

Journalisten-Workshop im PresseClub München, 25. Juli 2007

Studientypen und ihre wissenschaftliche Relevanz -
Probleme und Fehlerquellen bei der Durchführung und Auswertung
Teil 1 von 3

> **Teil 1: Welche Studientypen gibt es?**

Prof. Dr. Joerg Hasford, Institut für Medizinische
Informationsverarbeitung, Biometrie und Epidemiologie
der Ludwig-Maximilians-Universität München
E-Mail: has@ibe.med.uni-muenchen.de



GESELLSCHAFT FÜR ANGEWANDTE
VITAMINFORSCHUNG E.V.

Studientypen und ihre wissenschaftliche Relevanz - Probleme und Fehlerquellen bei der Durchführung und Auswertung

Prof. Dr. Joerg Hasford

IBE Institut für Medizinische Informationsverarbeitung,
Biometrie und Epidemiologie

Ludwig-Maximilians-Universität München

Email: has@ibe.med.uni-muenchen.de

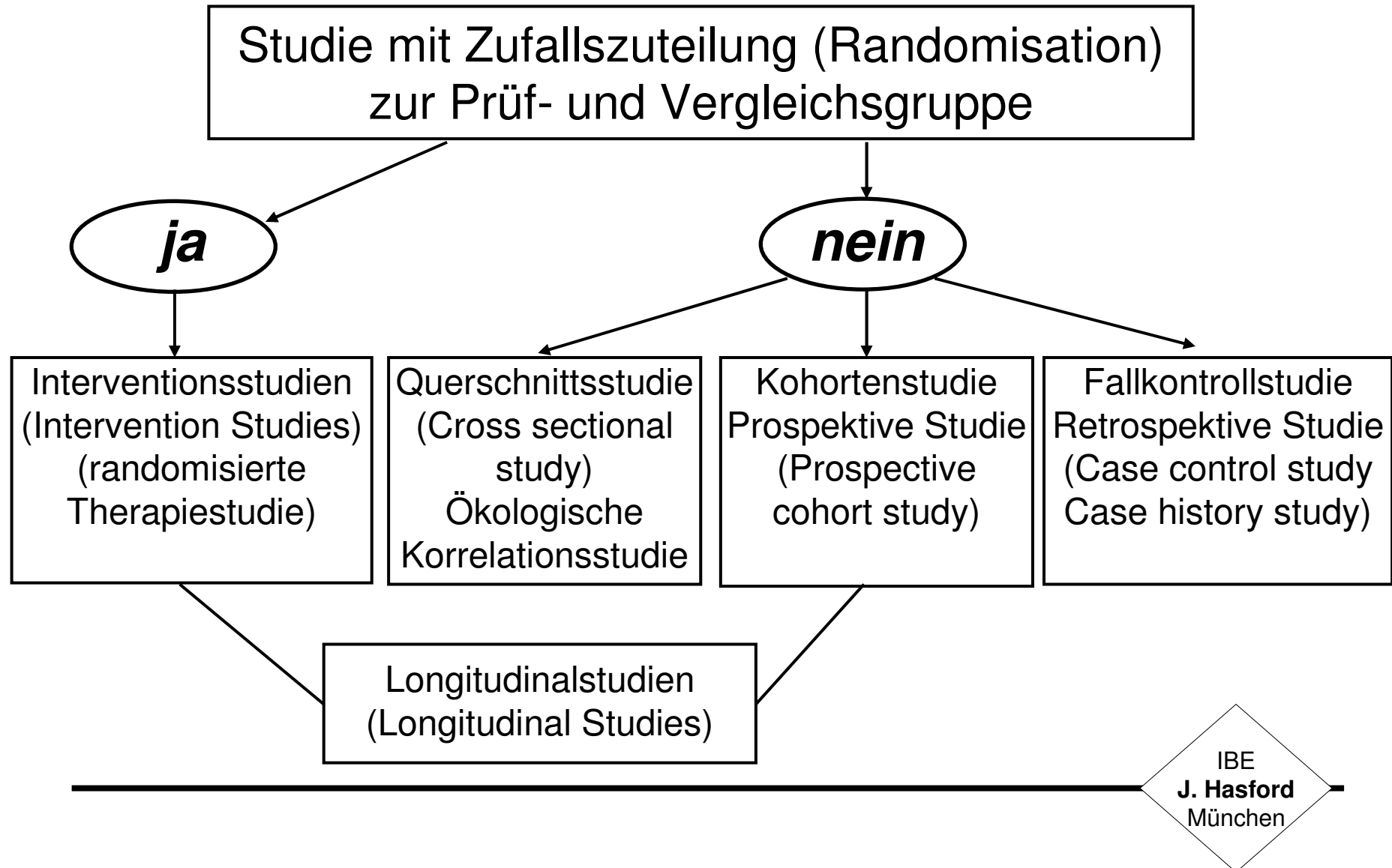


IBE
J. Hasford
München

Ziel aller Studien

Kausalketten zwischen
Arzneimittelanwendung / Ernährung /
Lebensmitteln und erwünschten /
unerwünschten Ereignissen und
Veränderungen zu untersuchen

Studientypen



Ökologische- / Korrelations-Studien

ANSATZ

Vergleich der Erkrankungsraten mit dem pro-Kopf-Verbrauch bestimmter Lebensmittel in dieser Population, z.B.

- Fleischverbrauch und die Inzidenz von Darmkrebs ($r = 0,85$ (m), $0,89$ (f))
- % der Kalorien in Fett und Inzidenz von Brustkrebs

Ökologische / Korrelations-Studien

SCHWÄCHEN

- Risikofaktoren für Erkrankung können variieren, sind aber oft unbekannt oder empirisch nicht verfügbar.
- basieren auf Populations- und nicht Individualdaten, z.B. Alkohol oder Fettkonsum und Brustkrebs
- fehlender Nachweis, dass es die Exponierten sind, die erkranken, d.h. kein Kausalschluss möglich
- keine Möglichkeit der unabhängigen Wiederholung der Studie

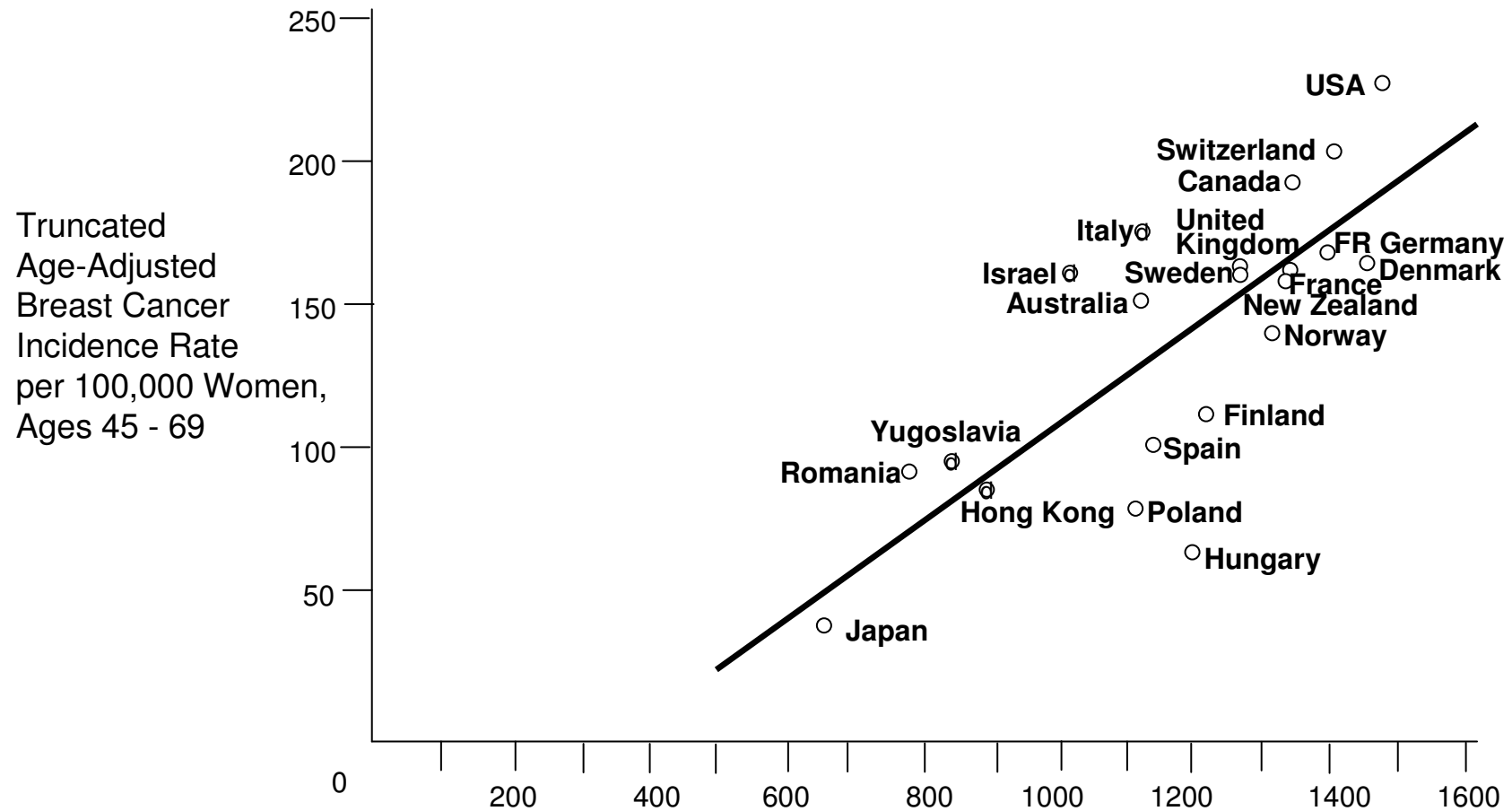


Figure 1. Breast cancer incidence rates from ref. 4 vs. per capita fat supply data from ref. 3 for 21 countries having good cancer registration

Ökologische / Korrelations-Studien

DATENBASIS

- „Schwund-Daten“, d.h. nationale quantitative Daten zur Erzeugung und Import eines Lebensmittels minus den Exportmengen, den Tierfuttermengen und sonstigem nicht alimentären Verbrauch
- Ernährungs-Survey-Daten, die oft auch nach Alter und Geschlecht stratifizierbar sind

Ökologische / Korrelations-Studien

STÄRKEN

- Unterschiede in der Ernährungsweise sind oft groß, z.B. mittlerer Fettverzehr
- Variabilität im pro-Kopf-Verbrauch innerhalb eines Landes ist meist über ein bis zwei Dekaden gering.
- große Fallzahlen (Populationen)
→ kleine Streuungen

Ökologische / Korrelations-Studien

POPULATIONEN MIT SPEZIELLEN ERNÄHRUNGSGEWOHNHEITEN

- oft religiöse oder ethnische Gruppierungen
- z.B. Siebenter Tags Adventisten sind weitgehend Vegetarier
 - Darmkrebs ↓
 - Brustkrebs ø
- oft auch sonst anderer Lebensstil, z.B. Alkohol, Rauchen
- Migrantenstudien
 - ➔ Einfluss genetischer Faktoren

Querschnittsstudien

Beschreiben Ernährung und Krankheiten in einer definierten Stichprobe zu einem festgelegten Zeitpunkt,

z.B.: Erhebung der Ernährung und der Zahngesundheit von Schulkindern am 15. Mai 2007, oder in der 20. Kalenderwoche 2007

Querschnittsstudien

STÄRKEN

- schnell durchführbar, vergleichsweise billig
- Schätzung der Prävalenz
- Identifizierung von Fällen für FKS

Querschnittsstudien

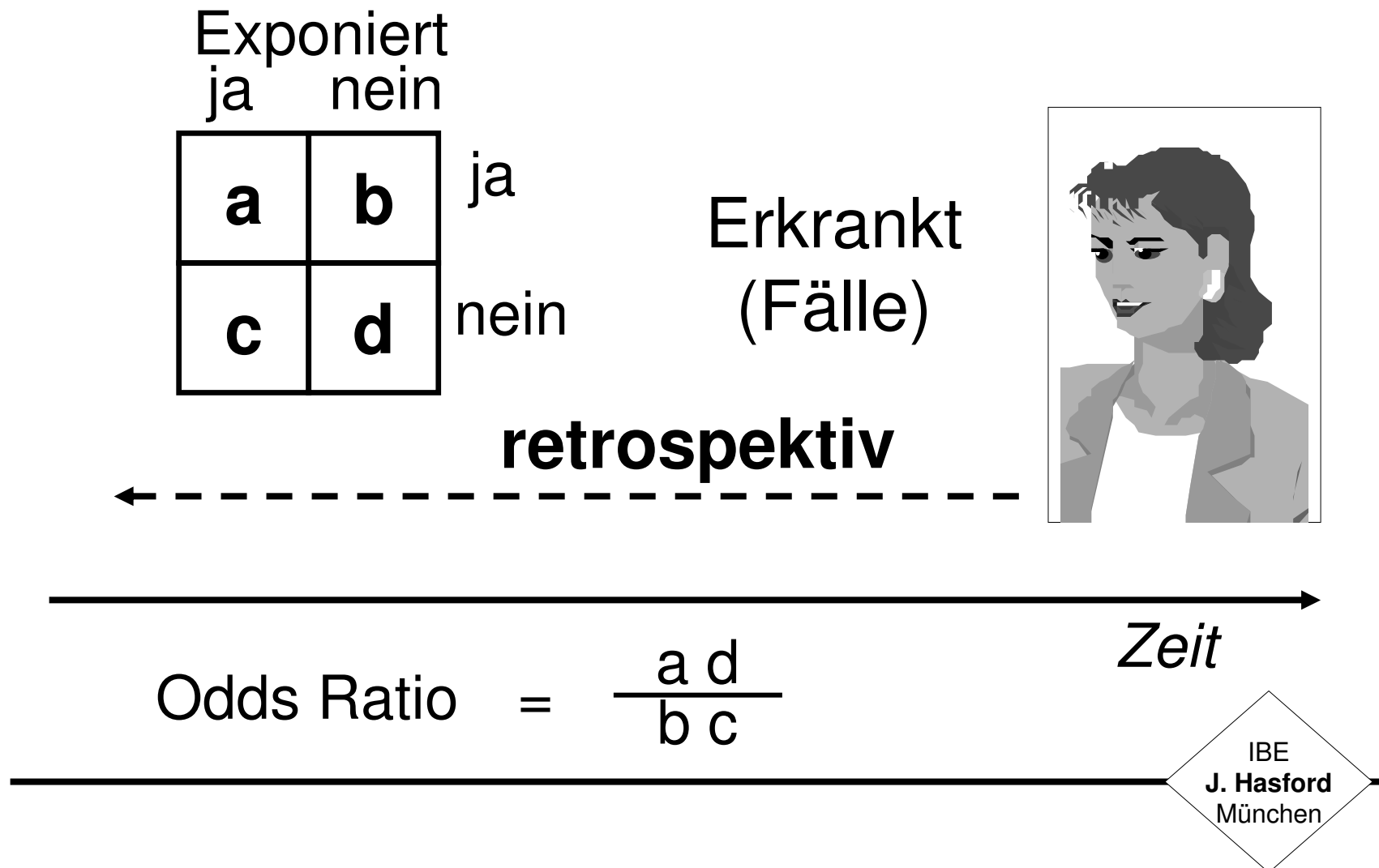
SCHWÄCHEN

- zeitliche Abfolge von Ursache und Wirkung nicht beurteilbar → keine Kausalschlüsse möglich
- keine sichere Erkennung von riskanten Lebensmitteln und Ernährungsweisen möglich
- für seltene Ernährungsweisen und Wirkungen nicht geeignet

Fallkontrollstudien

Retrospektive Beobachtungsstudie, bei der z.B. die Ernährung von Erkrankten mit der von Gesunden verglichen wird.

Fallkontrollstudien



The Odds Ratio

The odds or probability of being exposed vs. being unexposed to the drug for the cases is a / c .

The odds of being exposed vs. unexposed for the controls is b / d .

$$\Rightarrow \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a d}{b c}$$

Fallkontrollstudien

STÄRKEN

- personenbezogene Ernährungs- und Outcome-Daten
- geeignet für seltene Erkrankungen und lange Latenzzeiten
- vergleichsweise kostengünstig
- keine ethischen Probleme

Fallkontrollstudien

SCHWÄCHEN

- fast unvermeidliches, hohes Potenzial für Störfaktoren und Verzerrungen (Bias)
 - Erinnerungsvermögen (Recall) bias
 - Auswahl / Einschluss von Fällen und Kontrollen (Selection bias)
 - Confound by indication bias
- Teilnahmebereitschaft (Non participation bias)
- inkonsistente Ergebnisse häufig, wenn mehrere FKS durchgeführt wurden

Evaluation Criteria for Case-Control Studies

1. Selection method for cases and controls should be defined before the study starts.
2. Level of drug exposure should be established before data analysis.
3. Data collection should be unbiased by the use of blind interviewers relative to case versus control status of subjects, and the use of structured data-collection instruments.
4. Patient drug exposure recall should be equivalent in cases and controls.
5. Exclusion criteria should be unbiased for cases or controls.
6. Equal diagnostic examination based on explicit criteria should be implemented for cases and controls.

7. Similar intensity of medical surveillance for cases and controls before enrollment in the study should be observed.
8. Equal demographic susceptibility for the use of the drug by cases and controls should be assured.
9. Equal clinical susceptibility, or pharmacogenetic attributes, by cases and controls should be assured.
10. Berkson's bias should be avoided; cases and controls should arrive from a common defined population rather than selected from patients admitted to a hospital, because patients who are both exposed and diseased often are more likely to be hospitalized than those who are unexposed and diseased or exposed and non-diseased.

Kriterien für die Bewertung von Fallkontrollstudien

PLANUNG UND DESIGN

- ✓ eindeutige a priori Fragestellung und Hypothesen
- ✓ klare und einheitliche Ein- und Ausschlusskriterien für Fälle und Kontrollen
- ✓ Selektion der Fälle - Repräsentativität
- ✓ Selektion der Kontrollen - Repräsentativität
- ✓ Fallzahlschätzung
- ✓ Einsatz von Blindungstechniken
- ✓ Kontrolle von Bias

Kriterien für die Bewertung von Fallkontrollstudien

DURCHFÜHRUNG

- ✓ Responder-Raten von Fällen und Kontrollen
- ✓ Fehlklassifikationen durch Mängel in der Datenerhebung
- ✓ Confounding by indication
- ✓ Self-selection / diagnostic bias

Kriterien für die Bewertung von Fallkontrollstudien

AUSWERTUNG

- ✓ klare Trennung konfirmatorischer von explorativer Auswertung / multiple Tests
- ✓ Nachvollziehbarkeit
- ✓ Adäquate statistische Verfahren

Kohortenstudien

Prospektive Beobachtungsstudie mit oder ohne Vergleichsgruppe. Gruppen mit unterschiedlichen Expositionen (z.B. bei Ernährung, Arzneimitteln, Arbeitsplatzbelastungen) werden bezüglich Outcomes verglichen.

Kohortenstudie



		Erkrankt	
		ja	nein
Exponiert	ja	a	b
	nein	c	d

-----**prospektiv**----->

Relatives Risiko: $\frac{a}{(a + b)} : \frac{c}{(c + d)}$ *Zeit* →

Relatives Risiko: $\frac{a}{(a + b)} : \frac{c}{(c + d)}$

Zurechenbares Risiko: $\frac{a}{(a + b)} - \frac{c}{(c + d)}$

Das zurechenbare Risiko gibt den Teil der Erkrankungshäufigkeit an, der der Exposition zugerechnet werden kann.

ZR ist essentiell für die Risikobewertung.

Kohortenstudien

STÄRKEN

- Prospektiv (U → W)
- zeitlicher Ablauf der Sequenz: U → W
beurteilbar
- wiederholte standardisierte Messung der
Exposition personenbezogen möglich
- einheitliche Nachbeobachtung,
ggf. per Telefon und Fragebogen

Kohortenstudien

STÄRKEN

- geeignet auch für seltene Expositionen
- Analyse einer Vielzahl von Outcomes
- Risiko und Nutzen direkt bestimmbar
- Verknüpfung mit Krankheits- und Sterberegistern im Prinzip möglich

Kohortenstudien

SCHWÄCHEN

- (zu) hoher Aufwand für seltene Erkrankungen oder lange Latenzzeiten
- erhebliches Potenzial für Störfaktoren bei Auswahl / Einschluss von Teilnehmern (Selektion und non participation bias)
- Confounding by indication bias
- mangelhafte Vergleichbarkeit der Kontrollgruppe
- vorzeitige Ausscheider

Kohortenstudien

STATISTISCHE ANALYSE

- **DESKRIPTIV**

- relative Häufigkeiten → Inzidenzschätzung
- relatives Risiko (RR)
- Zurechenbares Risiko

- **TESTS**

- Chi²-Test
- Logrank-Test
- Logistische Regression
- Cox Modell

Kohortenstudien

BEKANNTE BEISPIELE

- Framingham Studie
- Nurses Health Study
- European Prospective Investigation into Cancer and Nutrition (EPIC)
([**www.iarc.fr/epic**](http://www.iarc.fr/epic))

Evaluation Criteria for Cohort Studies

1. Randomization in exposure; this will assure that exposed and not-exposed subjects have equal susceptibility for developing the ADR relative to factors other than drug exposure, However, it is not always feasible to randomize the exposure.
2. The exposed and not-exposed cohort should be comparable with respect to demographics and other background characteristics that may affect the likelihood of developing the ADR. This is assured by randomization when it is feasible.
3. The exposed and not-exposed cohort should be comparable with respect to clinical characteristics (pharmacogenetics) which may affect the likelihood of developing the ADR.
4. Drug exposure including compliance should be ascertained equally in both groups.

5. Medical surveillance for the ADR should be equal in the exposed and non-exposed cohorts.
6. The same diagnostic criteria and examination should be used in the exposed and non-exposed cohorts.
7. Dropout rates and characteristics of dropouts in the exposed and non-exposed cohorts should be similar.
8. The cohort should be representative of the population that normally uses the drug.
9. The cohort should have the characteristics of an inception cohort, the subjects are followed from the beginning of the drug exposure.

Checkliste zur Beurteilung von Kohortenstudien

Sind alle Teilnehmer

- am Anfang der Beobachtungszeit in die Kohorte aufgenommen worden ?
- am Anfang unter Risiko hinsichtlich der Entwicklung der Zielgröße(n) gewesen? (d.h. z.B. nachweislich frei auch von Frühstadien) ?
- ggf. zum gleichen Zeitpunkt im Krankheitsverlauf in die Kohorte aufgenommen worden (z.B. Prognosestudien) ?

Checkliste zur Beurteilung von Kohortenstudien

Sind alle Teilnehmer

- vollständig beobachtet worden
(➔ lost to follow-up) ?
- Wurde Beobachtungsgleichheit hergestellt ?
- Waren die Kohorten (d.h. deren prognostisches Profil) vergleichbar ?

Beobachtungsstudien vs. Experimentelle Studien

BEOBACHTUNGSSTUDIEN

- Hypothesen generieren
- keine Kausalschlüsse
- Identifizierung von Risikofaktoren und deren Gewicht

Beobachtungsstudien vs. Experimentelle Studien

RANDOMISIERTE STUDIEN

- Hypothesen testen
- Validierung von Risikofaktoren
- Wirksamkeitsnachweis /
Kausalschlüsse

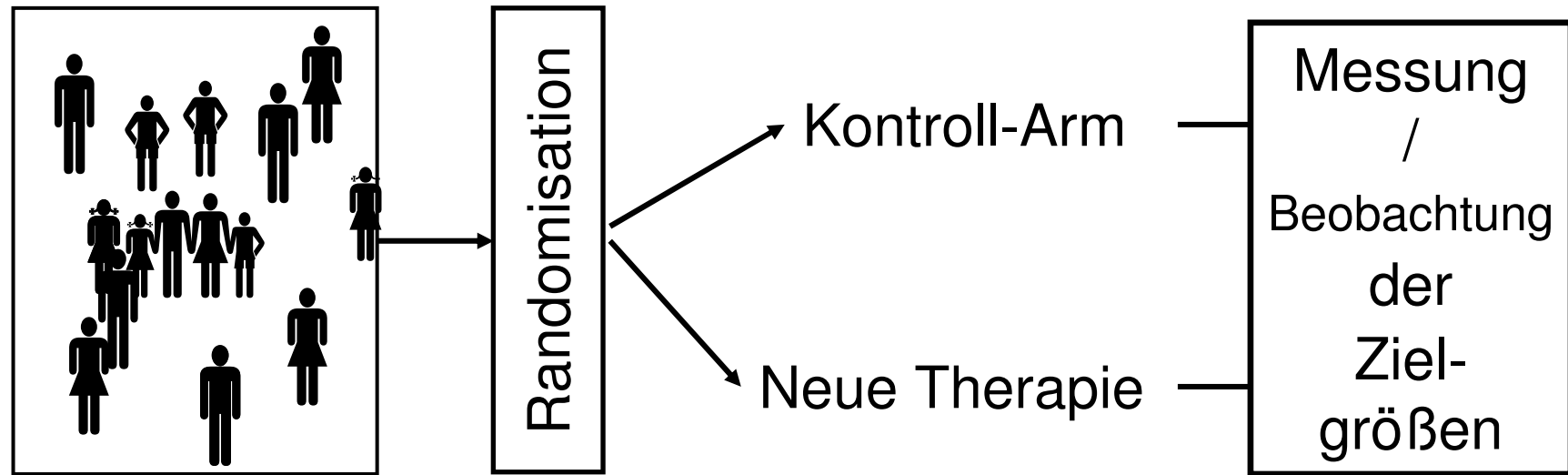
Causality Assessment in Observational Studies

No randomization → no proof of causality

A Relative Risk or Odds Ratio $\neq 1$ just indicates a potential association

- Exclude and / or assess bias
- Assess Causality using Bradford-Hill criteria
 - temporal sequence: cause → effect
 - strength of the association
 - dose-response relationship
 - reversibility
 - consistency
 - biological plausibility
 - analogy

Modell einer Randomisierten Studie



Definierte Zielpopulation

Einschlusskriterien

Ausschlusskriterien

I n t e r v e n t i o n

Auswertung

IBE
J. Hasford
München

Parallelgruppendesign

VORGEHENSWEISE

- Die bezüglich der Zielpopulation definierten Patienten werden
 - per Randomisation - in 2 oder mehr Behandlungsgruppen eingeteilt.
- Die Zielgrößen werden nach einer vorab definierten Zeit beobachtet.

Parallelgruppendesign

Behandlung	Zielvariable	
	Messung	
Kontroll-Arm	a	
Neue Therapie	c	

$$\text{Therapieeffekt} = c - a$$

Parallelgruppendesign

EIGENSCHAFTEN

- immer anwendbar, unabhängig von der Fragestellung (Efficacy, Safety, Nahrungsmittel, Sozialarbeit) und von der Skalierung des Zielkriteriums
- das am häufigsten eingesetzte Design klinischer Studien
- leicht interpretierbar
- einsetzbar in allen Phasen der klinischen Prüfung; Design der Wahl für Phase II und besonders für Phase III
- Auswertung zweier oder mehrerer unverbundener Stichproben
- Unterschieds- und Äquivalenzfragestellungen prüfbar

Vorteile der randomisierten Studie

- Strukturgleichheit dank Randomisation
- prospektive Schlussweise ($U \rightarrow W$)
- Beobachtungsgleichheit
- Experiment, Replizierbarkeit
- hohe Akzeptanz der Ergebnisse

Voraussetzungen für einen aussagekräftigen statistischen Vergleich

- ✓ Strukturgleichheit
- ✓ Beobachtungsgleichheit
- ✓ Interventionsgleichheit

Validität - Definitionen

Interne Validität

Die Ergebnisse treffen für die Patienten in der Studie tatsächlich zu.

Externe Validität

Die Ergebnisse der Studie können generalisiert werden, d.h. sie gelten auch für die Routineversorgung.

Kriterien für die Bewertung Randomisierter Studien

INTERNE VALIDITÄT

Planung

- Eindeutige a priori Fragestellung und Hypothesen
- Zentrale Randomisierung / Strukturgleichheit
- Beobachtungsgleichheit / Verblindung
- Genau beschriebene Therapien
- Fallzahlschätzung

Durchführung

- Qualitätskontrolle / -sicherung, GCP-Konformität
- Protokoll- und Patientencompliance
- Wenige lost to follow-up-Patienten

Auswertung

- ITT für Wirksamkeit, as treated für Verträglichkeit / UAW
- Analyse der Strukturgleichheit
- Beschreibung der vorzeitigen Ausscheider
- Adäquate statistische Verfahren
- Konfidenzintervalle
- Explorative von konfirmatorischen Analysen klar getrennt

Diskussion

- Bias, Protokollabweichungen
- Interpretation datennah

Kriterien für die Bewertung Randomisierter Studien

EXTERNE VALIDITÄT

- detaillierte Ein- und Ausschlusskriterien
- Repräsentativität von Patienten und studiendurchführenden Zentren
- Beschreibung der tatsächlich applizierten Therapie
- Dosierungen der Therapien adäquat
- Vergleichsgruppe adäquat behandelt
- Surrogatzielgröße oder Patienten relevante Zielgröße (Clinical Endpoint)?
- Intent to treat-Auswertung der Wirksamkeit?

Meta-Analyse: Definition

Die statistische Analyse der Ergebnisse verschiedener vergleichbarer Studien mit dem Ziel, die gesamte verfügbare Information zu einer validen Aussage über die Wirksamkeit einer Intervention methodisch-statistisch transparent und korrekt zusammenzufassen.

Basis:

- individuelle Patientendaten oder
- aggregierte Kennzahlen
(z.B. p-Werte, Heilungsraten etc.)
- statistische Weiterentwicklung des Review
- andere Ausdrücke: Pooling, Overview

Meta-Analyse: Ziele

- präzisere Schätzung der Effektgrößen
- größere Power der M.-A.:
kleinere Unterschiede sicher erkennbar
- Klärung widersprüchlicher Studienergebnisse
- bessere Bewertung der Verallgemeinbarkeit
- Analyse von a priori festgelegten Subgruppen
und neuer Fragestellungen

- empirische Basis für die Planung neuer Studien
- Evidence-based Medicine
- objektive, systematische und strukturierte Evaluation des State of the Art

Meta-Analyse: Arten

- Analyse von randomisierten Studien
- Analysen von Beobachtungs- und epidemiologischen Studien
- Analyse von Rohdaten
- Analyse von aggregierten Daten

Wissenschaftliche Ansprüche erfordern die Analyse von Rohdaten randomisierter Studien, da man nur dann alle interessierenden Effektmaße einheitlich berechnen kann.

Meta-Analyse: Vorgehensweise

- Festlegung der Fragestellung - Analyseplan
- Identifizierung **a l l e r** Studien 
- Zugang zu Prüfplänen und Rohdaten 
- Bewertung der Vergleichbarkeit (z.B. Design, EAK, Intervention, Zielgrößen) und Qualität 
- Analyse der einzelnen Studien - Rücksprache mit Studienleitern
- Meta-Analyse aller eingeschlossenen Studien
- Diskussion und Kommunikation

Auswahl der Studien

- sinnvoll: a-priori-Auswahl- bzw. Qualitätskriterien, anhand derer Studien in die Analyse einbezogen werden (z.B. Qualitäts-Score von Chalmers et al., Controlled Clinical Trials (1981))
- Idee für objektive Auswahl:
 - Beurteilung jeder Studie durch zwei unabhängige Personen
 - Verblindung bezüglich bestimmter Items (wie Autoren, Institution, Journal, Ergebnisse)

Auswahl der Studien

- **Einbeziehen nicht-publizierter Studien?**
 - Pro: auch nicht-signifikante Ergebnisse, größere Repräsentativität
 - Contra: sind u.U. von schlechterer Qualität (kein Peer-Reviewing)
- **Problem:**

Woher bekommt man Info zu nicht-publizierten Studien?

Störfaktor Veröffentlichung - Publication Bias

- Studien mit negativen Ergebnissen werden seltener und später publiziert als positive.
- Studien mit positiven Ergebnissen werden schneller und eher mehrfach publiziert.

Folge: Überschätzung des Effekts

„Negative Studie“

- H_0 beibehalten trotz adäquater Power, z.B. $\beta = 0.1$
- H_0 beibehalten bei zu kleinem Stichprobenumfang, z.B. $\beta = 0.4$

Störfaktor Veröffentlichung - Publication Bias

Beispiel: J.M. Stern, R.J. Simes:
BMJ 315:640-5 (1997)

Stichprobe: 197 RKS zwischen 1979 und 1988
der EK vorgelegt

- Publikationsrate 3,3 x höher für RKS mit
signifikantem Ergebnis
- 5-Jahres-Publikationsrate: 54% vs. 23%
- Mediane Zeit bis zur Publikation:
4,7 vs. 8 Jahre

Störfaktor Veröffentlichung - Publication Bias

- Meta-Analysen dürfen nicht nur publizierte Studien umfassen.
- Meta-Analysen sollten die Zeit bis zur Publikation berücksichtigen und eine entsprechende Deadline bezüglich des Starts aufzunehmender Studien vorsehen.
- Alle RKS sollten fortlaufend registriert werden
→ Cochrane Collaboration

Störfaktor: Mehrfach-Veröffentlichung

Beispiel: M.R. Tramer et al.:
BMJ 315:635 (1997)

Stichprobe: Alle zwischen 1991-1996
veröffentlichten RKS zu Ondansetron
bei postoperativer Emesis
84 Studien mit 11.980 Patienten

Ergebnisse:

- 14 / 84 Studien (17%) wurden mehrfach mit z.T. unterschiedlichen Autoren publiziert.
- 3335 / 11980 Patienten (28%) wurden mehrfach publiziert.
- Studien mit großen Wirksamkeitsunterschieden wurden deutlich häufiger mehrfach publiziert.
- Die Wirksamkeit von Ondansetron wird um 23% überschätzt.
- Number to treat:
4,9 Patienten (alle Studien)
vs. 9,5 Patienten (nur 1x publizierte Studien)

Kombinierbarkeit von Studien

- 'the apples-and-oranges problem'
- abhängig von Unterschieden
 - in den Behandlungen
 - in den Populationen (Ein- / Ausschlusskriterien)
 - in den klinischen Endpunkten
(„harte“ Endpunkte bevorzugen)
 - in der Definition von Krankheit / Exposition
 - im Studiendesign
 - in der Studienqualität

Meta-Analyse: Wesentliche Zielgrößen

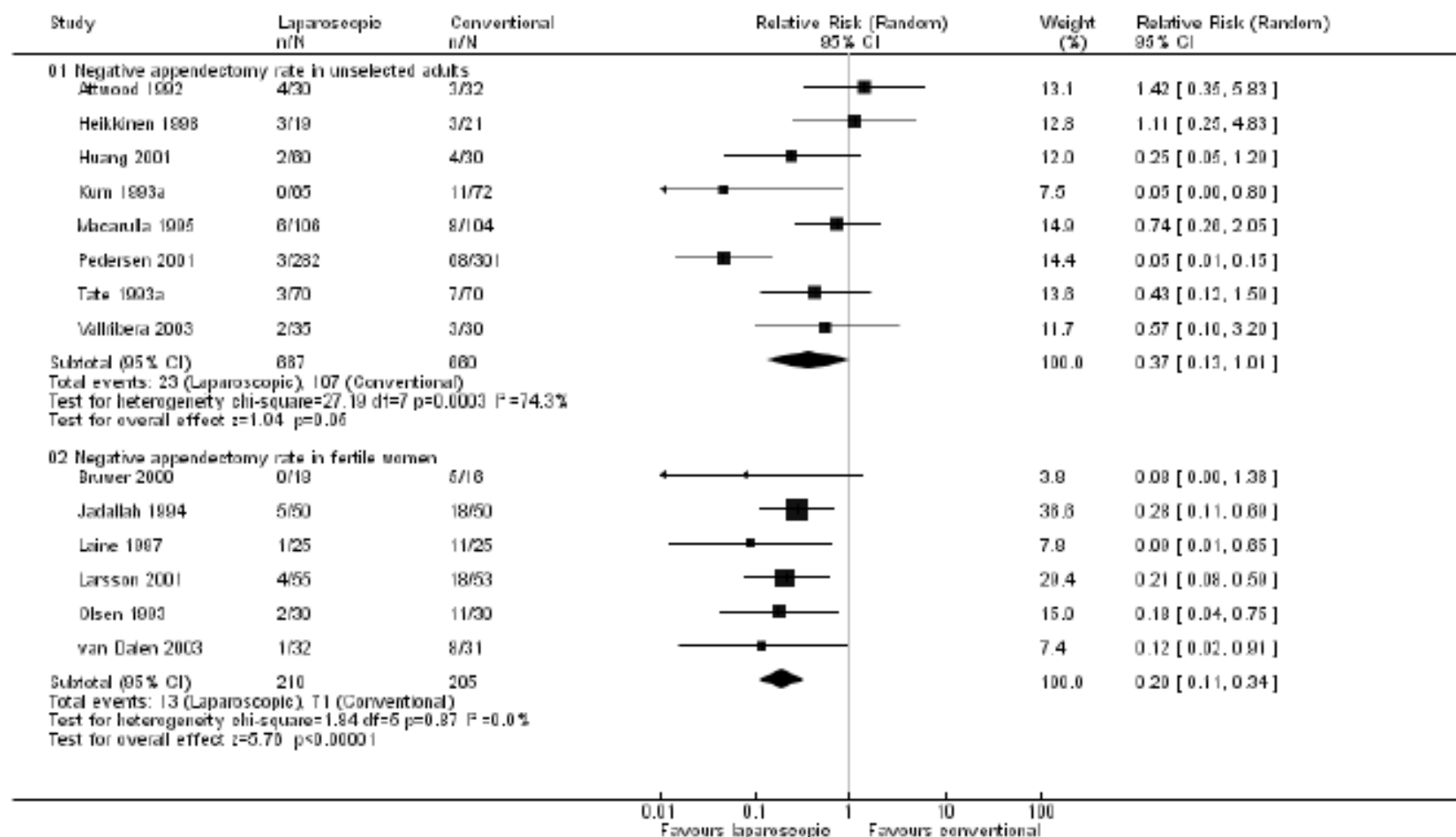
- Risikodifferenz
- Relatives Risiko
- Odds Ratio

Analysis 03.01. Comparison 03 Diagnostic laparoscopy (and open or laparoscopic appendectomy if necessary) versus immediate open appendectomy, Outcome 01 Negative appendectomy (rate) Show Statistical Analysis

Review: Laparoscopic versus open surgery for suspected appendicitis

Comparison: 03 Diagnostic laparoscopy (and open or laparoscopic appendectomy if necessary) versus immediate open appendectomy

Outcome: 01 Negative appendectomy (rate)



Meta-Analyse: Probleme

- Identifizierung aller Studien
- Publication Bias
- Citation / Language Bias
- Auswahl der Studien
- Kombinierbarkeit - Äpfel-Birnen
- Viele „Studien ohne Power“
- Komedikationen

Beurteilungskriterien für Meta-Analysen

- Wie umfassend waren die Bemühungen, alle Studien zu identifizieren ?
- Sind die Ein- und Ausschlusskriterien der Meta-Analyse und der Studien adäquat und detailliert beschrieben ?
- Sind die Studien und die Fallzahlen angegeben ?
- Sind die Zielgrößen adäquat und zwischen den Studien vergleichbar ?

- Wie ist die Qualität der Studien ?
- Variieren die Ergebnisse der einzelnen Studien stark ?
- Wurde die Stabilität der Ergebnisse durch Sensitivitätsanalysen geprüft ?
- Waren Kliniker involviert ?

Schlussfolgerungen

1. Meta-Analysen sind eine wertvolle Ergänzung des Methodenrepertoires zur Präzisierung von Effektgrößen und prognostischen Gruppen.
2. Meta-Analysen sind methodisch anspruchsvoll und zeitaufwändig.
3. Meta-Analysen sollten **a l l e** randomisierten Studien einbeziehen. Harte Zielgrößen sind von Vorteil.

4. Meta-Analysen von Studien mit ungenügender Power sind problematisch.
5. Meta-Analysen ersetzen keine große RS mit ausreichender Power.
6. Meta-Analysen, vor allem mit unpublizierten Studien, sind schwer zu bewerten.
7. Meta-Analysen erfordern eine enge Zusammenarbeit von Methodikern und Ärzten.