

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

QSB1

Epidemiologie, Medizinische Biometrie und Medizinische Informatik

Klinische Studien

Dr. rer. med. habil. Marita Ziepert /

Dr. rer. nat. Meinhard Mende

imise.

QSB1

Epidemiologie, Medizinische Biometrie und Medizinische Informatik

- Epidemiologie: VL 14
- Medizinische Biometrie: VL 1 bis VL 13
- Medizinische Informatik VL QSB1_1 bis 3

Übungen in Gruppen am Nachmittag

Lernziele

Am Ende dieser Lehreinheit

- verstehen Sie, warum es für klinische Studien in der Medizin statistische Methoden braucht,
- kennen Sie Klassifizierungsmerkmale und typische Designs von Studien,
- kennen Sie die Definition der klinischen Prüfung und wissen, dass diese Gesetzen wie z. B. dem AMG unterliegt,
- können Sie die Kriterien der Vergleichbarkeit anwenden und in klinischen Studien die Vergleichbarkeit herstellen,
- sind Sie mit dem Konzept intent-to-treat vertraut.

Vereinbarungen

- Wir lernen Prinzipien kennen, nach welchen Statistik funktioniert oder die wir bei der Anwendung beachten müssen. Diese sind durch ein großes P gekennzeichnet.

P

- Die Dame mit dem Fernglas zeigt an, wenn wir über den Tellerrand des unmittelbaren Stoffs hinaus schauen.



http://www.festartikelhaus.de/Deko-und-Gartenfiguren/Menschen/Oma-mit-Fernglas-K672::100684.html

- Cave canem! (Vorsicht, Hund) zeigt Ihnen an, wo gedankliche Fehlschlüsse und Stolperfallen möglich sind.

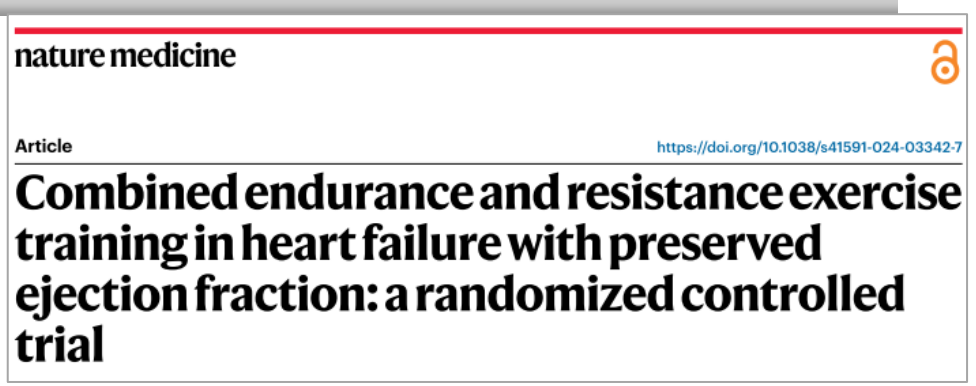
Jotquadrat: Grafitto in Bonn. commons. wikipedia.org



- Fachbegriffe sind rot gekennzeichnet.

1.1 Warum Statistik in der Medizin?

Beispiel 1.1 – Ex-DHF-Studie



Edelmann F et al. Nature Medicine 2025;31:306–314

Patients

- Probanden mit HFpEF, NYHA-Klasse II / III

Intervention

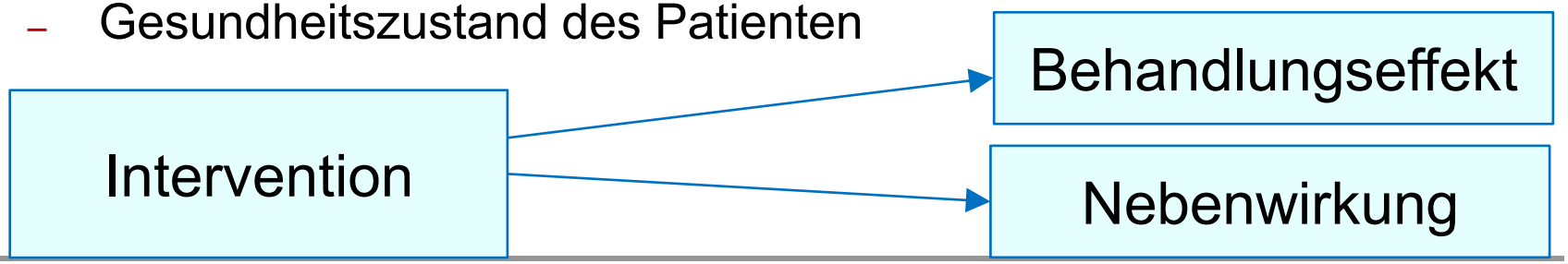
- Training + Usual Care (ET)

Comparison

- Usual Care (UC)

Outcome

- Gesundheitszustand des Patienten



Herausforderungen

1. Probandinnen und Probanden variieren in ihren sozio-demografischen und klinischen Merkmalen stark

   Repräsentative Stichprobe

2. Woran wird ein möglicher Erfolg der Intervention gemessen?

   Woran misst man einen möglichen Schaden für Patienten?

   Endpunkte

3. Studienergebnisse (Outcomes) sind nicht vorhersagbar (zufällig)

   Anwendung der mathematischen Statistik

Herausforderungen

4. Im Fokus steht eine mögliche Veränderung des Krankheits- / Gesundheitszustandes der Probanden.

   Follow-up (Nachbeobachtung)

5. Wie bekommen wir heraus, wie sich die Probanden ohne die Behandlung entwickelt hätten?

   Vergleichs- / Kontrollgruppe

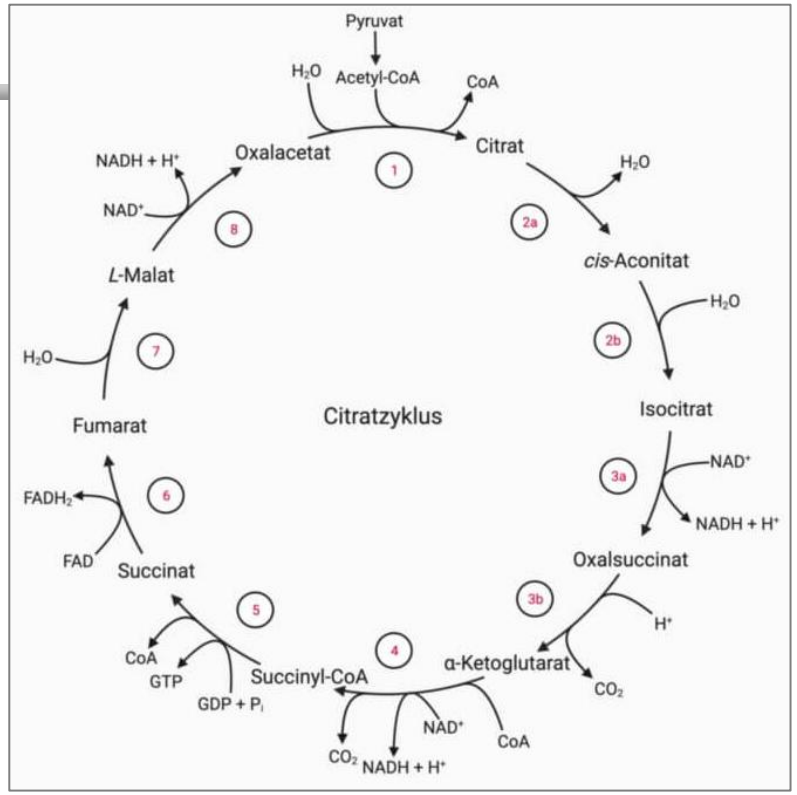
Der Mensch ist komplex

- Alle Lebensprozesse sind höchst komplex, hier Citratzyklus

  https://flexikon.doccheck.com/de/Citratzyklus (2025-10-28)

- exakte Vorhersagen sind oft unmöglich,

→ enorme Variabilität



Kein Teil des Universums ist so einfach, dass er ohne Abstraktion erkannt oder beherrscht werden kann.

(Rosenblueth A, Wiener N: The role of models in Science. Philos Science 1945;4:316)

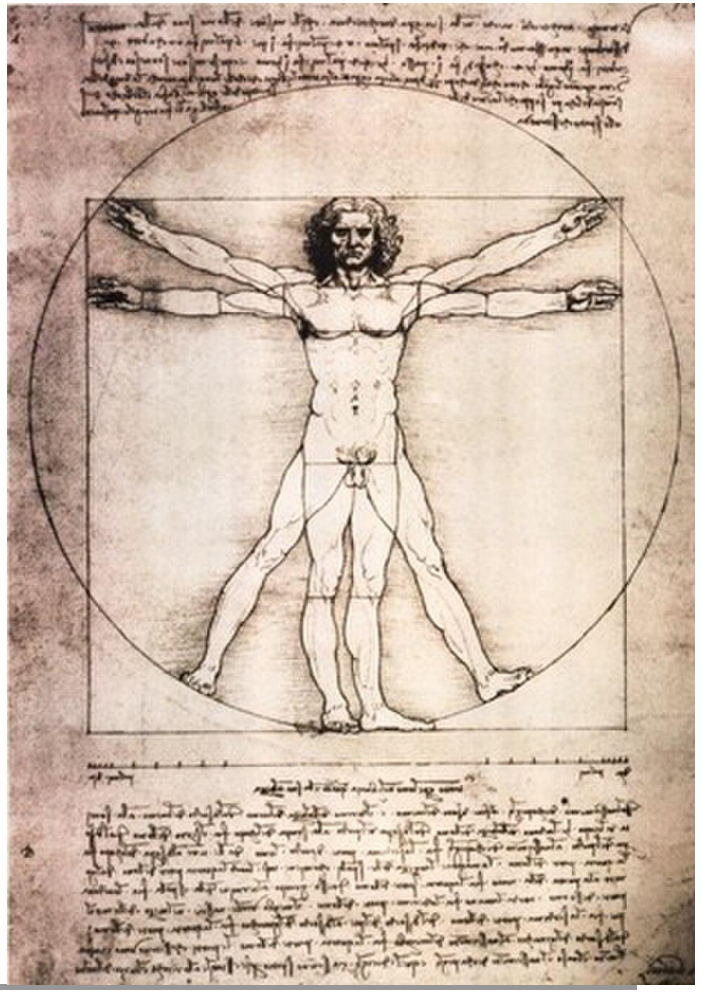
Ordnung und Chaos

Deterministisches Chaos

Ordnung und Chaos sind als zwei verschiedene Arten eines zugrundeliegenden Determinismus zu sehen. Und keine von ihnen existiert von allein.

Das typische System kann in verschiedenen Stadien existieren, einige geordnet, andere chaotisch.

Stewart L. Spielt Gott Roulette? dt. bei Birkhäuser, Basel 1990



https://commons.wikimedia.org/w/index.php?search=mensch+da+vinci&title=Special%3AMediaSearch&type=image

Zufall

Kennzeichen des Lebens ist, dass das Ergebnis vieler Prozesse nicht vorhersagen lässt: zufälliger Ausgang

Zufall

- accidente (it.)
- coincidenza (it.)
- zu – fallen (von oben auf eine niedrigere Ebene)
- was fällt uns zu?

Der Mensch – ein Zufallsprodukt?!

Evolutionsforscher Axel Meyer:

Selbstverständlich arbeitete die Evolution nicht auf die Entwicklung des Menschen hin. Wir sind ein Zufallsprodukt, wie jede andere Tierart auch. *(2019-02-06)*

https://www.profil.at/wissenschaft/evolutionsforscher-meyer-wir-zufallsprodukt-10632231

Die Thora:

Und Gott schuf den Menschen als sein Bild, als Bild Gottes schuf er ihn; als männlich und weiblich schuf er sie.

(Genesis 1, 27; Elberfelder Übersetzung)

Zufall bis ins Kleinste

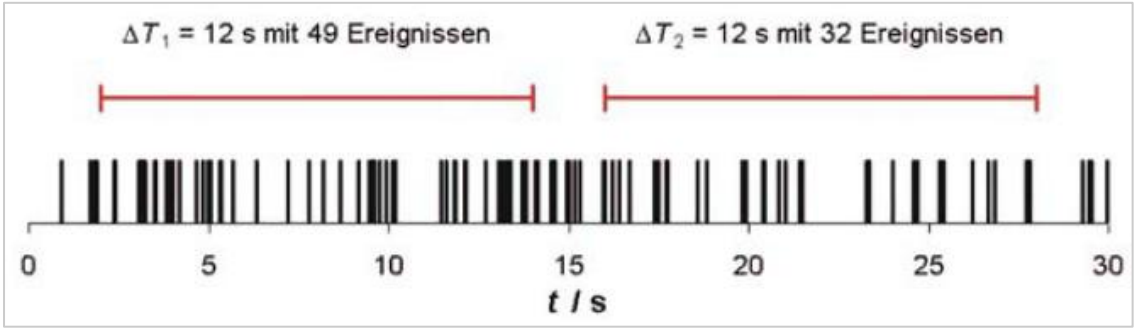
- Prozesse, deren Ergebnis wir nicht vorhersagen können

Zufallsexperiment

$$^{90}_{38}Sr \rightarrow {}^{90}_{39}Y \rightarrow {}^{90}_{40}Zr$$

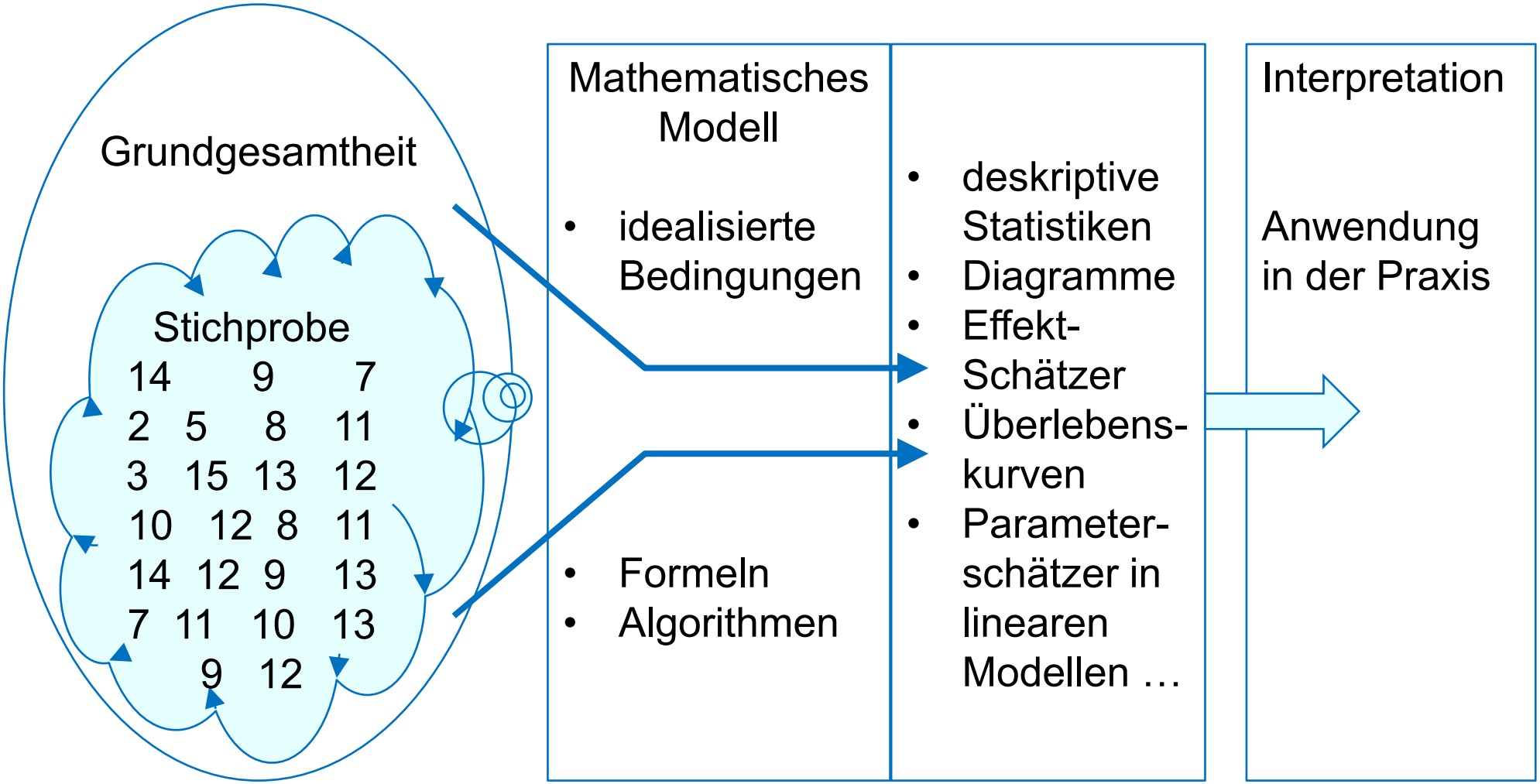
FH Hannover, Physikalisches Praktikum, Versuch 4.3,
https://schrewe.wp.hs-hannover.de/
public_www/Physik_Labor_Elektotechnik/4.3_Radioaktiver_Zerfall_0711.pdf

- Gemessenes Zufallsergebnis



Realisierung

Mathematische Statistik



1.1.2 Endpunkte

Endpunkt: Merkmal / Kriterium, das

- den Nutzen (oder Schaden) einer Intervention für den Patienten treffend und objektiv charakterisiert,
- relativ einfach und ohne Gefährdung (nicht invasiv) von allen Probanden zu erfassen ist,
- möglichst in der klinischen Routine erfasst wird
- und statistisch auswertbar ist

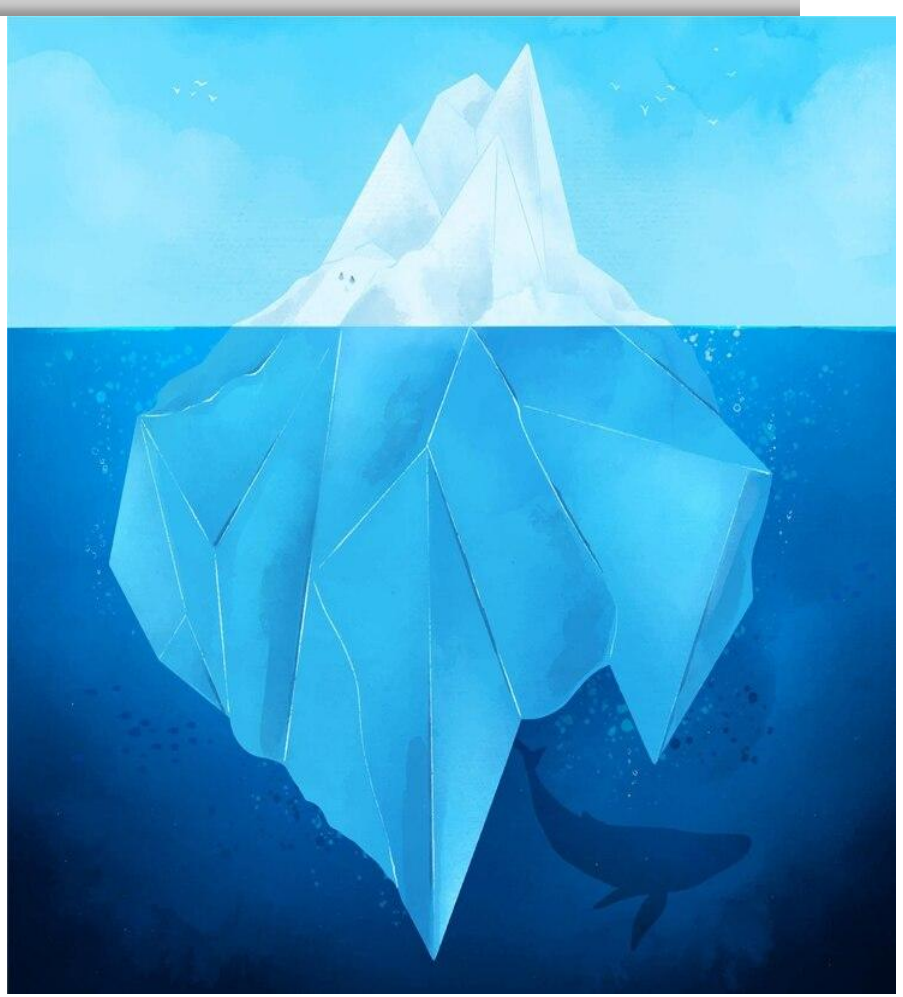
- In Klinischen Studien werden
  - primäre Endpunkte
  - sekundäre Endpunkte unterschieden

Messen

Mögliches Motto der Wissenschaft

Man muss messen, was messbar ist, und messbar machen, was zunächst nicht messbar ist.
(Martin 1868)

- Körpergröße (cm) / Alter (J.)
- peak VO_2
- NYHA-Klasse (I, II, III, IV)
- Hospitalisierung (ja / nein)
- Lebensqualität (?)
- …



Freie Clipart von pikisuperstar

Biometrie – Leben messen?

Um das Leben zu erforschen, zerlegen wir Lebewesen in ihre Bestandteile, wir zerstören sie, um herauszufinden, was Leben ist.

Prof. Dr. Wolfgang Alt (persönliche Kommunikation 2008)

Quantifizierung nicht direkt messbarer Merkmale:

Lebensqualität

- Fragebögen (z. B. SF-36, MLHFQ, …)

Schmerzen

- Visuelle Analogskala

Vitalität

- über Surrogatendpunkte (z. B. Vitalkapazität oder Herzfrequenzvariabilität)

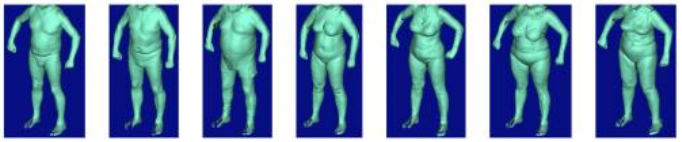
Biometrie = Leben messen

Beispiel 1.2 LIFE-Projekt:

- Ziel: Ursachen für Zivilisationskrankheiten erforschen
- Studie:

  von ca. 10 000 Leipzigern wurden viele Merkmale erfasst (soziodemographische, anthropometrische Merkmale, Labor, Echo und EKG, Fragebögen, Ganzkörperscanner, …

  Löffler-Wirth H et al. PLoS ONE 11(7): e0159887



- Es wäre aber vermessen zu denken, dass wir dann den Menschen verstanden hätten.

1.1.3 Kausalität

- sagt aus, dass eine Ursache eine Wirkung nach sich zieht.
- Kinder fragen: "Warum …?"

Wissenschaft möchte kausale Zusammenhänge aufklären

Kausalität spricht z. B. so:

- Laparoskopische Prostatektomie mit Roboter bewirkt im Mittel eine frühere Wiederherstellung der Harnkontinenz.
- CO_2-Ausstoß durch Industrie, Haushalte und Verkehr erhöhen langfristig die mittlere Temperatur der Atmosphäre
- Alkohol in der Schwangerschaft schadet dem Embryo.

Assoziation

- Statistik findet Zusammenhänge (Assoziation)

Assoziation spricht so:

- Einige Patienten mit HFpEF fühlen sich nach regelmäßigem Training besser als Patienten mit Standardbehandlung.
- Unter Diabetikern ist Bluthochdruck häufiger als unter Probanden ohne Diabetes mellitus.

Jeder Mensch neigt dazu, Beobachtungen kausal zu interpretieren!

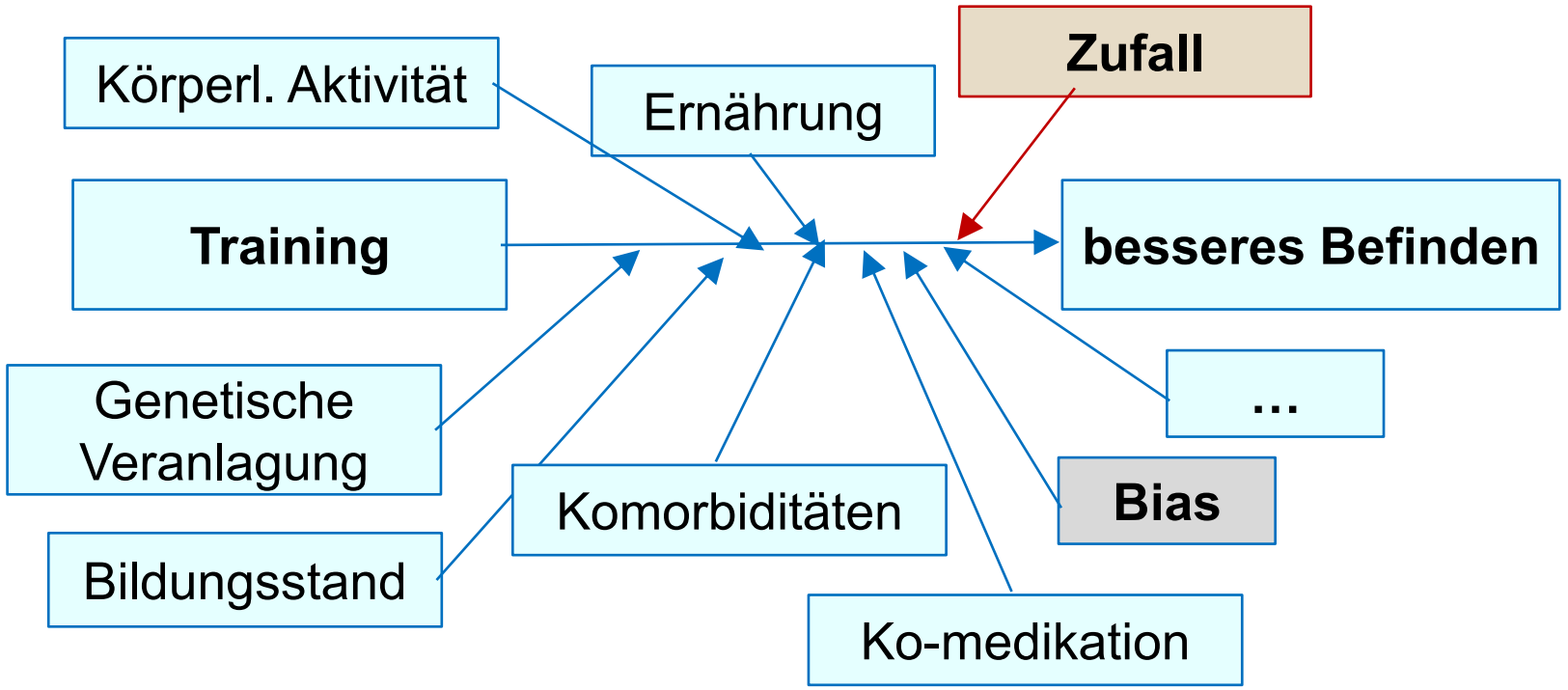
Die einfache Antwort auf diese Frage lautet, dass man, wenn man sich auf seine Intuition verlässt, häufig irrt, indem man ein Zufallsereignis als ein systematisches fehlinterpretiert.

Quelle: Daniel Kahnemann: Schnelles Denken, langsames Denken. Dt. Ausgabe: Siedler Verlag München 2011, S. 149



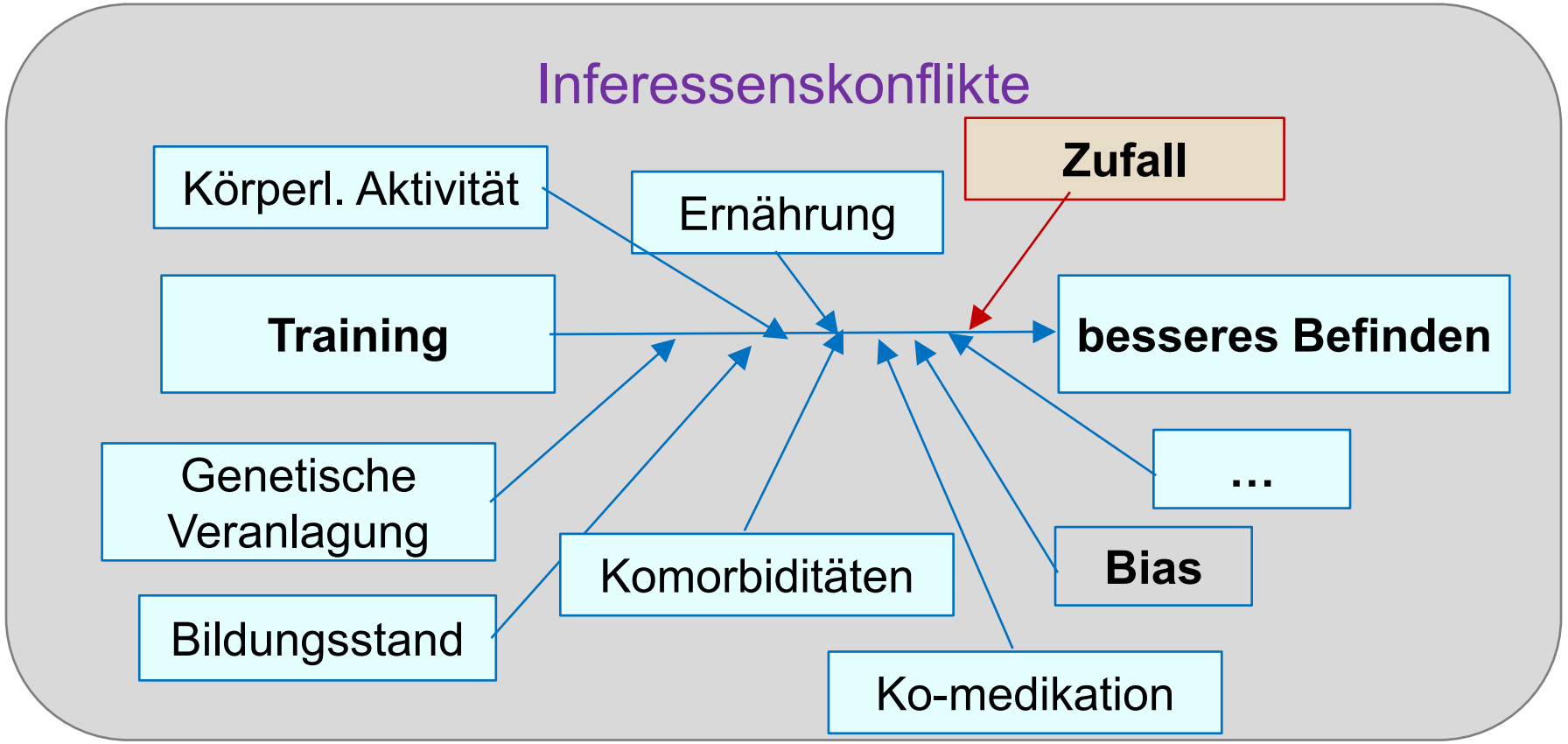
Erklärungen für den Behandlungseffekt

Alle Lebewesen sind (höchst) komplexe Systeme, in denen kausale Zusammenhänge vielfältig modifiziert oder sogar überdeckt werden.



Anwendung der Statistik ist nicht neutral

Interessen beeinflussen das Design, die Durchführung und Auswertung von Studien:



Den Zufall *herausrechnen*

Statistik: Mittelwertbildung → mittlere Ereignisrate schätzen.

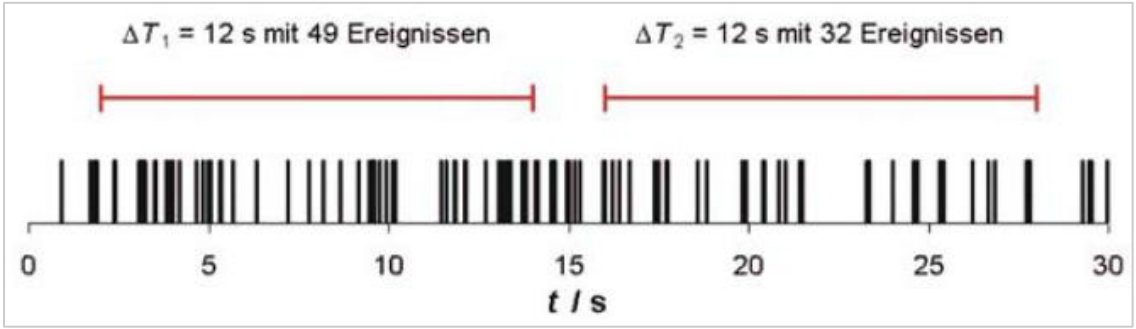
Zufallsexperiment

$$^{90}_{38}Sr \rightarrow {}^{90}_{39}Y \rightarrow {}^{90}_{40}Zr$$

FH Hannover, Physikalisches Praktikum, Versuch 4.3,
https://schrewe.wp.hs-hannover.de/
public_www/Physik_Labor_Elektotechnik/4.3_Radioaktiver_Zerfall_0711.pdf

- Gemessenes Zufallsergebnis

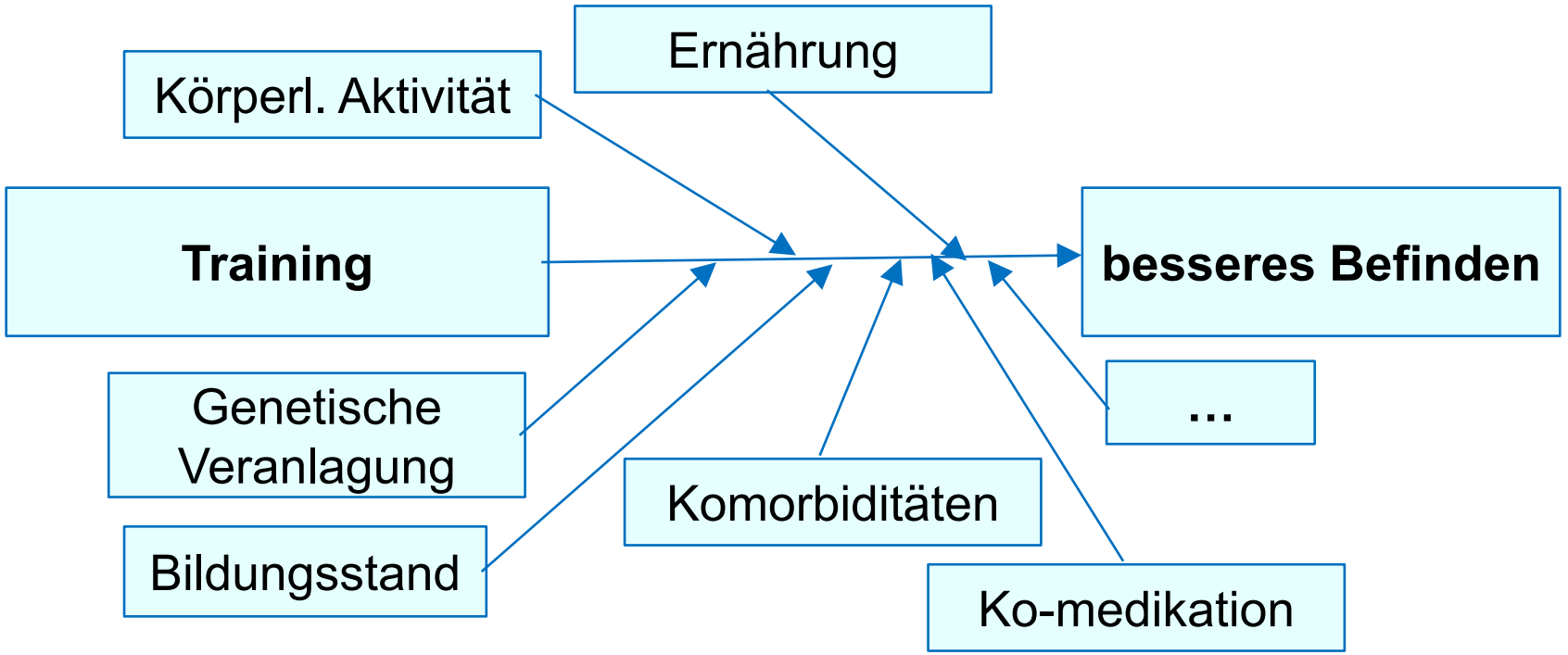
Realisierung



geschätzte Rate λ

Störfaktoren (Confounder)

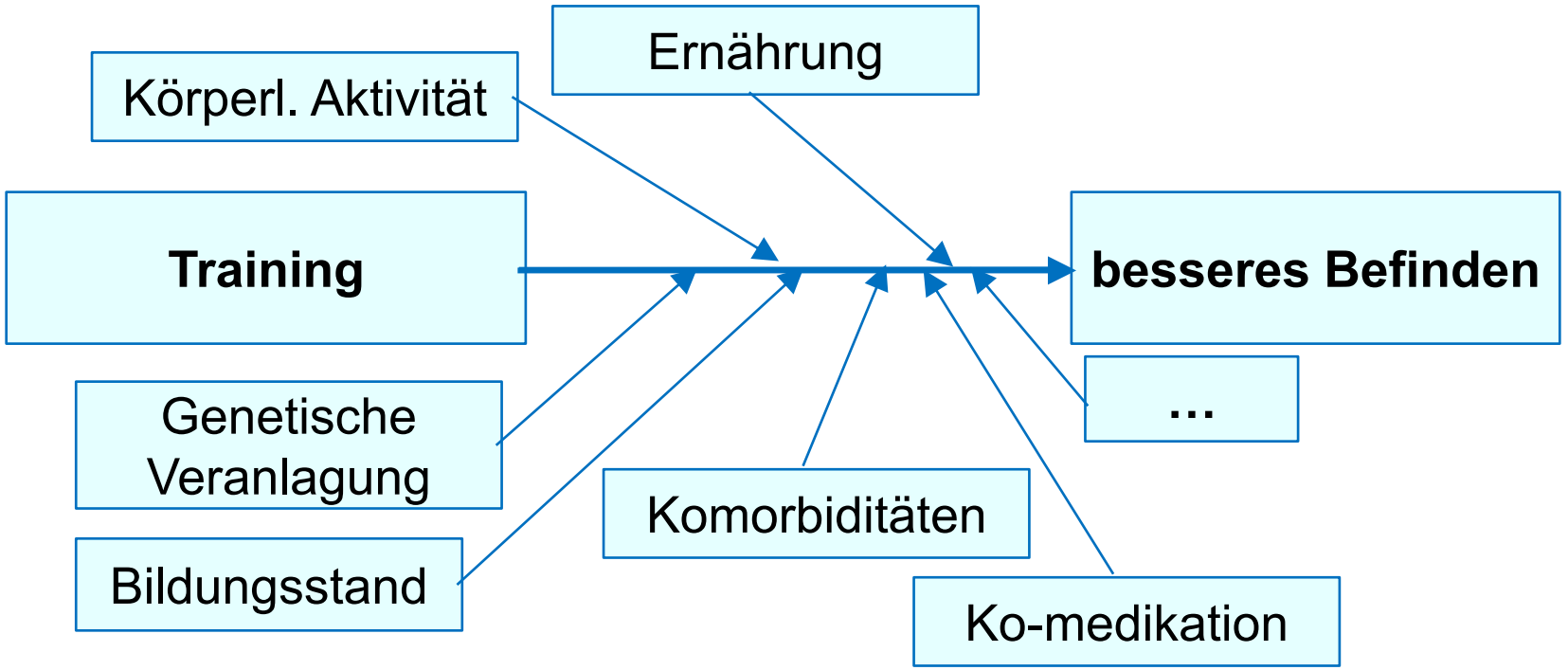
Alle Lebewesen sind (höchst) komplexe Systeme, in denen kausale Zusammenhänge vielfältig modifiziert, vorgespiegelt oder überdeckt werden.



Störfaktoren eliminieren

Gelänge es, alle Störfaktoren (Confounder) auszuschalten …

(confound: durcheinanderbringen, vermengen, verwirren, …)



Randomisierte, kontrollierte klinische Studie

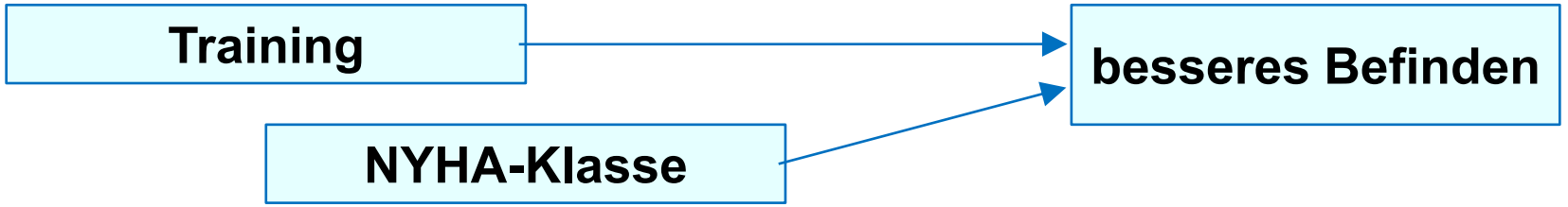
→ … könnte man einen beobachteten Effekt
allein auf die Behandlung zurückführen



Randomisierte, Kontrollierte Klinische Studien (RCTs) werden so designt, dass beobachtete Effekte (weitgehend) auf die Intervention zurückgeführt werden können.

1.1.4 Wissenschaft benötigt Modelle

- Modelle: vereinfachte Abbildungen von Prozessen, die für bestimmte Zwecke ersonnen wurden (nach: H. Schaefer. Modelle in der Medizin. Springer 1992)



- Modell wird durch klar definierte Bedingungen spezifiziert (z. B. Normalverteilung)
  - Das ermöglicht die Anwendung mathematischer Formeln.
  - Ergebnisse gelten nur im Rahmen des Modells
  - Robustheit: Anwendbarkeit des Modells, auch wenn nicht alle Bedingungen erfüllt sind

1.2 Charakteristika von Studien

1.2 Studien charakterisieren

Studiendesign

Struktur einer problemgeleiteten, systematischen Beobachtung zur Beantwortung einer spezifischen Fragestellung (Studie). Das Studiendesign hängt von der Fragestellung ab und z.B. auch davon, ob es sich um ein qualitatives oder ein quantitatives Problem handelt.

https://www.ebm-netzwerk.de/de/medien/pdf/ebm-glossar.pdf/@@download
(2026-08-19)

Design 1: Vergleich mit historischer Kontrolle

Führe mit einer Anzahl von HFpEF-Patienten regelmäßig überwachtes Kraft- und Ausdauertraining durch und vergleiche die Veränderung des Gesundheitszustandes nach 1 Jahr mit Patienten aus dem Register

- Vorteil: relativ einfach und preisgünstig zu realisieren
- Nachteile:
  - Kontrollpatienten können sich in wesentlichen Parametern deutlich unterscheiden (z. B. Alter, Komorbiditäten, …)

  → Kohorten u. U. kaum vergleichbar,
  - Endpunkt ggf. im historischen Register nicht adäquat erfasst

Design 2: Prospektive Behandlungsstudie

Sammle in deiner Klinik sukzessive Patienten, die regelmäßig überwachtes Training durchgeführt haben und vergleiche die Veränderung des Gesundheitszustandes mit der von Patienten mit Standardbehandlung.

Vorteile:

- prospektiver Ansatz erlaubt Kontrolle aller Arbeitsschritte
- Prozeduren entsprechen dem normalen Klinikablauf

Nachteile:

- Willkürliche Zuordnung zu den beiden Gruppen könnte eine Behandlungsmethode bevorteilen (allocation bias)
- Umfeld der Patienten (Herkunft / Begleitbehandlung) schwer kontrollierbar

Design 3: Randomisierte, kontrollierte Interventionsstudie

Weise Patienten mit genau definierten Eigenschaften zufällig den Studienarmen A) Training (ET) bzw. B) Usual Care (UC) zu,

- gestalte alle Studienbedingungen (außer der Intervention) gleich
- und erfasse den Endpunkt nach 1 Jahr so genau wie möglich

Vorteile:

- Randomisation schafft Vergleichbarkeit zum Anfang, die durch
- Verblindung, Arbeit nach Protokoll und Arbeitsanweisungen im Studienverlauf aufrechterhalten wird

Nachteile:

- RCT ist aufwändig und stellt Herausforderungen an den Klinikalltag .
- Probanden evtl. nicht repräsentativ

Klassifizierung von Studien I

Klassifizierung von Studien durch Begriffspaare:

- **Beobachtung** oder **Experiment**
- **retrospektiv** vs. **prospektiv**
- **Längsschnitt** oder **Querschnitt**
- **nicht kontrolliert** vs. **kontrolliert**

Klassifizierung I: Beobachtung / Experiment

Beobachtung

- Keine Intervention
- Eigenschaften werden beobachtet und dokumentiert
- keine Schlussfolgerungen auf kausale Zusammenhänge möglich (Was ist Ursache und was ist Wirkung)



Experiment

- Intervention

  (z.B. neues Medikament)
- Eigenschaften werden zugewiesen
- Schlussfolgerungen auf kausale Zusammenhänge prinzipiell möglich



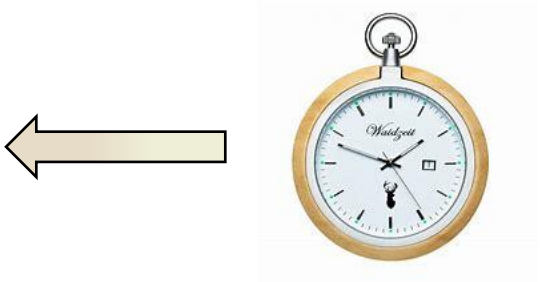
Klassifizierung II: Retrospektiv / Prospektiv

Retrospektiv

= zeitlich zurückschauend

⇨ Daten liegen bei Beginn der Studie bereits vor

- eingeschränkte Kontrollen von möglichen Störgrößen

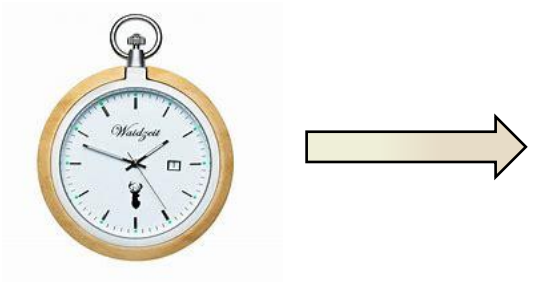


Prospektiv

= zeitlich nach vorn schauend

⇨ Daten entstehen nach Beginn der Studie

- mögliche Störgrößen können kontrolliert werden



Klassifizierung III: Längsschnitt / Querschnitt



:

Längsschnitt

(Longitudinal)

- wiederholte Erfassung eines Merkmals
- dient der Verlaufsbeschreibung
- Beispiel: Heft der Zensurnoten in der Schule

Querschnitt

(Cross-sectional)



- einmalige Erfassung von Merkmalen zu einem Zeitpunkt
- dient der Zustandsbeschreibung
- Beispiel: Bundestagswahl

Um therapeutische / Neben-Wirkungen bestimmen zu können, bedarf es Nachbeobachtungen (Follow-ups)

B 1.2 LIFE-Projekt: zunächst Querschnitt

- Fünf Jahre nach Beginn der LIFE-Studie in Leipzig wurden einige Ergebnisse vorgestellt:



Leipzig › Lokales › LIFE-Studie: Leipziger schlafen schlecht und werden immer dicker

PARTNER IM RND REDAKTIONSNETZWERK DEUTSCHLAND

Lokales / Forschungsprojekt

18:30 Uhr / 24.09.2015

LIFE-Studie: Leipziger schlafen schlecht und werden immer dicker

Es ist eine der größten Studien ihrer Art: Das Leipziger Forschungszentrum für Zivilisationskrankheiten (LIFE) hat in den vergangenen fünf Jahren rund 10.000 Menschen untersucht und befragt. Am Donnerstag wurden die Ergebnisse vorgestellt.

- Was ist hier falsch formuliert?

Klassifizierung IV: Kontrolliert / nicht kontrolliert

Kontrolliert (mit Kontrolle)

- Experimentell
  - Mit vergleichbarer Kontrollgruppe
- Beobachtung
  - Fall-Kontroll-Studie

Nicht kontrolliert (ohne Kontrolle)

- Experimentell
  - Pilotstudien zur Prüfung der Machbarkeit und Hypothesengenerierung
- Beobachtung
  - Registerdaten,
  - Fallsammlungen

Fall-Kontroll-Studie I

Retrospektive kontrollierte Beobachtungsstudie:

- Untersucht Zusammenhänge zwischen Exposition und Erkrankung
- Kranken werden vergleichbare Gesunde zugeordnet (matching) und retrospektiv die zu untersuchende Exposition in beiden Gruppen ermittelt.



Beispiel: Zusammenhang
Rauchen und Bronchialkarzinom



(Doll u. Hill, BMJ 1950;4682:739-48)

Fall-Kontroll-Studie

Vorteile:

- i. Allg. einfacher Zugang zu Fällen,
- kostengünstig,
- lange Beobachtungszeiten entfallen

Probleme:

- Auswahl der Kontrollen oft schwierig,
- kein sicherer Kausalitätsnachweis,
- Gefahr von Bias: Strukturgleichheit nicht gegeben

Klassifizierung von Studien V

Kohortenstudie:

- Prospektive longitudinale Beobachtungsstudie
- eine definierte Patientengruppe (Kohorte) wird über einen bestimmten Zeitraum beobachtet, um zu sehen, wie viele Patienten ein vorher definiertes Ereignis entwickeln (Ermittlung der Inzidenz einer Erkrankung)
- Beispiel:

  Framingham Heart Study (seit 1948)

  6000 Bürger der Kleinstadt Framingham (MA, USA) werden seit 1948 alle 2 Jahre untersucht

  Ziel: Ermittlung von Faktoren, die mit der Entstehung von Atherosklerose und KHK assoziiert sind

Aufgabe 1

Klassifizieren Sie die folgenden Studien nach diesen 4 Kriterien!

Beobachtung	oder	Experiment
retrospektiv	vs.	prospektiv
Längsschnitt	oder	Querschnitt
nicht kontrolliert	vs.	kontrolliert

1. **DIAST-CHF** Studienziele:

– Wann und wie entwickelt sich diastolische Herzinsuffizienz bei Risikopatienten

– Wie hoch ist die Letalität? …

z. B.
ESC Heart Fail.
2015;2(2):76-84

knapp 2000 Patienten mit Zeichen und Symptomen der Herzinsuffizienz, aber erhaltener Pumpfunktion wurden bis 10 Jahre nachverfolgt und viele Merkmale erhoben;

Aufgabe 2

Klassifizieren Sie die folgenden Studien nach diesen 4 Kriterien!

Beobachtung	oder	Experiment
retrospektiv	vs.	prospektiv
Längsschnitt	oder	Querschnitt
nicht kontrolliert	vs.	kontrolliert

2. Schülersportstudie

Studienfragen:

- Verändert sich durch häufigeren Sportunterricht die Fitness der Schüler in 2, 3, … Jahren?
- Kann regelmäßiger Sport Adipositas vermeiden?
- Hat regelmäßiger Sportunterricht langfristig positive Auswirkungen …

Schulklassen wurden 2 Gruppen zugeordnet: A) Erweiterter Sportunterricht / B) Normaler Sportunterricht; jährlich wurden viele Tests und Untersuchungen durchgeführt.

z. B.
Müller UM et al.
Circ J 2017; 81: 717–725

Aufgabe 3

Klassifizieren Sie die folgende Studie nach diesen 4 Kriterien!

Beobachtung	oder	Experiment
retrospektiv	vs.	prospektiv
Längsschnitt	oder	Querschnitt
nicht kontrolliert	vs.	kontrolliert

3. Endokarditis nach AK-Ersatz: Fragestellung
   - Wie verhält sich die ein-Jahres Mortalität von Probanden mit Endokarditis nach AK-Ersatz, je nachdem, ob sie nur antibiotisch oder zusätzlich chirurgisch behandelt wurden?
   - Welche Merkmale triggern das Letalitätsrisiko? …

Mangner N et al.
J Am Heart Assoc.
2018;7:e010027

Datenbank des HZL: 20 mit Klappen-OP vs. 44 mit antibiotischer Behandlung allein

1.3 Klinische Prüfung

1.3.1 Definition Klinische Studie (Prüfung)

§4 Abs. 23 AMG:

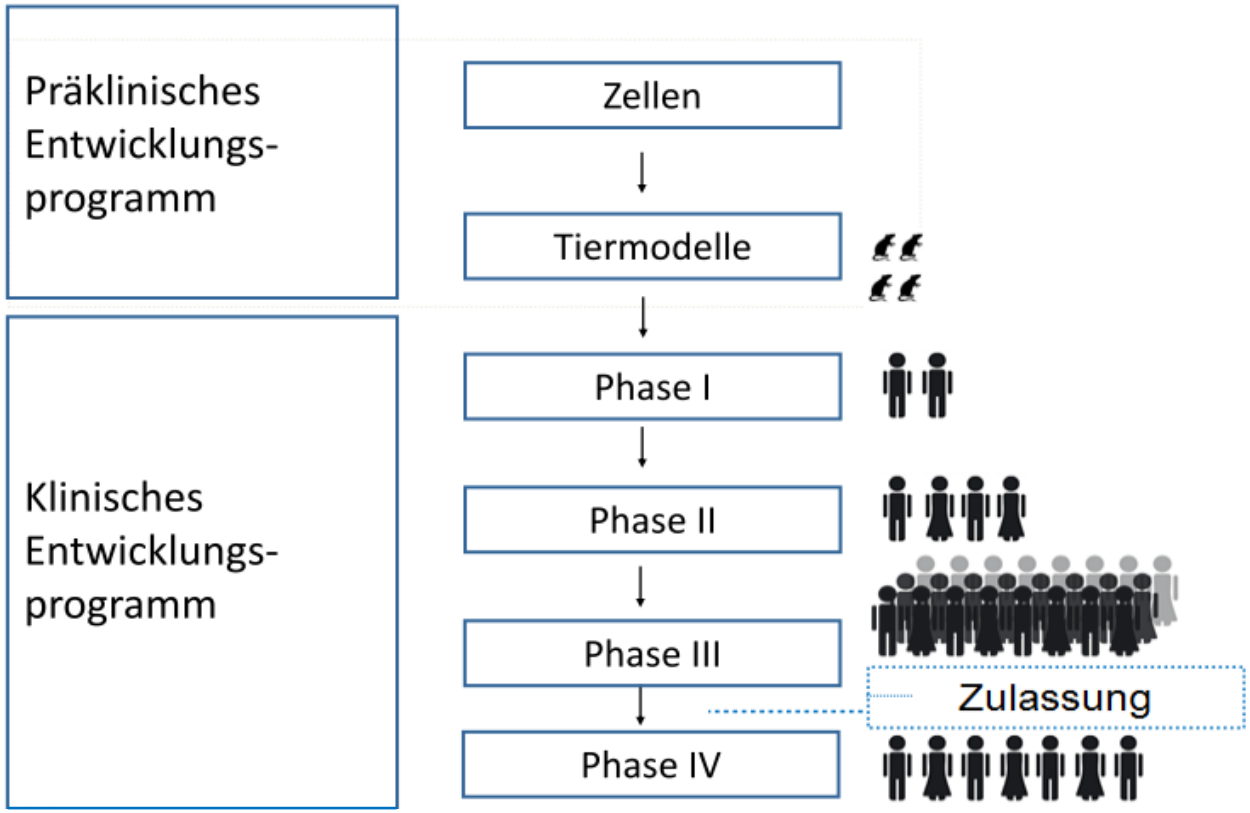
Eine klinische Prüfung bei Menschen ist jede am Menschen durchgeführte Untersuchung, die dazu bestimmt ist,

- klinische oder pharmakologische Wirkungen von Arzneimitteln zu erforschen oder nachzuweisen
- oder Nebenwirkungen festzustellen
- oder die Resorption, die Verteilung, den Stoffwechsel oder die Ausscheidung zu untersuchen, mit dem Ziel,
- sich von der Unbedenklichkeit der Arzneimittel zu überzeugen.

https://dag-hszt.de/files/downloads/Rechtsfragen%20der%20Neulandmedizin%20Klinische%20Pr%C3%BCfung%20und%20Heilversuch%20EK%2031.%20M%C3%A4rz%202020%20Vortagsversion-1.pdf (2026-08-21)

Hinweis: übertragbar auf alle Arten von planbaren Interventionen

Phasen der Klinischen Prüfung



Th. Sudhop, Einführung in die Planung und Durchführung klinischer Prüfungen, BfArM, 18.10.2016

Phase I und Phase II Studien

Phase I: Ist es sicher genug, um fortzufahren?

- Erstanwendung am gesunden Menschen (n < 20)
- Prüfen der Sicherheit und Verträglichkeit
- Pharmakokinetik und Pharmakodynamik

Phase II: Gibt es Hinweise auf positive Effekte?

- Erstanwendung am kranken Menschen (mehrere Indikationen)
- Prüfung der Wirksamkeit und Nebenwirkungen bei wenigen Patienten (therapeutisch explorativ)
- Untersuchung zu Dosis- / Wirkungsbeziehung, Dosisabschätzung

Phase III und Phase IV Studien

Phase III: Ist die neue Behandlung besser als der Standard ?

- (Randomisierte) kontrollierte klinische Therapiestudie
- Wirksamkeitsnachweis der neuen Behandlung gegen Standardtherapie
- Sicherheitsanalysen
- Grundlage für Nutzen/Risikoabwägung für die Zulassung

Phase IV: Ist die Behandlung in der Versorgung sicher ?

- Prüfung der Sicherheit und Wirksamkeit unter Praxisbedingungen
- Sammlung von Informationen nach der Zulassung
- Entdeckung seltener Nebenwirkungen, gezielte Untersuchung bestimmter Subpopulationen, Lebensqualität

Ablauf einer Phase-III-Studie: Ex-DHF

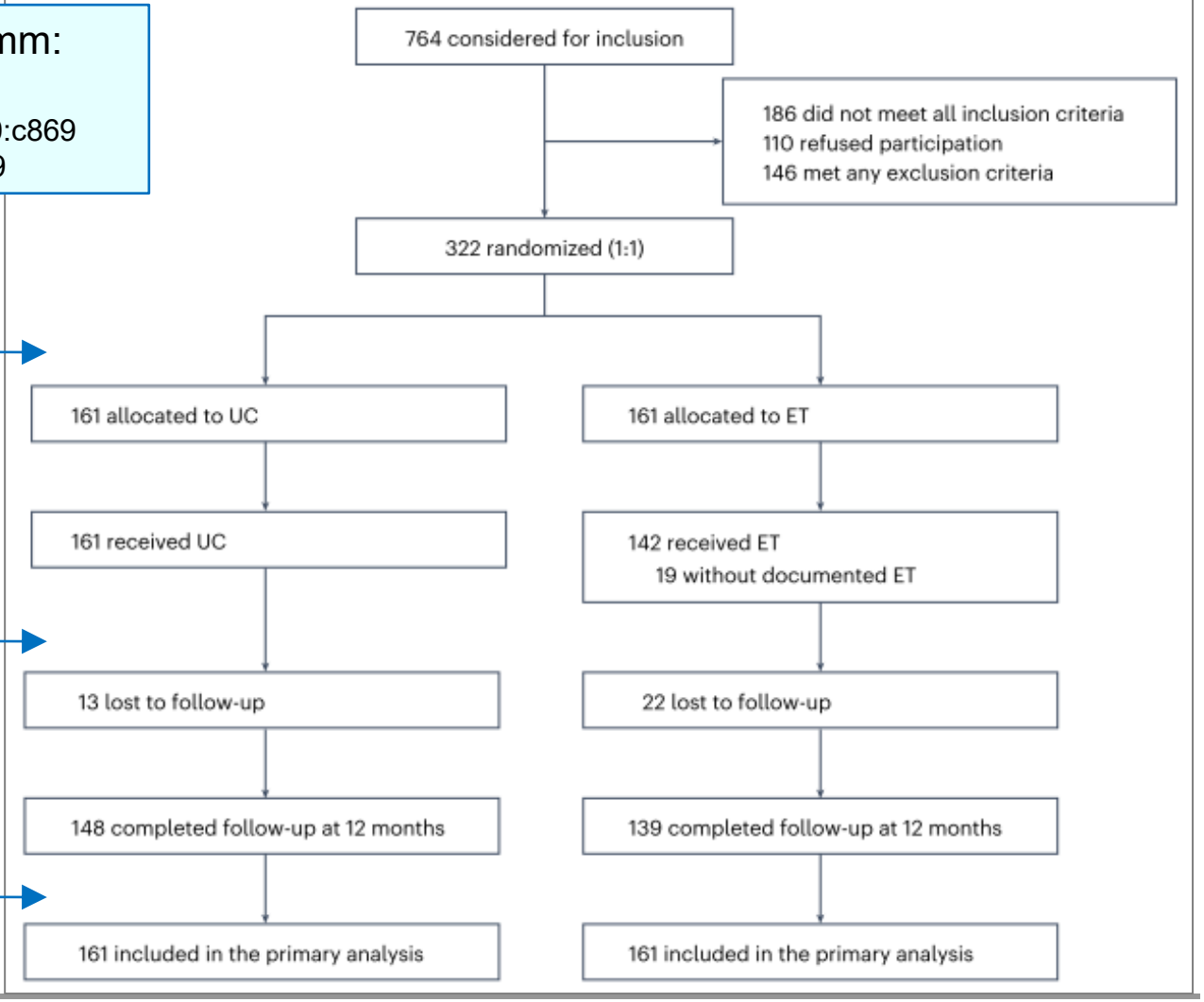
CONSORT-Diagramm:

Moher et al. BMJ 2010;340:c869
doi: 10.1136/bmj.c869

Randomisation

Follow-up

Analysis



Studienfrage

Präzisierung der Studienfrage:

- detailliert formulieren und begründen
- Kenntnis des aktuellen Forschungstandes und der internationalen Literatur

Eur J Heart Failure 2017;19:1067-1074

Study design

The primary objective of the Ex-DHF trial is to determine whether ET on top of usual care is superior to usual care alone in improving a clinical composite outcome score in HFpEF patients. The score includes all-cause mortality, hospitalizations, symptoms, global self-assessment, exercise capacity, and diastolic function.

Primäres Ziel / sekundäre Ziele:

The Ex-DHF trial also evaluates whether ET improves maximal and submaximal exercise tolerance, HRQoL, echocardiographic parameters of diastolic and systolic function, left ventricular geometry and dimensions, and markers of neurohumoral activation in HFpEF patients.

Patienten- / Probandenauswahl 1

Einschlusskriterien:

- charakterisieren die Merkmale von Patienten, die für die therapeutische Prüfung in Betracht kommen.

Inclusion criteria

- Stable symptomatic heart failure with preserved ejection fraction (diagnosis according to the recommendations of the European Society of Cardiology):[21]
  - NYHA class II–III, peakVO_2 <25 mL/kg/min
  - LVEF ≥50%
  - E/e′ >15 **or** 15 ≥E/e′ >8 and any of the following: NT-proBNP >220 ng/L **or** atrial fibrillation
- Age ≥18 years
- Symptom severity and heart failure medication were stable during the last 4 weeks
- General ability of the patient to declare willingness to participate in this trial
- Written informed consent

Patienten- / Probandenauswahl 2

Ausschlusskriterien:

- charakterisieren die Merkmale, die eine Aufnahme von Patienten in die Prüfung selbst dann verhindern, wenn die Einschlusskriterien erfüllt sind.

Exclusion criteria

- Non-cardiac causes for heart failure-like symptoms
  - Chronic obstructive pulmonary disease GOLD stages ≥II
  - Anaemia (haemoglobin <11 mg/dL)
  - Significant renal dysfunction (eGFR <30 mL/min/1.73 m^2 indexed to BSA)
  - Significant peripheral artery disease (Fontaine ≥IIb)
  - Musculoskeletal disease that contributes to reduced exercise performance
  - Specific cardiomyopathy (e.g. amyloidosis)
  - Haemodynamically significant valvular disorders
- Significant coronary artery disease (current angina pectoris CCS ≥II or positive stress test, myocardial infarction or coronary artery bypass graft within the last 3 months)
- Any inability or contraindication to participate in CPET or in an exercise programme (e.g. physiological, mental) or to supply essential information (e.g. questionnaire, diary)
- Ineffective control of resting blood pressure (≥140/90 mmHg or ≥160/100 mmHg with ≥3 antihypertensive drugs) or of resting heart rate (≥100 b.p.m.)
- Expected low compliance (e.g. by travel distance to trial site; planned absences longer than 4 weeks during follow-up) or ongoing drug abuse
- Pregnant or nursing women
- Concomitant participation in other interventional clinical trials

1.3.2 Gleichheitsprinzipien

Ceteris paribus (*unter sonst gleichen Bedingungen,*
F.Trixler, https://www.nano.geo.uni-muenchen.de/lectures/wissenschaftliches_Arbeiten_Veroeffentlichen.pdf)

Prinzip in der Wissenschaft und Ökonomie,

- eine Variable zu verändern und
- alle anderen Variablen konstant zu lassen

In interventionellen, vergleichenden klinischen Studien

- unterscheidet sich nur die Intervention und
- alle anderen Umstände werden (möglichst) gleich gehalten
- damit man aus statistischer Assoziation auf die kausale Wirkung der Intervention schließen kann (vgl. Folien 23 – 25)

Vergleichbarkeit I

Vergleichende Untersuchungen haben das Ziel, Versuchs- und Kontrollgruppen hinsichtlich aller Eigenschaften, außer der zu untersuchenden Einflussgröße vergleichbar zu machen bzw. die Vergleichbarkeit zu erhalten.

- Dennoch können sich, trotz aller Bemühungen zwei Arten von Abweichungen ergeben:

1. Zufällige Abweichungen

liegen immer auf Grund der Variabilität der Probanden und der Umgebung vor

- diese mitteln sich bei vielen Probanden heraus, sodass
- mit Hilfe statistischer Verfahren Aussagen möglich werden.

Vergleichbarkeit II

2. Systematische Abweichungen:

sind Abweichungen, welche Ergebnisse in eine Richtung verzerren (engl.: bias)

- Diese mitteln sich nicht heraus und können die Interpretation empfindlich stören.
- Verzerrungsquellen kann man nicht per se mit höherer Fallzahl beseitigen.
- Bias kann nur prospektiv durch Maßnahmen der Versuchs-planung
  - vermieden /
  - vermindert
  - bzw. kontrolliert werden.

Bias

- systematischer Fehler, Vorurteil, Verzerrung …

Biasarten:

- Selektion bias
- Detection bias
  (häufig verändert der Messprozess das zu Untersuchende selbst)
- Publication bias …

Bias kann häufig nicht durch biometrische Maßnahmen (z. B. Erhöhung der Fallzahl) eliminiert werden, sondern bedarf z. B. ein verbessertes Studiendesign.

Kriterien der Vergleichbarkeit

1. Strukturgleichheit
2. Behandlungsgleichheit
3. Beobachtungsgleichheit
4. Auswertungsgleichheit

Strukturgleichheit I

Strukturgleichheit:

- Die zu vergleichenden Therapiegruppen unterscheiden sich vor Beginn einer Studie in allen wichtigen Eigenschaften und Störgrößen nicht oder nur zufällig.
- Strukturgleichheit erreicht man durch

Randomisation

R. A. Fisher 1925. Statistical Methods for Research Workers

- d. h., Randomisation schafft im Mittel Gleichheit hinsichtlich aller, auch unbekannter Confounder.
- In der Randomisation hilft der Zufall dem Wissenschaftler



Strukturgleichheit II

Wichtige ethische Voraussetzungen für die Randomisation bei klinischen Prüfungen:

1. Es gibt keine Beweise für die Unterlegenheit einer Behandlungsmodalität (Prinzip der vergleichbaren Unsicherheit).
2. Die Behandlungen sind aufgrund ihrer Applikationsweise zumutbar.
3. Der Patient wird aufgeklärt und stimmt der Verfahrensweise schriftlich zu:

Informed consent

Randomisation I

Randomisationstechniken:

Wir betrachten eine Gruppe von Patienten:

- 2 Behandlungsgruppen A und B im Verhältnis 1 : 1
- Einfache Randomisation:
  - Sukzessives wiederholtes Anwenden eines Zufallsgenerators (z.B. Münze: Wappen – Gruppe 1, Zahl – Gruppe 2)

Nr.	1	2	3	4	5	6	7	8	9	10
Gruppe	2	2	2	1	2	1	1	2	2	1

Nachteil:
Mögliche Imbalancen bei kleinen Fallzahlen

Randomisation II

- Blockrandomisation:

  gewährleistet gleiche Fallzahl in den zu vergleichenden Therapiegruppen (Balanciertheit), z. B.

  - zufällige Auswahl aus Blöcken, z. B. AABB, BAAB, ABAB, BABA, …
  - Nachteil: Gruppe teilweise vorhersagbar

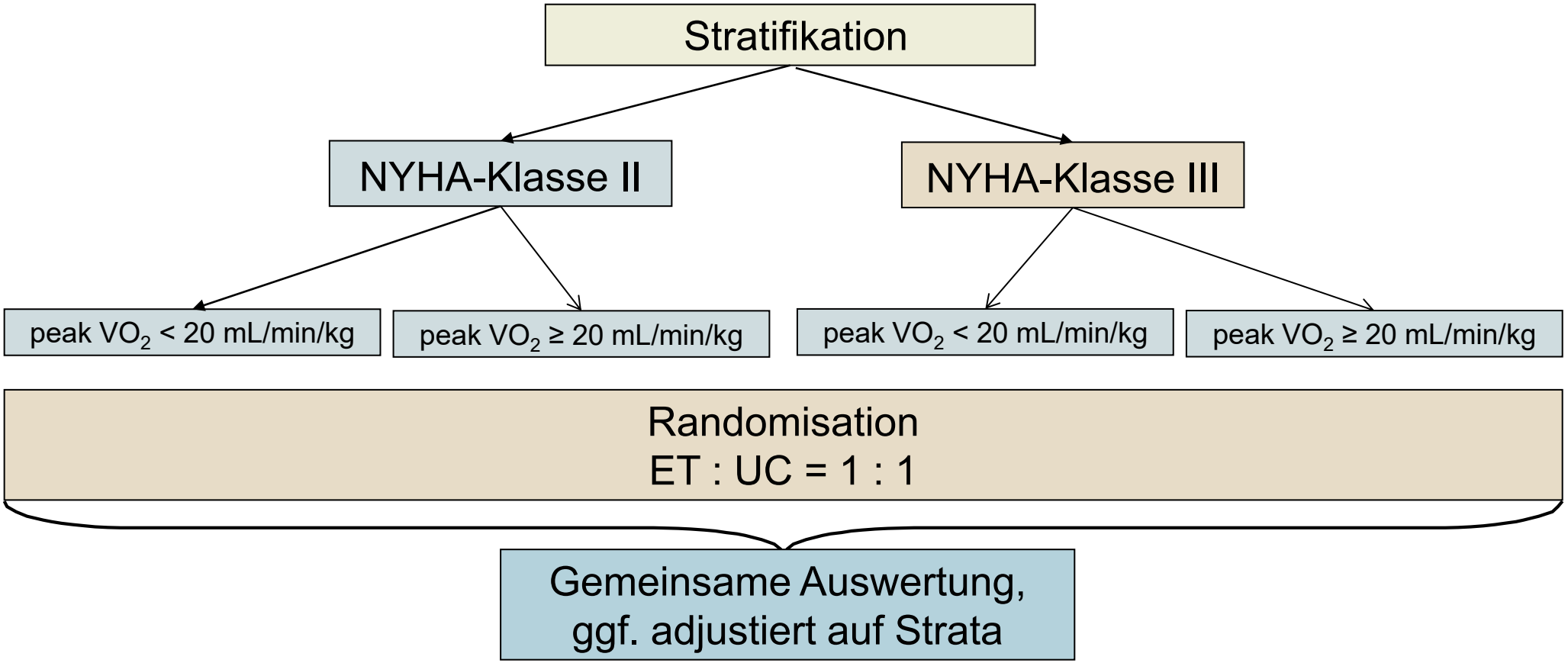
- Stratifizierte Randomisation:

  gewährleistet annähernd Gleichverteilung von Fallzahlen und Störgrößen

  - Entsprechend der Kategorien bekannter Störgrößen (z. B. Tumor-grad) werden Schichten (Strata) gebildet und
  - Patienten werden innerhalb der Schichten randomisiert

Beispiel: Stratifizierte Randomisation I

Beispiel: Ex-DHF-Studie



Beispiel: Stratifizierte Randomisation II

In der Ex-DHF-Studie wurde mit Hilfe eines Minimierungs-verfahrens mit Zufallskomponente stratifiziert randomisiert.

	ET	UC	Total
NYHA class			
II	139 (86%)	138 (86%)	277 (86%)
III	22 (14%)	23 (14%)	45 (14%)
Total	161 (100%)	161 (100%)	322 (100%)
peak VO_2			
< 20 mL/min/kg	53 (33%)	52 (32%)	105 (33%)
≥ 20 mL/min/kg	108 (67%)	109 (68%)	217 (67%)
Total	161 (100%)	161 (100%)	322 (100%)

Kontrollgruppe

Um herauszufinden, wie sich Probanden entwickeln, welche die experimentelle Intervention nicht bekommen, benötigt man eine Kontrollgruppe, wenn möglich Placebo:

therapeutische Intervention, die von der aktiven Therapie nach Art der Verabreichung, Aussehen, Farbe, Geschmack und Geruch nicht zu unterscheiden ist, aber keinen spezifischen bekannten Wirkmechanismus hat (IQWiG-Glossar, https://www.iqwig.de/sonstiges/glossar/placebo.html)

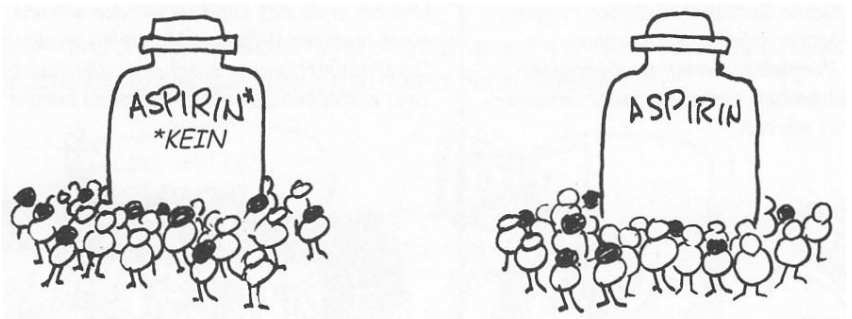
- Weil es ethisch oft nicht vertretbar ist, Patienten *nicht* zu behandeln, werden Patienten der Kontrollgruppe mit Standard-behandlung (nach Leitlinie / usual care / treatment as usual) behandelt.

Kontrollgruppe: Placebo

- Unter Placeboeffekt versteht man die objektive, erfolgreiche Behandlung mit einem therapeutischen Manöver, in dem kein Wirkstoff zum Einsatz kommt.
- Der Effekt beruht im wesentlichen auf Fremd- und Autosuggestion.
- Er ist ein wichtiges Behandlungsprinzip insbesondere bei Erkrankungen mit psychogenem Anteil.

zum Beispiel bei:

- Schmerztherapie
- Asthma
- Schlaflosigkeit
- Herzrhythmusstörung



Gonick & Smith: Statistik in Cartoons

Behandlungsgleichheit I

Behandlungsgleichheit:

- Probanden / Patienten in den zu vergleichenden Gruppen werden außer in Bezug auf die zu vergleichenden Therapien (möglichst) gleich behandelt
- Maßnahmen für Behandlungsgleichheit

⇨

- Verblindung
- Einsatz von Placebo
- Standardisiertes Vorgehen bei der Behandlung

Behandlungsgleichheit II

Was kann im Rahmen der Patientenbehandlung den Therapieerfolg beeinflussen?

zum Beispiel:

- Begleittherapie
- Prophylaktische Maßnahmen
- Therapiemodifikation
- Technische Abläufe
- Zuwendung des Arztes und des Pflegepersonals



Unterscheiden sich diese Behandlungsmaßnahmen in den zu vergleichenden Therapiearmen, kommt es zu Verzerrungen

Behandlungsgleichheit III

Placebo (bei Medikamentenstudien) und Sham-Prozeduren ermöglichen die Schätzung eines Placebo-Effekts. Damit lassen sich die Interventionen verblinden.

Verblindung

- für Patient: Einfach-Blind-Versuch
- für Patient und Arzt: Doppelt-Blind-Versuch
- für Patient, Arzt und Biometriker Dreifach-Blind-Versuch

Probleme:

- Sham-Prozeduren / Placebos werden nicht durch die Kassen bezahlt
- Verblindung (besonders bei invasiven Therapien) nur selten möglich

Beobachtungsgleichheit I

Beobachtungsgleichheit:

Alle Merkmale, insbesondere der Behandlungserfolg, müssen an allen Patienten objektiv unter gleichen Bedingungen - insbesondere unabhängig von der bei dem einzelnen Patienten angewandten Therapie - erfasst werden.

(Skript Biomathe Uni Münster, https://www.medizin.uni-muenster.de/fileadmin/einrichtung/ibkf/lehre/skripte/biomathe/bio/script9.html, 2024-11-02)

- Wie sichert man Beobachtungsgleichheit?

⇨

- Blindversuche
- Standardisierte Beobachtung und Dokumentation
- Reviewbeurteilung

Beobachtungsgleichheit II

Ziel von Behandlungs- und Beobachtungsgleichheit:

- Ungleichheiten, die (außer der Prüfintervention) im Studienablauf zwischen den Probanden auftreten, sollen
  - vermieden bzw.
  - vermindert bzw.
  - zumindest erfasst werden.

Beispiel:

- Ex-DHF-Studie (Edelmann F et al. Nature Medicine 2025; https://doi.org/10.1038/s41591-024-03342-7)
- Herzecho: Messungen lokal und zentral im Corelab

Auswertungsgleichheit I

Auswertungsgleichheit

die Daten aller Behandlungsgruppen werden hinsichtlich der Wirksamkeits- und Nebenwirkungsparameter gleichartig behandelt

- Zudem soll die Berücksichtigung von Therapieabbrüchen oder von Therapieabweichungen bei der Auswertung in allen Behandlungsgruppen gleichartig sein

Maßnahme:

- Im Statistischen Analyseplan werden vor Studienabschluss alle Rahmenbedingungen für die Auswertung festgelegt.

Auswertungsgleichheit II

Auswertung besonderer Fälle:

Die Wertung besonders verlaufender Fälle wie

- Protokollabweichungen
- Abbruch der Therapie durch den Patienten
- Nebenwirkungsbedingte Therapieänderungen
- Verlust aus dem Follow-up-Prozess (Drop out)



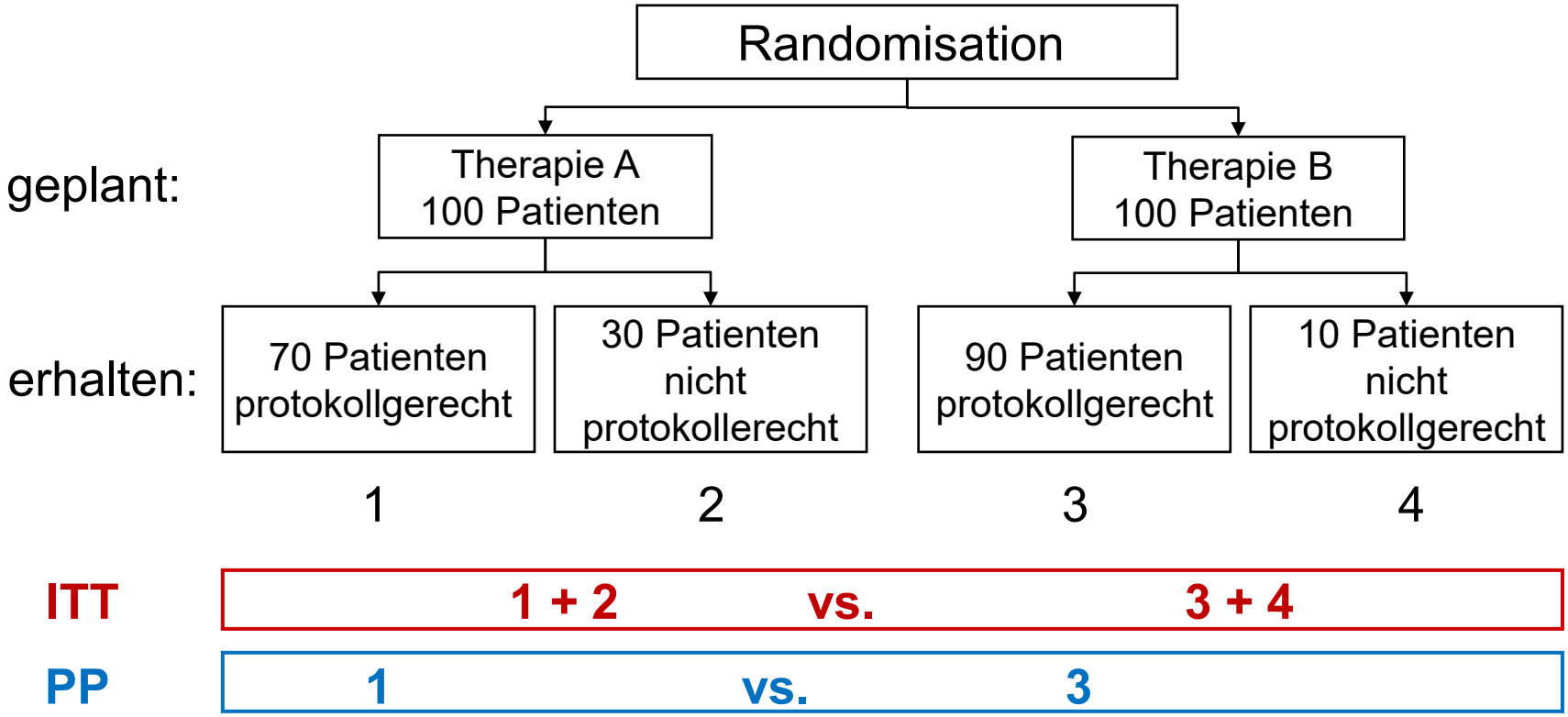
muss einheitlich im Studienprotokoll / statistischen Analyseplan festgelegt werden.

Auswertungsgleichheit III

- Die Intent-to-Treat Analyse (ITT) ist eine wichtige Vorgabe zur Herstellung der Auswertungsgleichheit

  (vgl. ICH.E9 (R1), Nov 2019)
- Die Auswertung erfolgt entsprechend der
  - Randomisation,
  - d. h. nach der geplanten Therapie,
- ungeachtet dessen, welche Therapie die Patienten tatsächlich erhalten haben

1.3.3 Auswertungspopulationen I



Intent-to-treat (ITT): Auswertung nach Randomisation

Per Protocol (PP): Auswertung nach Protokoll

Auswertungspopulationen II

Intent-to-treat (ITT) vs. Per Protocol (PP):

Studie mit 200 Patienten, davon brechen 40 Patienten wegen Nebenwirkungen die Therapie ab (30 in Arm A und 10 in Arm B):

	Arm A 100 Patienten		**Arm B** 100 Patienten	
	ITT	**PP**	**ITT**	**PP**
Erfolg	60	60	54	54
Misserfolg	10	10	36	36
Abbruch	30		10	
	100	**70**	**100**	**90**
Erfolgsraten	60/100 = 60 %	60/70 = 86 %	54/100 = 54 %	54/90 = 60 %

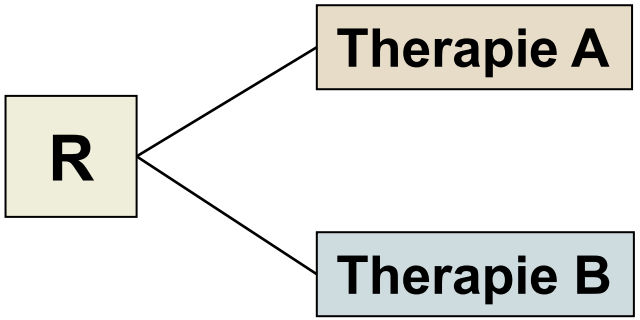
Signifikanz: p(ITT) = 0,39 p(PP) = 0,0004

Beispiel: EX-DHF

- Abweichungen:
  - zufällig: kaum problematisch
  - systematisch: Verzerrungen, die die Studienergebnisse beeinflussen können (Bias)
- Realität:
  - nur 45% der Probanden im ET-Arm nahmen an mindestens 2/3 der geplanten Trainingseinheiten teil

1.3.4 Studiendesign I

Parallelgruppendesign:

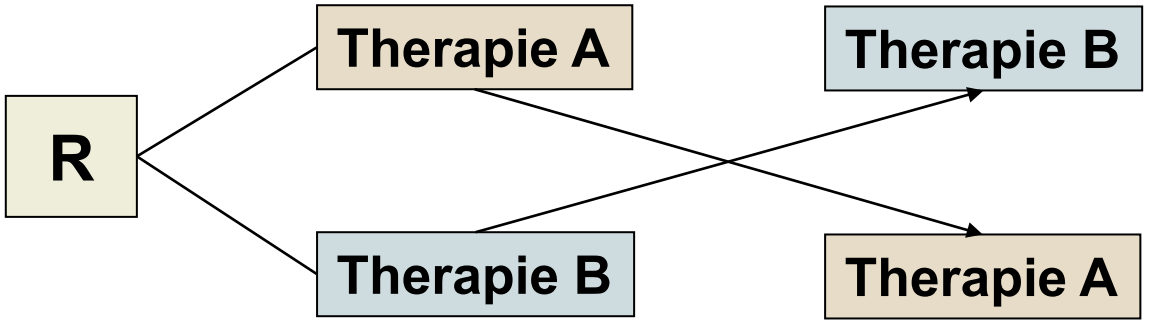


Mögliche Fragestellungen:

- Test auf Unterschied
- Test auf Äquivalenz
- Test auf Nichtunterlegenheit

Studiendesign II

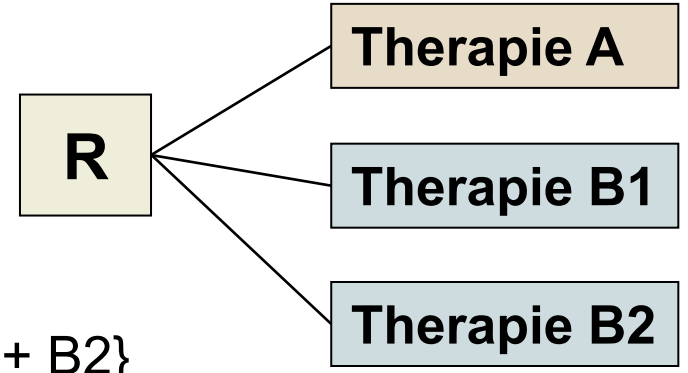
Cross-Over Design:



Ziel:	Effektiver Vergleich von Therapien durch Ausschalten der interindividuellen Variabilität
Voraussetzung:	**kein** Carry-over-Effekt
Carry-over-Effekt:	Carry-over ist das Nachwirken einer Therapie von einer Behandlungsperiode bis in eine nachfolgende Behandlungsperiode

Studiendesign III

Orthogonales Design



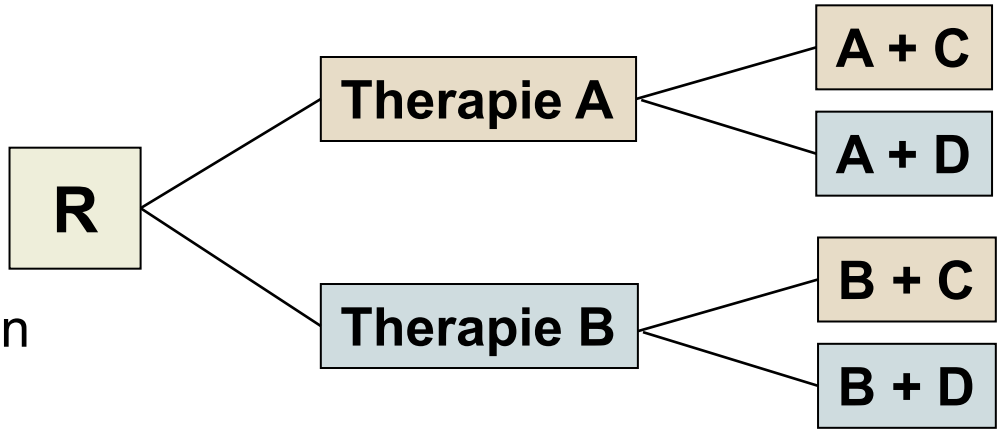
- B1 und B2 können z. B. verschiedene Dosierungen eines neuen Medikaments sein
- Verglichen wird zunächst A vs. {B1 + B2}
  - Ist ein Therapieunterschied nachweisbar, vergleicht man B1 vs. B2

- Hinweis: Orthogonale Designs haben mathematisch gute Eigenschaften. Z. B. wird – trotz mehrfachen Testens (VL 9 / 10) der Fehler I. Art eingehalten.

Studiendesign IV

Zweifaktorielles Design

- geeignet zum Vergleich von Kombinationstherapien (Onkologie!)



- Man vergleicht
  - A vs. B, aber auch
  - C vs. D
- Das *spart* Fallzahl, solange man Interaktionseffekte ausschließen kann.
- Um diese nachzuweisen, braucht es höhere Fallzahlen.

1.3.5 Validität von Studien 1

Bewertung, in welcher Güte Schlussfolgerungen aus Studien gezogen werden können:

Wie professionell / sachlich richtig wurde die Studie in den Zentren durchgeführt?

- Interne Validität:
  - Methodisch korrekte Planung, Durchführung und Analyse, so dass beobachtete Effekte auch der zu untersuchenden Intervention zugeschrieben werden können

  handwerkliche Qualität, Glaubwürdigkeit der Ergebnisse

  - Struktur-, Behandlungs-, Beobachtungs- und Auswertungsgleichheit sind Kriterien der internen Validität

Validität von Studien 2

Bewertung, in welcher Güte Schlussfolgerungen aus Studien gezogen werden können:

Wie sehr sind die Ergebnisse einer Studie verallgemeinerbar (andere Kliniken, andere Länder, andere Patienten, …)?

- Externe Validität (Generalisierbarkeit):
  - Die Studienergebnisse lassen sich unverzerrt unter Praxisbedingungen reproduzieren:
  - Validierung der Ergebnisse anhand unabhängiger Datensätze
  - Übertragbarkeit der Ergebnisse



UNIVERSITÄT LEIPZIG

Medizinische Fakultät

Vielen Dank!

Dr. rer. nat. Meinhard Mende

Senior-Wissenschaftler, Institut für Medizinische Informatik, Statistik und Epidemiologie und Zentrum für Klinische Studien

Härtelstraße 16-18, 04107 Leipzig

Meinhard.mende@zks.uni-leipzig.de