

Quellen, Internet, Software



- British Medical Journal Statistics at Square One
- Deutsches Ärzteblatt Bewertung wissenschaftlicher Publikationen
- Deutsche Medizinische Wochenschrift Statistik-Serie
- Jumbo Münster
- Graphpad Quickcalcs
- R, Stata, SPSS, SAS, Statistica, GraphPad, MedCalc
- nQuery Advisor, East, PASS, StudySize

09.02.2017

hanno.ulmer@i-med.ac.at

Seite 4

Internet Software - Public Domain



Grafstat, Das Fragebogenprogramm
<http://www.grafstat.de/>

Der onlineFragebogen
<https://www.socisurvey.de/>

Open source survey application
<http://www.limesurvey.org/>

Für klinische Studien
<http://project-redcap.org/>

Datenformate
<http://www.cdisc.org/>

5

Häufig verwendete statistische Methoden



■ Studienplanung

- Fragestellung – Hypothesenformulierung
- Auswahl des Studiendesign
- Fallzahlschätzung

■ Statistische Auswertung

- Deskriptive Statistik
- Parameterschätzung mittels **Konfidenzintervalle**
- Hypothesenprüfung mittels Signifikanztests (**P-Werte**)
- Regressionsanalysen z.B. Cox Modell (**Hazard ratios**)

09.02.2017

hanno.ulmer@i-med.ac.at

6

Fragestellungen und bevorzugtes Studiendesign

Therapie	Randomised Controlled Trial (RCT), Verblindung (drei-,zwei-,einfach,offen)
Diagnose	Querschnittsstudie mit Vergleich mit Goldstandard, Verblindung
Ursachen- forschung	Kohortenstudie Case-control study
Prognose	Kohortenstudie

09.02.2017

7

PICO



- **P** = Population, Studienobjekte
Auswahl bzgl. Alter, Geschlecht,
Krankheitsstadium...
- **I** = Intervention(en)
zu untersuchende Behandlungsform,
Risikofaktoren...
- **C** = Comparison Intervention
zu vergleichende Behandlungsform, Goldstandard,
keine Risikofaktoren, kein Vergleich...
- **O** = Outcome(s) of interest
Effekt der Intervention

8

„Wie lese ich eine Studie in 10 Minuten“



- Grundlegender Aufbau eines Studienberichts (papers):

IMRaD Schema
- Weitere Strukturierungen:
 - CONSORT Statement für Interventionsstudien (RCTs)
 - STROBE Statement für Beobachtungsstudien (epidemiologische Studien:
Kohortenstudie, Fall-Kontroll-Studie u. Querschnittstudie)
 - STARD Statement für diagnostische Studien

09.02.2017

hanno.ulmer@i-med.ac.at

9



First European Multicenter Results With a New Transcutaneous Bone Conduction Hearing Implant System: Short-Term Safety and Efficacy

*Georg Sprinzl, †Thomas Lenarz, ‡Arneborg Ernst, §Rudolf Hagen,
*Astrid Wolf-Magele, †Hamidreza Mojallal, ‡Ingo Todt, §Robert Mlynski,
and ||Mario D. Wolfgramm

*Department of Otorhinolaryngology, Medical University Innsbruck, Austria; †Department of Otorhinolaryngology, Hannover Medical School, Germany; ‡Clinic for Ears, Nose, and Throat, Unfallkrankenhaus Berlin, Germany; §Department of Otorhinolaryngology, Plastic, Aesthetic and Reconstructive Head and Neck Surgery, University Clinic Würzburg, Germany; and ||Vibrant MED-EL, Innsbruck, Austria



TABLE 1. Demographic data and medical parameter disease factors of the 12 study participants

Demographics				Disease factors and medical history						
Subject no.	Age at surgery	Sex	Study site	Implanted ear	No. previous ear surgeries	Duration of HL (yr)	Type of HL	Etiology	PTA _a DC implanted ear (dB HL)	PTA _a AC implanted ear (dB HL)
1	69	M	Berlin	R	2	60	CHL	Cholesteatoma	5	45
2	69	F	Berlin	R	4	60	CHL	Cholesteatoma	19	46
3	44	F	Berlin	R	2	9	Mixed	Otosclerosis	35	50
4	28	M	Hannover	R	2	15	CHL	COM	6	30
5	65	F	Hannover	R	1	2	CHL	Glioma tumor	6	66
6	65	F	Hannover	L	1	1	Mixed	Chronic mastoiditis	14	53
7	63	F	Hannover	L	3	22	Mixed	COM	18	67
8	35	M	Würzburg	R	5	35	CHL	Cholesteatoma	8	49
9	20	F	Innsbruck	L	2	20	CHL	Atresia auris	11	73
10	19	F	Innsbruck	R	2	19	Mixed	Cholesteatoma	21	61
11	28	F	Innsbruck	R	0	28	Mixed	Atresia auris	25	83
12	27	F	Innsbruck	R	1	27	CHL	Atresia auris	15	73



Study Design

The study was a prospective, single-subject repeated-measures design, in which each subject served as his/her own control. Performance on audiometric tests preoperatively was compared with the aided 3 month postoperative condition using the Bonebridge. This type of design has been applied frequently to the evaluation of implantable hearing devices in multicenter clinical trials (35–37). It minimizes the effect of variability inherent to the population to the evaluation of treatment outcomes. Standardized evaluation methods were used to assure the reliability of the data across different investigational centers.

Medizinproduktstudie



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Statistical analyses were performed using IBM SPSS Statistics 19 (IBM, Armonik, NY, USA). One-way repeated-measure ANOVAs with time as factor were performed (significance was accepted at $p \leq 0.05$) and followed by post hoc pairwise comparisons to examine significant differences between the single test intervals. For each ANOVA, Mauchly's test of sphericity was applied. If sphericity could not be assumed, a Greenhouse-Geisser correction was used as part of the ANOVA. P-values of the pairwise comparisons were adjusted with the Holm-Sidak method. Box-Whisker Plots represent the whole data set. Whiskers extend to the maximum value within 1.5 times the interquartile range (IQR) above the third quartile or the minimum value within 1.5 times the IQR below the first quartile. Values outside this range are considered to be outliers, depicted as individual dots. Tukey box-whisker plots were generated using GraphPad Prism 5 (<http://www.graphpad.com>).

hanno.ulmer@i-med.ac.at

13

Medizinproduktstudie



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

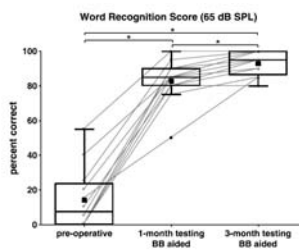


FIG. 3. Word recognition scores in quiet (Freiburger monosyllables) for the implanted ear: preoperative, 1-month postoperative and 3-months postoperative. Both postoperative scores are significantly improved from preoperative scores ($p < 0.001$) and from each other ($p = 0.010$), $n = 12$, BB = Bonebridge.

hanno.ulmer@i-med.ac.at

14

Arzneimittelstudie



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

THE NEW ENGLAND JOURNAL OF MEDICINE

ORIGINAL ARTICLE

Placebo-Controlled Trial of Cytisine for Smoking Cessation

Robert West, Ph.D., Witold Zatonski, M.D., Magdalena Cedzynska, M.A., Dorota Lewandowska, Ph.D., M.D., Joanna Pazik, Ph.D., M.D., Paul Aveyard, Ph.D., M.D., and John Stapleton, M.Sc.

hanno.ulmer@i-med.ac.at

15

Allgemeines zur Studienplanung



Dr. Hanno Ulmer
hanno.ulmer@imed.ac.at
Innsbruck, Oktober 2010

Department für Medizinische Statistik, Informatik und Gesundheitsökonomie, Medizinische Universität Innsbruck

Richtlinien und Empfehlungen



- **Deklaration von Helsinki (World Medical Association)**
- Ethische Gesichtspunkte
- **GCP ... Good Clinical Practice**
 - International anerkannte formale Kriterien
- **ICH ... International Committee on Harmonization**
 - Ergänzt durch nationale Gesetzgebung (z.B. Arzneimittelgesetz)
- **Consort, Stard, Strobe, Remark-Statements**
 - Studienberichterstellung
- **SOPs für Klinische Studien**
<http://www.tmf-ev.de/Produkte/SOP.aspx>

Modul 2.02

hanno.ulmer@i-med.ac.at

20

Deklaration von Helsinki



Regelt seit 1964 ethische Prinzipien für die medizinische Forschung am Menschen
Revision 1983, 1989, 1996, 2000
Download: World Medical Association: <http://www.wma.net>
Inoffizielle Übersetzung: <http://www.bundesaerztekammer.de>

Ethische Grundsätze als Leitlinie für Personen, die in der medizinischen Forschung am Menschen tätig sind, incl. identifizierbaren menschlichen Daten und Materialien

Ethikkommission, Informed Consent

Modul 2.02

hanno.ulmer@i-med.ac.at

21

International Conference of Harmonisation (ICH)



Mitglieder:

- Kommission der Europäischen Gemeinschaft
- European Federation of Pharmaceutical Industries and Associations (EFPIA)
- Ministry of Health, Labor and Welfare (MHLW)
- Japan Pharmaceutical Manufacturers Association (JPMA)
- US Food and Drug Administration (FDA)
- Pharmaceutical Research and Manufacturers of America (PhRMA)

09.02.2017

hanno.ulmer@i-med.ac.at

22

Richtlinie ICH E9 (Statistical Principles for Clinical Trials, 1998)



- Harmonisierung statistischer Vorgehensweisen
 - Biostatistical Methodology in Clinical Trials in Application for Marketing Authorisations (Committee for Proprietary Medicinal Products, 1994)
 - Guidelines on Statistical Analysis of Clinical Studies (Japanese Ministry of Health and Welfare, 1992)
 - Guideline for the Format and Content of Clinical and Statistical Sections of a New Drug Application (U.S. Food and Drug Administration, 1988)

09.02.2017

hanno.ulmer@i-med.ac.at

23

Weitere wichtige Richtlinien



- ICH E6 Guideline for good clinical practice
 - Beschreibung der wichtigsten Elemente
 - Richtlinie muß für Zulassungsstudien befolgt werden
 - Bezieht sich auf Aspekte zur Durchführung klinischer Studien
 - Beschreibt die Rolle der Ethikkommission
 - Beschreibt Aufgaben des Prüfarztes, des Sponsors und des Monitors

09.02.2017

hanno.ulmer@i-med.ac.at

24

CONSORT-Richtlinien



- **CONSORT (Consolidated Standards of Reporting Trials)**
- 22 internationale Standards und Richtlinien zur Darstellung von Ergebnissen in wissenschaftlichen Arbeiten
 - Studienplanung
 - Durchführung
 - Statistische Analyse
 - Interpretation
- Flussdiagramm über die Rekrutierung der Patienten
- Transparenz einer Studie gewährleisten
- www.consort-statement.org/
 - Revidierte Version 2001

Vorschlag „Studienprotokoll – Statistik“



ad 9 Statistik

9.1 Fallzahlplanung

Die Fallzahl wird in der Regel berechnet aus der primären Zielvariablen, dem klinisch relevanten Unterschied, der in der Studie nachgewiesen werden soll, dem statistischen Verfahren, das dazu verwendet wird, sowie aus dem Fehler 1. Art α und der Power $1-\beta$, die erzielt werden soll. Die Fallzahl ist außerdem abhängig von der Anzahl der Gruppen, die verglichen werden sollen. Angaben zu diesen Parametern sowie zum Verfahren der Fallzahlplanung sind erforderlich.

9.2 Randomisierung

Bei der Beschreibung der Randomisierung sind Angaben erforderlich zur Art der Randomisierung, zur Anzahl der Gruppen und der erforderlichen Strata und ggf. zum organisatorischen Ablauf der Randomisierung. Dabei sind besonders bei verblindeten Studien Maßnahmen aufzulisten, die die Verblindung gewährleisten.

9.3 Statistische Methoden

Zielgrößen, Definition von Auswertungskollektiven, Datenanalyse, Zwischenauswertungen, Verweis auf ICH-GCP E9: Statistical Principles for Clinical Trials

Randomisierte Klinische Studien (Randomized Clinical Trials)



Dr. Hanno Ulmer
hanno.ulmer@imed.ac.at
Innsbruck, Oktober 2010

Department für Medizinische Statistik, Informatik und Gesundheitsökonomie, Medizinische Universität Innsbruck

Hierarchie von Medizinischen Studien



09.02.2017

28

RCT: Randomisierte kontrollierte Studie



- **Klinische Studien sind ein Experiment**
 - Einflussfaktor wird gesteuert
 - Alle anderen Faktoren sollen möglichst konstant gehalten werden
- **Randomisierung**
 - Behandlungszuteilung kann nicht vorhergesagt werden
 - Ausschluss von Verzerrungen durch Selektion
- **Stratifizierung – homogene Untergruppen**
 - Alter, Geschlecht, Schweregrad

Modul 2.02

hanno.ulmer@i-med.ac.at

29

RCTs



- **Kontrollgruppe**
 - Placebo / Standardtherapie
 - Statistische Vergleichbarkeit der wesentlichen Merkmale
 - Strukturgleichheit
- **Verblindung**
 - Offen
 - Einfach verblindet
 - Patient weiss nicht, welche Therapie er bekommt
 - Doppel-blind
 - Patient und Arzt wissen nicht, welche Therapie Patient bekommt

Modul 2.02

hanno.ulmer@i-med.ac.at

30

Einige einfache Studienpläne



- **Parallelgruppen**
 - Zwei oder mehrere unabhängige, aber vergleichbare Gruppen
 - Werden zeitlich parallel behandelt
 - Mit unterschiedlichen Therapieformen
- **Cross-Over**
 - Eine Gruppe mit zwei oder mehreren aufeinander folgenden Therapien behandeln
 - Jeder Proband erhält jede Therapie aber in unterschiedlicher Reihenfolge
 - Randomisierung der Reihenfolge
 - z.B.: ABC, ACB, BAC, BCA, CAB, CBA
 - Carry-over-Effekte beachten
- **Faktorielle Pläne**
 - Kombination von zwei oder mehreren Einflussfaktoren
 - Wechselwirkungen überprüfbar

Modul 2.02

hanno.ulmer@i-med.ac.at

31

Parallelgruppenstudie



- z.B. Untersuchung der Wirksamkeit eines neuen blutdrucksenkenden Medikaments im Vergleich zu einer Standardtherapie (=Kontrolle)
 - Manchmal Placebo
- **Nullhypothese H_0** : die beiden Therapien sind im Mittel gleich wirksam
 - z.B. die Änderung des diastolischen Blutdrucks ist im Mittel in beiden Gruppen gleich hoch
 - Die Ungültigkeit der Nullhypothese ist zu beweisen!
- **Alternativhypothese H_1** : die beiden Therapien sind im Mittel unterschiedlich stark wirksam
- **Voraussetzung**: Die beiden Gruppen stimmen in den wesentlichen Merkmalen überein – siehe Randomisierung
 - Nur die Therapien sind unterschiedlich

Modul 2.02

hanno.ulmer@i-med.ac.at

32

Verblindung



- **Ausschaltung von Verzerrungen/systematischen Fehlern (Bias) aufgrund von Vorinformationen**
 - Selektions-Bias, Beobachter-Bias
- **Ausprägungen des Einflussfaktors sind für den Beobachter unbekannt**
 - **Doppel-Blindstudie**
 - Patient und Arzt wissen nicht, welche Therapieform angewendet wird
 - Nicht immer möglich
- **Beschreibung der Art der Verblindung**
 - äußere Form, Geschmack, Farbe
 - sichtbare Unterschiede bei Behandlung (z.B. Farbe des Urins)
- **Beurteilung des Outcomes**
 - unabhängige Beurteilung durch Dritte, Rater Blinding

Modul 2.02

hanno.ulmer@i-med.ac.at

33

Randomisierung



- Zufällige Zuteilung eines Patienten zu einer von zwei oder mehreren Therapien
 - Um in der Praxis eine Balance zwischen den Gruppen zu erreichen, werden kleinere Blöcke mit $m=4,6,8,\dots$ Personen randomisiert
 - Wenn m zu klein ist, besteht die Gefahr, dass der Randomisierungcode bekannt wird

Modul 2.02

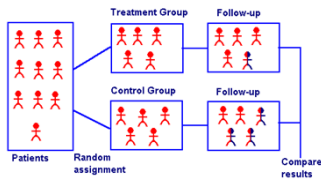
hanno.ulmer@i-med.ac.at

34

Randomisierung



- Zufällige Zuordnung eines Subjektes/Objektes einer Stichprobe zu einer der Gruppen des Einflussfaktors
 - z.B. mit Hilfe eines Zufallsgenerators
- ZIEL:
 - Vermeidung eines Selektions-Bias
 - Jedes Objekt hat die gleiche Chance zufällig einer Gruppe zugeteilt zu werden
 - Strukturgleichheit (Zusätzliche Stratifizierung ist möglich)

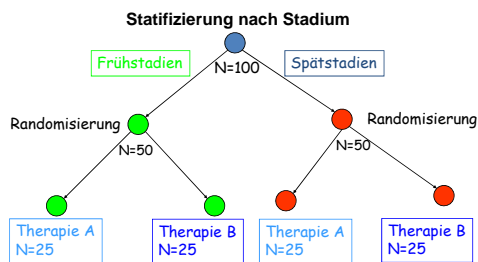


Modul 2.02

hanno.ulmer@i-med.ac.at

35

Randomisierung / Stratifizierung



Modul 2.02

hanno.ulmer@i-med.ac.at

36

Vergleichbarkeit der Gruppen



- **Bezüglich der**
 - Gemessenen Merkmale
 - Dokumentieren
 - Bei der Auswertung berücksichtigen
 - Nicht-gemessenen Merkmale
 - Unbekannten Merkmale
 - Können möglicherweise zu Verzerrungen führen
- **Randomisierung**
- **Stratifizierung**
- **Paarbildung**
 - Zwillinge, Matching

Patientenflussdiagramm

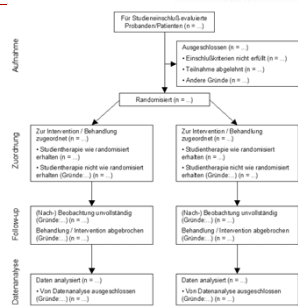


- **Intention-To-Treat**

versus

- **Per-Protocol**

- **Vergleichbarkeit der Gruppen:**
 - Strukturgleichheit
 - Beobachtungsgleichheit
 - Behandlungsgleichheit



Mögliche Verzerrungen (=Bias) in RCTs



- **Bias = systematische Abweichung, Verzerrung**
- **Selection bias**
 - Vermeidung durch Randomisierung
- **Erhebungsfehler**
 - Subjektive Aspekte
 - Vermeidung durch Verblindung
- **Drop-out und Responder bias**
- **Publication bias**
 - Language bias, Zeitschriften – Proceedings,
 - Fast nur positive Resultate, ...
- **Interpretations- und Wahrnehmungsbias**

Fallzahl



- Fixe, berechnete Fallzahl
 - Basierend auf Annahmen über Unterschiede/Behandlungseffekte
 - Fehler 1. und 2. Art (z.B. Signifikanzniveau=0.05, Power=80%)
- Sequentielle Pläne
 - 1 bis K geplante Zwischenauswertungen
 - Interimsanalysen, Adaptive Studienpläne
- Berücksichtigung der Ausfallsrate bei statistischen Analyse
 - **Intention-to-treat** Analyse: alle **eingeschlossenen** Fälle
 - **Per Protokoll** Analyse: alle **abgeschlossenen** Fälle

Checklist für Fallzahlschätzung (Testproblem)



- Fehler 1. Art (üblicherweise 0,05)
- Fehler 2. Art (0,1 oder 0,2)
- Auswahl des Hauptzielkriteriums
- Zu erwartender Unterschied und Angabe eines Variationsmaßes
- Begründung dafür – Literatur oder Vorstudie
- Auswahl des statistischen Tests
- Falls mehrere Hypothesen formuliert werden, Korrektur des Fehler 1. Art oder Hierarchisierung der Hypothesen
- Drop-Out Rate berücksichtigen

Fallzahlschätzung



- Je größer die Fallzahl desto geringer die Wahrscheinlichkeit für den Fehler 2. Art, desto höher die Präzision der Schätzung
- Fallzahlschätzung bedingt bereits die Definition des Testproblems durch Operationalisierung der Fragestellung, durch Formulierung von Null- und Alternativhypothese
- Fallzahlschätzung bedingt das Festlegen einer speziellen Alternative. „Es gibt keinen Unterschied zwischen den Gruppen“ genügt nicht. Die Größe des zu erwartenden Unterschieds muss festgelegt werden.
- Fallzahlschätzung bedingt bereits die Auswahl des statistischen Tests, die Definition von Fehler 1. und 2. Art

Vergleich von Häufigkeiten zweier unverbundener Stichproben



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

In einer Karzinomstudie erhofft man sich, die bisher mit der Standardtherapie (S) erreichte Rezidivrate von 50% mit einer neuen Therapie (T) auf 40% zu senken.

Für eine kontrollierte klinische Studie mit den beiden Therapiegruppen S und T und dem Zielkriterium 'Rezidivrate' sowie den weiteren Festlegungen (zweiseitiger Test, Irrtumswahrscheinlichkeit 5%, Power 80%) ergeben sich folgende Parameter für die Fallzahlberechnung:
 $p_1 = 0.5$, $p_2 = 0.4$, $\alpha = 0.05$, $1 - \beta = 0.80$.

ergibt eine Fallzahl von 388 Patienten für jede Behandlungsgruppe.

Vergleich von Mittelwerten zweier unverbundener Stichproben



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

In einer Hypertoniestudie erhofft man sich, die bisher mit der Standardtherapie (S) erreichte durchschnittliche Blutdrucksenkung von 15 mm mit einer neuen Therapie (T) auf 20 mm zu senken.

Für eine kontrollierte klinische Studie mit den beiden Therapiegruppen S und T und dem Zielkriterium 'Senkung des Blutdrucks' sowie den weiteren Festlegungen (Standardabweichung = 15, zweiseitiger Test, Irrtumswahrscheinlichkeit 5%, Power 80%) ergeben sich folgende Parameter für die Fallzahlberechnung:
 $\mu_1 = 15$, $\mu_2 = 20$, $\sigma = 15$, $\alpha = 0.05$, $1 - \beta = 0.80$

ergibt eine Fallzahl von 142 Patienten für jede Behandlungsgruppe.

Vergleich von Mittelwerten zweier verbundener Stichproben



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

In einer Phase-II-Studie soll überprüft werden, ob sich ein neues Medikament zur Blutdrucksenkung eignet. Geeignet ist das Medikament dann, wenn bei Hypertonikern mit einem durchschnittlichen systolischen Blutdruck von 150 mm eine Senkung um mindestens 10 mm erreicht wird.

Für diese klinische Studie lautet das Zielkriterium 'Differenz des Blutdrucks vor und nach Behandlung'. Mit den weiteren Festlegungen (Standardabweichung = 15, zweiseitiger Test, Irrtumswahrscheinlichkeit 5%, Power 80%) ergeben sich folgende Parameter für die Fallzahlberechnung:
 $\mu_1 = 150$, $\mu_2 = 140$, $\sigma = 15$, $\alpha = 0.05$, $1 - \beta = 0.80$

ergibt eine Fallzahl von 20 Patienten.

Vergleich von Überlebenszeiten zweier unverbundener Stichproben



In einer Karzinomstudie erhofft man sich, die bisher mit der Standardtherapie (S) erreichte mediane Überlebenszeit von 36 Monaten mit einer neuen Therapie (T) auf 48 Monate zu erhöhen.

Für eine kontrollierte klinische Studie mit den beiden Therapiegruppen S und T und dem Zielkriterium 'Überlebenszeit' sowie den weiteren Festlegungen (Rekrutierungszeit=24 Monate, Nachbeobachtungszeit = 36 Monate, zweiseitiger Test, Irrtumswahrscheinlichkeit 5%, Power 80%)

ergibt eine Fallzahl von 349 Patienten für jede Behandlungsgruppe.

Zusammenfassung: Studiendesigns wichtige Begriffe (1)



- Arzneimittel vs. Medizinprodukt
- Prüfsubstanz vs. Kontrollsubstanz
- Kontrolliert vs. Unkontrolliert
- Monozentrisch vs. Multizentrisch
- Offen vs. Einfachblind vs. Doppelblind
- Randomisiert vs. stratifiziert randomisiert vs. nicht Randomisiert
- Intention-to-treat versus per-protokol
- Parallelgruppen vs. Crossover vs. Faktoriell
- Superiority vs. Non-Inferiority
- Efficacy vs. Safety
- Enrichment Designs, Biomarker, Withdrawal Trials

09.02.2017

47

Zusammenfassung: Studiendesigns wichtige Begriffe (2)



- Retrospektiv vs. prospektiv
- Nicht-interventionelle Studie (NIS)
- Beobachtungsstudien, z.B. epidemiologische Studien wie Kohorten- und Fall-Kontroll Studien
- Hauptzielgröße („Primärer Endpunkt“): klinische Relevanz, Studiendesign + Fallzahlberechnung (Power)
- Nebenzielgrößen („Sekundäre Endpunkte“)
- Subgruppenanalysen, post-hoc Analysen
- Interimsanalysen
- Signifikanztests, Konfidenzintervalle
- Bias, Confounding

09.02.2017

48
