

# Evaluation Randomisiert-Kontrollierter Studien und Experimente mit R

## Randomisiert-Kontrollierte Studien: Aufbau & Hintergrund

---

Prof. Dr. David Ebert & Mathias Harrer

Graduiertenseminar TUM-FGZ

Psychology & Digital Mental Health Care, Technische Universität München

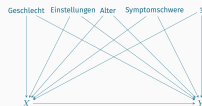
## Block I | R Entdecken

- ✓ Einführung in R & RStudio
- ✓ R Basics: Objektklassen, Funktionen, Operatoren, ...



## Block II | Einführung & Hintergrund

- ✓ Methodischer Hintergrund von randomisierten Studien
- ✓ Import und Manipulation von Studiendaten mit R
- ✓ Schätzung fehlender Werte



## Block III | Analysieren & Verstehen

- ✓ Statistische Wirksamkeitsanalyse
- ✓ Alphakorrektur, Effektstärkenberechnung, Visualisierung, ...



# **Randomisiert-Kontrollierte Studien: Kernmerkmale**

---

*"Randomized controlled trials are the **most rigorous** way of determining whether a **cause-effect relation** exists between **treatment and outcome** and for assessing the **cost effectiveness** [sic] of a treatment."*

– Sibbald & Roland (1998)

Randomisiert-kontrollierte Studien werden häufig als **"Goldstandard"** zum Nachweis der Wirksamkeit einer Intervention angesehen.

(Akobeng, 2005; Backmann, 2017; Hariton & Locascio, 2018; Lilienfeld et al., 2018)

Dieser Status wird auch **teils hinterfragt und relativiert.**

(z.B. Cartwright, 2007, 2011; Grossman & Mackenzie, 2005; Kaptchuk, 2001; Pearl & Mackenzie, 2018, Kapitel 4).

### Kernmerkmale Randomisiert-Kontrollierter Studien

- **Zufällige Zuordnung** zu Experimental- und Kontrollbedingungen.
- Teilnehmende und Studienpersonal wissen nicht, wem welche Bedingung zugeordnet wurde ("**doppelte Verblindung**"). Dies ist oftmals aber nicht möglich!
- Alle Gruppen erhalten die gleiche Behandlung, bis auf die Experimentalgruppe
- Teilnehmende werden als Teil der Gruppe analysiert, zu der sie zugeordnet wurden; unabhängig davon, ob sie diese Behandlung tatsächlich (vollständig) erhalten haben ("once randomized, always analyzed"; **Intention-to-treat-Ansatz**)
- Die Analyse fokussiert auf die Schätzung der **Größe der Unterschiede zwischen den Gruppen** bezüglich eines vordefinierten Outcomes ab. Dieses Outcome ist zeitlich und hinsichtlich des genutzten Instruments genau definiert.

siehe Sibbald & Roland (1998).

## RCTs & Kausale Inferenz

---

## Konfundierung

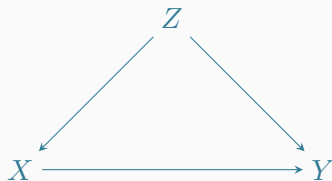
Ziel eines Experimentes ist es, **kausale Schlussfolgerungen** ziehen zu können (z.B. “Behandlung  $X$  führt zu Outcome  $Y$ ”).

Das Experimentaldesign soll es erlauben, von beobachteten Zusammenhängen ausgehend ( $X \leftrightarrow Y$ ) Ursache-Wirkungs-Beziehungen abzuleiten ( $X \rightarrow Y$ ; “kausale Inferenz”).

Dieser Inferenzschluss ist gefährdet, wenn die Wirkungsbeziehung  $X \rightarrow Y$  durch Drittfaktoren  $Z$  **konfundiert** wird.



Kausalmechanismen können durch **Kausaldiagramme** (“directed acyclic graphs”; DAGs) formalisiert werden:



siehe Pearl (2009), Kapitel 1.2.1; Greenland et al. (1999).

### Potential Outcomes Model of Causation (Rubin, 1974)

Der **kausale Effekt einer Behandlung**  $\tau$  auf Person  $i$  ist die **Differenz** zum Messzeitpunkt  $t$  zwischen

- $Y$ , vorausgesetzt, dass  $i$  die Behandlung erhält ( $X_1$ ); und
- $Y$ , vorausgesetzt, dass  $i$  die Behandlung nicht erhält ( $X_0$ ):

$$\tau_{i,t} = Y_{i,t}(X_1) - Y_{i,t}(X_0)$$

**Problem:** Ein Individuum  $i$  kann nie gleichzeitig  $X_1$  und  $X_0$  erhalten;  $\tau_{i,t}$  basiert also gewissermaßen auf **"kontrafaktischen"** (*counterfactual*) Zuständen. Die Schätzung von  $\tau_{i,t}$  setzt daher die **"Austauschbarkeit"** (Exchangeability) von  $i$  voraus (Greenland & Robins, 1986, 2009).

→ Wie kann **"Austauschbarkeit"** erreicht werden, insbesondere wenn unbekannte Einflussfaktoren  $Z$  das Outcome  $Y$  ebenfalls beeinflussen?

## "The Skillful Interrogation of Nature": R. A. Fisher & die Erfindung des randomisierten Experiments (Lehmann, 2011, Kapitel 5)



aus Pearl & MacKenzie (2018), Kapitel 4.

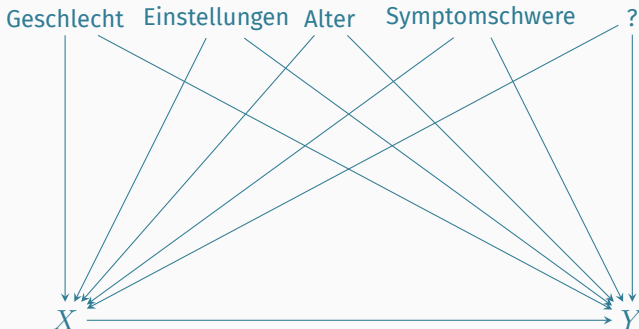
In den 1920er Jahren beginnt Fisher in der Rothamsted Forschungsstation mit randomisierten Experimenten zum Effekt von Düngemitteln.

Seine Einsicht: durch die Randomisierung wird der **systematische** Einfluss **sämtlicher** Konfundierungsvariablen ausgeschaltet; Unterschiede in den Gruppen kommen nur **zufällig** zustande.

**”The Skillful Interrogation of Nature”: R. A. Fisher & die Erfindung des randomisierten Experiments** (Lehmann, 2011, Kapitel 5)

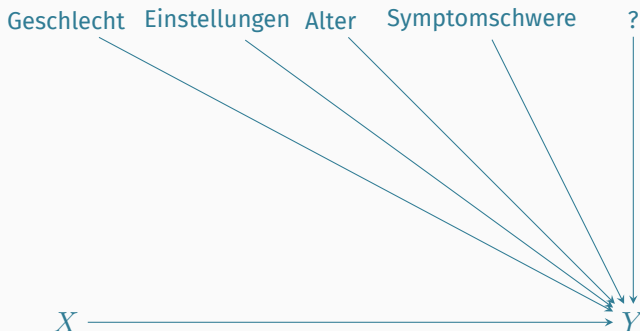
- Wird ein randomisiertes Experiment unendlich oft wiederholt, **“canceln” sich die Zufallseffekte gegenseitig aus**. Es entstehen dadurch zwei (hypothetische) Kollektive, die mit Blick auf alle (bekannten und unbekannten) Einflussfaktoren **“austauschbar”** sind.
- Die Randomisierung **ersetzt** zahllose **unbekannte und bekannte Einflussvariablen** durch eine **einzige**: den unsystematischen Einfluss des **Zufalls**, der sich statistisch **quantifizieren** lässt (z.B. durch Konfidenzintervalle).

**"The Skillful Interrogation of Nature": R. A. Fisher & die Erfindung des randomisierten Experiments** (Lehmann, 2011, Kapitel 5)




**Vor Randomisierung von  $X$ .**

**"The Skillful Interrogation of Nature": R. A. Fisher & die Erfindung des randomisierten Experiments** (Lehmann, 2011, Kapitel 5)



**Nach Randomisierung von  $X$ :** Confounders bleiben als **prognostische** Variablen erhalten.

# Praxis-Teil



```
128 }
129
130 }
131
132 mail{
133   background: url(../img/mailico.png) no-repeat center;
134   display: inline-block;
135   width: 12px;
136   height: 14px;
137   float: left;
138   margin: 2px 7px 0 0;
139 }
140 phone{
141   background: url(../img/phoneico.png) no-repeat center;
142   display: inline-block;
143   width: 20px;
144   height: 18px;
145   float: left;
146   margin: 2px 7px 0 0;
147 }
```

### Anmerkung zu Randomisierungstests

- Zur Überprüfung, ob die Randomisierung "funktioniert" hat, werden immer noch häufig Randomisierungstests durchgeführt (z.B. ein  $t$ -Test zum Vergleich der Gruppen zu Baseline).
- Von solchen Tests wird in der methodischen Literatur weithin abgeraten ("*an unhealthy research behavior that is hard to eradicate*").
- **Grund:** die Idee von Randomisierung ist nicht, jedes mal perfekt balancierte Gruppen zu erhalten, sondern über alle Randomisierungen hinweg → zufällige Gruppenunterschiede sind im Einzelfall selbst bei perfekter Randomisierung erwartbar!
- A priori für wichtig angesehene prognostische Variablen sollten ohnehin in der späteren Analyse kontrolliert werden.

(Austin et al., 2010; De Boer et al., 2015)



*“[S]ignificance testing of baseline differences in randomized controlled trials (RCTs) should not be performed, because it is **superfluous** and can **mislead investigators and their readers.**”*

– Moher et al. (2012), CONSORT 2010 statement

*“[T]he test actually assesses the probability (the ‘P value’) that the observed difference [...] could have occurred by chance when in reality there was no difference. [...] [P]erforming a significance test to compare baseline variables is to assess the **probability of something having occurred by chance** when we know that it **did occur by chance**. Such a procedure is **clearly absurd.**”*

– Altman (1985)

## Studienhypothesen in RCTs

---

## Konfirmatorische Hypothesen

**Superiority Trial (Einseitige Testung):** ist die Experimentalgruppe der Kontrollgruppe *überlegen*?

$$H_0 : \mu_1 \leq \mu_0 \Rightarrow \tau \leq 0$$

$$H_1 : \mu_1 > \mu_0 \Rightarrow \tau > 0$$

**Non-Inferiority Trial:** ist die Experimentalgruppe *nicht unakzeptabel schlechter* als die Kontrollgruppe?

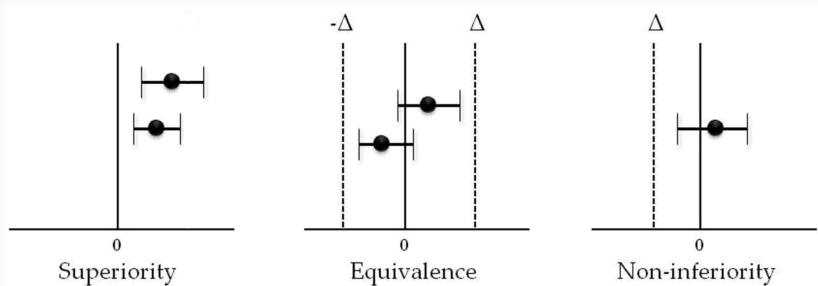
$$H_0 : \mu_1 < \mu_0 - \Delta \Rightarrow \tau < 0 - \Delta$$

$$H_1 : \mu_1 \geq \mu_0 - \Delta \Rightarrow \tau \geq 0 - \Delta$$

**Equivalence Trial:** ist die Experimentalgruppe mit der Kontrollgruppe *vergleichbar*?

$$H_0 : |\mu_1 - \mu_2| > \Delta \Rightarrow |\tau| > \Delta$$

$$H_1 : |\mu_1 - \mu_2| \leq \Delta \Rightarrow |\tau| \leq \Delta$$




adaptiert nach Schumi & Wittes (2011).

Fragen?

Anmerkungen?

Kommentare?



```
128 }
129
130 }
131
132 .mail{
133     background: url(../img/mailico.png) no-repeat center;
134     display: inline-block;
135     width: 120px;
136     height: 140px;
137     float: left;
138     margin: 2px 7px 0 0;
139 }
140 .phone{
141     background: url(../img/phoneico.png) no-repeat center;
142     display: inline-block;
143     width: 20px;
144     height: 18px;
145     float: left;
146     margin: 2px 7px 0 0;
147 }
```

## Referenzen

---

- Akobeng, A. K. (2005). Understanding randomised controlled trials. *Archives of Disease in Childhood*, 90(8), 840–844.
- Altman, D. G. (1985). Comparability of randomised groups. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 34(1), 125–136.
- Austin, P. C., Manca, A., Zwarenstein, M., Juurlink, D. N., & Stanbrook, M. B. (2010). A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: A review of trials published in leading medical journals. *Journal of Clinical Epidemiology*, 63(2), 142–153.
- Backmann, M. (2017). What's in a gold standard? In defence of randomised controlled trials. *Medicine, Health Care and Philosophy*, 20(4), 513–523.
- Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties*, 2(1), 11–20.
- Cartwright, N. (2011). Predicting what will happen when we act. What counts for warrant? *Preventive Medicine*, 53(4-5), 221–224.
- De Boer, M. R., Waterlander, W. E., Kuijper, L. D., Steenhuis, I. H., & Twisk, J. W. (2015). Testing for baseline differences in randomized controlled trials: An unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 1–8.

- Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, 37–48.
- Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, 15(3), 413–419.
- Greenland, S., & Robins, J. M. (2009). Identifiability, exchangeability and confounding revisited. *Epidemiologic Perspectives & Innovations*, 6(1), 1–9.
- Grossman, J., & Mackenzie, F. J. (2005). The randomized controlled trial: Gold standard, or merely standard? *Perspectives in Biology and Medicine*, 48(4), 516–534.
- Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics and Gynaecology*, 125(13), 1716.
- Kaptchuk, T. J. (2001). The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology*, 54(6), 541–549.
- Lehmann, E. L. (2011). *Fisher, neyman, and the creation of classical statistics*. Springer Science & Business Media.
- Lilienfeld, S. O., McKay, D., & Hollon, S. D. (2018). Why randomised controlled trials of psychological treatments are still essential. *The Lancet Psychiatry*, 5(7), 536–538.



- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P., Elbourne, D., Egger, M., & Altman, D. G. (2012). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery*, 10(1), 28–55.
- Pearl, J. (2009). *Causality*. Cambridge university press.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic books.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688.
- Schumi, J., & Wittes, J. T. (2011). Through the looking glass: Understanding non-inferiority. *Trials*, 12(1), 1–12.
- Sibbald, B., & Roland, M. (1998). Understanding controlled trials. Why are randomised controlled trials important? *BMJ: British Medical Journal*, 316(7126), 201.