
- Typically, we will **not** do any proofs.
- We will have a look at mathematical derivations, details etc. if it is helpful to understand the underlying models.

Software

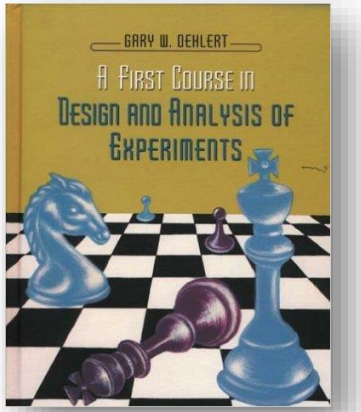
- We use the statistical computing software **R**.
- We will only do things “by hand” if it is helpful for your understanding.
- I will try to show you in class how the presented models can be fitted in R.
- There will be an introduction to R in today’s exercise class.
- If you have never used R before, the initial effort is **large**.
- Nevertheless, it will pay off as you will probably need it for your
  - master thesis
  - future research projects
  - future job
- Hence, learn it **now**.

Topics

- Principles of experimental design
- Completely randomized designs
- Specific differences (contrasts)
- Factorial treatment structure
- Complete block designs
- Random effects
- Mixed effects
- Split plot designs
- Incomplete block designs
- Fractional factorials
- Power analysis



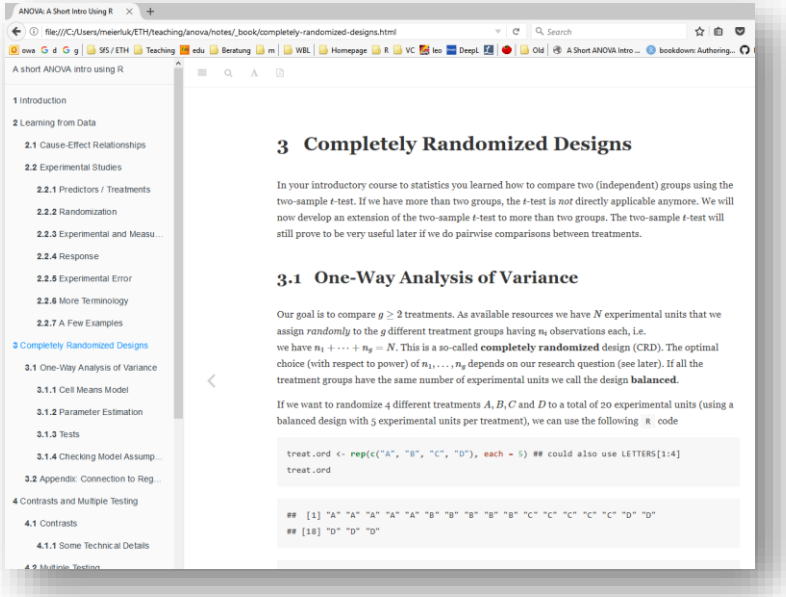
Book



- We mostly follow the book ***A first course in Design and Analysis of Experiments*** by *Gary Oehlert*.
- Book is out of print (although mostly good) but PDF can be downloaded for **free** at http://users.stat.umn.edu/~gary/Book.html
- Book has about 600 pages but we will **not** do all chapters / details.
- I will try to give you a detailed chapter list what we will discuss next week (in case you like to prepare for class).

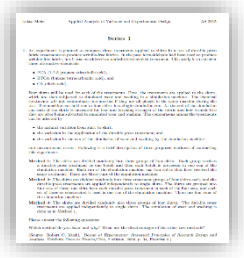
Lecture Notes

- There are lecture notes.
- There is both a PDF and an HTML version available.
- Lecture notes also include R code.
- This is still in development, **please report errors**!



Exercises (CAS / DAS: weekly, as usual)

- **Every other week** there will be a **2 hour exercise session**.
- Today: Introduction to R.
- Bring your **own notebook**.
- Work on the current exercise series.
- **Ask questions**, **discuss!**
- No “classical” exercise session in the sense that you get hints and then try to solve it at home.



Exercise sheet published

Exercise class takes place

Hand in on Thu, 12:00 in HG J 68

Get back corrected version in **lecture**

Week

n

$n+1$

$n+2$

Exam (CAS / DAS: on the computer)

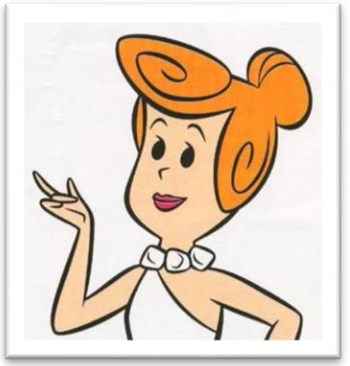
- Exam will be on **paper**.
- We will not ask you technical R-details, but you should be able to
  - **interpret** R-output
  - **know the syntax** of (e.g.) `lmer` to fit an appropriate model (e.g., split-plot model)
  - etc.
- **Style**: Multiple Choice, old exams will be made available on the website.

Introductory Example

Salk Vaccine Field Trial (Freedman et al, 2007)

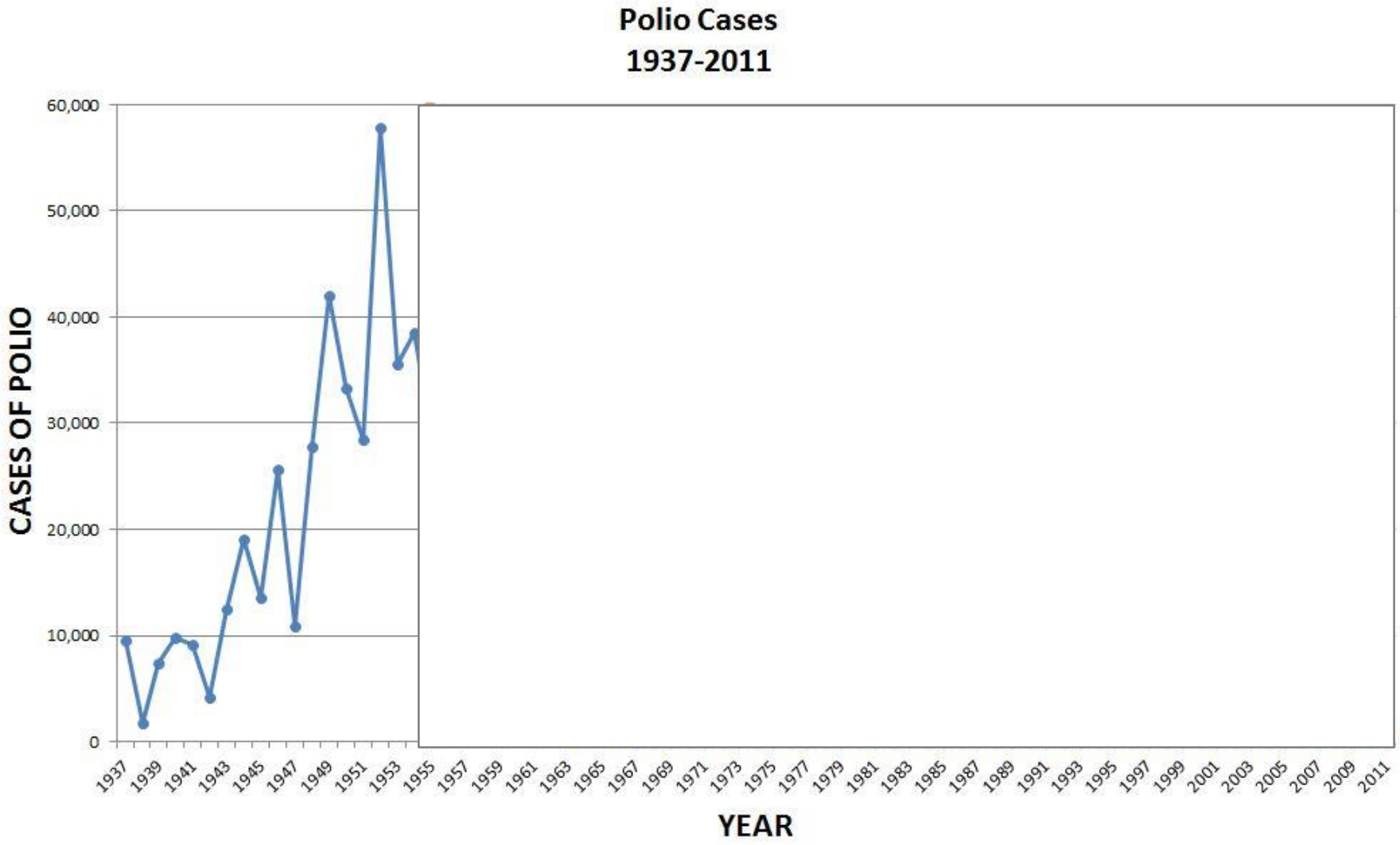
- **Polio** caused hundreds of thousands victims (mainly **children**) in the first half of the twentieth century.
- By about 1950, several **vaccines** had been discovered, among others the one from **Jonas Salk** (the most promising).
- In the lab, everything looked good so far.
- By 1954 the public health service was ready to try the vaccine in the real word (i.e., outside the lab on patients).
- How should they “measure” the **effectiveness** of the vaccine **in the real world**?

Salk Vaccine Field Trial (Freedman et al, 2007)



- We love our children and polio is bad, so let us give the vaccine to a very **large number of children this year!**
- We can determine the **incidence rate** of polio **this year** and compare it to the rate of **last year**.
- This does not sound very complicated.
- Unfortunately, this is **not** a good idea because polio is an **epidemic** disease.
- Incidence rate can vary **substantially** from year to year.

Salk Vaccine Field Trial (Freedman et al, 2007)



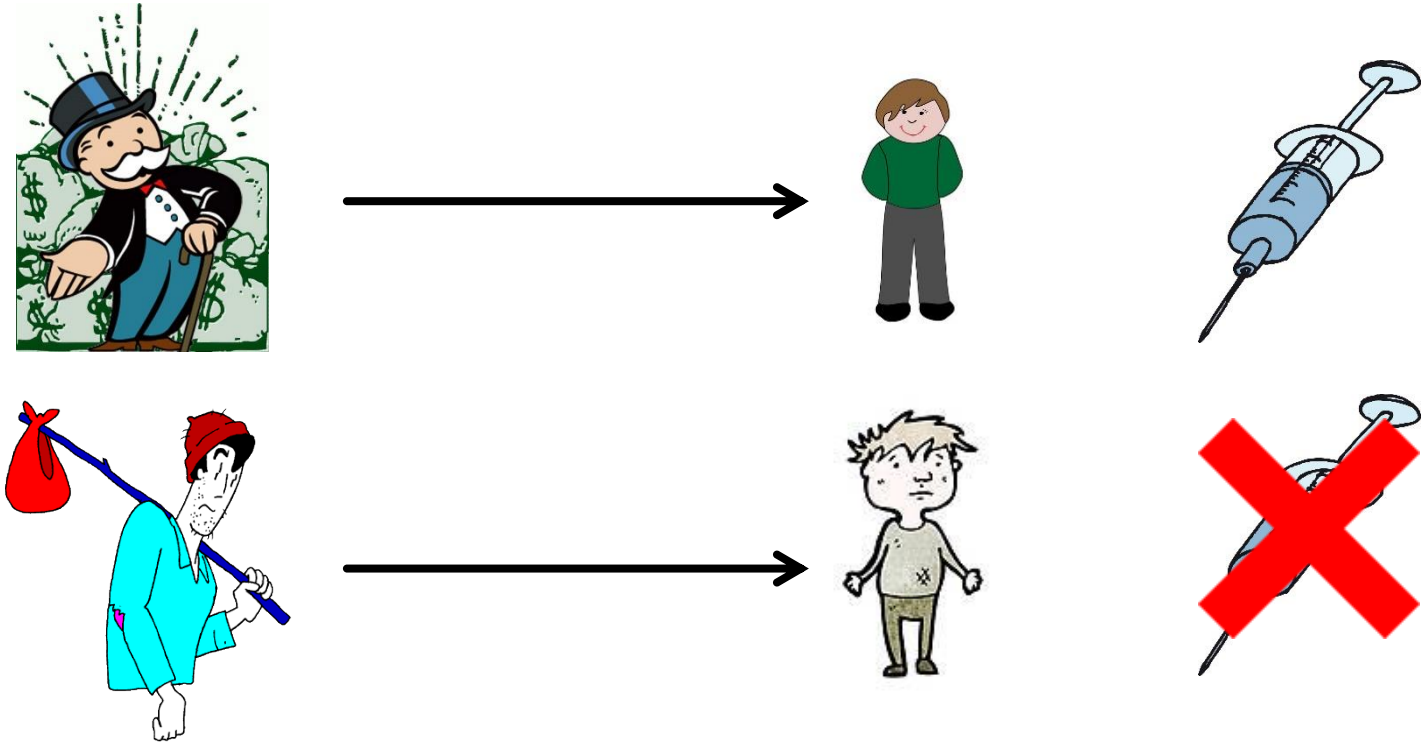
http://vaccines.procon.org/view.additional-resource.php?resourceID=005964

Salk Vaccine Field Trial (Freedman et al, 2007)

- Whatever effect we see, we **cannot** say whether it was the effect of the year, of the vaccine, or a combination of the two.
- We say that the two effects are **confounded** (mixed up).
- Therefore, we need to leave some children **unvaccinated** this year and use them as a **control group**.
- This will allow us to measure the effectiveness of the vaccine by **comparing** the **rates** at which the children get polio in the two groups (**treatment** vs. **control**).

Salk Vaccine Field Trial (Freedman et al, 2007)

- Of course, parents' permission is required for vaccination.
- One possibility would be to build treatment and control groups based on the **parents' decision**.
- However, **higher-income** parents would more likely **consent** to treatment than lower-income parents.



Salk Vaccine Field Trial (Freedman et al, 2007)

- In addition, children of higher-income parents are more vulnerable to polio (effect of hygiene).
- Hence, this design is **biased** against the vaccine (the family background is confounded with the effect of the vaccine).
- We need a control and a treatment group that come from the **same population**.
- Here: Only consider children whose parents consented to vaccination.
- Every child should have a 50% chance of being put in the control or the treatment group (**randomization**).

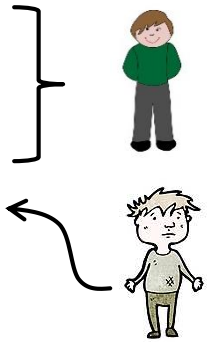
Salk Vaccine Field Trial (Freedman et al, 2007)

- Children in the control group were given a **placebo** and **no one** was told whether they were in the control or the treatment group.
- Reason: Want to make sure that the effect was due to the vaccine and not due to the “idea of getting treatment”.
- In addition, doctors (who had to decide whether a child contracted polio during the experiment) were **not** told whether a child got the real vaccine or the placebo.
- Together, this is called **double-blinding**.
- Hence we have a so called **randomized controlled double-blind experiment**.

Salk Vaccine Field Trial (Freedman et al, 2007)

- Results:

	Group size	Rate (= per 100’000)
Treatment	200’000	28
Control	200’000	71
No consent	350’000	46



- Highly **significant difference** between rates (e.g., use Fisher’s exact test; we will not discuss it in this course).
- This field trial already illustrated many concepts of experimental design.
- We will now have a more detailed look at some of these aspects.

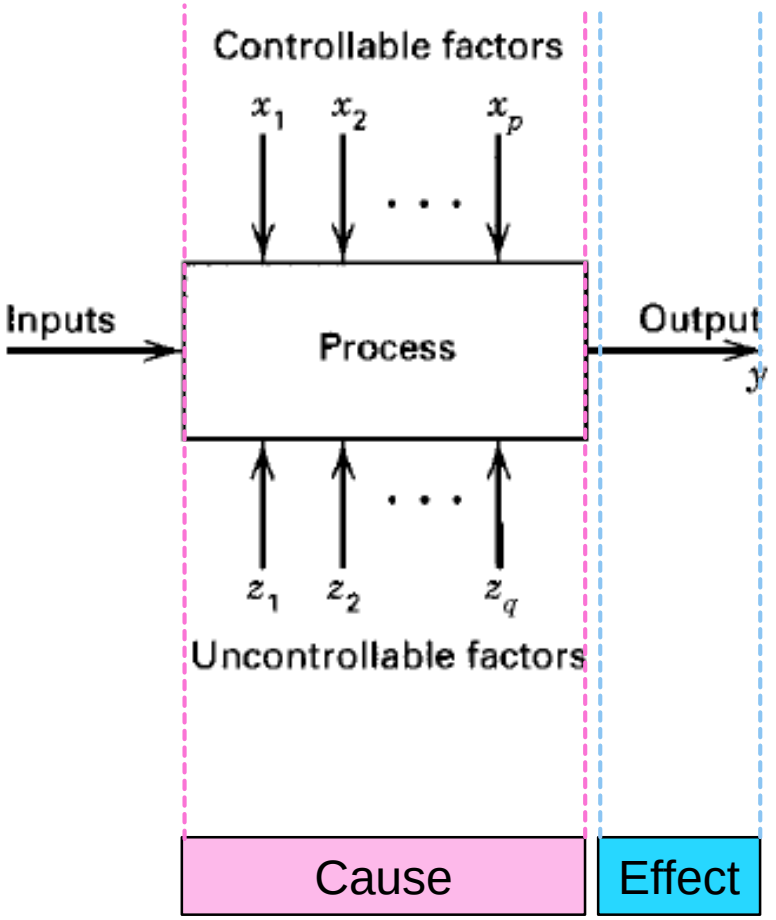
Why Experiment or Collect Data?

Cause and Effect of a Process or System

Terminology

Cause and Effect

Typically, data is collected to discover a **cause - effect relationship** of a (complex) "**process**" or a "**system**".



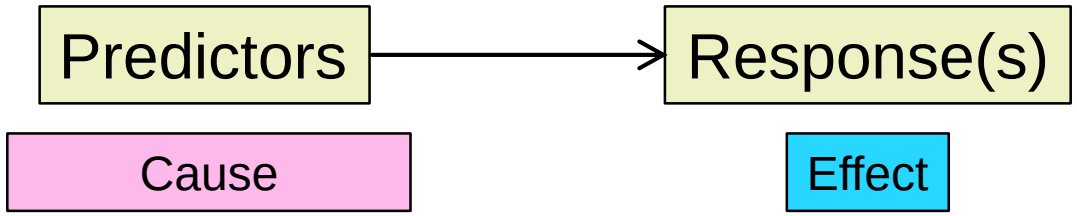
From Montgomery, 1991

Typical Questions

- What is the influence of different fertilizers on biomass?
- Is a new drug an effective cure for a disease? How do side-effects depend on dose?
- How do people rate different recipes of chocolate chip cookies?
- How do the settings of a chemical process influence yield?
- See more examples later.

Predictors and Response

- We also call the input factors **explanatory variables** or **predictors** and the output the **response**.
- Hence, we want to understand the relationship



- Ideally, we want to establish a **causal** relationship, i.e. we want to find out the **effect on the response** if we make an **intervention on a predictor**.
- Making an intervention means actively setting a predictor to some value.
- Typically, a lot of predictors are involved.

Different Kinds of Predictors

One distinguishes between **predictors** that

1) are of **primary interest** and that can be (ideally) **varied according** to our „wishes“: the conditions we want to compare, or the „**treatments**“.
2) are **systematically recorded** such that potential effects can be later eliminated in our calculations („controlling for…”).
3) can be **kept constant** and whose effects can therefore be eliminated.
4) we can **neither record nor keep constant**.

2) to 4) are also called **nuisance variables**.

Examples of Nuisance Variables

- In ecological or agronomical studies:
  - Soil properties (2)
  - Weather (2)
  - Material (2, 3)
  - Personnel (2, 3)
  - ...
- Measurements on humans:
  - Age (2, 3)
  - Weight (2, 3)
  - Potential diseases (2, 3, 4)
  - Stress-level (2, 3, 4)
  - Fitness (2, 3, 4)
  - Genotype (2, 4)
  - …

Response

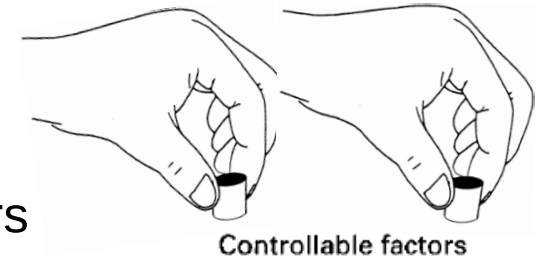
- The **response** should be chosen such that it reflects **useful information** about the process under study.
- The response is what **you** measure to **judge** what happened in the process.
- It is **your responsibility** that the response is a reasonable quantity to study your research hypothesis.
- If not directly measurable, use a **surrogate response** (e.g., use CD4 counts as surrogate for HIV progression).
- Hypothetical example: amount of sleep after taking tranquilizer
  - Measure hours that person was sleeping.
  - Measure number of coffees that person is drinking in the morning.
  - …

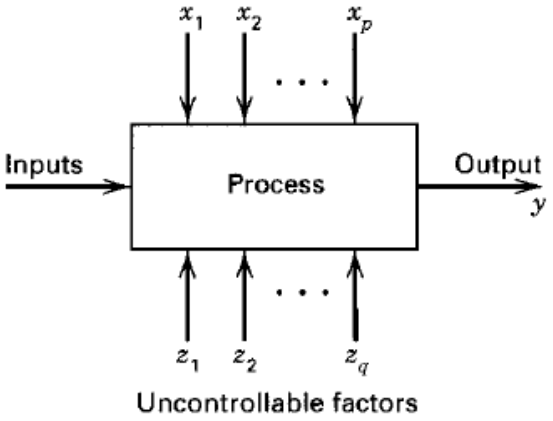
Observational Studies

- Overview
- Association vs. Causation
- Confounding

Experimental Study

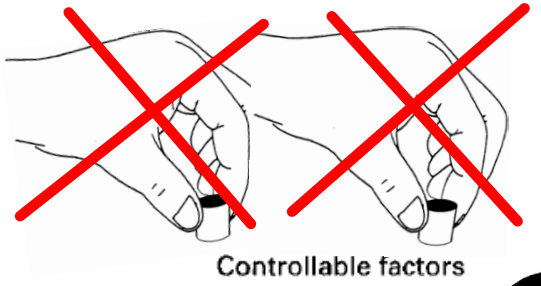
- Can **control** (some) predictors

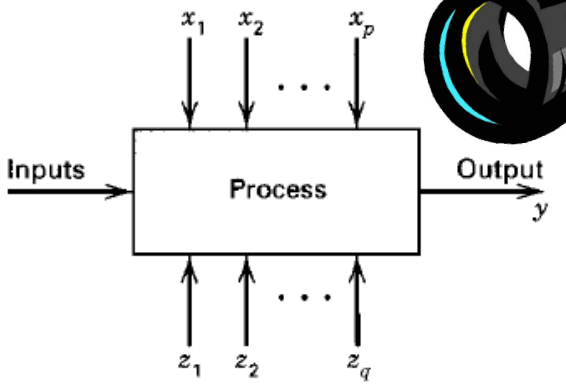


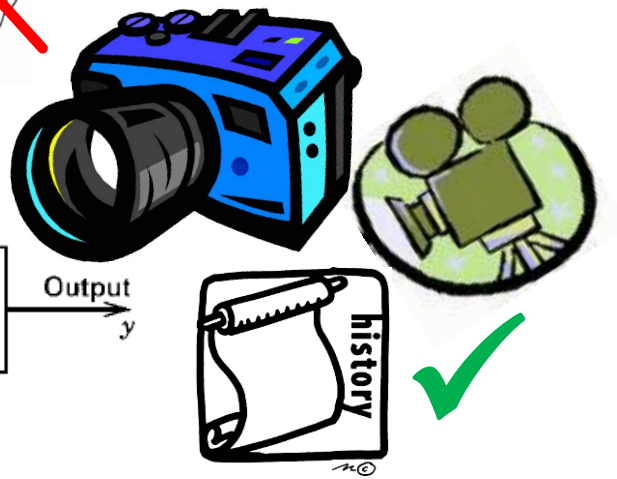


Observational Study

- Cross-sectional study
- Cohort study
- Case-control study







Observational Study

- Observation of subjects / objects in an **existing (uncontrolled) situation**.
- Examples
  - Consumer behaviour in different countries
  - Epidemiological studies
  - Air quality in ETH Mensa at different times and days
  - Heavy metal pollution in soil at various locations

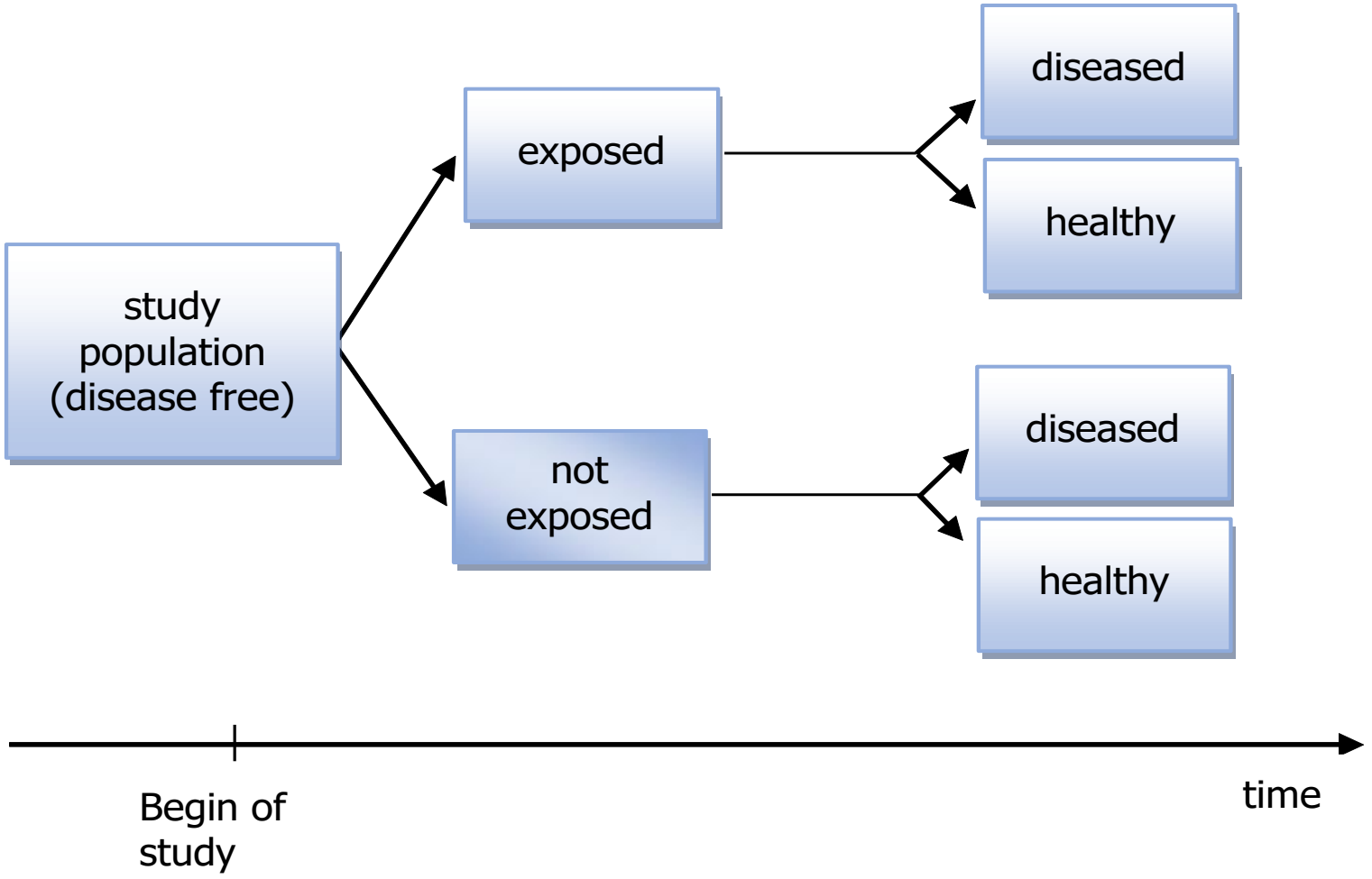
Different Types of Observational Studies

- **Cross-sectional study**
  - “Snapshot” of population at a given time-point.
- Prospective: **Cohort study**
  - What will happen if…?
  - Determining the risk (e.g. lung cancer) of exposed (smokers) vs. non-exposed (non-smokers) subjects (people).
- Retrospective: **Case-control study**
  - Why did it develop this way?
  - Comparison of habits of healthy vs. non-healthy persons.

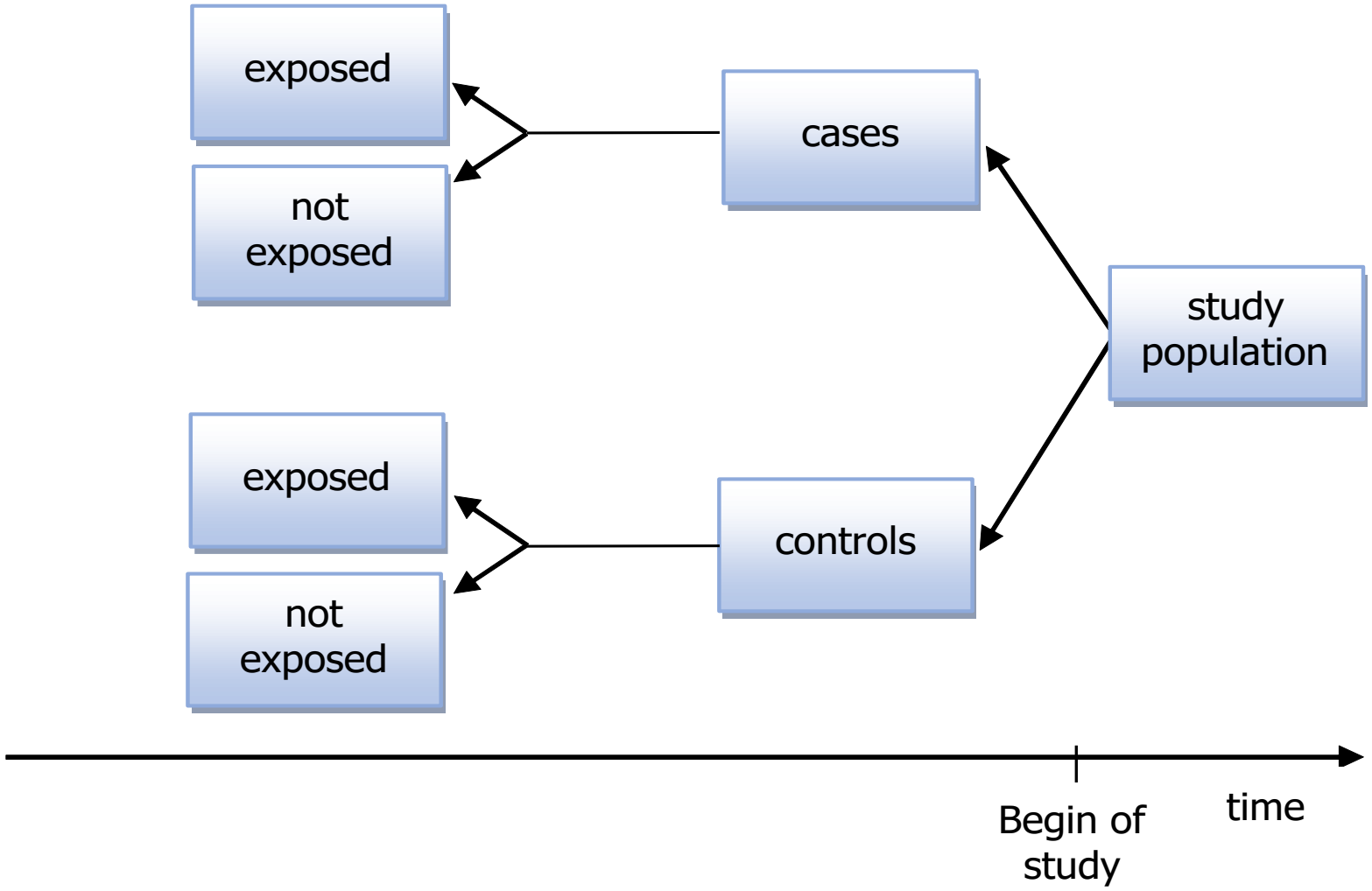
Example for Cross-Sectional Study (Roth, 2014)

- **Consumer behavior survey**
- **Response**: Consumption of meat per household and year.
- **Predictors** according to different categories:
  - (1) Regions
  - (2) Age, profession, education of leading person, household size, income, number and ages of children, …
  - (3) Method of collecting data, measurement method.
  - (4) Genotype, social environment, health status, …

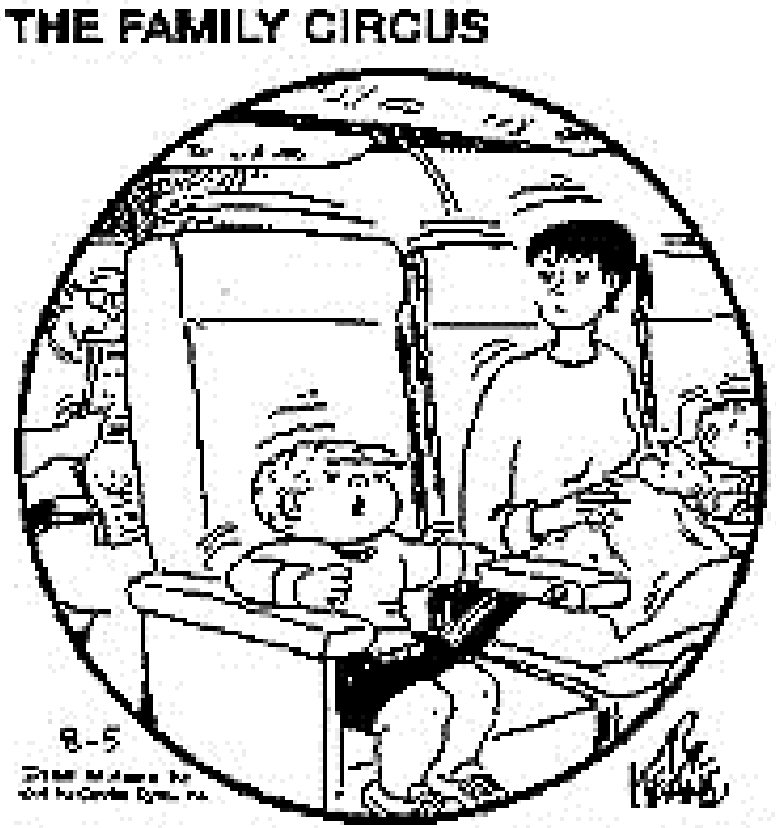
Cohort Study (prospective)



Case-Control Study (retrospective)



Causality and Observational Studies

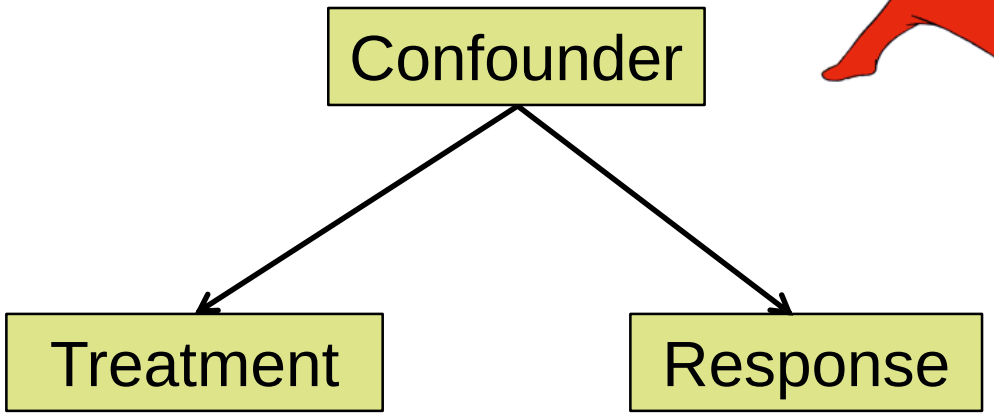


"I wish they didn't turn on that seatbelt sign so much! Every time they do so, it get's bumpy"

Causality and Observational Studies

- In an observational study we have **no control (or no idea)** of the mechanism that assigned the “subjects” to the different “treatment” groups.
- It might very well be the case that some (hidden) predictors influence both the treatment “assignment” and the response, i.e. we have **confounders**.
- Let’s have a look at them in more detail.

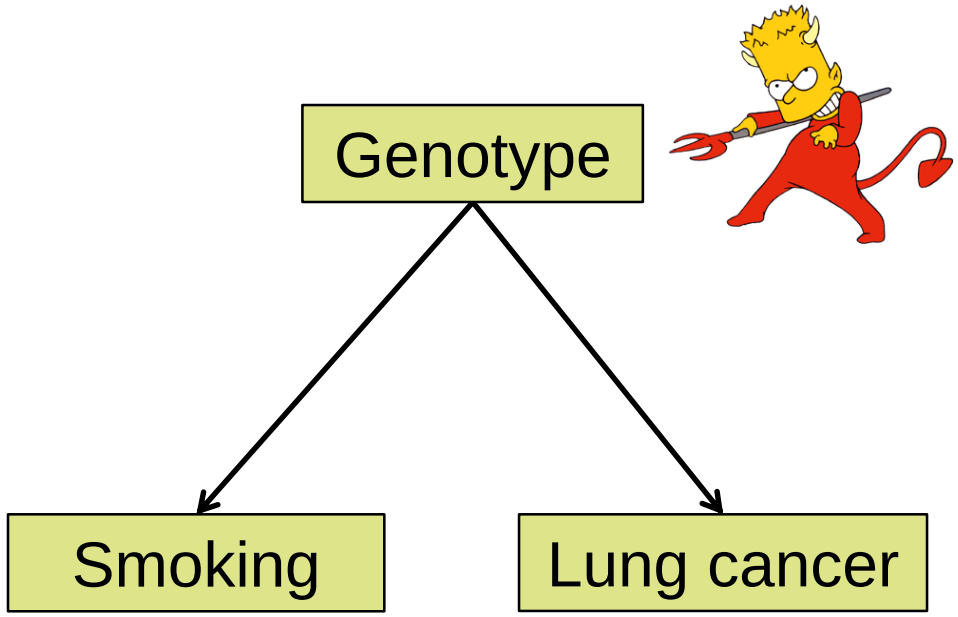
Confounder



- In an observational study you would see an **association** between treatment and response, although there is **no** underlying cause–effect relationship.
- “Solution” in observational studies: Record potential confounders, use them in models later on.
- But: What about **hidden** confounders?

Early Research Regarding Smoking and Lung Cancer

- Argument of (famous) R.A. Fisher working for the tobacco industry: “There might be a **common cause** involved” (i.e., a **confounder**).



- Here: Experiment (intervention!) not feasible due to **ethical issues**. Any volunteers?

Spurious Associations: Widespread Phenomenon

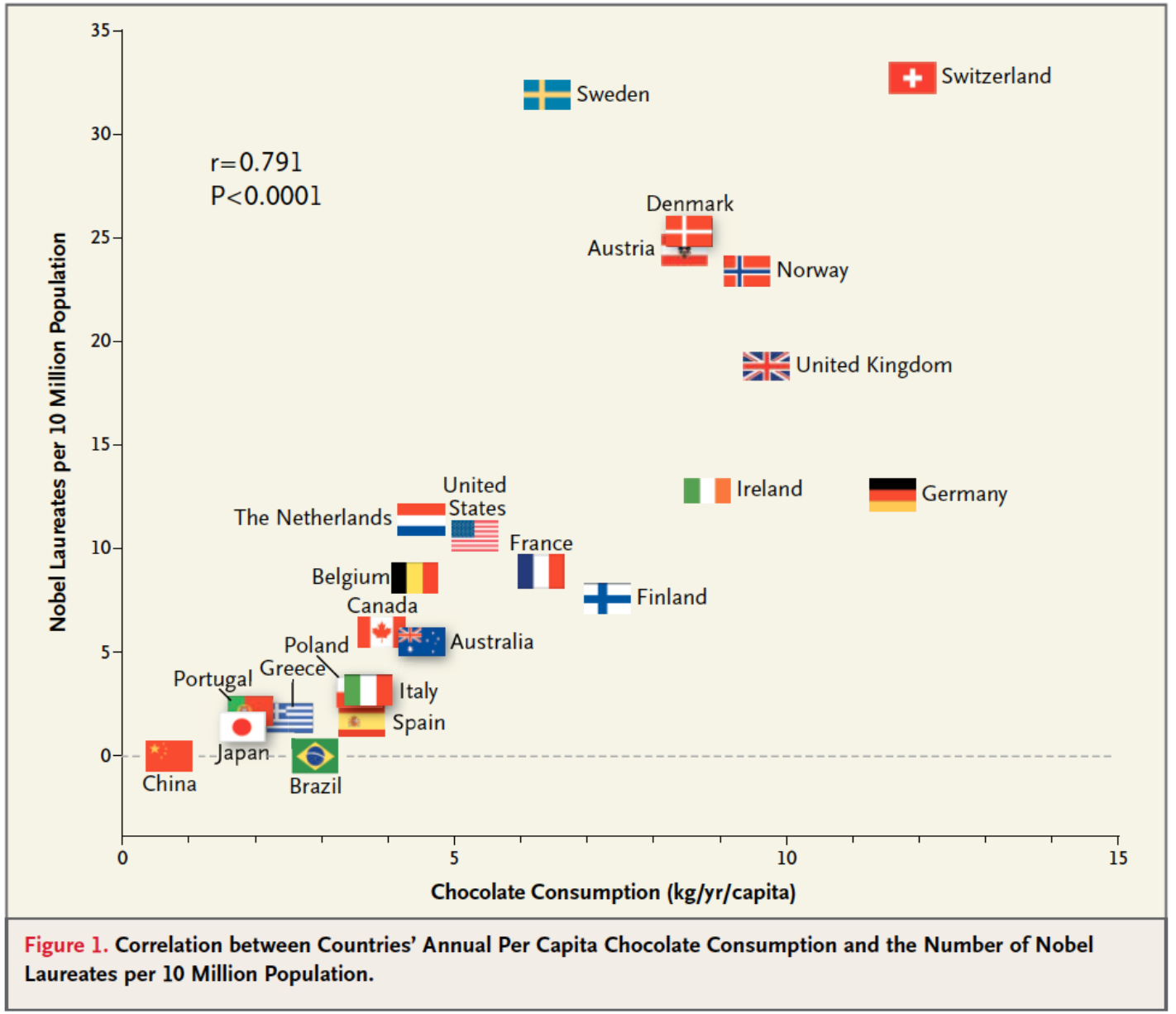
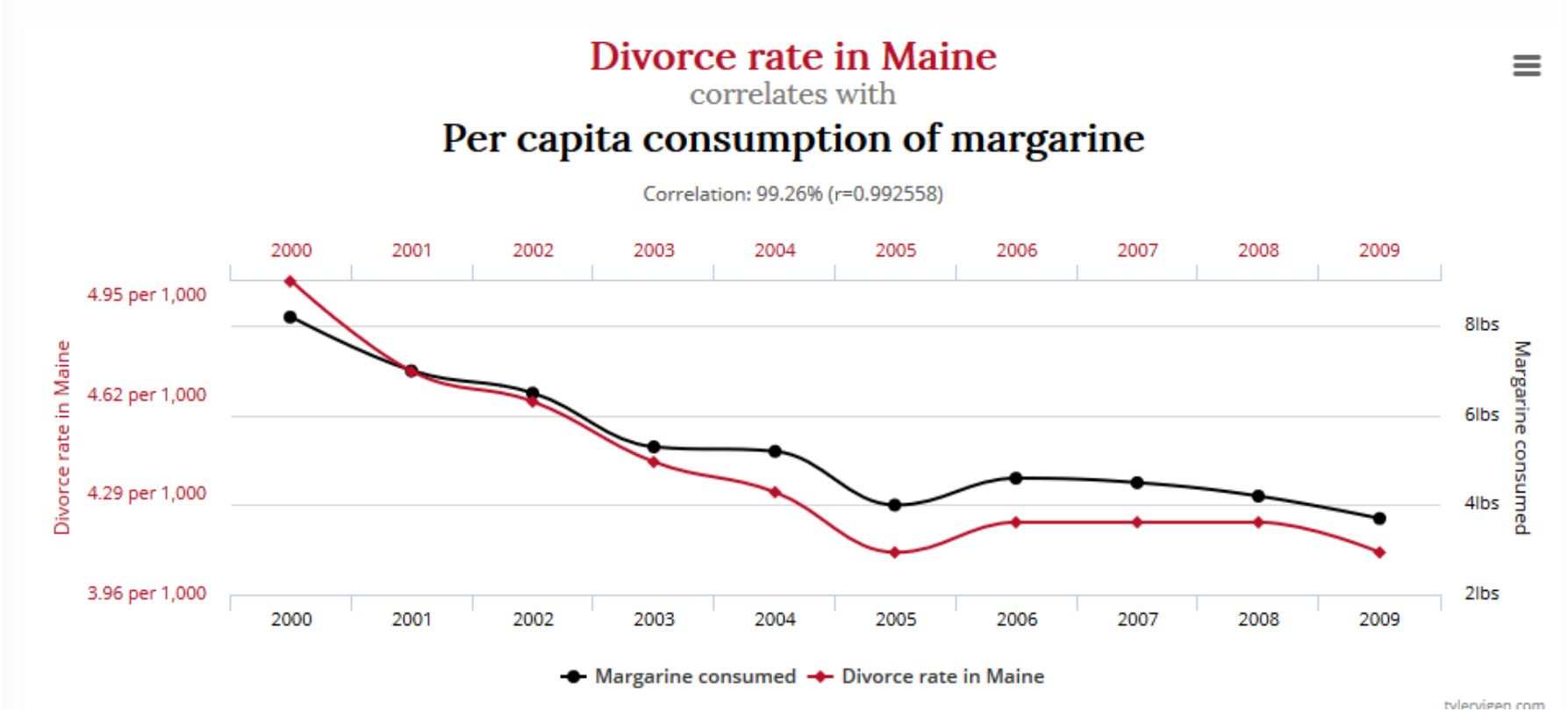


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

New England Journal of Medicine

Spurious Associations: if you search long enough…

See http://www.tylervigen.com/spurious-correlations



Experimental Studies

- Ingredients
- Terminology
- Randomization and Blocking
- Comparison to Observational Studies

Experimental Study

What is an **experiment**?

From Montgomery (1991):

*“Literally, an experiment is a test. A **designed experiment** is a test or series of tests in which **purposeful changes** are made to the **input variables** of a **process or system** so that we may **observe** and **identify** the **reasons** for **changes in the output response**.”*

Experimental Study

- Observation of „subjects“ or „objects“ in a **controlled setting** (according to your „wishlist“).
- Examples:
  - Salk vaccine trial, other clinical trials.
  - Field test to compare different fertilizers and / or harvesting methods.
  - Infection tests in greenhouse.
  - Psychological or pedagogical experiments.
  - Different settings to optimize yield of a food production process.
  - Determining the lifetime of objects under different “stress scenarios” in the lab.

Ingredients of an Experimental Study

An experimental study consists of

- Different **treatments** (the **interventions** you perform on the system), e.g. different kinds of fertilizers.
- **Experimental units**, the “things” (“subjects”, “objects”) to which we apply the treatments by randomization, e.g. plots of land receiving fertilizer.
- **Method that assigns treatments to experimental units**
  - Randomization
  - Restricted randomization (blocking)
- **Response(s)**, e.g. biomass of plants.

More on Experimental Units

- **Experimental unit**
  - The “things” to which we apply the treatments by randomization
  - Rule: An experimental unit should be able to receive any treatment (**independently** of the others).

- **Measurement unit**
  - Actual “object” on which the response is measured.
  - Potentially: measurement unit ≠ experimental unit (!)

- The measurement made on the **whole** experimental unit (potentially a sum or an average of the measurements on the measurement units) will be the **basis of the analysis of the experiment**.

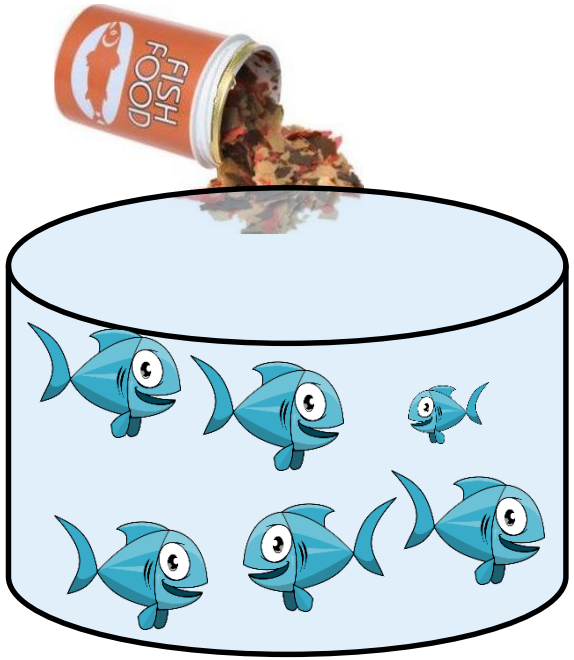
Experimental vs. Measurement Unit: Example

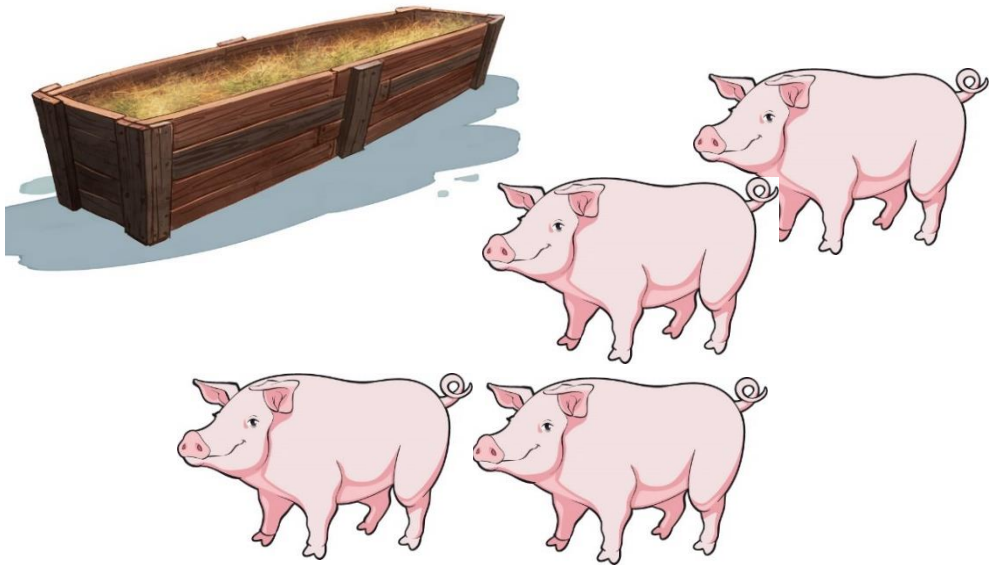
From Oehlert (2010):

- **Six classrooms** of **25 first graders each** are assigned at random to **two** different reading programs.
- Evaluation is at the end of the school year through a common **reading exam**.
- Are there $6 \times 25 = 150$ or 6 experimental units?
- Experimental unit $=$
  Measurement unit $=$

Experimental vs. Measurement Unit: Example

Similar problems:





Randomization

- We have seen: **Confounding** can be very problematic.
- How can we **protect** ourselves from known (or even worse: unknown) confounders?
- Use randomization!
- Randomization means: The **allocation** of the **experimental units** to the different **treatments** is **random**.
- Ensures that potential confounders are “averaged out”.

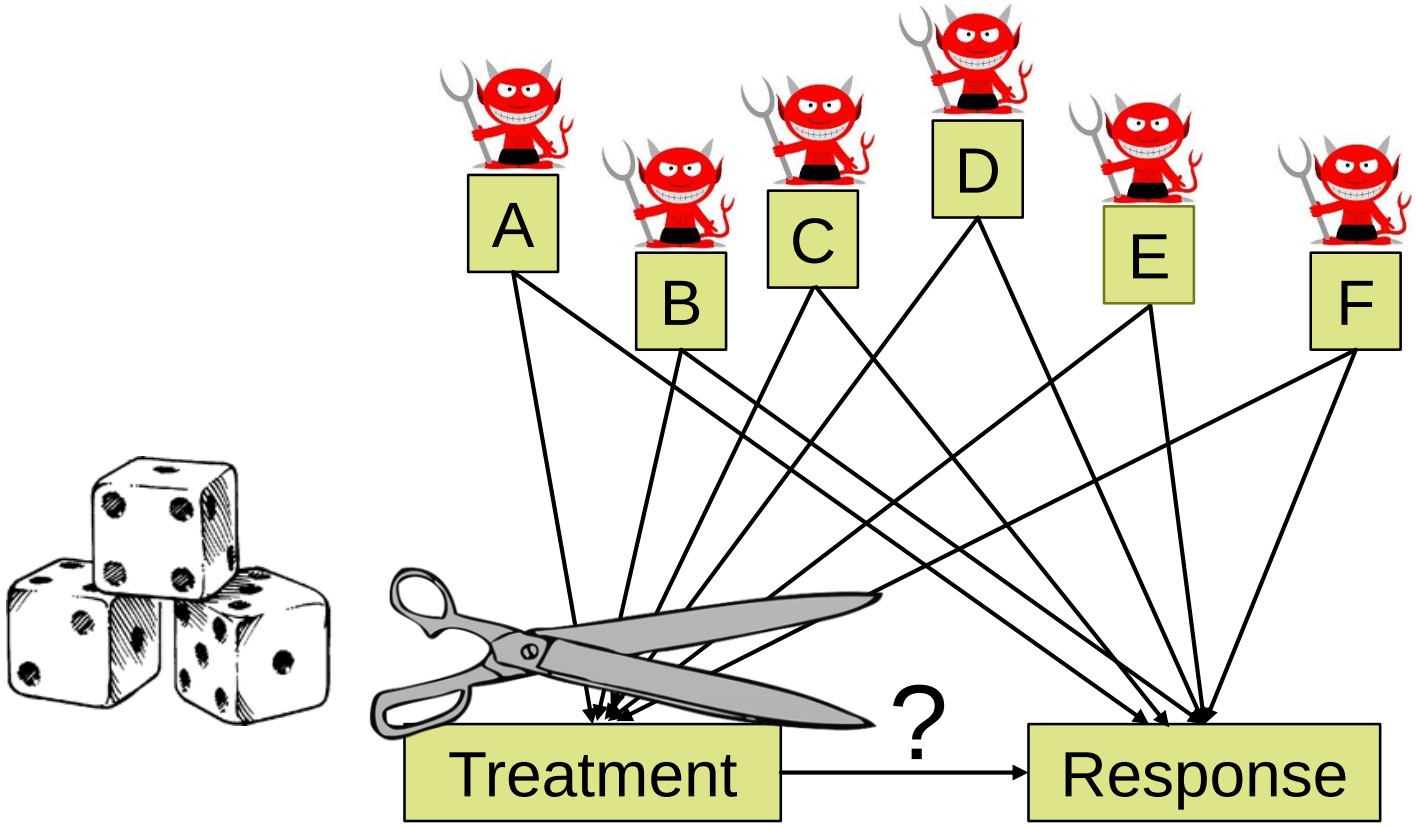
Randomization: Example (from Oehlert, 2010)

- Want to compare new drug treatment to surgery with respect to five-year survival.
- We have a total of 100 patients.
- We know: surgery might be problematic for patients with severe disease.
- Tempting to put these in drug group (→ **confounds** patient status with treatment).
- Better: make up basket with 50 red and 50 white balls (or toss a coin). Draw ball for each patient. Red means surgery, white drug.

Why is Randomization so Powerful?

- Whatever feature of the experimental units are associated with our response, randomization ensures that approximately half of the patients with this feature is being put in each of the treatment groups.
- Here: Approximately half of the “strong” patients get the drug etc.
- Randomization ensures that the only **systematic difference between the groups** is the **treatment**.
- This is why a (properly) randomized experiment allows us to make a statement about a **causal effect**.

Randomization Protects us from Confounders



Randomization

- Cochran and Cox (1957):

  *“Randomization is somewhat analogous to **insurance**, in that it is a **precaution against disturbances** that may or may not occur and that may or may not be serious if they do occur. It is **generally advisable to take the trouble to randomize** even when it is not expected that there will be any serious bias from failure to randomize. The experimenter is thus **protected against unusual events** that upset his expectations.”*

- Oehlert (2010):

  *“Randomization generally **costs little** in time and trouble, but it can **save us from disaster**.”*

Randomizing other Things

- We can and (should) also randomize (or use blocking)
  - **Order** in which experimental units are used (if not used simultaneously).
  - **Locations** at which experimental units are used (if not all at the same location).
  - If using multiple **measuring instruments**: randomize which units are measured on which instruments.
  - …

Blocking, a Restricted Randomization Scheme

- In the preceding experiment we would better consider
  - age
  - gender,
  - health status
  - etc.

  and do the randomization and comparison "**within" homogeneous groups**.
- This strategy is known as **blocking**.
- A **block** is a **subset** of the **experimental units** that is **more homogenous** than the entire set.
- We already **know** that the response of different blocks can be (substantially) different.
- **Blocking increases precision** of an experiment, because we use subsets of homogeneous units.

Randomization and Blocking

General rule is:

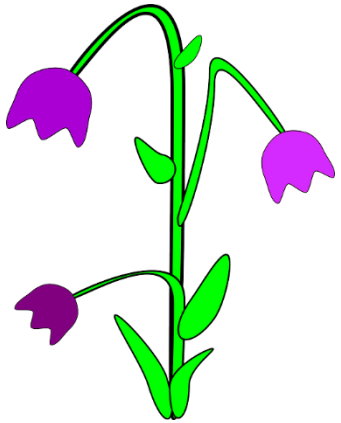
“Block what you can; randomize what you cannot.”

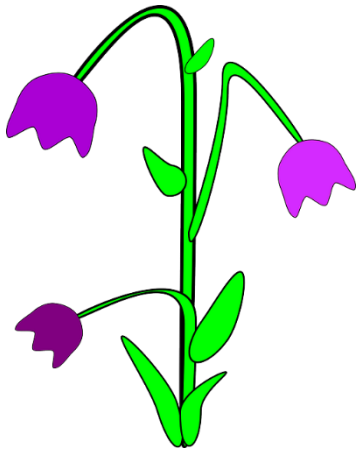
Experimental Error

- Different experimental units will give different responses to the **same** treatment.
- Applying the same treatment to the same experimental unit (if possible) will result in different responses.
- Experiments must be designed such that we have an estimate of this so called **experimental error**.
- This is achieved by using **replicates**, i.e. applying the **same** treatment to **multiple** experimental units.
- If we have no idea of the experimental error, we **cannot** compare treatments (i.e., no statistical inference is possible)!

Example: Missing Replicates

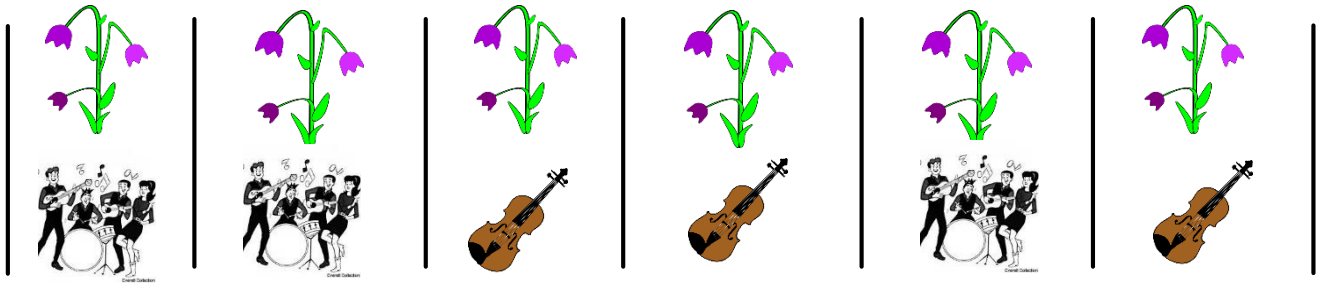
- As seen on Swiss TV…
- Plant 1: Treatment with Music A
- Plant 2: Treatment with Music B





Example: Missing Replicates

- Measure biomass after 4 weeks.
- Is the potential difference that we see due to the treatments (music) or is this just natural variation from plant to plant?
- Unfortunately, the experiment **doesn't give us any information** about the variation from plant to plant.
- We would need replicates: multiple plants receiving the same treatment!



Some More Terminology

- **Blinding** (see also Salk vaccine field trial)
  - **Blinding: Evaluators** don't know which treatment is given to which experimental unit.
  - With humans (patients): **double-blinding**: Neither the **evaluators** nor the **patient** know the assignment.
- **Insurance** against (unintentional) **bias** (e.g., due to expectations).

Some More Terminology

- **Control treatment**
  - “Standard” treatment used as a **baseline for comparison** with other treatments.
  - “Null” treatment (no treatment at all).
  - Important, still often forgotten (see next slide).
- **Placebo**
  - Null treatment in case that simply the **act of applying a treatment** (whatever) has an effect.
  - Often used with humans, but can also be useful in other settings.

Why are Controls Important? (partly based on a true story)

- Meet Mike, physiotherapist who developed a new (costly) therapy.
- Mike: “On average, my new daily therapy reduces the pain score of my patients by 30% one month after knee surgery.”

Why are Controls Important?

However…

- People **not getting any treatment at all** have a reduction of about 60% of their pain score (on average)!
- Want to make an appointment?
- Not always as obvious as here…
- You should always ask:

“How does it compare to the standard / null treatment?”

Guidelines for Designing Experiments (Montgomery, 1991)

- Statement of problem / hypotheses
- Select response variable
- Determine sources of variation in response (predictors):
  - factors of interest
  - nuisance factors (blocking, randomization)
  - factors that can be held constant
- Choose a proper design and randomization scheme

Comparison Experimental vs. Observational Study

	Experimental Study	*Observational Study*
Situation	Controlled: "The settings you wish are the ones you get"	Given: "What you observe is what you get"
Analysis	Typically easy	Difficult
Interpretation	Causal (if properly set up)	Association, (causal with extra assumptions and effort)

"If your experiment needs statistics, you ought to do a better experiment", Ernest Rutherford

Statistical Methodology Point of View