



# Experimentelles Design: Gute Prinzipien

# Was Ihre Professoren wichtig finden...

Umfrage zur Vorbereitung der Vorlesung (15 Professoren von BIO/HST):  
Ist ein Thema (wichtig / nice to have / unwichtig) ?

Rang 1: ANOVA (15/0/0)

Rang 2: je 14/1/0

- Lin. Regression
- Poweranalyse
- **Versuchsplanung (Experimental Design)**

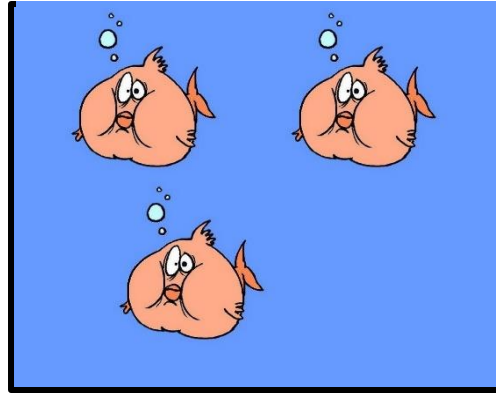
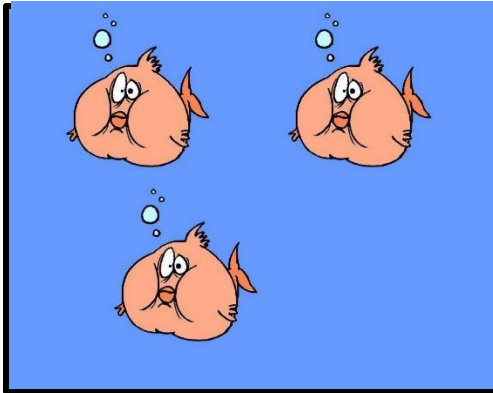
Zum Vergleich: Rang 9 - Mixed Models (5/4/6)

# Versuchsplanung – Experimental design

"It does not matter very much if you cannot do your own advanced statistical analysis. If your experiment is properly designed, you will often be able to find somebody to help you with the statistics. However, if your experiment is not properly designed, or not thoroughly randomized, or lacking adequate controls, then no matter how good you are at statistics, some (or possibly even all) of your experimental effort will have been wasted. No amount of high-powered statistical analysis can turn a bad experiment into a good one."

From: Crawley, M. J. (2005) Statistics – An Introduction Using R. Wiley, Chichester, UK.

# Studien gehen häufig schief...



Stichprobengrösse: 2 oder 6 ?

## ...und sollten deshalb sorgfältig geplant werden

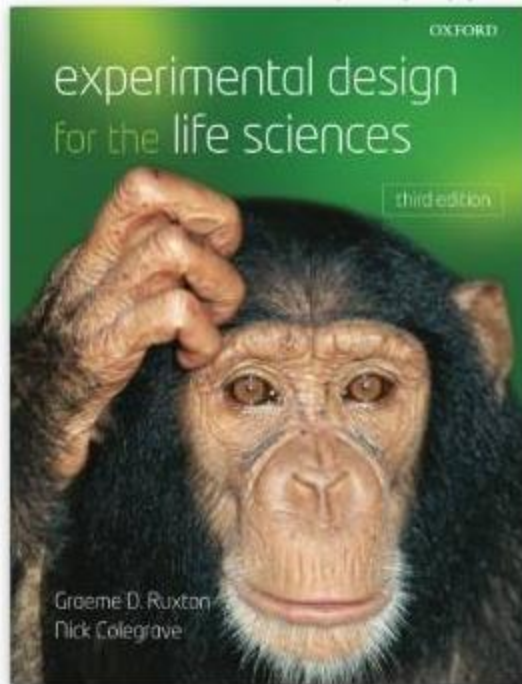
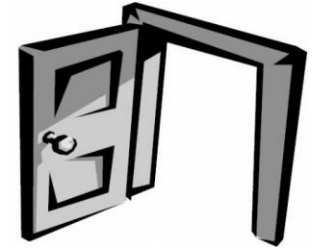
- Effizienz: Zeit und Kosten
- Ethik: Kein unnötiges Leid



**«Educated Guess»**

# Literatur: **Experimental Design**

“Türöffner“



**Taschenbuch:** 196 Seiten

**Verlag:** Oxford University Press, USA; Auflage: 0003 (28. November 2010)

**Sprache:** Englisch

**ISBN-10:** 0199569126

**ISBN-13:** 978-0199569120

**Buch nur als Ergänzung**

Via ETH Bibliothek erhältlich (leider nicht online)

# Plan für die Vorlesung

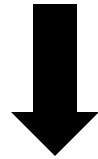
- **Hypothese** formulieren
  - Testbare Hypothese
  - Welche stat. Auswertung / Stichprobengrösse
  - Pilotstudie
  - Endlich: Daten erheben und auswerten
- **Prinzipien** für gutes experimentelles Design
  - Kontrolle
  - Experiment: Kausalität; Beobachtung: Assoziation
  - Replikate
  - Zufall / Randomisierung



**KERZ** – damit Ihnen ein Licht aufgeht 😊

# Wissenschaftliches Prinzip

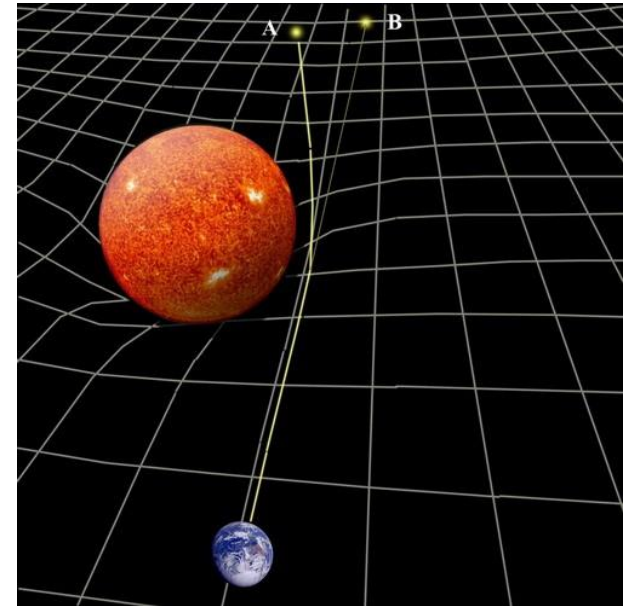
Hypothese



Zwingende, testbare Vorhersage  
der Hypothese



Realität





# Hypothesen: Wissenschaftliches Vorgehen

- Bsp: Verhalten von Schimpansen
- Variante 1:
  - 100 h Videoaufnahme im Zoo vom Schimpansengehege
  - Da müssen interessante Schlussfolgerungen drin sein
  - Zur Statistik-Beratung: “Welche mächtige Methode soll ich verwenden, damit ich einen signifikanten p-Wert bekomme ?”
  - Beratungsteam:



# Testbare Hypothesen

- Beobachtung: Aktivität der Schimpansen variiert am Tag
- **Hypothese:** “Aktivität hängt vom Futterzeitplan ab.”
- Entsprechende testbare Vorhersage:  
«Die Zeit, die ein Schimpanse in Bewegung verbringt ist in der Stunde vor dem Füttern grösser als in jeder anderen Stunde am Tag.»
- **Nullhypothese:** “Aktivität hängt **nicht** vom Futterzeitplan ab.”
- Entsprechende testbare Vorhersage:  
«Bzgl. der Bewegungszeit gibt es keinen Unterschied zwischen der Stunde vor dem Füttern und jeder anderen Stunde am Tag.»

## Plane stat. Auswertung

- Welche Daten genau werden erhoben?  
Z.B.: Videoaufnahme über 100 h in 5 Zoo's  
Für jeden Zoo: Durchschn. Bewegungszeit in der Std. vor Futter und durchschn. Bewegungszeit in übriger Zeit
- Welche Auswertungsmethode wird angewendet?  
Gepaarter t-Test (für jeden Zoo: Eine durchschn. Zeit für die Stunde vor Futter und eine durchschn. Zeit für die übrige Zeit)

# Pilotstudie

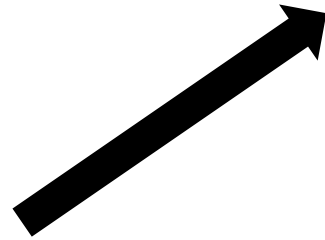
- Beobachte in drei Zoos für je einen Tag
- Erfahrungen mit Datenerhebung (Wo stellt man Kamera auf, was kann schief gehen, ...)
- Stichprobengrösse: Wie viele Zoo's müssen wir untersuchen?  
→ Power Analyse beim gepaarten t-Test (Varianz aus Pilotstudie)
- Haben wir die richtigen Daten für die Auswertung erhoben?

## Also: Erst denken, dann handeln

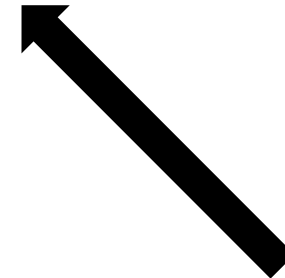
**Hypothese** formulieren

- Testbare Hypothese
- Welche stat. Auswertung / Stichprobengrösse
- Pilotstudie
- Endlich: Daten erheben und auswerten

## Praktisches Beispiel: Arbeitsweg



# Womit ?



# Hypothese 1: “Laufen ist die beste Methode.”

- “Beste” ist subjektiv; Personen haben unterschiedliche Vorlieben

Laufen ist sicher gesund, aber...

- es ist anstrengend
- im Tram kann man lesen
- die Polybahn ästhetisch ansprechender
- ...

Zu subjektiv

## Hypothese 2: “Laufen ist die schnellste Methode.”

- Macht es einen Unterschied, ob man müde oder ausgeruht ist ?
- Wenn man an der Polybahn nicht warten muss, ist man evtl. schneller; sonst langsamer
- Wenn das Tram gerade kommt, ist man evtl. schneller, sonst langsamer
- Je nach Uhrzeit bleibt das Tram am Zentral stecken...

Zu unpräzise



## Hypothese 3: “Laufen ist im Mittel die schnellste Methode.”

Im Mittel über was?

- Alle Tage, Werktage, Wochenenden ?
- Alle Uhrzeiten (z.B. nachts um 3:00) ?

Zu unpräzise

## Hypothese 4: “Wenn man werktags (planmässig) mit dem Zug um 7.52 ankommt, ist zügiges Laufen im Mittel die schnellste Methode an die ETH zu kommen.”

- Damit könnten wir ein Experiment designen (s. später)
- Mögliche Auswertung:  
 $H_0$ : «Alle drei Methoden sind im Mittel gleich schnell»  
→ 1-weg ANOVA  
Tukey-HSD um signifikante Unterschiede zwischen Methoden zu quantifizieren
- Die Hypothese könnte noch präziser sein (Sommer, Winter; Semester oder Ferien) → Kompromiss

Präzision



Pragmatik

# Plan für die Vorlesung



## ■ Hypothese formulieren

- Testbare Hypothese
- Welche stat. Auswertung / Stichprobengrösse
- Pilotstudie
- Endlich: Daten erheben und auswerten

## ■ Prinzipien für gutes experimentelles Design

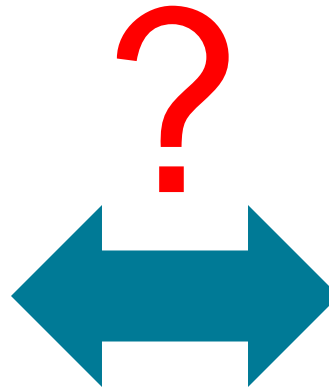
- Kontrolle
- Experiment: Kausalität; Beobachtung: Assoziation
- Replikate
- Zufall / Randomisierung



**KERZ** – damit Ihnen ein Licht aufgeht 😊

# Ursache und Wirkung

Opfer durch Ertrinken

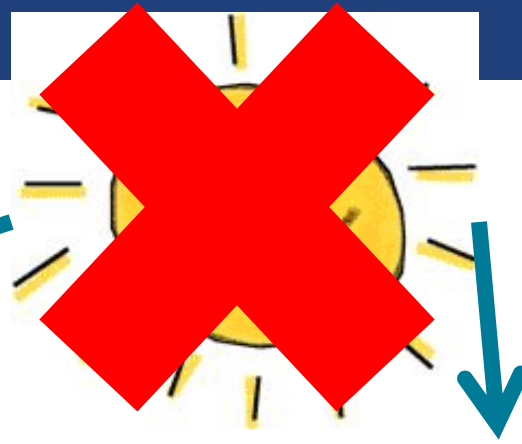


Eisverkauf



# Ursache und Wirkung

Opfer durch Ertrinken



Eisverkauf



## «Spurious Correlation»

- Oftmals gibt es zwischen zwei Variablen eine Korrelation, aber keinen Kausalzusammenhang  
→ «Spurious» Correlation
- Z.B.: Schokoladenkonsum und Nobelpreiszahl sind korreliert; damit ist aber noch nicht bewiesen, dass es einen Kausalzusammenhang gibt

# Mehr «Spurious Correlation»

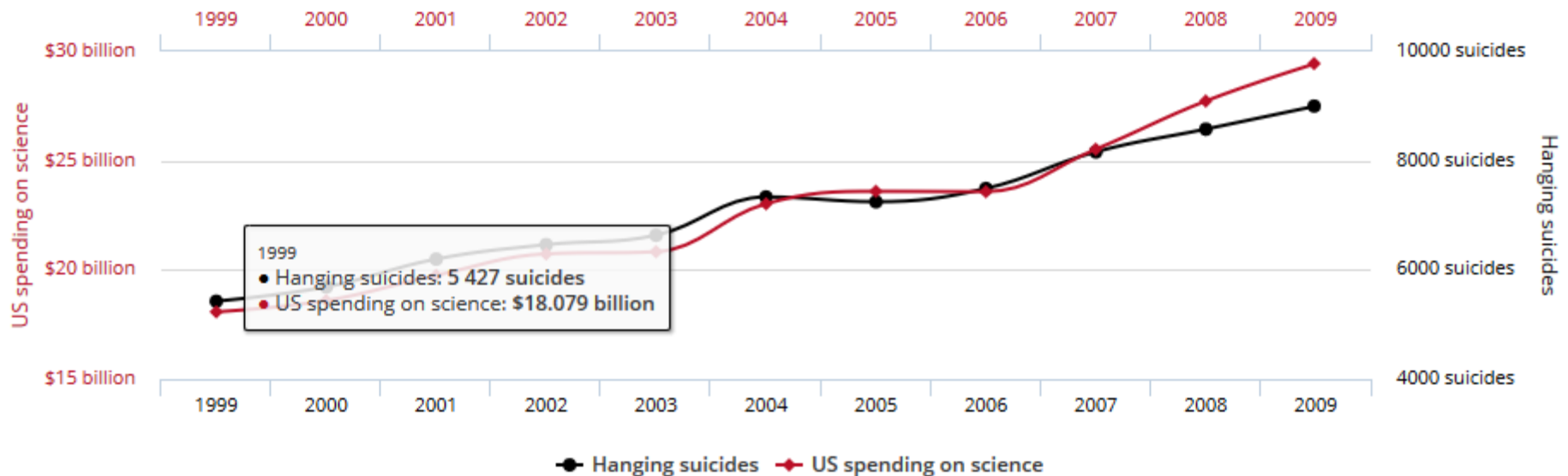
Für lange Winterabende:

<http://www.tylervigen.com/spurious-correlations>

# Bsp 1: Wissenschaft treibt in den Selbstmord...

**US spending on science, space, and technology**  
correlates with  
**Suicides by hanging, strangulation and suffocation**

Correlation: 99.79% ( $r=0.99789126$ )



tylervigen.com

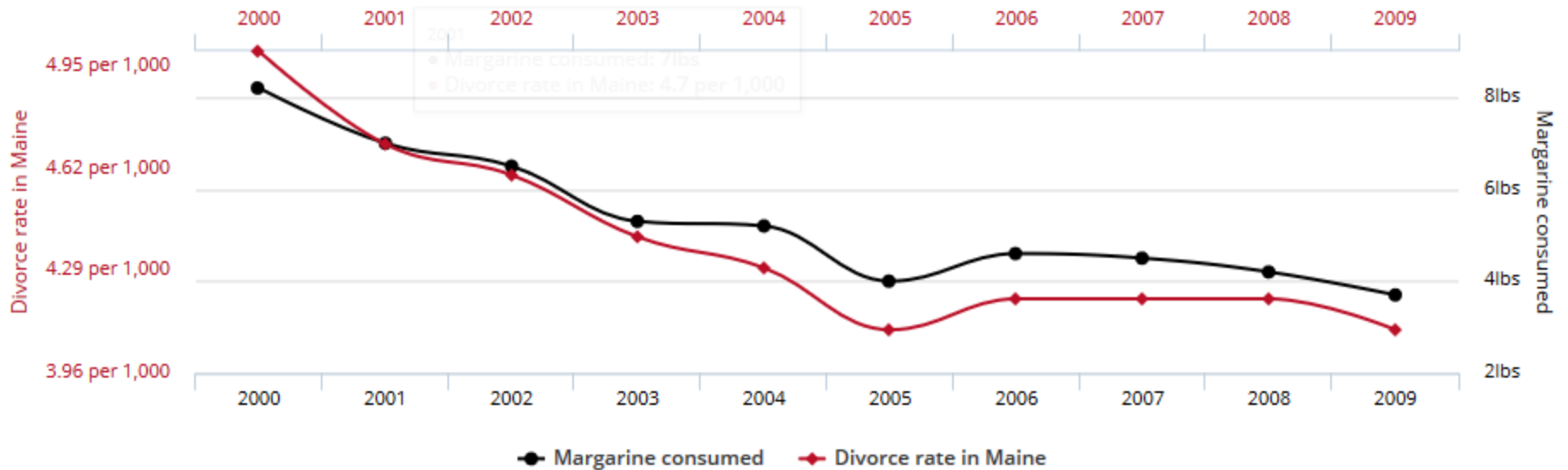
Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention



## Bsp 2: Scheidungsgrund

### Divorce rate in Maine correlates with Per capita consumption of margarine

Correlation: 99.26% ( $r=0.992558$ )

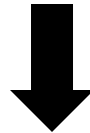


tylervigen.com

Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

# Fazit

Mit Korrelation alleine  
ist es kaum möglich  
Kausalzusammenhänge  
zu beweisen



Experimente

# Kausaler Zusammenhang

≠

# Korrelation

# Prinzipien für gutes experimentelles Design

- Kontrolle
  - Experiment: Kausalität; Beobachtung: Assoziation
  - Replikate
  - Zufall / Randomisierung
- 
- KERZ – damit Ihnen ein Licht aufgeht 😊



Randomized Controlled Experiment  
with Replicates (RCT)



# Kausaleffekt finden



Experiment

# Kausaleffekt finden



# Kausaleffekt finden



# Kausaleffekt finden



In der Realität sind alle Felder leicht unterschiedlich:  
Wie wäre das Ergebnis auf einem anderen Feld ?

## Brauchen Replikate.



Experiment

# Kausaleffekt finden



# Kausaleffekt finden



# Kausaleffekt finden



# Kausaleffekt finden



Dünger besser als kein Dünger?

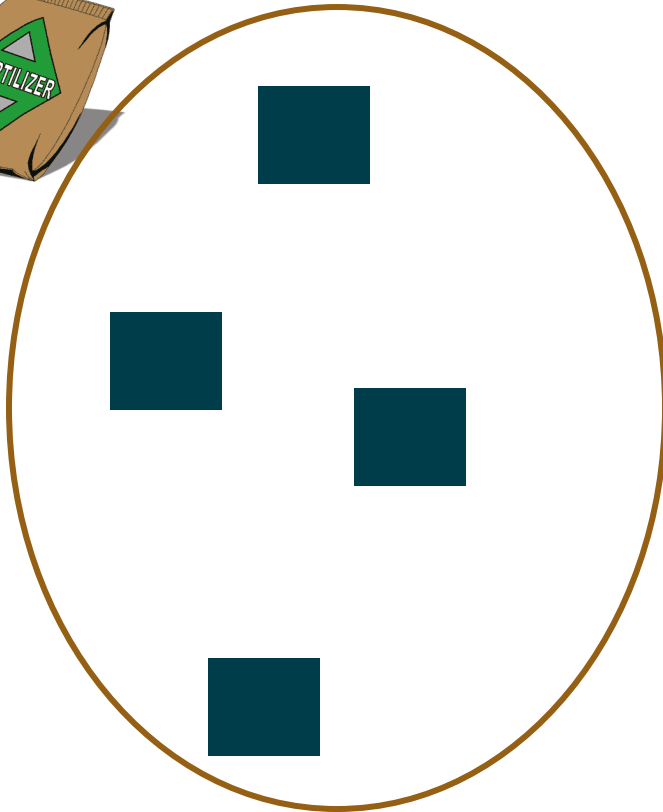
Keine Ahnung!

Wie viele rote Blumen hätte es ohne Dünger gegeben?

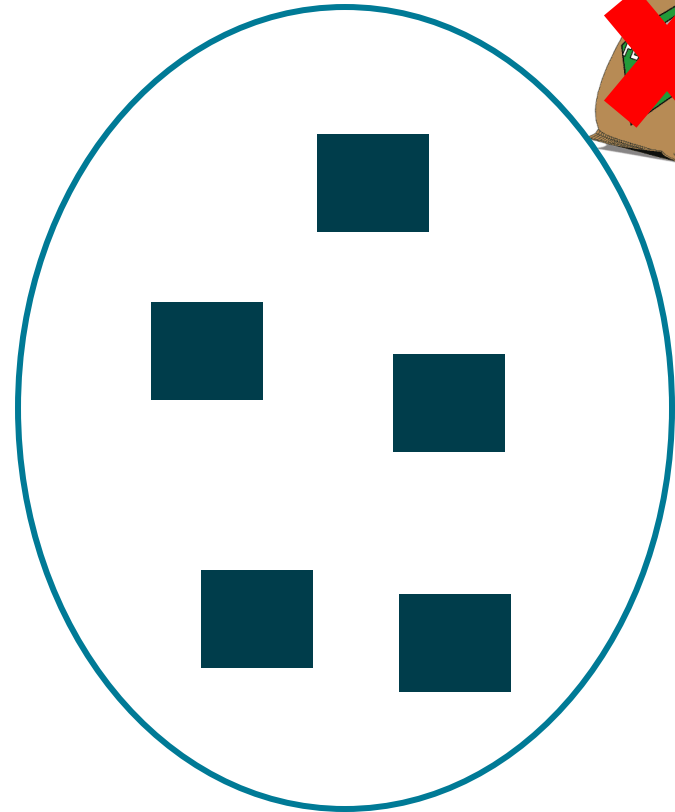
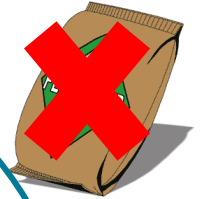
## Brauchen eine Kontrollgruppe

Experiment

# Kausaleffekt finden



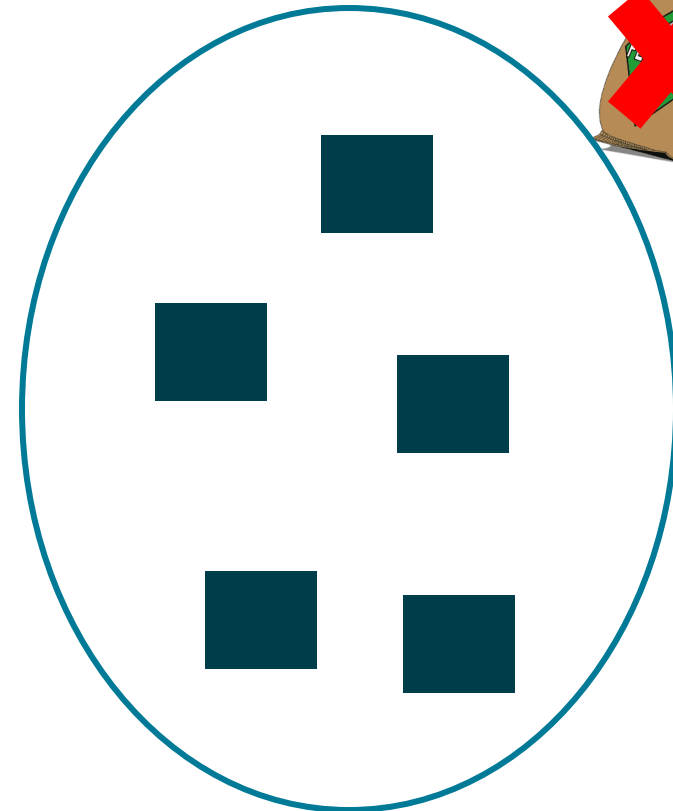
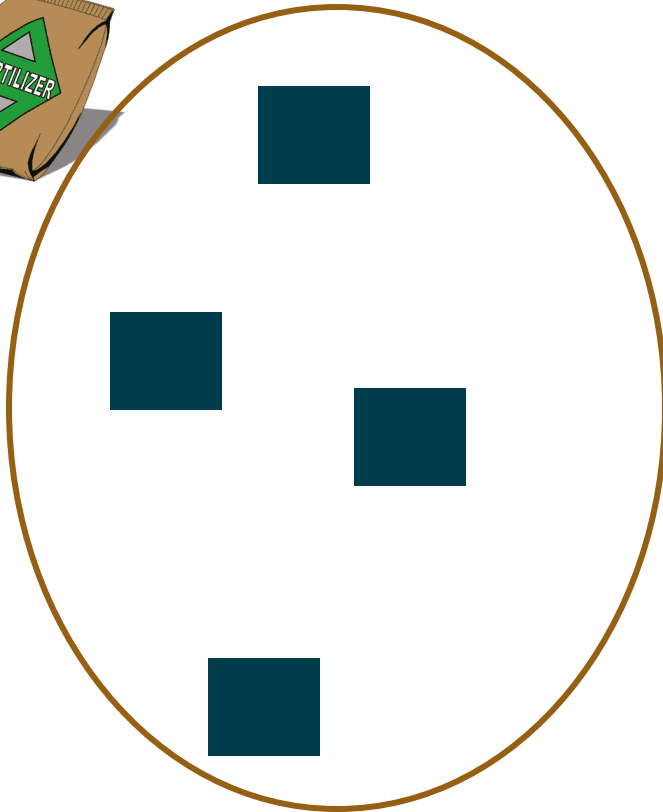
Behandlungsgruppe



Kontrollgruppe

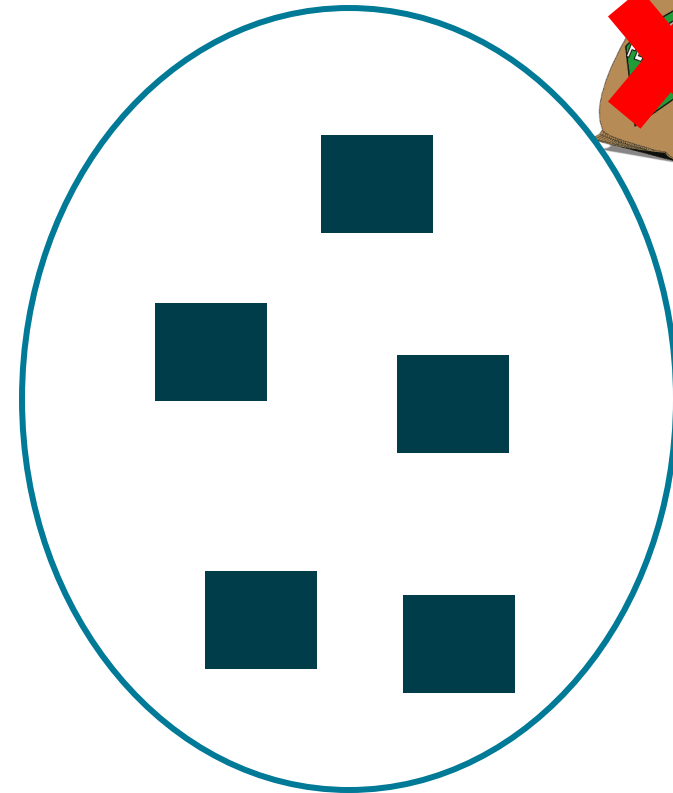
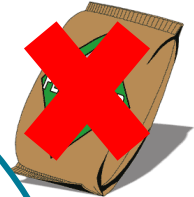
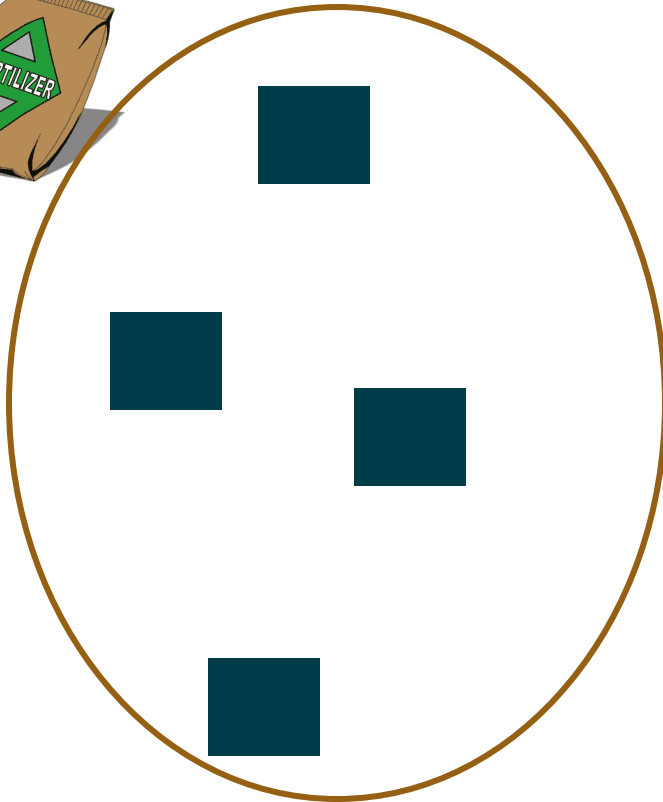
## Experiment

## Kausaleffekt finden



Zwei Gruppen von Feldern in allem gleich  
(Bodenqualität, Wasser, Sonnenlicht, ...)

# Kausaleffekt finden

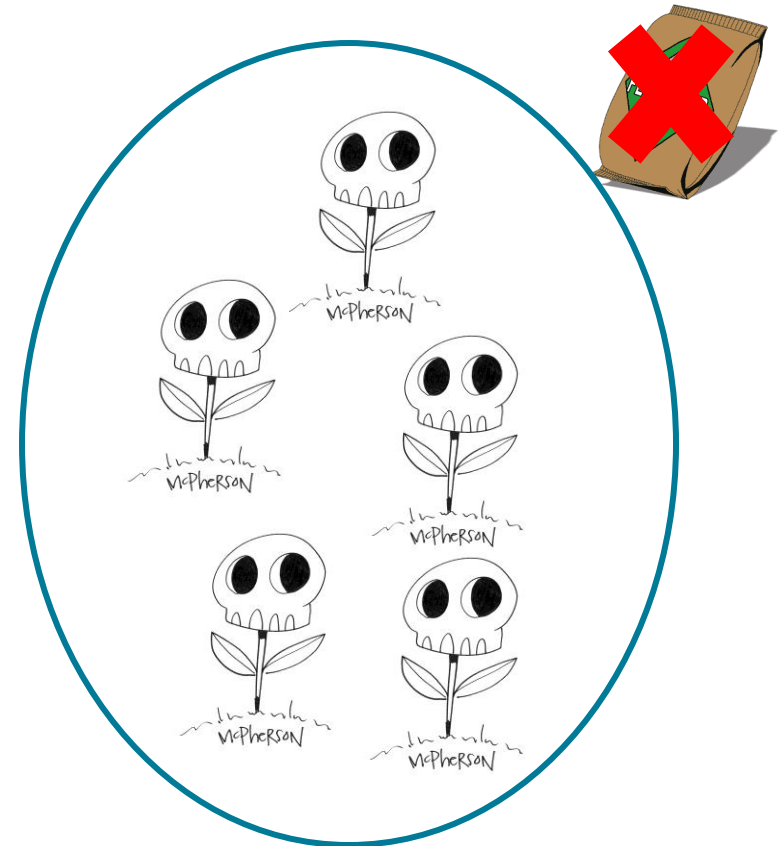
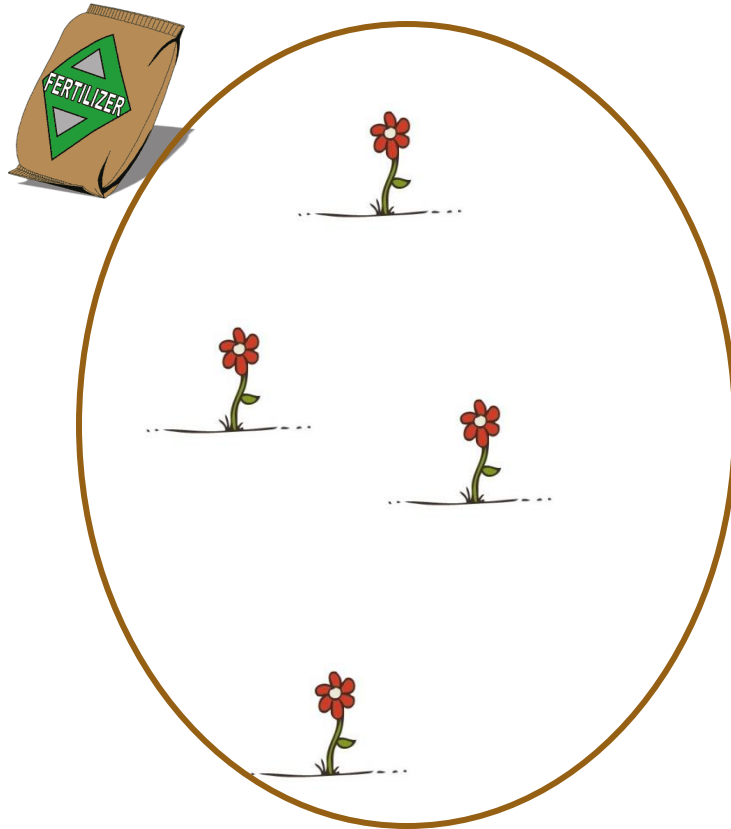


Zwei Gruppen von Feldern in allem gleich:  
(Bodenqualität, Wasser, Sonnenlicht, ...)

**Praxis: Randomisieren =  
Zufällige Zuordnung der Felder**

Experiment

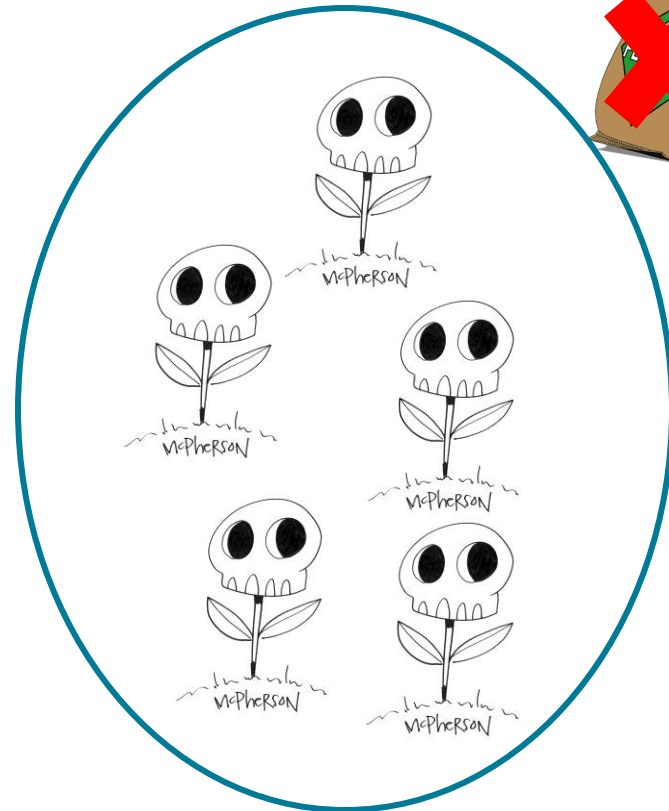
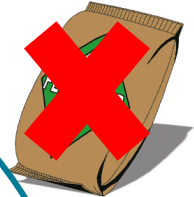
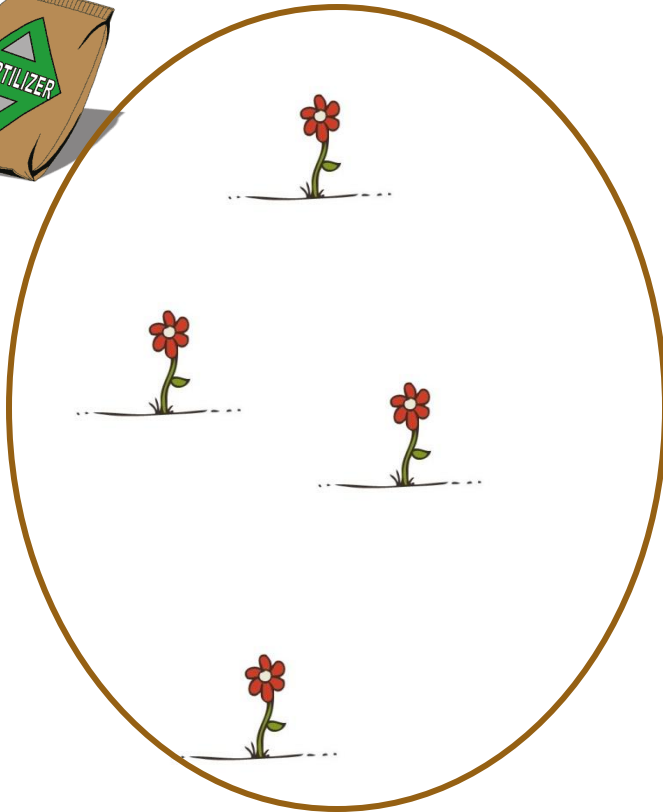
# Kausaleffekt finden





## Experiment

## Kausaleffekt finden



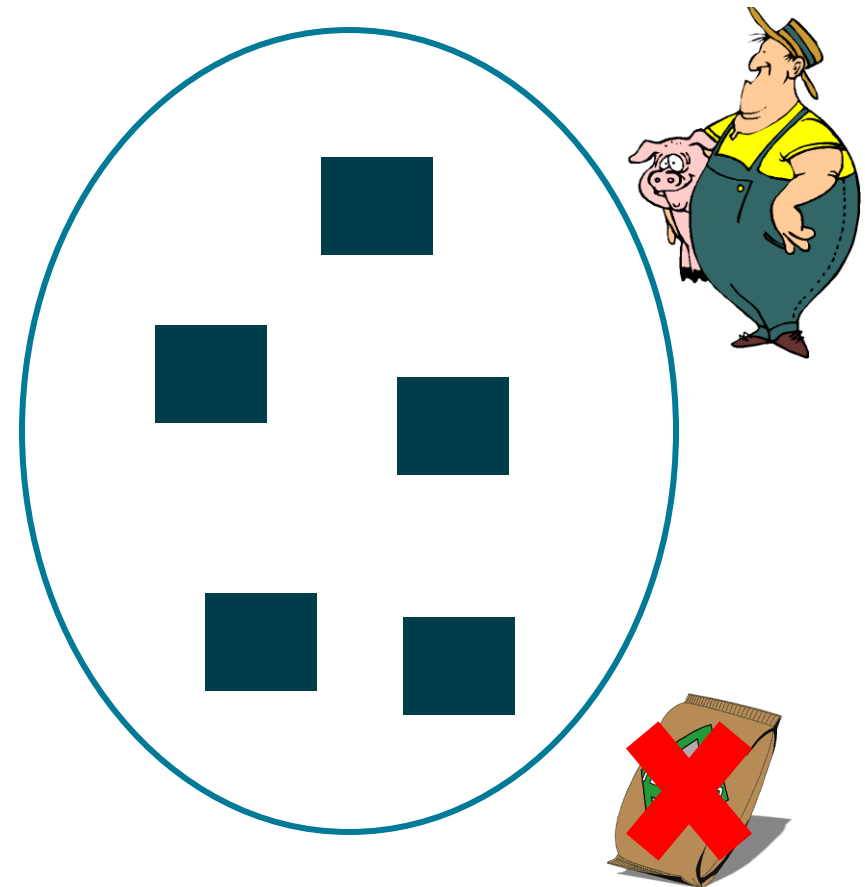
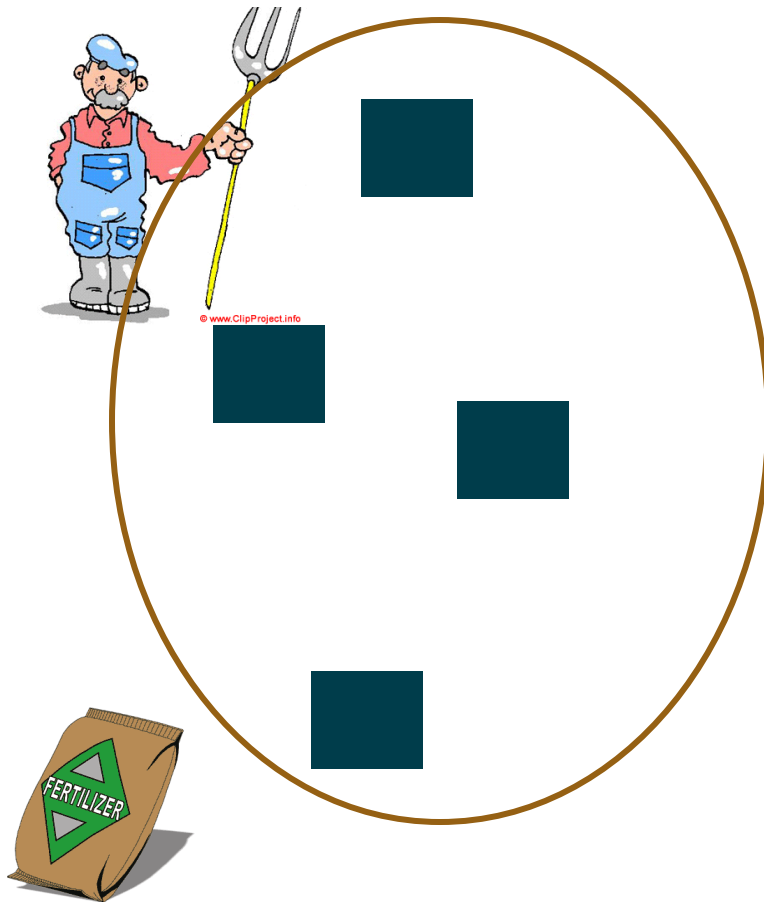
Ergebnis ist wegen Dünger,  
weil alles andere gleich war

# Manchmal sind randomisierte, kontrollierte Experimente nicht machbar

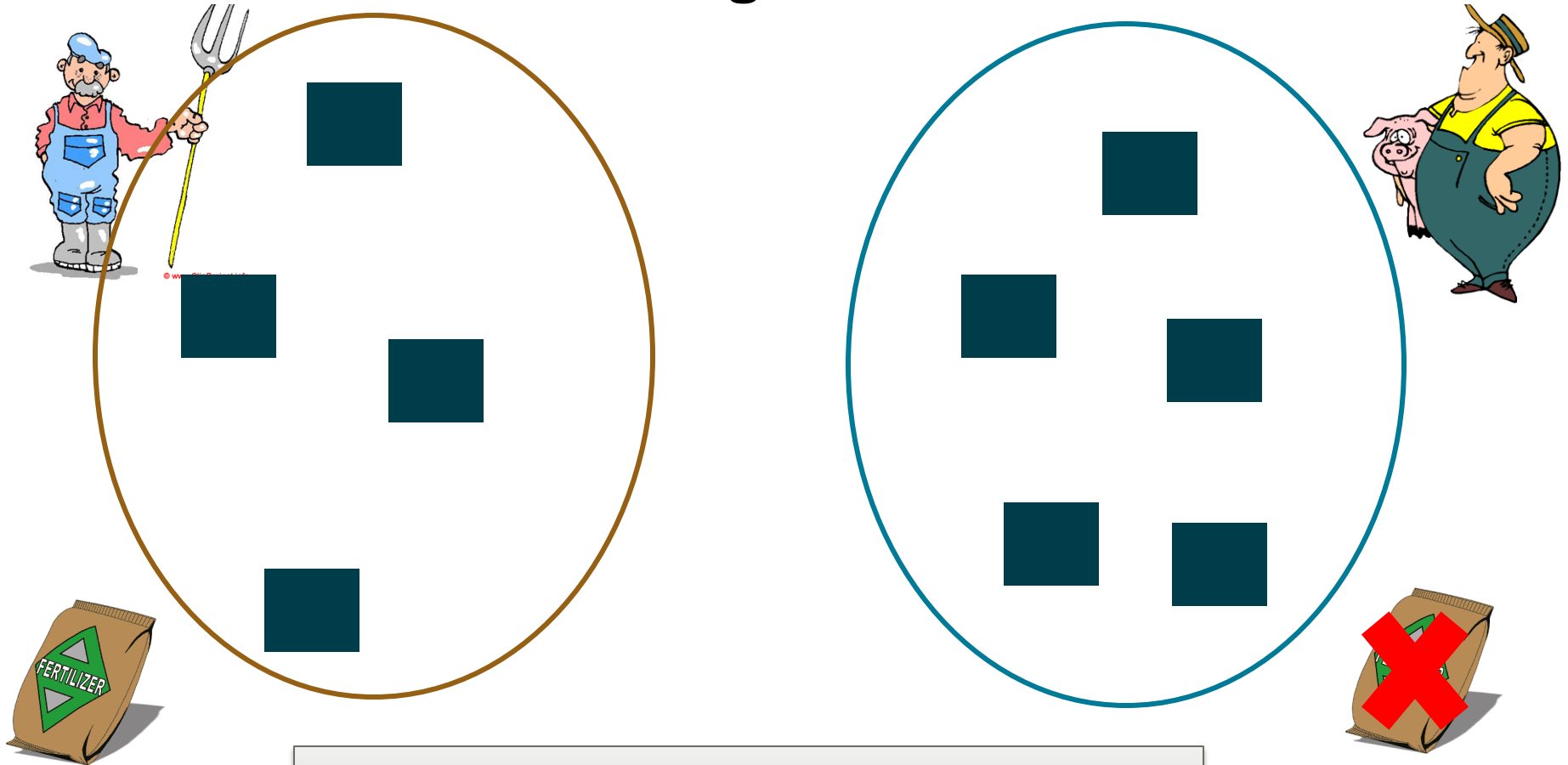
- zu teuer, zu zeitaufwändig (Genexpressionsdaten)
- unethisch, nicht machbar (HIV Behandlung, Rauchen)

# Falls Experiment nicht machbar...

... mache Beobachtungen.

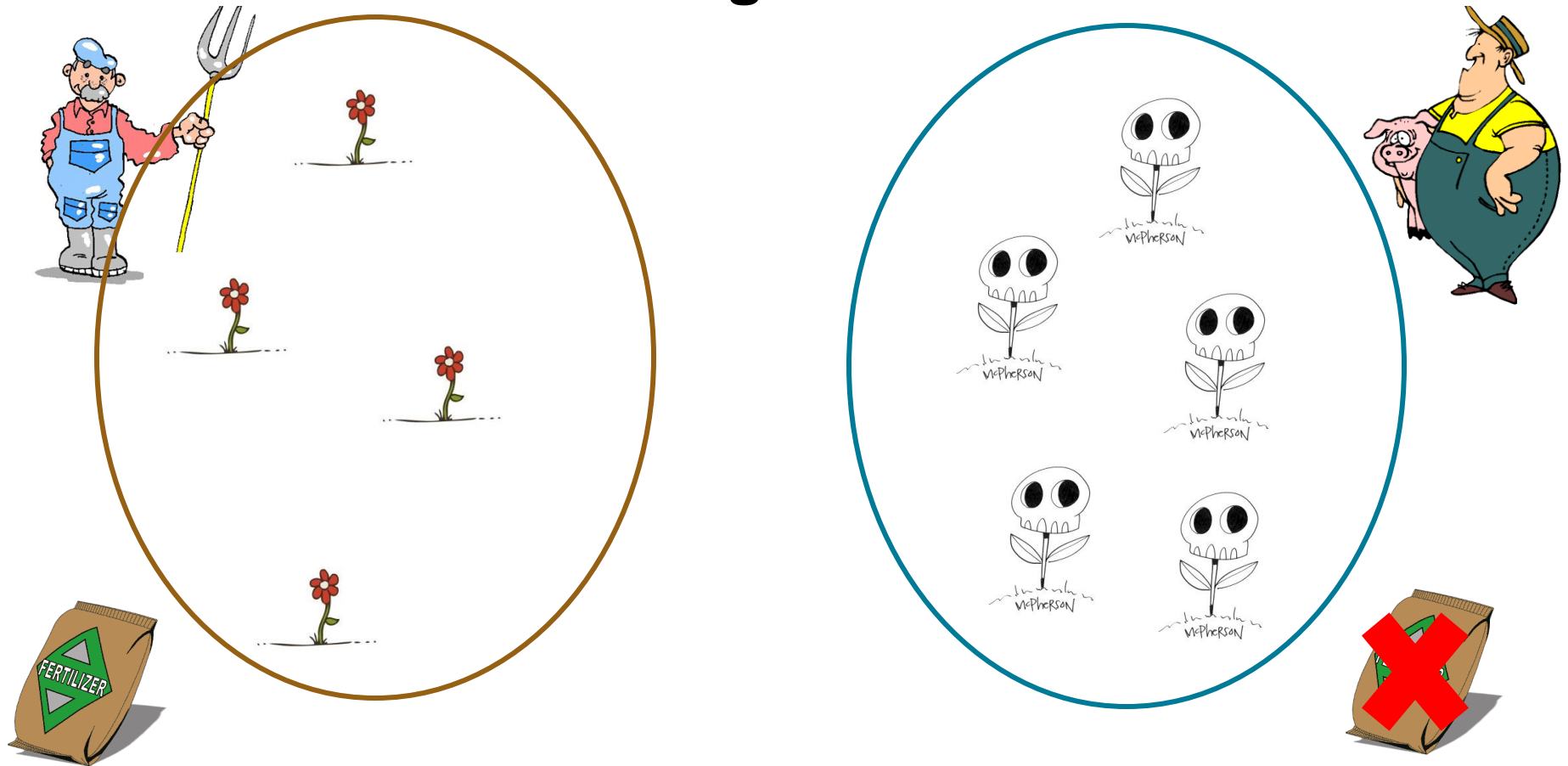


... mache Beobachtungen.

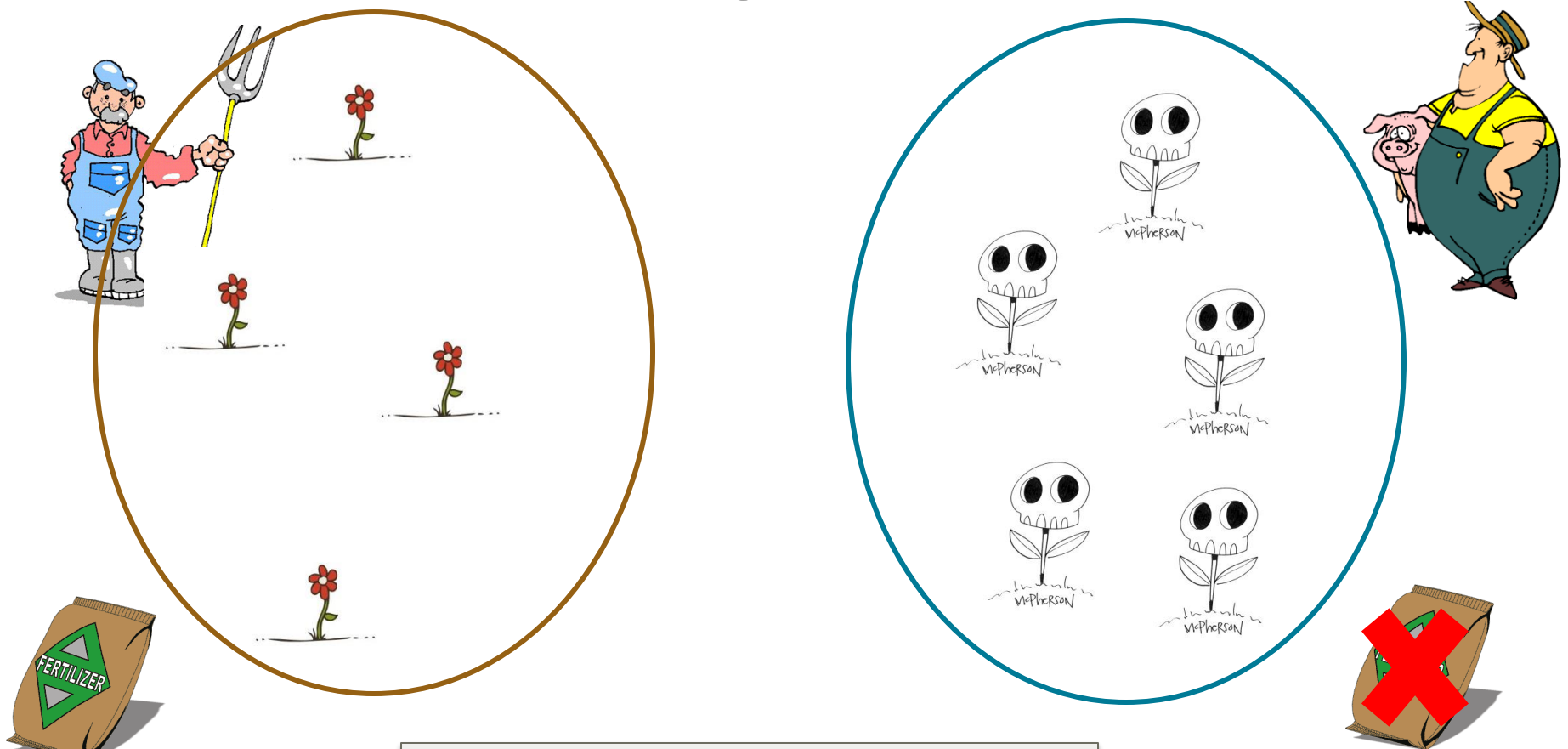


Es ist nicht garantiert, dass beide Gruppen  
in allen Aspekten gleich sind

... mache Beobachtungen.

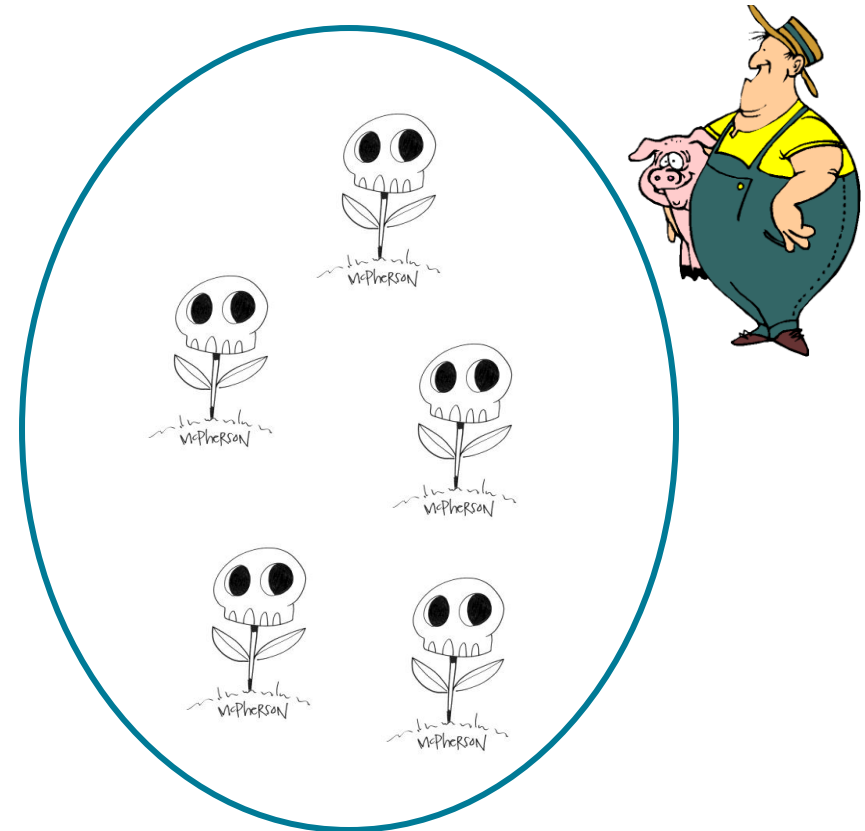
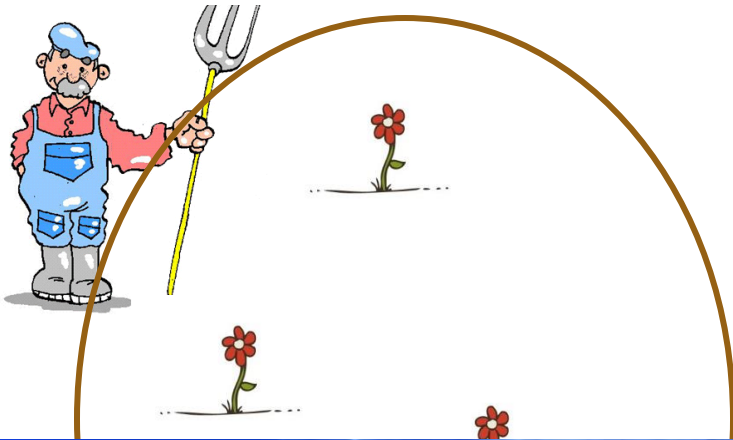


... mache Beobachtungen.



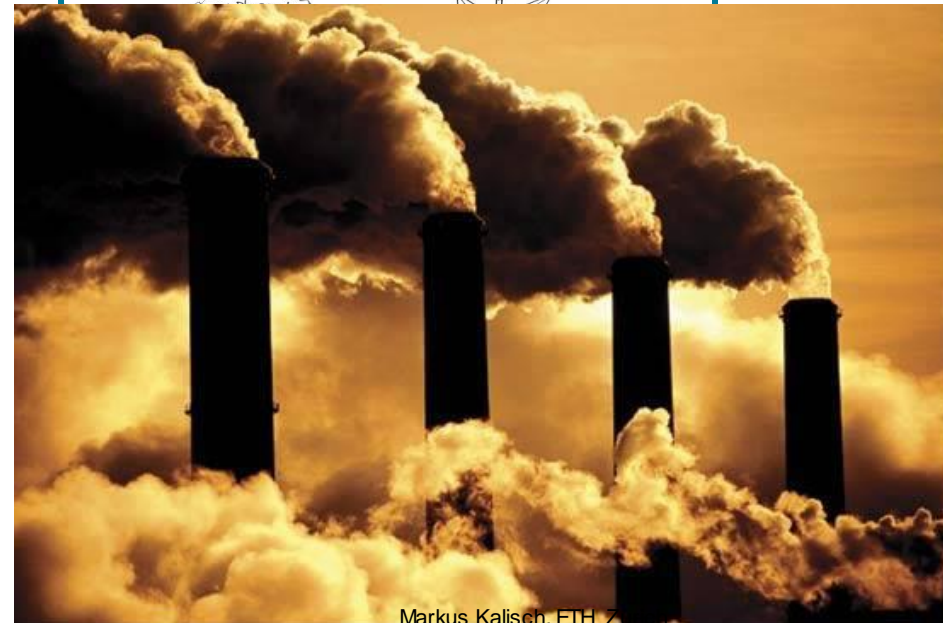
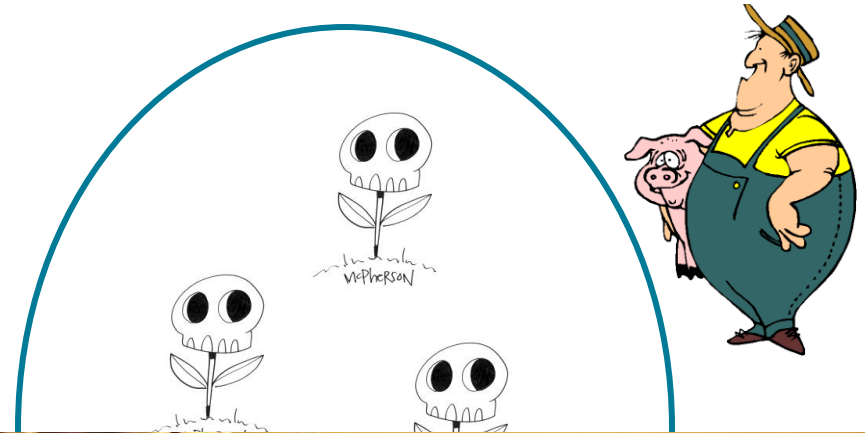
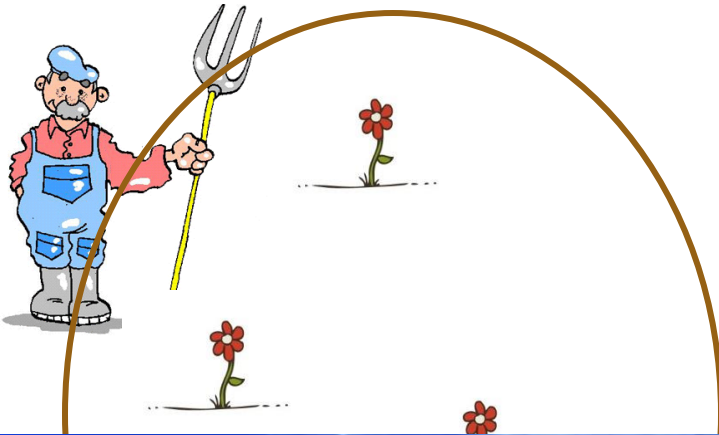
Ist das Ergebnis wegen Dünger?  
Keine Ahnung!

... mache Beobachtungen.

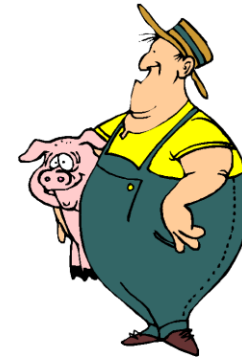




... mache Beobachtungen.

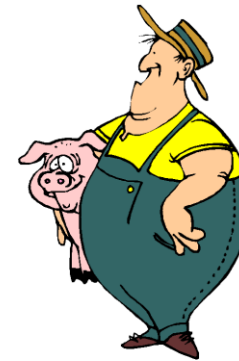


**Besser: Vergleiche Bauern, die in möglichst vielen Punkten übereinstimmen.**



Aber: Wir können nie sicher sein, dass es nicht doch noch irgendwelche **relevanten Unterschiede zwischen den Gruppen** gibt.

Beobachtungsstudie

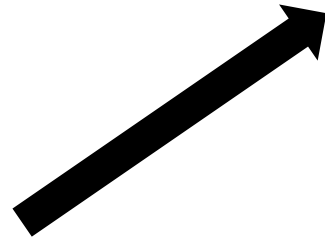


# Zusammenfassung

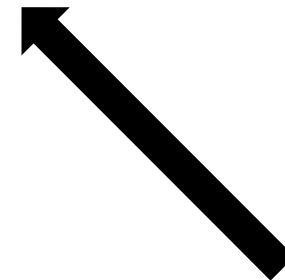
- **Randomisierte, kontrollierte Experiment mit Replikaten:** Beste Möglichkeit, Daten zu sammeln (“Goldstandard”)
- **Beobachtungsstudie:** Man muss skeptisch sein – kam der Effekt (viele schöne Blumen) durch die Behandlung (Dünger), oder durch einen Umstand, der in beiden Gruppen unterschiedlich war (Luftqualität)?



## Praktisches Beispiel: Arbeitsweg



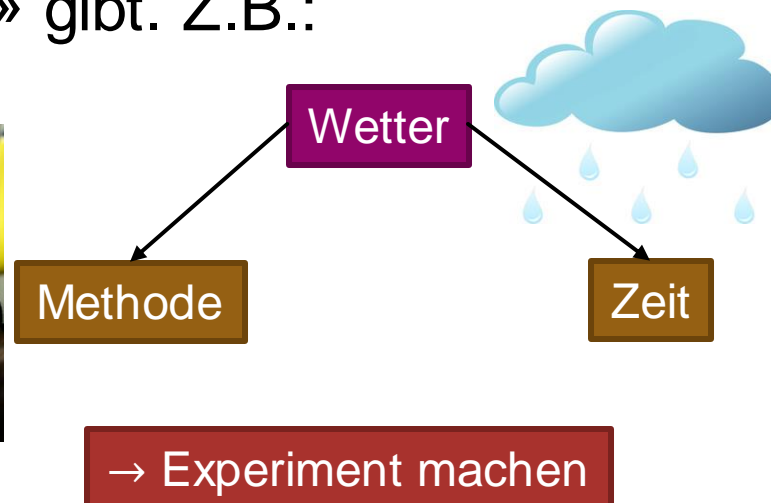
# Womit ?





# Arbeitsweg: Beobachtung oder Experiment ?

- Beobachtung: Evtl. Daten von Smartphone von den letzten Monaten ?
- Wurden alle drei Methoden häufig (genug) verwendet ?
- Wir **können nicht sicher sein**, dass es keinen «Confounder» gibt. Z.B.:



## Arbeitsweg: Replikate

- Pilotstudie: Vernünftige Stichprobengrösse
- Oder Überlegung:
  - Typische Schwankung der Zeit ca. 2 Minuten = 120 Sek
  - Wie viele Beobachtungen braucht man, damit Schwankung des Mittelwerts  $\sim 30$  Sek  
( $=1/4$  der Originalstreuung)
  - $\sqrt{n}$ -Gesetz: 16 Beobachtungen
  - Also je Methode 16 Beobachtungen

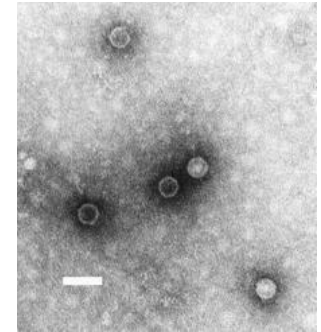
## Arbeitsweg: Kontrolle

- Tram und Polybahn dienen als Kontrolle zur Methode Laufen.
- Mögliche Auswertung:  
1-weg ANOVA  
Paarweise Vergleiche oder Kontrast «Laufen» vs. «Rest»



# Beispiel: Polio

Poliovirus  
Hochansteckend  
v.a. Kinder



Q: Wikipedia - Poliovirus

## Outcomes of poliovirus infection

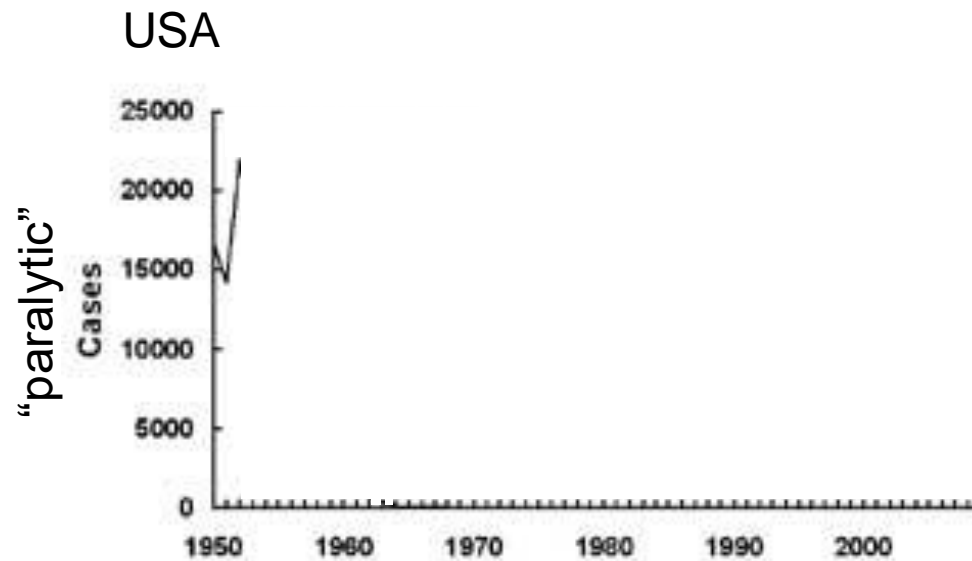
Outcome	Proportion of cases <sup>[1]</sup>
No symptoms	90–95%
Minor illness	4–8%
Nonparalytic aseptic meningitis	1–2%
Paralytic poliomyelitis	0.1–0.5%
— Spinal polio	79% of paralytic cases
— Bulbospinal polio	19% of paralytic cases
— Bulbar polio	2% of paralytic cases

Q: Wikipedia - Poliomyelitis



Q: <http://www.polioeradication.org/Polioandprevention.aspx>

## Anfang 20. Jh.: Polio-Epidemie in Entwicklungsländern



<http://www.cdc.gov/vaccines/pubs/pinkbook/polio.html>

# 1954: Impfstoff ?

- Jonas Salk:  
Vielversprechender Impfstoff im Labor
- Nutzen vs. Risiken ?

→ **Experiment**



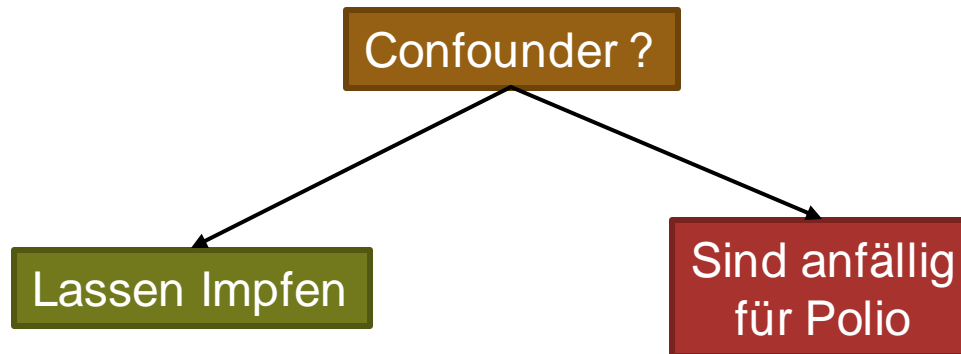
Q: Wikipedia – Jonas Salk

# Ist Impfstoff wirksam ?

- Mögliches Experiment: 1954 werden alle Kinder geimpft
- **Problem:** Polio ist von Jahr zu Jahr sehr unterschiedlich stark verbreitet
  - 1954 könnte ein Jahr mit natürlicherweise wenig Polio sein
  - wenige Poliofälle, obwohl Impfstoff evtl. gar nicht wirkt
- Brauchen **Kontrollgruppe**

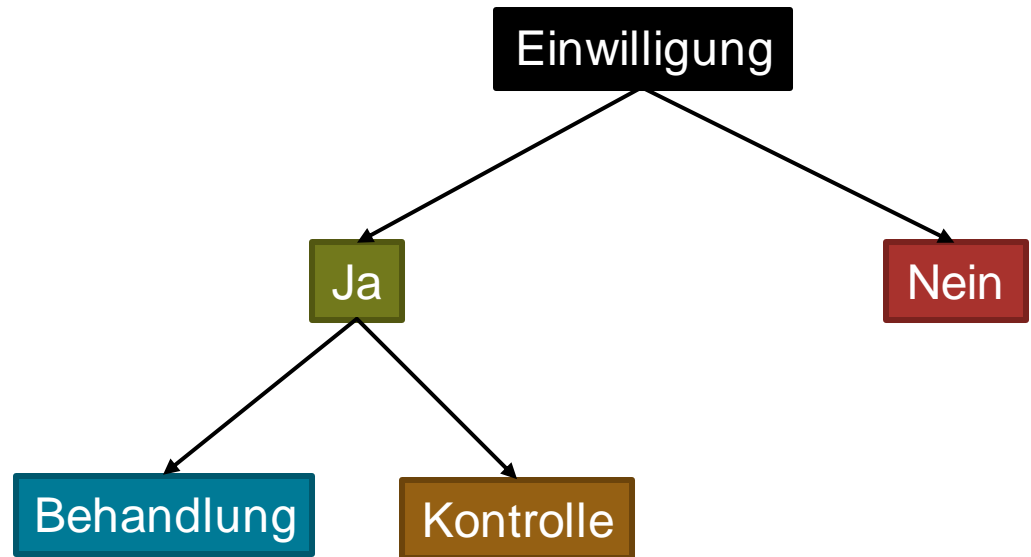
## Welche Kontrollgruppe ?

- Eltern müssen mit Impfung einverstanden sein  
→ Möglichkeit 1:  
Kinder, die geimpft werden dürfen → Behandlungsgruppe  
Kinder, die nicht geimpft werden dürfen → Kontrolle
- **Problem:**



# Welche Kontrollgruppe ?

- Eltern müssen mit Impfung einverstanden sein
- Möglichkeit 2: **Randomisierung**



# Doppelblind

- Patienten wussten nicht, ob in Behandlungs- oder Kontrollgruppe
- Behandelnde Ärzte wussten nicht, ob in Behandlungs- oder Kontrollgruppe



# Das grösste medizinische Experiment aller Zeiten

Randomisiertes, kontrolliertes Experiment bei Kindern in 1. bis 3. Schulklasse

[<http://www.stat.luc.edu/StatisticsfortheSciences/MeierPolio.htm>]

	Anzahl Kinder	Polio bekommen
Behandlung	200.000	56
Kontrolle	200.000	142
Verweigert	350.000	161

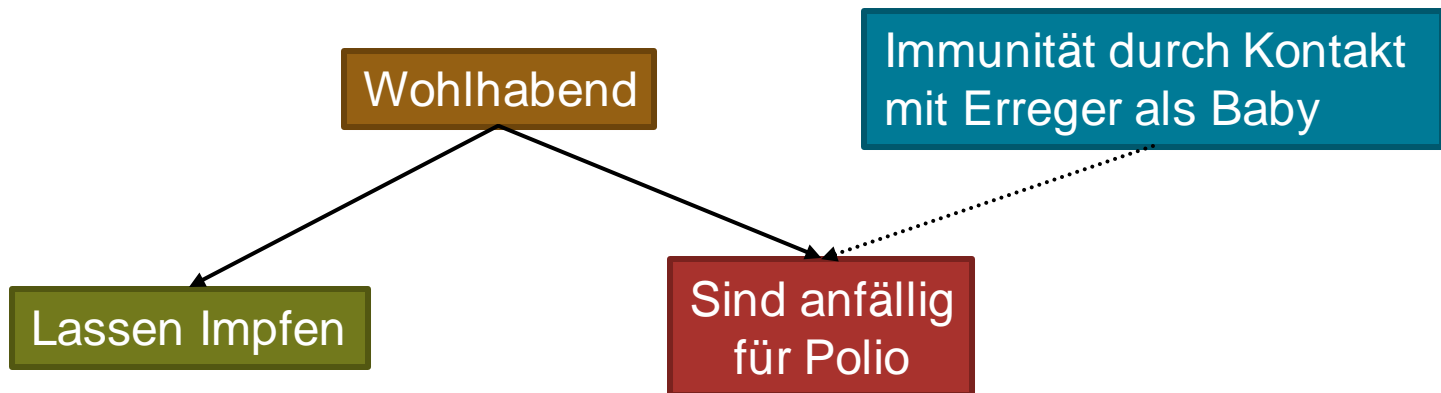
- Kontrollgruppe hatte mehr Polio-Fälle: Könnte das Zufall sein?
- KAUM: p-Wert = 0.000000001 (z.B. Fisher-Test)



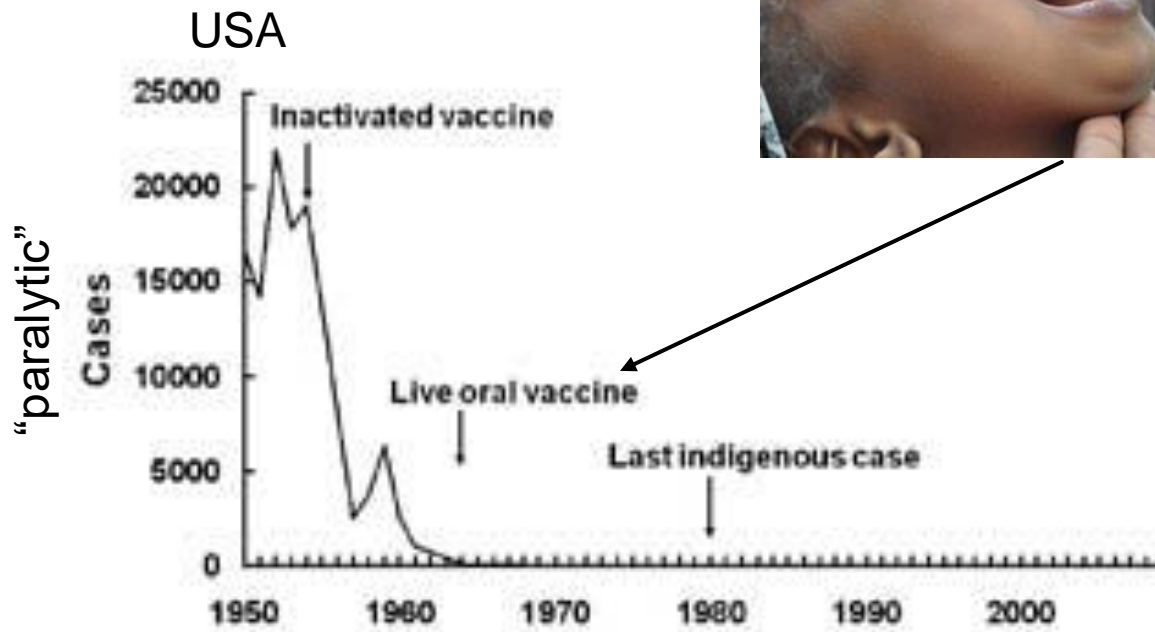
# Hygieneeffekt: Zum Glück hatten wir eine gute Kontrollgruppe

	Anzahl Kinder	Polio bekommen
Behandlung	200.000	56
Kontrolle	200.000	142
Verweigert	350.000	161

Hygieneeffekt ist sichtbar  
Verweigerer haben überraschend kleine Polio-Rate



# Polio nach 1954



<http://www.cdc.gov/vaccines/pubs/pinkbook/polio.html>

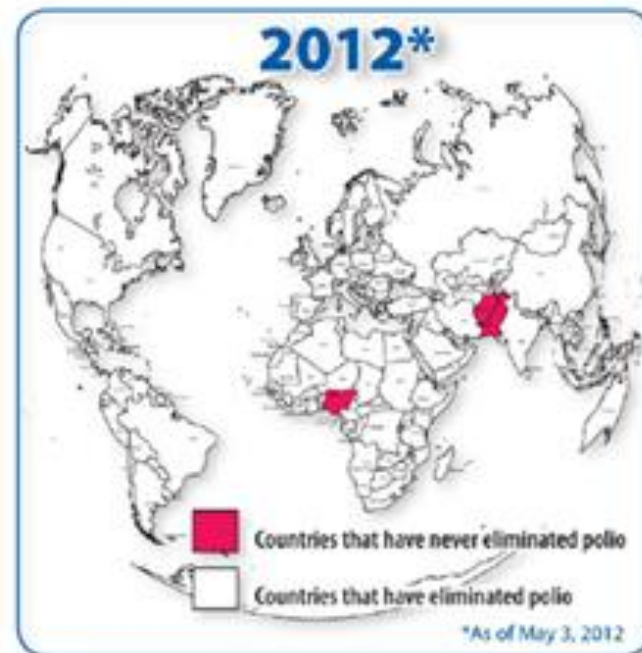


# Weltweit: “End Polio Now!”

- 1988: <http://www.polioeradication.org/>
- Aktuelles Ziel: 2018



~350 000  
Fälle



2013:  
**312 Fälle**  
Endemisch:  
Nigeria  
Pakistan  
Afghanistan

<http://blogs.cdc.gov/genomics/2013/01/24/connection-between-polio-eradication-and-primary-immunodeficiency/>

# Polio: End Game

- 2016: Nur 37 Fälle (endemisch)
- Typ 1,2 & 3 Virus:  
Typ 1 & 2 gelten als ausgerottet



Quelle: [http://polioeradication.org/wp-content/uploads/2017/08/AR2016\\_EN.pdf](http://polioeradication.org/wp-content/uploads/2017/08/AR2016_EN.pdf)

# More recent: RCT in policy making & developmental economics

- Malaria-Netze (<http://www.rollbackmalaria.org/>):  
<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD000363.pub2/full>
- Linsen als Anreiz zur Impfung
- Wie verbessert man Schulbildung möglichst effektiv ?



TED

<http://www.povertyactionlab.org/duflo>

# Zusammenfassung

- Hypothese
- Kontrolle
- Experiment: Kausalität; Beobachtung: Assoziation
- Replikate
- Zufall / Randomisierung
- KERZ – damit Ihnen ein Licht aufgeht 😊

