Systematische Verzerrungen und ihre Vermeidung in klinischen Studien

Dr. Dirk Hasenclever

IMISE, Leipzig

Gefahren für die Glaubwürdigkeit klinischer Studien

- **Schlechte Datenqualität**: Falsche, gefälschte, verfälschte, gezielt unvollständige Daten
- kontrolliert durch **Professionelle Qualitätssicherung, Datenmanagement und -monitoring**

Systematische Verzerrungen kontrolliert durch **Biasdiagnostik & Versuchsplanung**	**Zufallsbefunde** kontrolliert durch **Statistik**

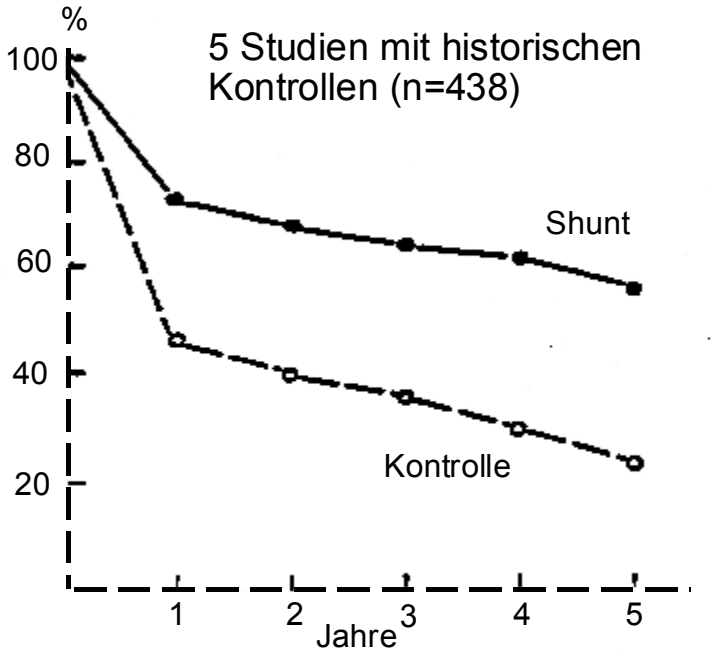
Ergebnisse von 20 unkontrollierten Studien
5 FU bei Kolonkarzinom

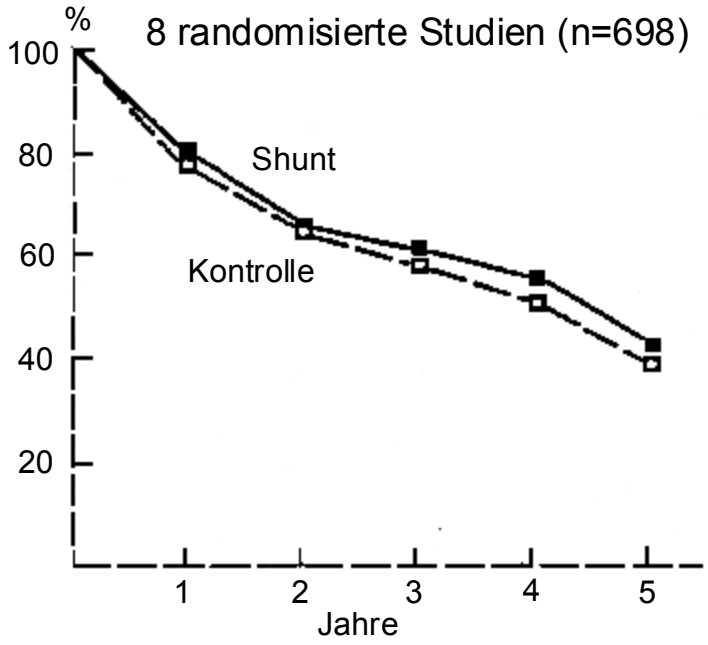
Anzahl der Patienten in der Studie	Patienten mit Tumor-regression (in %)
13	85
19	63
47	55
17	47
13	46
12	42
37	41
22	36
37	35
12	33
150	31
48	27
183	23
30	20
141	17
144	15
87	11
22	9
11	9

(Moertler & Reitemeier, 1969)

Historische Kontrollen versus Randomisierte Kontrollen

Vergleiche der Überlebenszeit bei Leberzirrhose mit Ösophagus-Varizen mit und ohne Pfortader-Shunt- Operation (nach Sacks, 1982)





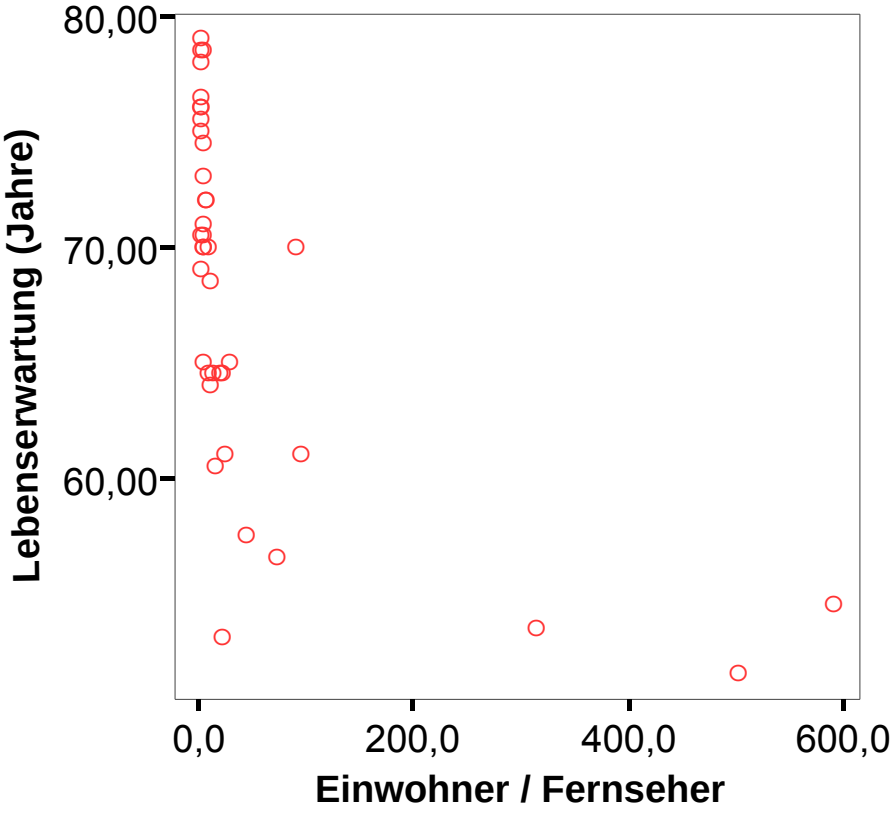
Probleme retrospektiver und/oder unkontrollierter Studien

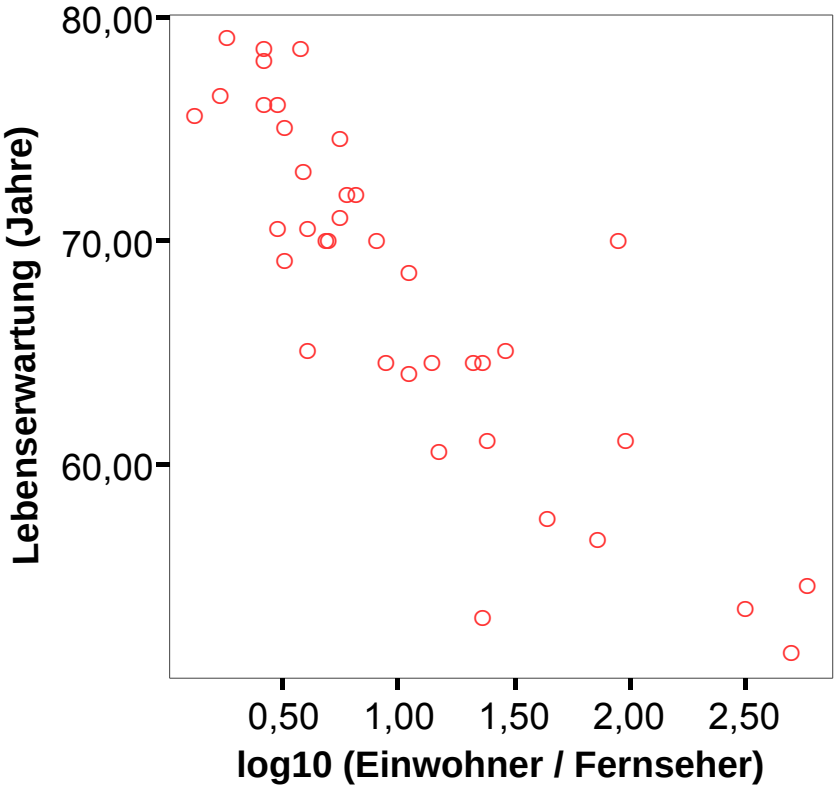
- **Datenqualität („Schweizer Käse“)**
  - **Dokumentation in Krankenakten unstrukturiert und notorisch lückenhaft**
- **Vergleichbarkeit der Patientengruppen („Äpfel und Birnen“)**
  - **Alter,Stadium, Vorgeschichte**
  - **....**
- **Unklare Selektion für gewählte Therapie („Daumen des Oberarztes“)**
  - **Arzt wählt Therapie für seinen Patienten aus**
- **Therapiedurchführung („alltäglicher Schlendrian“)**
  - **Unkontrollierte Protokoladhärenz**
  - **Supportive Maßnahmen**
- **Beobachtungsungleichkeit („schaun ‘mer ‘mal“)**
  - **Keine gleichartigen Bewertungskriterien bei „weiche“ Endpunkten**
  - **Subjektive Einschätzungen**
  - **Differente Endpunktdefinitionen**
- **Behandlung von Sonderfällen („Wie hätten Sie‘s denn gerne?“)**
  - **Studienabbrechern**
  - **Therapieversagern**
  - **“missing values”**

Fernsehen verlängert das Leben!

LAND	Lebenserwartung	Einwohner/TV	LAND	Lebenserwartung	Einwohner/TV
Argentinien	71	4	Morokko	65	21
Bangladesh	54	315	Burma	55	592
Brasilien	65	4	Pakistan	57	73
Kanada	77	2	Peru	65	14
China	70	8	Philippinen	65	9
Kolombien	71	6	Polen	73	4
Ägypten	61	15	Rumänien	72	6
Äthiopien	52	503	Russland	69	3
Frankreich	78	3	Süd-Afrika	64	11
Deutschland	76	3	Spanien	79	3
Indien	58	44	Sudan	53	23
Indonesien	61	24	Taiwan	75	3
Iran	65	23	Thailand	69	11
Italien	79	4	Türkei	70	5
Japan	79	2	Ukraine	71	3
Kenya	61	96	Großbritann	76	3
Nord-Korea	70	90	USA	76	1
Süd-Korea	70	5	Venezuela	75	6
Mexiko	72	7	Vietnam	65	29

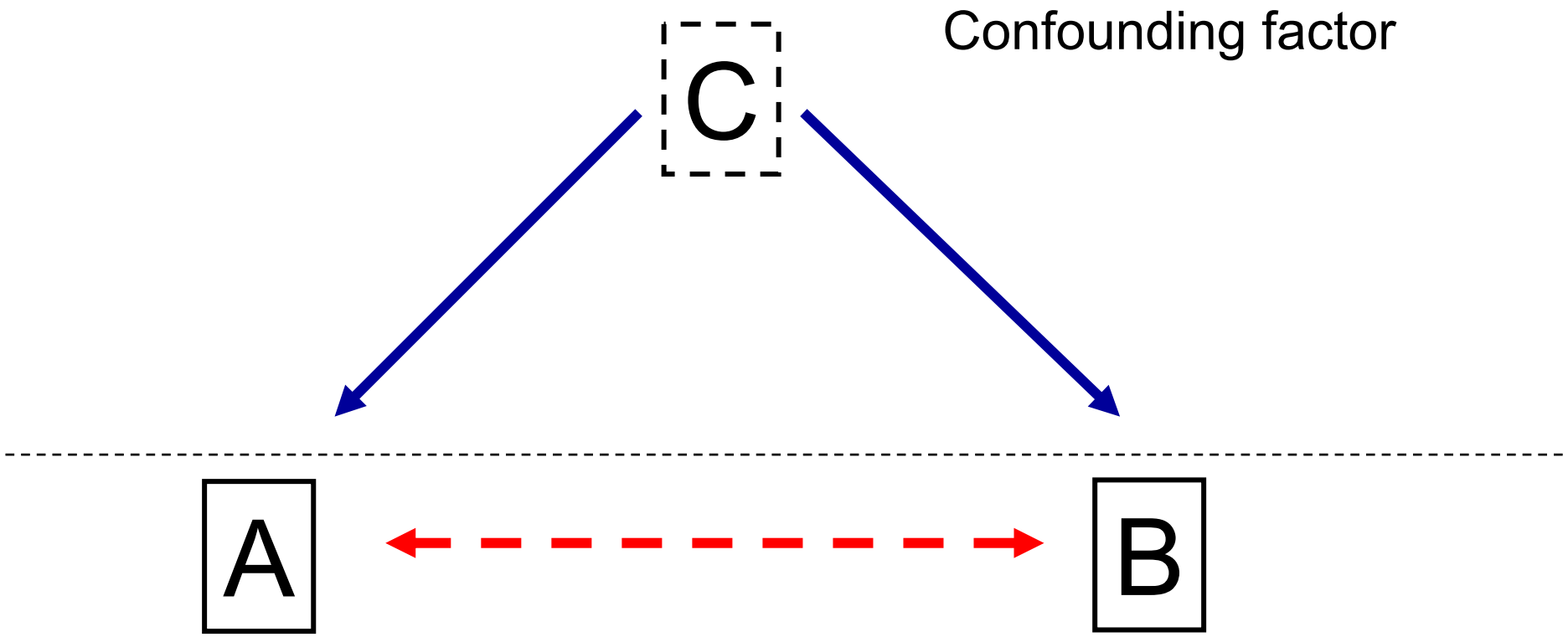
Fernsehen verlängert das Leben!(?)





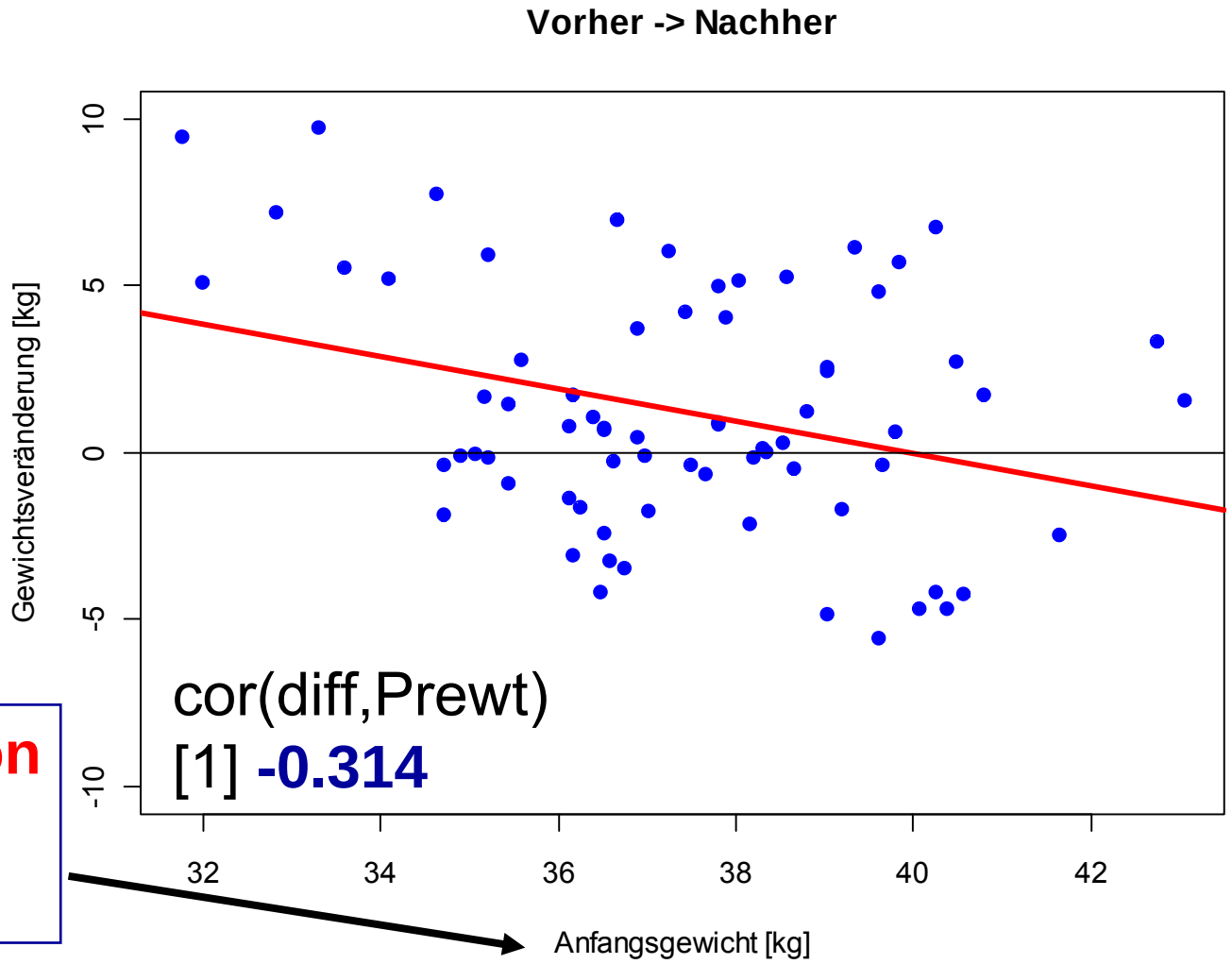
Confounding factor

Scheinkorrelation



Regression to the mean

Patientinnen
mit Magersucht



Qua Konstruktion abhängig
VORSICHT

Beispiel systematische Verzerrung
Simpson‘s Paradoxon

- Änderung der Assoziation zwischen zwei Merkmalen, wenn bzgl. eines dritten Merkmales stratifiziert wird

- Doktorarbeit: Vergleich zweier Therapien bei gastro-intestinalen Karzinomen

(Quelle: Guggenmoos-Holzmann I, Wernecke: Medizinische Statistik, 1996)

	mehr als 5 Jahre überlebt		
	ja	**nein**	
Therapie A	30 (30%)	70	100
Therapie B	80 (40%)	120	200
Total	110 (37%)	190	300

➡ Therapie B um 10% besser als Therapie A

Simpson‘s Paradoxon

Stratifikation nach Tumorlokalisation:

Magen

Therapie B um 5% schlechter als Therapie A

	mehr als 5 Jahre überlebt		
	ja	nein	
Therapie A	20 (25%)	60	80
Therapie B	8 (20%)	32	40
Total	28 (23%)	92	120

Kolon/Rektum

Therapie B um 5% schlechter als Therapie A

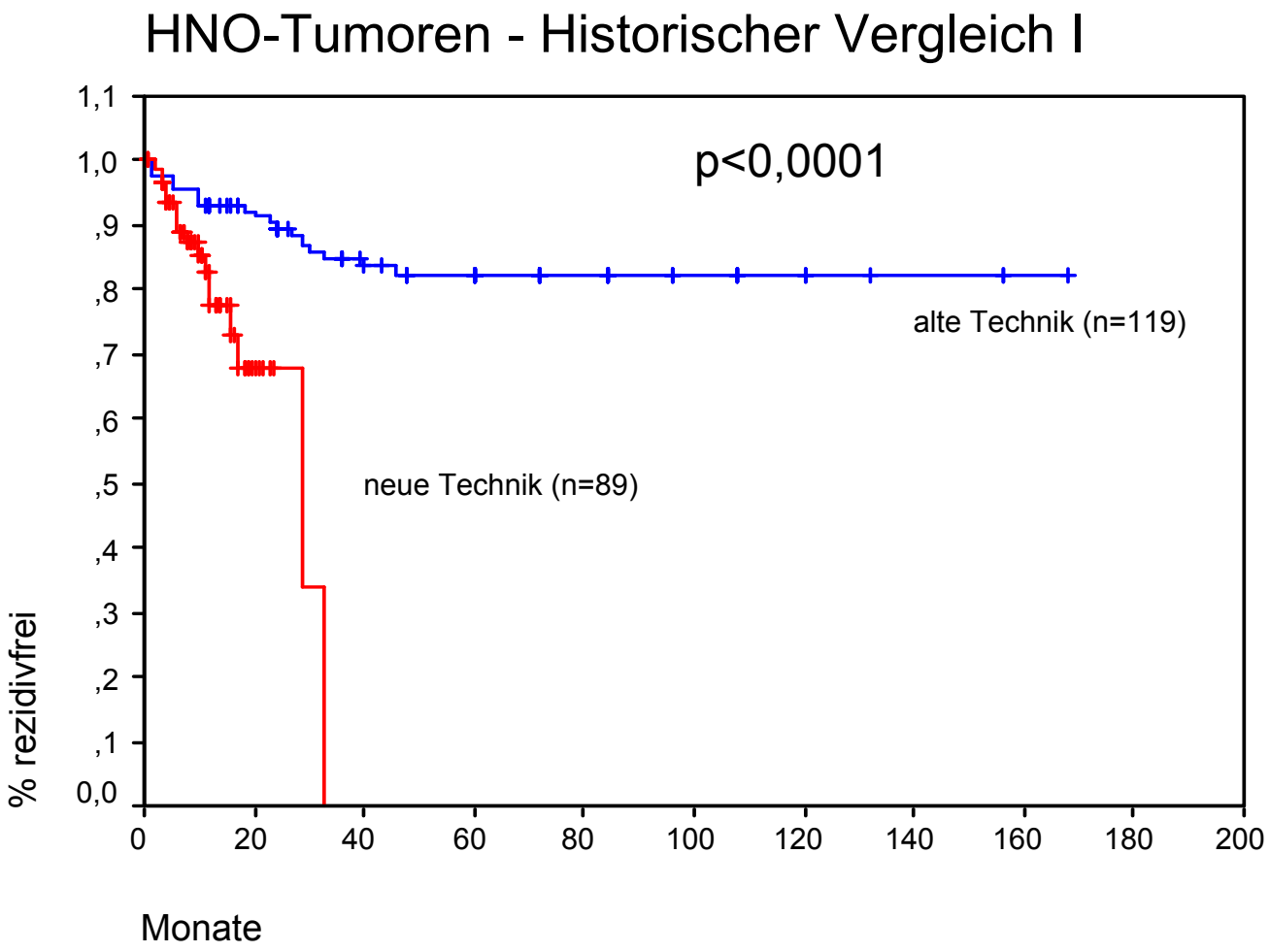
	mehr als 5 Jahre überlebt		
	ja	nein	
Therapie A	10 (50%)	10	20
Therapie B	72 (45%)	88	160
Total	82 (46%)	98	180

Simpson‘s Paradoxon

- **Heterogenität**
  - Magenkarzinome haben unabhängig von der Therapie eine deutlich schlechtere Prognose als Kolon/Rektum-Karzinome.

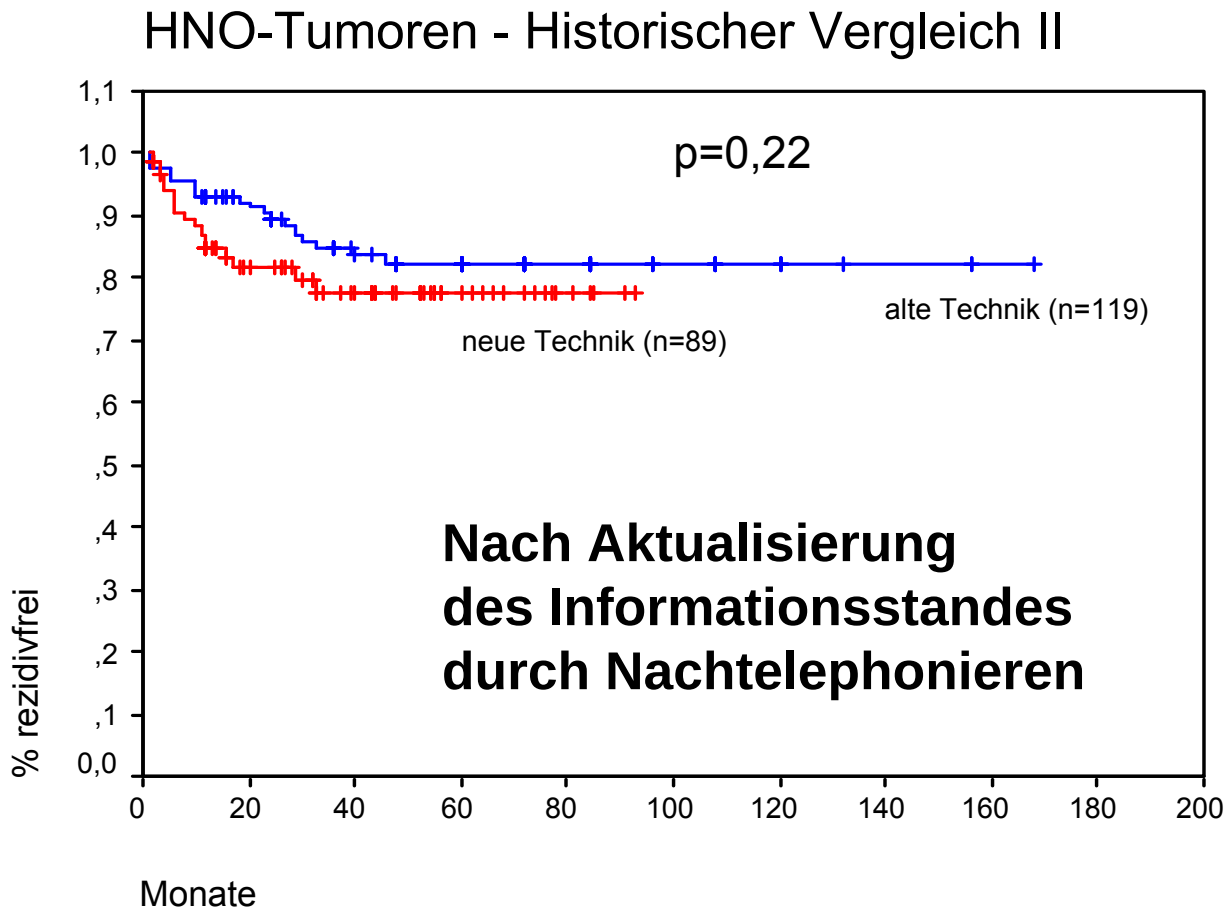
- **Imbalance in Therapiezuweisung**
  - Therapie B wurde häufiger in der prognostisch günstigeren Untergruppe eingesetzt als Therapie A.

„Muß der Klinikschef ins Gefängnis?“



Vergleich der neuen Bestrahlungs-Technik bei HNO-Tumoren mit der alten Technik bzgl. rezidivfreien Überlebens nach Krankenaktenlage

Informative Zensur – durch Zusammenhang zwischen Informationsfluss und Outcome



Verzerrungsquellen und Fehlschlüsse

- **Scheinkorrelation**: Dritte Größe (confounder) induziert eine Korrelation. Korrelation ≠ Kausalität
- **Selektion**
- **Informativer Drop-Out im Zeitverlauf**
- **Natürlicher Verlauf** (suggeriert Wirksamkeit einer Intervention wenn Kontrollgruppe fehlt.)
- **Auswertungsartefakte:** ("Korrelation" in Plot Differenz zum Ausgangswert gegen Ausgangswert)
- **Suggestivfragen**
- **Fehleranfällige Messverfahren**
- **…**

Studienplanung: Herstellen der Vergleichbarkeit!

Zeitpunkt	Heterogenitätsquellen	Ziel
bei Eintritt in die Studie	Patientcharakteristika (bekannte und latente Faktoren, die den Therapieerfolg beeinflussen)	Strukturgleichheit
während der Durchführung der Studie	Behandlung (Strategien bei Unverträglichkeit der Therapie, Zeitschema, supportive Maßnahmen, Pflege)	Beobachtungs- und Behandlungsgleich-heit
	Beobachtung (Zeitschema, Beurteilungskriterien)	
nach Beendigung der Studie	Auswertung (Zählweise nicht protokollgerecht behandelter Patienten)	Auswertungsgleich-heit

Randomisieren

Standardisieren
Qualitäts-kontrolle

Valide Mess-größen
Verblinden
Unabhängiger Review

Vorab-Analyseplan

Randomisation

Randomisation:

Zufällige Zuordnung der Patienten

zu den Therapiegruppen

- Bekannte und unbekannte Störgrößen werden dabei, gleichmäßig auf die Therapiearme verteilt,
  falls die Fallzahl genügend groß ist.
- Jeder Patient hat die gleiche Chance in einen der Therapiearme zu kommen.
- Randomisationsergebnis darf für den Arzt nicht vorhersagbar sein.
- Studienstandard: Zentrale Randomisierung im Studiensekretariat oder via Internet

Ursprung der Randomisationsidee

(nach I. Chalmers)

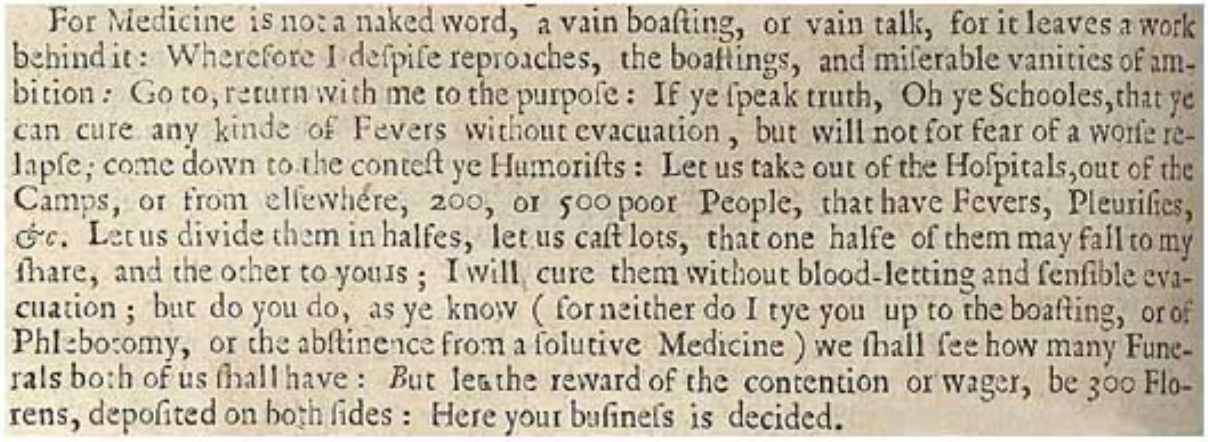
Van Helmant (Holland, 1662):

„Laß uns aus dem Krankenhaus Leute mit Fieber nehmen, laß uns Lose ziehen, so daß eine Hälfte mir zufällt und die andere Dir.

Ich werde sie ohne Aderlaß kurieren, aber Du verfährst wie Dir beliebt ...

Wir werden sehen, wie viele Begräbnisse wir beide haben werden."

van Helmont, JB. Oriatrike, or physick refined: the common errors therein refuted and the whole are reformed and rectified. Lodowick-Loyd: London, 1662, p.526



For Medicine is not a naked word, a vain boasting, or vain talk, for it leaves a work behind it: Wherefore I despise reproaches, the boastings, and miserable vanities of ambition: Go to, return with me to the purpose: If ye speak truth, Oh ye Schooles, that ye can cure any kinde of Fevers without evacuation, but will not for fear of a worse relapse; come down to the contest ye Humorists: Let us take out of the Hospitals, out of the Camps, or from elsewhere, 200, or 500 poor People, that have Fevers, Pleurisies, &c. Let us divide them in halfes, let us cast lots, that one halfe of them may fall to my share, and the other to yours; I will cure them without blood-letting and sensible evacuation; but do you do, as ye know (for neither do I tye you up to the boasting, or of Phlebotomy, or the abstinence from a solutive Medicine) we shall see how many Funerals both of us shall have: But let the reward of the contention or wager, be 300 Florens, deposited on both sides: Here your business is decided.

Voraussetzung der Randomisation

- Es gibt keine Beweise für die Unterlegenheit einer Behandlungsmodalität

  (Prinzip der vergleichbaren Unsicherheit)

- Die Behandlungen sind aufgrund ihrer Applikationsweise zumutbar

- Der Patient wird aufgeklärt und stimmt schriftlich der Verfahrensweise zu.

Techniken zur Randomisation

- **einfache Randomisation**

  Zufallszahlen:Beispiel: Zahlen 0-4: Therapie A
  Zahlen 5-9: Therapie B
  Nachteil: Imbalancen (ungleiche Fallzahl in den Armen) möglich

- **Blockrandomisation** (Random Permuted Block Design)
  Beispiel: Therapie A und B, Blöcke der Länge 4, mögliche Permutationen

AABB	ABAB	ABBA	BAAB	BABA	BBAA
1	2	3	4	5	6

  • Auswürfeln der Blockreihenfolge zur Erstellung einer Randomisationsliste:

BABA	ABBA	ABBA	AABB
5	3	3	1

Stratifizierte Randomisation

- Sind Störgrößen bekannt, werden entsprechend der Kategorien dieser Störgrößen **Schichten (Strata)** gebildet.

- Die Randomisation erfolgt dann innerhalb der Schichten.

- Adaptive Randomisationsalgorithmen

  Bei jeder Aufnahme eines Patienten wird die **aktuelle Verteilung bekannter Störgrößen auf Therapiearme** berücksichtigt (z.B. Minimierungsverfahren nach Pocock, 1975)

Grundidee der Minimierungsmethode

- Definiere ein Mass für Imbalance mit geeigneter Berücksichtigung der Stratifikationsfaktoren

- Bevorzuge Allokation, die globale Imbalance minimiert (deterministisch oder mit Zufallskomponente).

Implementation der Minimierungsmethode am KKSL ("Balanciere die Randverteilungen")

Zu randomisieren sei ein Patient mit Risikoprofil (m_1,..m_r).
Für jeden Randomisationsarm i (i = 1,...,t ; t ≥ 2) berechne einen Score:

$$score_i = \left[\left(\sum_{j=1}^{r} w_j \cdot n_{(S_j = m_j), i} \right) + w_0 \cdot n_i \right] / V_i$$

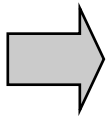
V_i	Randomisationsanteil von Arm i
S_j	Stratifikationsmerkmal j (j = 1,...,r)
r	Anzahl der Stratifikationsmerkmale
m_j	Wert von S_j für den zu randomisierenden Patienten
$n_{(S_j = m_j), i}$	aktuelle Anzahl der Patienten in Arm i, die im Stratifikationsmerkmal j den Wert m_j haben
w_j	Gewicht für Stratifikationsmerkmal j
w_0	Gewicht für Gesamtarmbilanz
n_i	aktuelle Anzahl der Patienten in Arm I

Der Arm mit dem kleinsten Score wird mit Wahrscheinlichkeit p genommen, mit Wahrscheinlichkeit 1-p wird unter den restlichen Armen „gewürfelt".
(Falls p=1 "deterministische Randomisierung")

Behandlungs- und Beobachtungsgleichheit

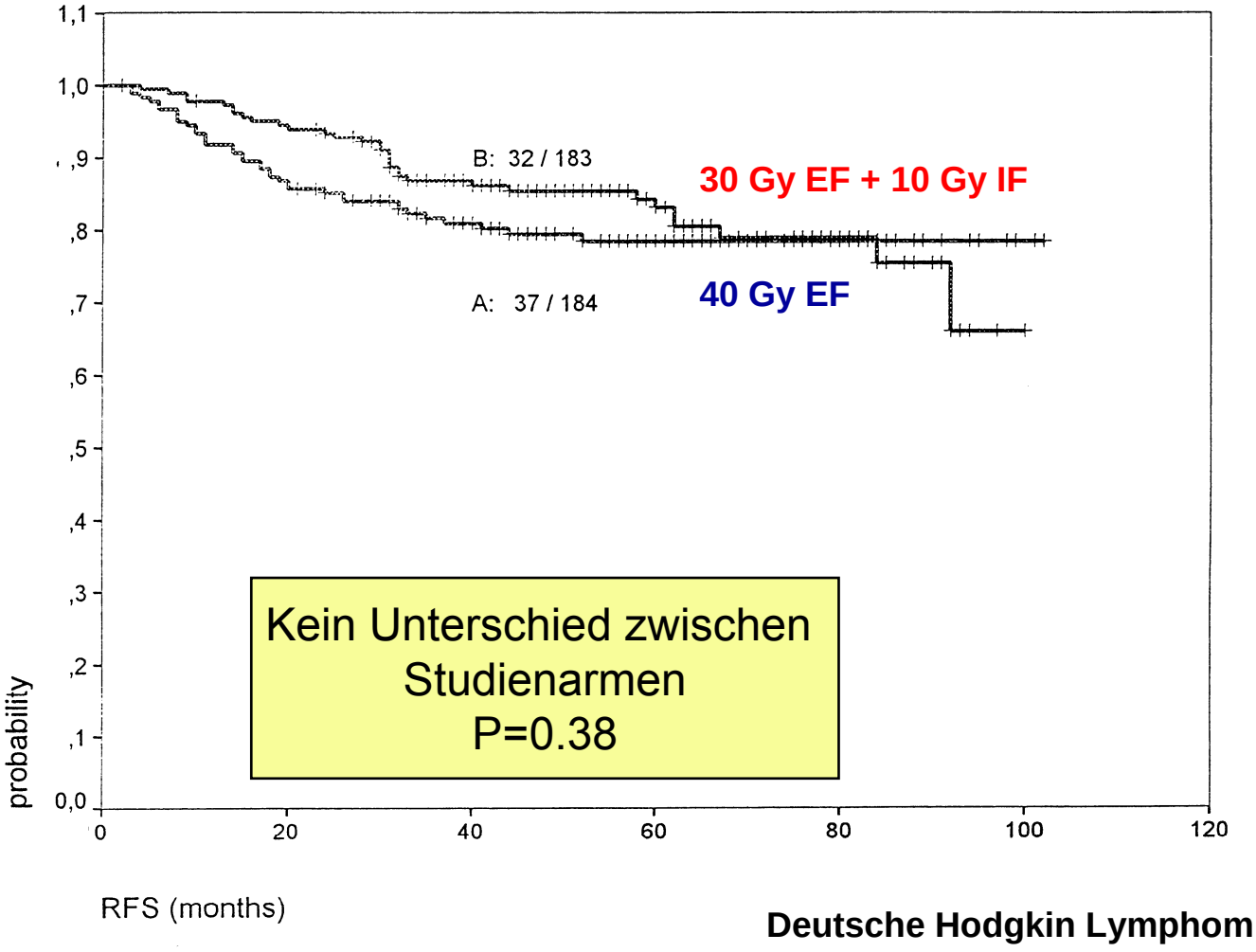
- Außer bezüglich der zu untersuchenden Therapien müssen Patienten in allen anderen Aspekten gleich behandelt oder beobachtet werden.

- Wie erreicht man **Behandlungs- und Beobachtungs gleichheit**?



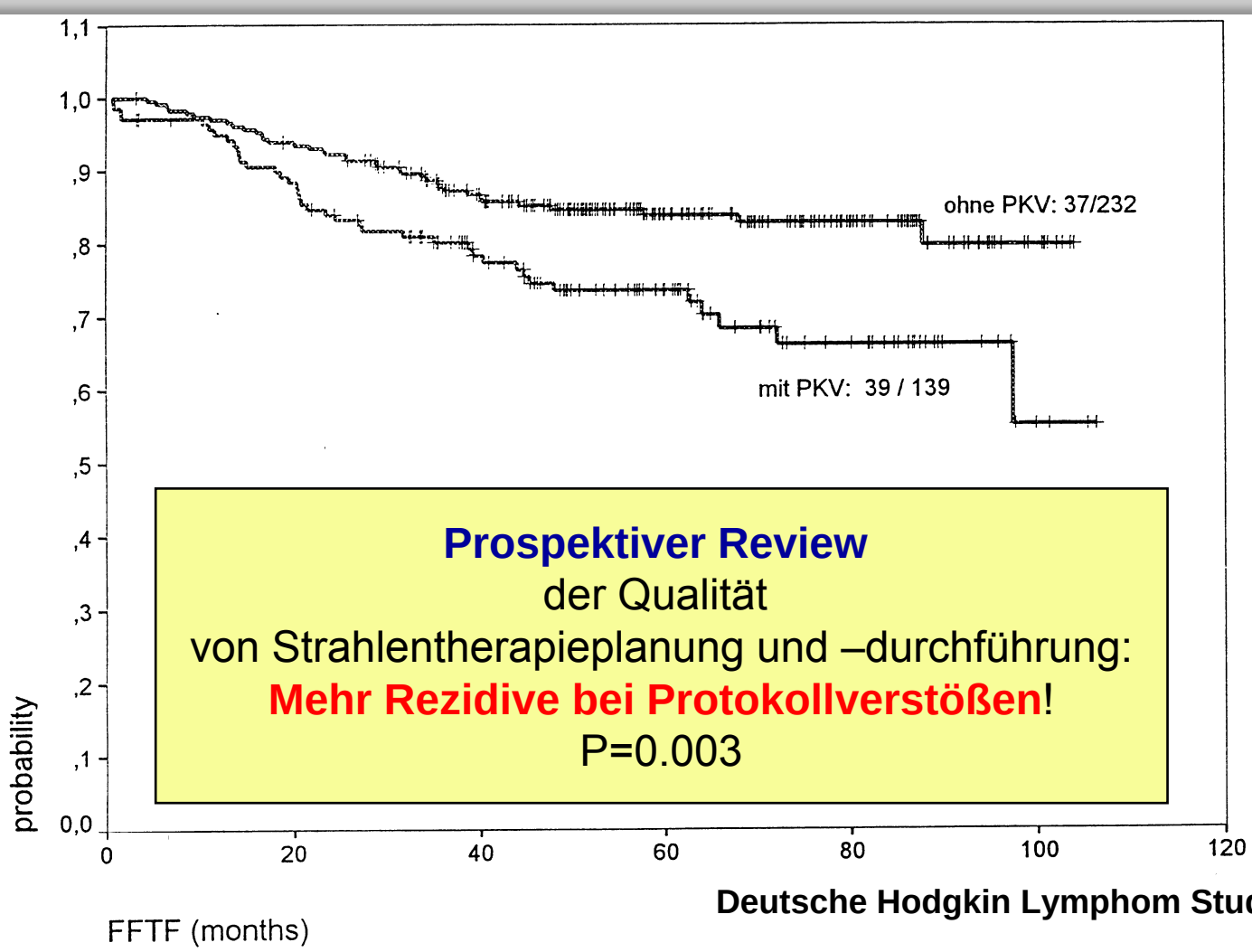
- Behandlung nach **standardisierten** Protokollen und strikte **Kontrolle der Protokolladhärenz**
- **Verblindung** und Einsatz von **Placebos**

Beispiel: Fehlende Protokolladhärenz
HD 4 RFS Armvergleich PS I, II ohne RF



Deutsche Hodgkin Lymphom Studiengruppe

Beispiel: Durchführungsprobleme
HD4 FFTF Protokollverstoß während Bestrahlung



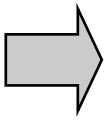
Prospektiver Review
der Qualität
von Strahlentherapieplanung und –durchführung:
Mehr Rezidive bei Protokollverstößen!
P=0.003

Deutsche Hodgkin Lymphom Studiengruppe

Verblindung bei subjektiven Beurteilungen

- Beurteilungen besonders von Endpunkten durch Patient und Arzt
  können **verzerrt** werden
  durch die **Kenntnis des Therapiearms** :

  - Beurteilung und Ausmessen von Bildgebung
  - Einschätzung Allgemeinzustand
  - Einschätzung von Schmerz
  - …



Verblindung von Patient und/oder Arzt
oder Einsatz **unabhängige Beurteiler**

Blindversuch

Versuch	Zuordnung der Patienten zu den Therapiearmen unbekannt für		
	Patient	behandelnden Arzt	Beurteiler/ Biometriker
einfach blind	✵		
doppel blind	✵	✵	
dreifach blind	✵	✵	✵

Offener Versuch: Die Zuordnung zu den Behandlungsgruppen ist allen bekannt.

Blindversuch

Blindversuche sind bei leichter Erkennbarkeit der angewandten Therapie

nicht möglich bzw. schwierig durchzuführen.

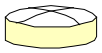
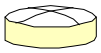
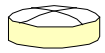
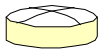
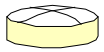
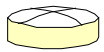
Zum Beispiel:

- unterschiedliche Nebenwirkung
- unterschiedliche Applikationsformen
- Vergleich eines chirurgischen Eingriffs mit einer medikamentösen Behandlung

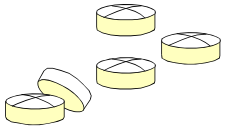
Blindversuch

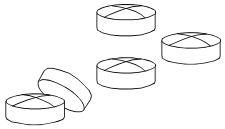
„double dummy“-Technik

Zum Beispiel: Bei unterschiedlichem Zeitplan der Applikation

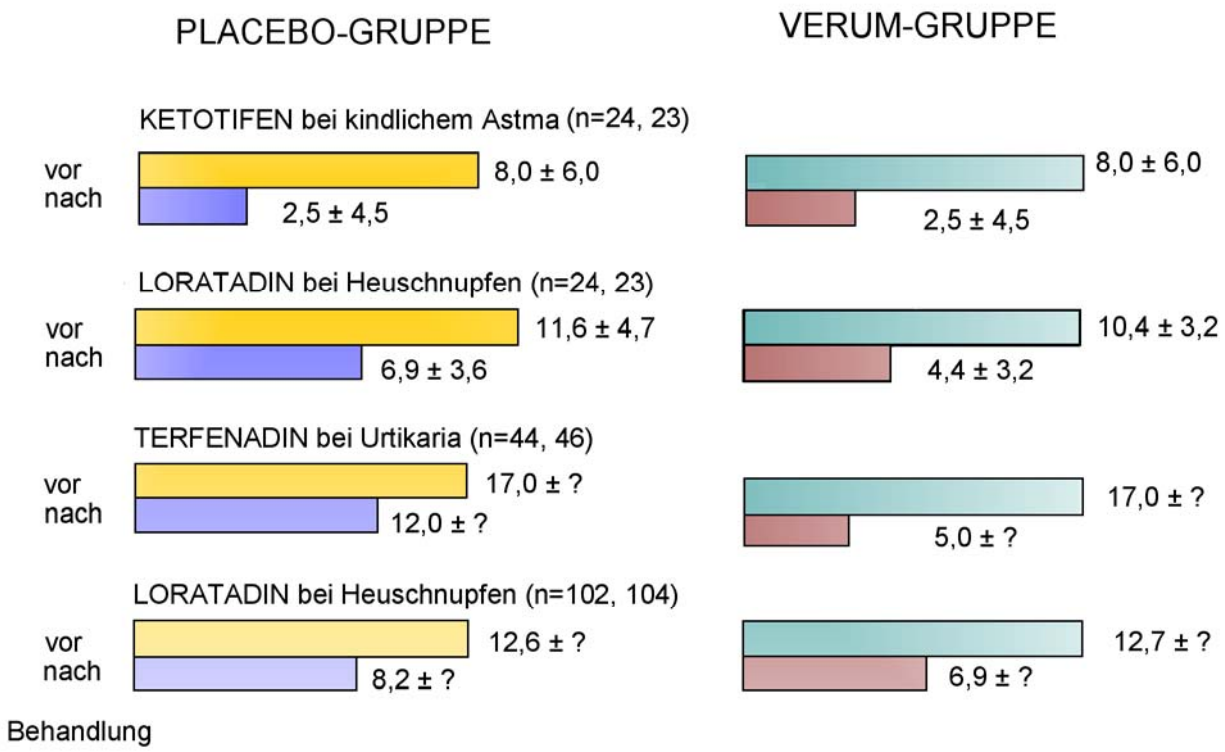
	Tag 1	Tag 2	Tag 3	Tag 4	Tag 5
Therapie A					
Therapie B					

Zum Beispiel: Bei unterschiedlichen Applikationsformen

Therapie A:  + 

Therapie B:  + 

Placebo-Effekte bei Allergiebehandlungen



Symptomscores vor und nach verschiedenen Behandlungen bei Allergien (nach Schäfer, 1991)

Nachweis therapeutischer Wirksamkeit erst gegeben, wenn die Therapie wirksamer als Placebobehandlung ist.

Auswertungsgleichheit
Intent to Treat Analyse

- Eine wichtige Technik zur Herstellung von Auswertungsgleichheit ist die Auswertung
  - **Strikt nach der Randomisationsergebnis**
  - d.h. nach der geplanten Therapie (engl.: **intent to treat**)
- ungeachtet dessen, welche Therapie die Patienten tatsächlich erhalten haben

Grund:

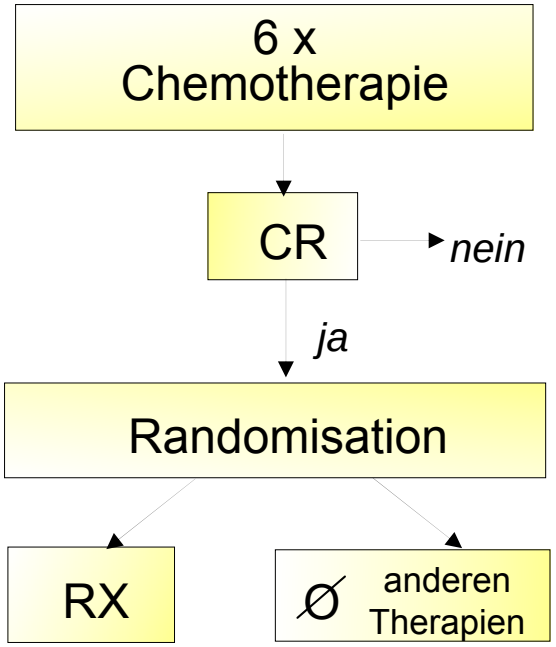
Wenn es gravierende Abweichungen zwischen der geplanten und der tatsächlich gegebenen Therapie gibt, sind Selektionen zu vermuten, welche die zum Randomisationszeitpunkt vorliegende Vergleichbarkeit zerstören können.

Intent to treat versus per protocol
Beispiel: SWOG-Studie zum NHL

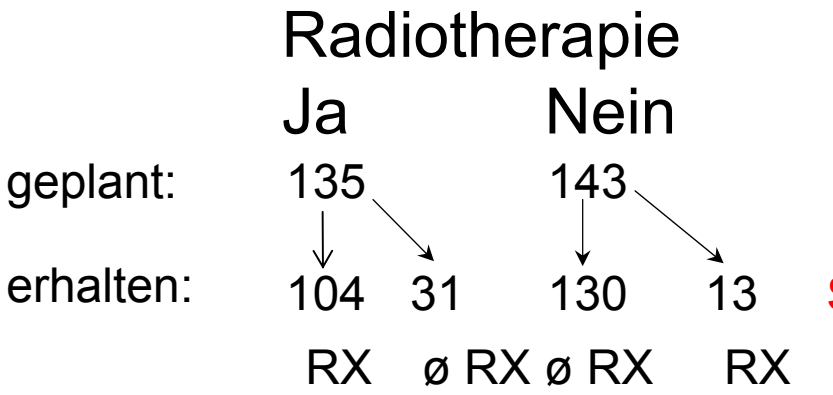
geplante Therapie *erhaltene Therapie*

SWOG-Studie zum NHL

(Ann. Intern. Med. 120, 1994, 903-912)



Intent to treat versus per protocol
Beispiel: SWOG-Studie zum NHL (2)



SELEKTION!

5-Jahres-Rezidivrate

ITT:	geplant:	21 %	32 %	**p=0.09**	**n.s.**
PP:	erhalten:	15 %	33 %	**p=0.002**	**!!**

Vorgehensweise bei der Analyse

■ Werden Patienten anders behandelt als geplant, wird die Aussagekraft der Studie geschwächt!

Zuerst immer

“intent-to-treat”-Analyse

- Auswertung nach Randomisationsarm
  - ➔ dabei werden Verzerrungen durch Selektionseffekt vermieden
  - ➔ unterschätzt dabei werden eventuell vorhandene Unterschiede

Deshalb Stabilitätsüberprüfung mit:

Per Protocol Analyse

- Auswertung nach Behandlungsarm
  - ➔ dabei wird der möglicherweise vorhandene Unterschied eher überschätzt (Selektionseffekt)

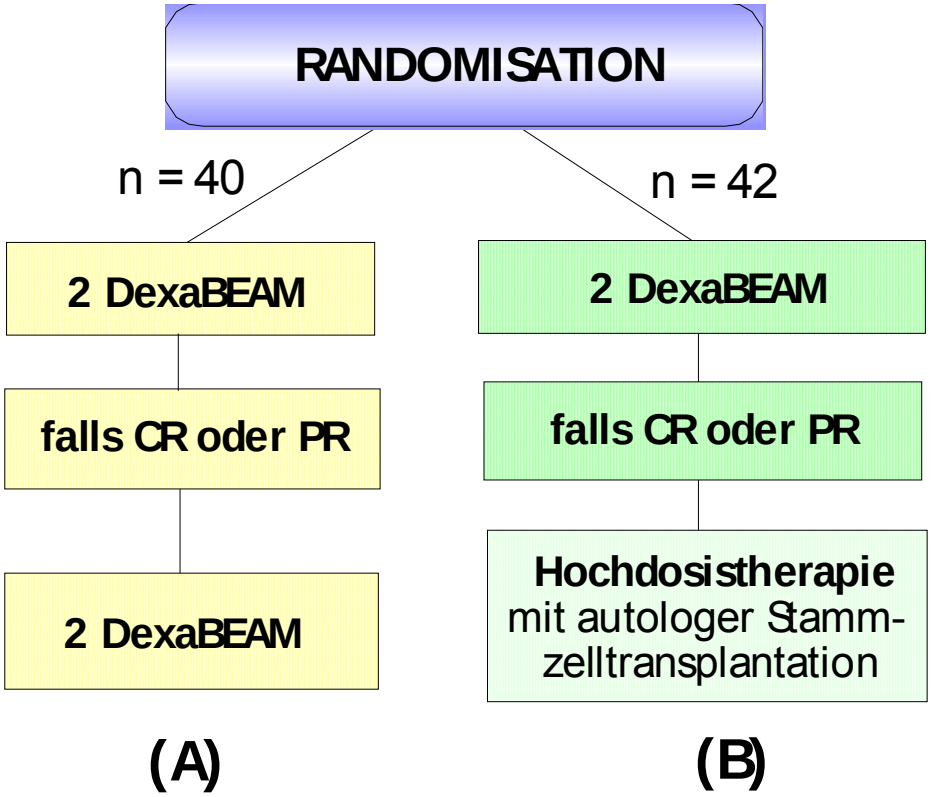
MERKE

■ Sind die Ergebnisse der “intent-to treat-Analyse” und der “per protocol Analyse” unterschiedlich, muss darüber unbedingt diskutiert werden!

Beispiel: HDR1-Studie

HDR1-Studie der Deutschen Hodgkin-Lymphom-Studiengruppe

Kollektiv: Patienten mit Rezidiv eines Hodgkin Lymphoms



Non PR nach 2 Zyklen:

Arm A: 8
Arm B: 20

Beispiel: HDR1 Studie

Mögliche Erklärungen:

- Zufallseffekt
  - mehr frühe Rezidive mit schlechtem Ansprechen in Arm B

- Verzögerung in Arm B durch Überweisung zum Transplantations-zentrum
  - erneutes Tumorwachstum in der Pause

  ➔ **BEHANDLUNGSGLEICHHEIT VERLETZT**

- Zweituntersuchung am Transplantationszentrum
  - schärfere Beurteilung des Ansprechens

  ➔ **BEOBACHTUNGSGLEICHHEIT VERLETZT**

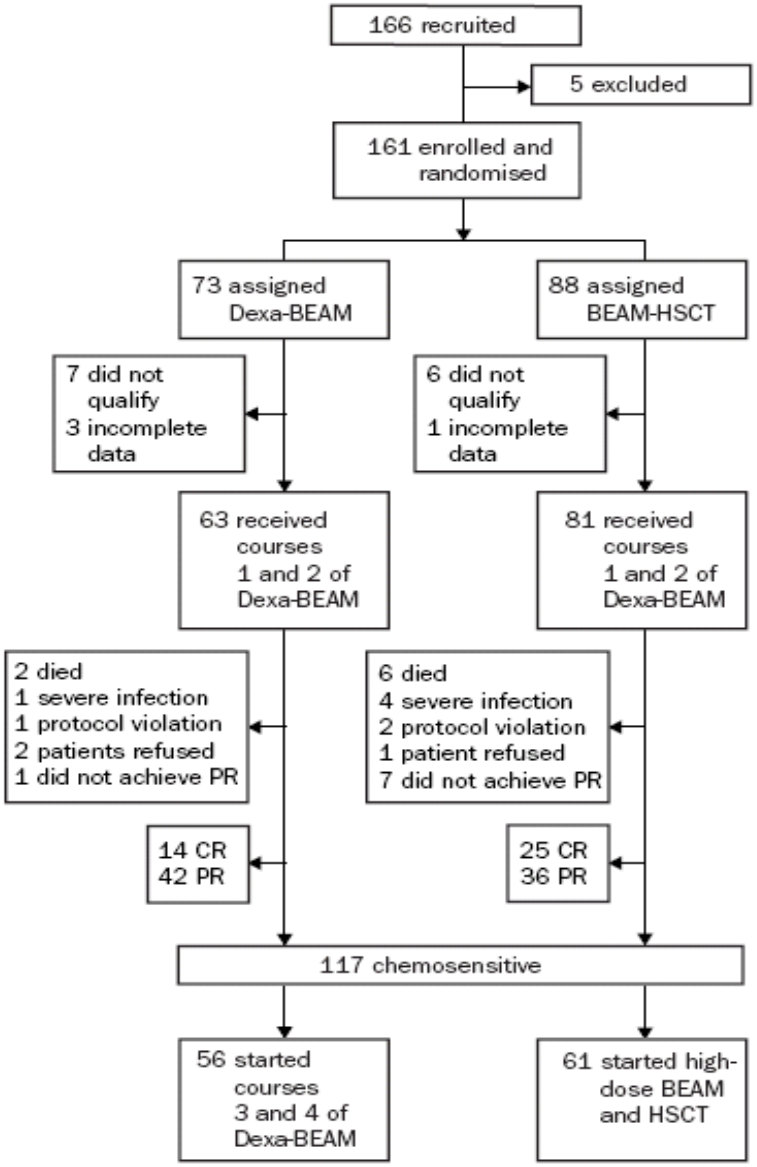
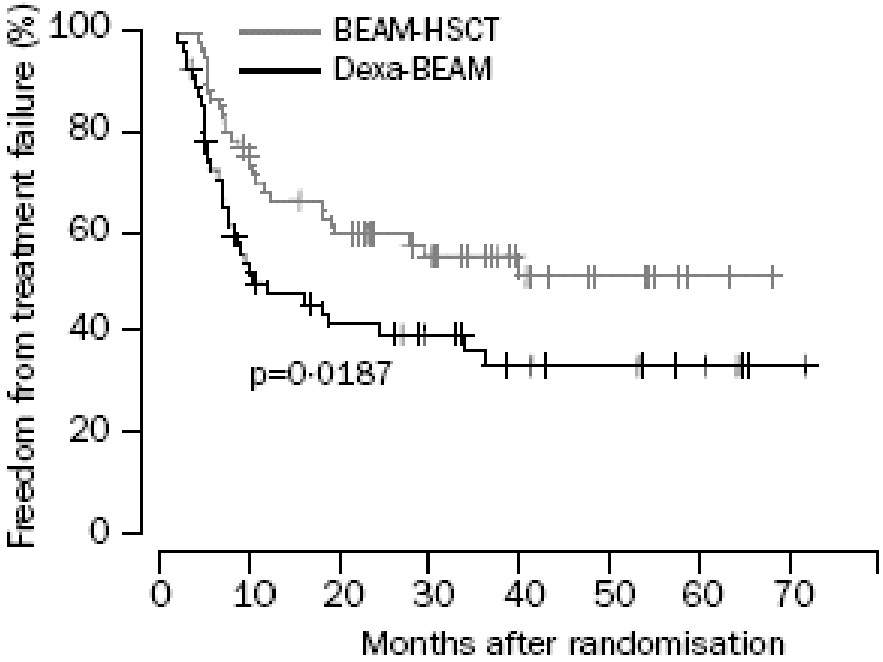


Figure 2: **Trial profile**
CR=complete remission; PR=partial remission.



Number of patients

BEAM-HSCT	61	43	34	25	13	8	7	0
Dexa-BEAM	56	27	20	15	10	8	5	1

Figure 3: **Freedom from treatment failure for patients with relapsed chemosensitive Hodgkin's disease**

Schmitz N et al 2002 Lancet

Bis zur nächsten Übung

- Auf der Vorlesungseite zum ‘runterladen: Bitte lesen!

  - Analyzing the Same Data in Two Ways: A Demonstration Model to Illustrate the Reporting and Misreporting of Clinical Trials