
Daten richtig interpretieren!

Verzerrungsquellen und ihre Vermeidung

Dr. Dirk Hasenclever

Institut für Medizinische Informatik, Statistik &
Epidemiologie (IMISE)

Zentrum für Klinische Studien Leipzig (ZKS-KKS)

dirk.hasenclever@imise.uni-leipzig.de

Numeracy

■ Seit Beginn der Bundesliga im Jahre 1963 gab es genau 50...

- o A) Eigentore
- o B) Verschiedene Deutsche Meister
- o C) Vereine in der Bundesliga
- o D) Trainerwechsel während der Saison

Fernsehquiz
Pilawa Oktober 2009

Lernziel Numeracy: Umgang mit Daten

- **Numeracy** is the ability to reason with numbers and other mathematical concepts.
- To be numerically literate, a person has to be comfortable with
 - logic and reasoning.
- Some of the areas that are involved in numeracy include:
 - basic numbers,
 - orders of magnitude,
 - geometry,
 - algebra,
 - probability and statistics.

Was ist Evidenz?

- **Deutsch**
- "Evident" =
"ohne weiteres einleuchtend",
"augenscheinlich", "klar",
"offensichtlich".
- "Evidens" (Latein) synonym mit
"perspicuus", = "durchsichtig".
- Eine Evidenz ist eine **unmittelbar
einleuchtende Erkenntnis, eine
Einsicht ohne methodische
Vermittlung.**
- **Englisch**
- "Evidence"
heißt nach dem Oxford
Dictionary of Current English
**"available facts, circumstances,
etc. indicating whether or not a
thing is true or valid"**.

Klinische Studien bilden Grundlage für Evidence-based Medicine

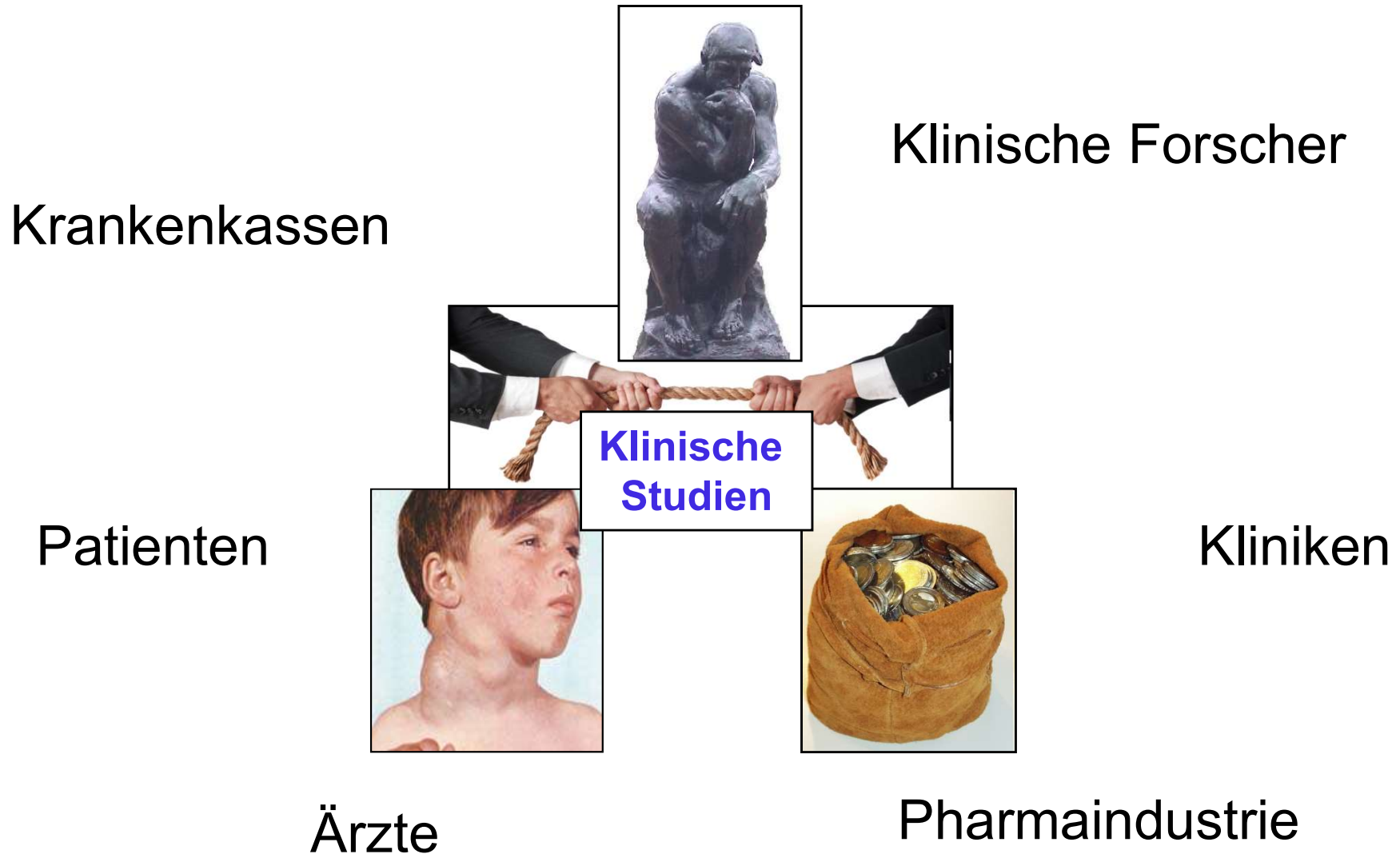
Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients.

The practice of evidence-based medicine means **integrating** individual clinical expertise with the best available external clinical evidence from systematic research.

By **individual clinical expertise** we mean the proficiency and judgement that individual clinicians acquire through clinical experience and clinical practice. ...

By **best available external clinical evidence** we mean clinically relevant research, often from the basic sciences of medicine, but especially from patient centred clinical research into the accuracy and precision of diagnostic tests ..., the power of prognostic markers, and the efficacy and safety of therapeutic, rehabilitative, and preventive regimens.

Medizinisch-wissenschaftlicher Fortschritt & Rationaler Ausgleich im Gesundheitssystem



Zielsetzungen

■ Grundverständnis für Biometrie & Statistik

- **Biometrie:**

- Analyse von Messprozessen
- Diagnostik systematischer Verzerrungsquellen
- Versuchsplanung zur Vermeidung von Verzerrungen

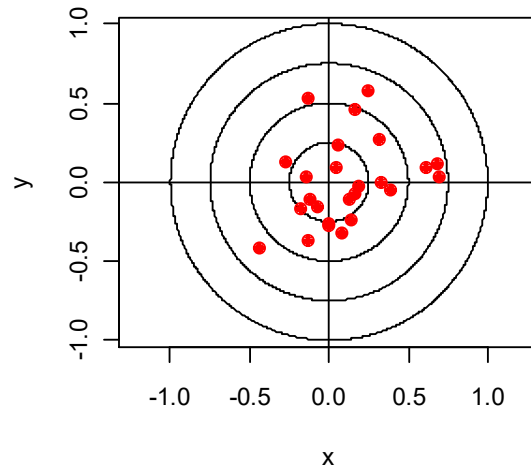
- **Statistik:**

- Umgang mit zufälligen Messfehlern und biologischer Varianz in Stichproben

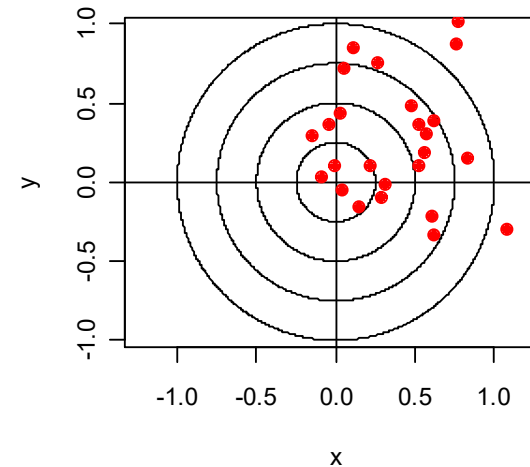
■ Grundverständnis Datenanalyse

Fehler ist zusammengesetzt aus Verzerrung (Bias) und Zufalls-Streuung

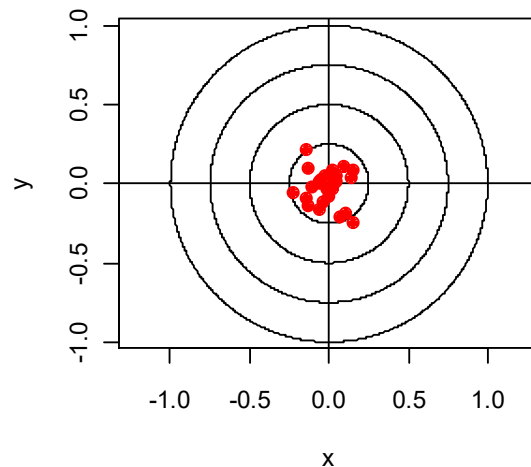
Bias=0 var=hoch



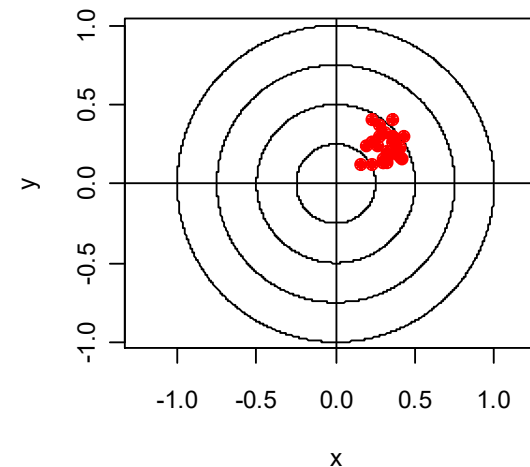
Bias=deutlich var=hoch



Bias=0 var=niedrig



Bias=deutlich var=niedrig



Wann sind Daten interpretierbar und glaubwürdig?

- Keine **systematischer Verzerrungen**
 - im **Messprozess für Einzeldaten** und
 - im Prozess der **Stichprobenerhebung**
- Keine **Zufallsbefunde**
- **(Keine interessensgeleitete Einflußnahme)**

Doctor admits *Lancet* study is fiction

A Norwegian researcher dreamed up the lives and lifestyles of some 900 people — and used them in a study on cancer. Then, last October, Jon Sudbø had his results published in *The Lancet*.

The revelation comes hard on the heels of the Woo Suk Hwang scandal, in which several important advances in human cloning reported by the South Korean researcher turned out to be faked.

The blatant nature of Sudbø's fiction emphasizes questions already being asked about the effectiveness of peer review, even in top journals, and about who should be responsible for catching fraud (see page 243).

The latest deception was discovered by

Camilla Stoltenberg, a director of epidemiology at the Norwegian Institute of Public Health in Oslo, who was catching up on her literature reading over Christmas. Sudbø's paper claims to analyse a public-health database and show that taking anti-inflammatory drugs can reduce the incidence of mouth cancer (J. Sudbø *et al. Lancet* 366, 1359–1366; 2005). "I was surprised because it refers to the Cohort of Norway, for which I am responsible," Stoltenberg says. She knew that this could not have been the source of the lifestyle data as the paper claimed.

Last week, Sudbø, who is based at the Norwegian Radium Hospital in Oslo,

is also not clear what his 13 co-authors knew about the fraud — the paper identifies three others as contributing equally to the research, and among the other co-authors are Sudbø's wife and his identical twin. None of the authors could be reached for comment.

The hospital has asked that Sudbø's other work be examined in an independent investigation, to be set up this week by Anders Ekbohm, an epidemiologist at the Karolinska Institute in Stockholm, Sweden. But the case is already set to change research policy in Norway. The country's health minister, Sylvia Brustad, announced on 16 January that previously stalled reforms on medical research will probably be law by

the autumn. The new rules would put more responsibility for catching fraud on the shoulders of the institutions where the research was done.

Richard Horton, editor of *The Lancet*, insists his journal is not at fault. "This is all so similar to the Hwang thing that we have just been through," he says wearily. "Peer review is a great system for detecting badly done research, but if you have an investigator determined to fabricate an entire study, it is not possible to pick it up." The mechanism of peer review at his journal is currently being examined as part of the largest study ever conducted into the process (see page 252). ■

Emma Marris

Betrug

Nature 17.01.2006

NORWEGIAN RADIUM HOSPITAL



Novel approach:
Jon Sudbø
simply invented
his test subjects.

admitted that the data had not come from that database or any other, but from thin air.

Many details of this case still need to be worked out. There is some indication that Sudbø may have mental health problems. It

Irreführung durch Datenfälschung: Hochdosischemotherapie beim Brustkrebs

Bezwoda WR, Seymour L, Dansey RD.

High-dose chemotherapy with hematopoietic rescue as primary treatment for metastatic breast cancer: a randomized trial. JCO 1995 10:2483-9

1995: **Bezwoda : Hochdosis bringt Überlebensvorteil!**
1995-2000: **Fünf Studien können Resultat nicht reproduzieren!**
2000: **Onsite Review der Daten: Lancet 355: 999-1003 (2000):**

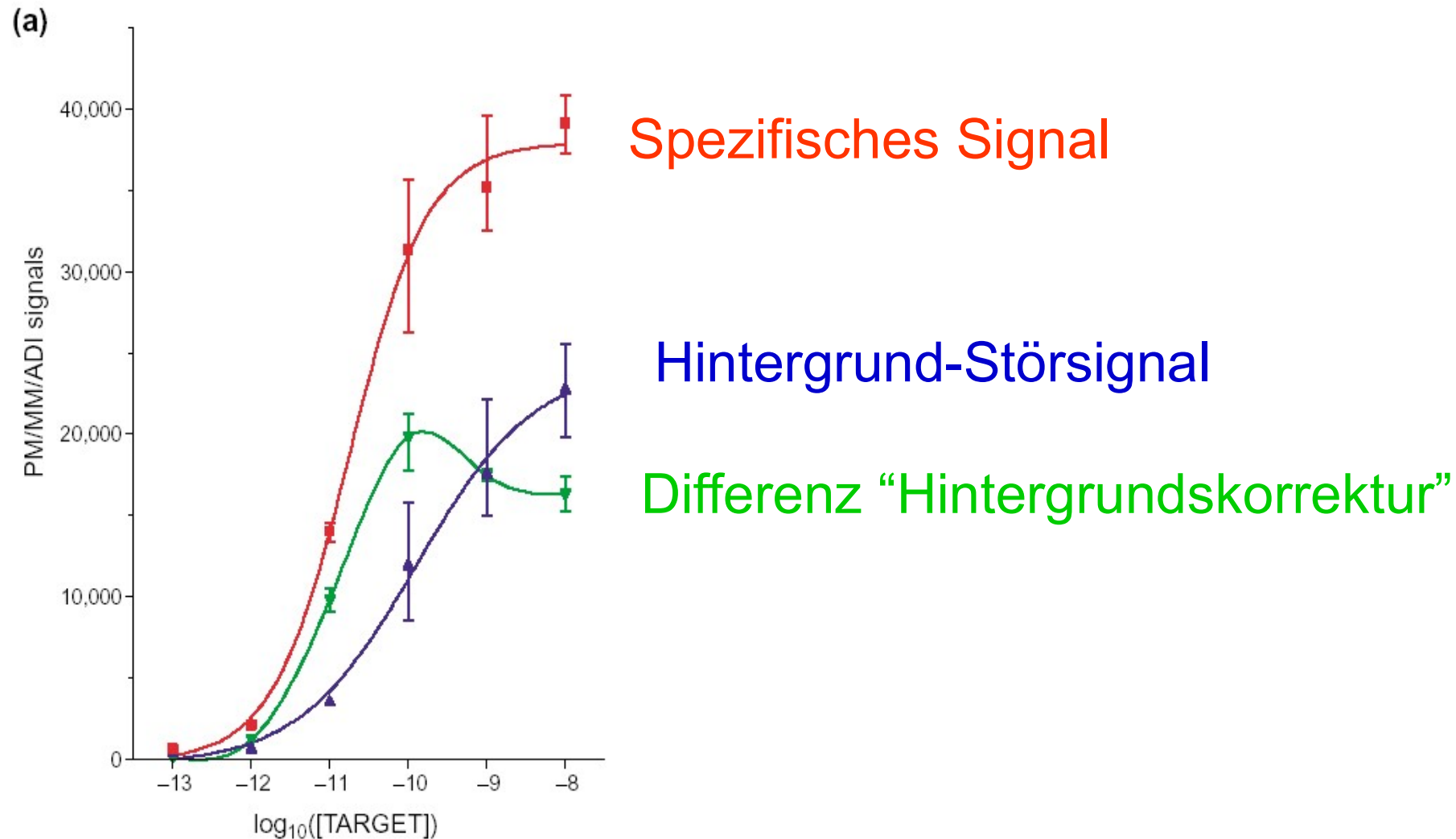
“The review of the Bezwoda study did not confirm the validity of what was done and what was reported.

The multiple discrepancies were of sufficient magnitude to invalidate the study results.

Bezwoda’s admission that he misrepresented the control group discredits his published findings.

Therefore, the study results should not be used as the basis for future randomised trials.”

Systematische Verzerrung im Messverfahren I



Chudin 2001

PM red, AvDiff green, MM blue (gemittelt über Probe-Paare); ranges

Systematische Verzerrung im Messverfahren II

- Erhebung von Nebenwirkungen einer Chemotherapie
 - Methode 1: Auswertung der spontanen Klartextnotizen in der Krankenakte.
 - Methode 2: Strukturierter Fragebogen (CTC-Kriterien)

Grade	0	1	2	3	4	Grade
Haemoglobin (g/dl) (mmol/l)	≥ 11.0 ≥ 6.8	10.0 - < 11.0 6.2 - < 6.8	8.0 - < 10.0 5 - < 6.2	6.5 - < 8.0 4 - < 5	< 6.5 < 4	Grade
WBC (* 10 ⁹ /l) (... /μl)	≥ 4.0 ≥ 4000	3.0 - < 4.0 3000 - < 4000	2.0 - < 3.0 2000 - < 3000	1.0 - < 2.0 1000 - < 2000	< 1.0 < 1000	Grade
Neutrophils (* 10 ⁹ /l) (... /μl)	≥ 2.0 ≥ 2000	1.5 - < 2.0 1500 - < 2000	1.0 - < 1.5 1000 - < 1500	0.5 - < 1.0 500 - < 1000	< 0.5 < 500	Grade
Platelets (* 10 ⁹ /l) (... /μl)	≥ 100.0 ≥ 100000	75.0 - < 100.0 75000 - < 100000	50.0 - < 75.0 50000 - < 75000	10.0 - < 50.0 10000 - < 50000	< 10.0 < 10000	Grade
Creatinine	within normal range (WNR)	N - 1.5 x N	> 1.5 x N - 3.0 x N	> 3.0 x N - 6.0 x N	> 6.0 x N	Grade
Bilirubin	within normal range (WNR)	N - 1.5 x N	> 1.5 x N - 3.0 x N	> 3.0 x N - 10.0 x N	> 10.0 x N	Grade
Liver enzymes (GOT/GPT/GGT)	within normal range (WNR)	N - 2.5 x N	> 2.5 x N - 5.0 x N	> 5.0 x N - 20.0 x N	> 20.0 x N	Grade
Fever	none	38.0 - 39.0 °C	> 39.0 - 40.0 °C	> 40.0 °C for less than 24 hours	> 40.0 °C for more than 24 hours or accompanied by hypotension	Grade
Infection	none	mild	moderate	severe	life-threatening	Grade

Systematische Verzerrung im Prozess der Stichprobenerhebung I

Die Komplikationsrate von Hausgeburten
ist signifikant niedriger als
die Komplikationsrate von Geburten im Krankenhaus!
(?!)

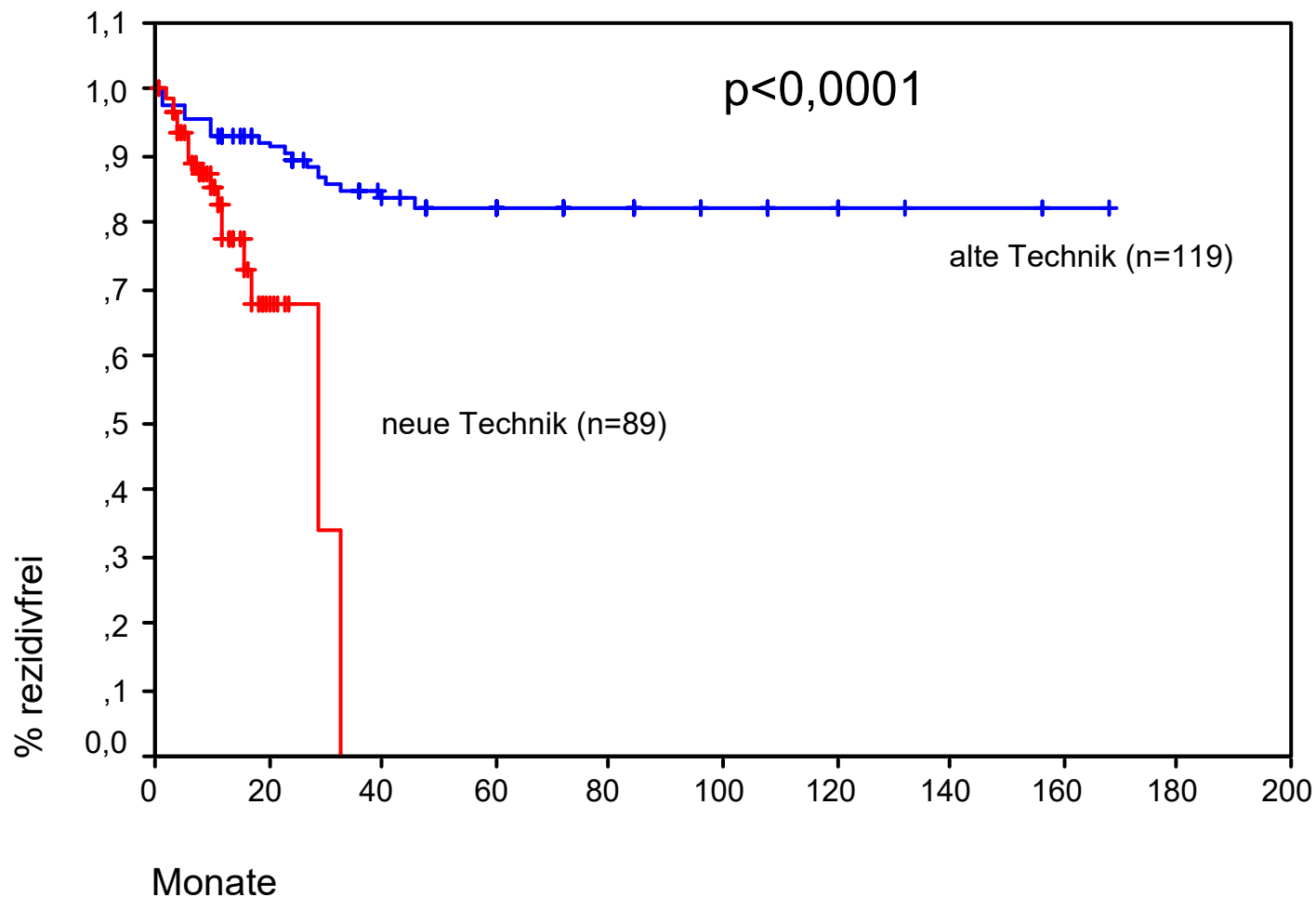
*Leipziger Volkszeitung
Panorama-Seite ~Frühjahr 1997*

Systematische Verzerrung im Prozess der Stichprobenerhebung II

- Klinker beurteilen Laborwerte und notieren in der Krankenakte
 - **Hb=oB** ohne Befund falls der Wert im Normalbereich liegt
 - Den konkreten Wert z.B. **Hb=9.3 mg/dl** wenn der Wert auffällig ist.
- Bei Datenabstraktion aus Krankenakten entsteht so aus “oB” ein **Fehlwert (“missing value”)**
- Wertet man naiv die nicht fehlenden Laborwerte aus, entsteht ein völlig verzerrtes Bild
- **Zufällig fehlende Werte sind bedauerlich, aber nicht direkt irreführend,**
- **Selektiv fehlende Werte verzerren.**

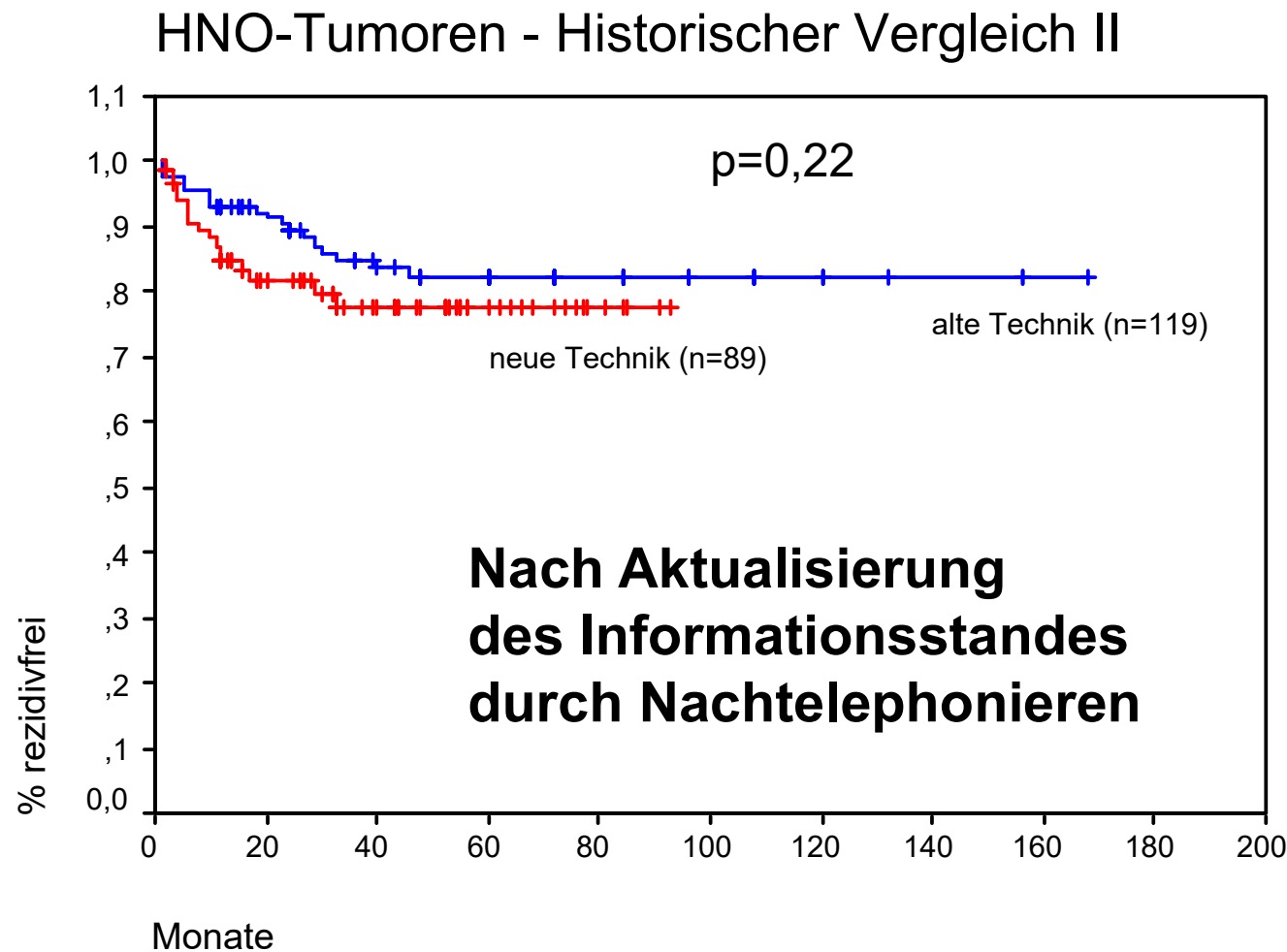
„Muß der Klinikscheff ins Gefängnis?“

HNO-Tumoren - Historischer Vergleich I



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

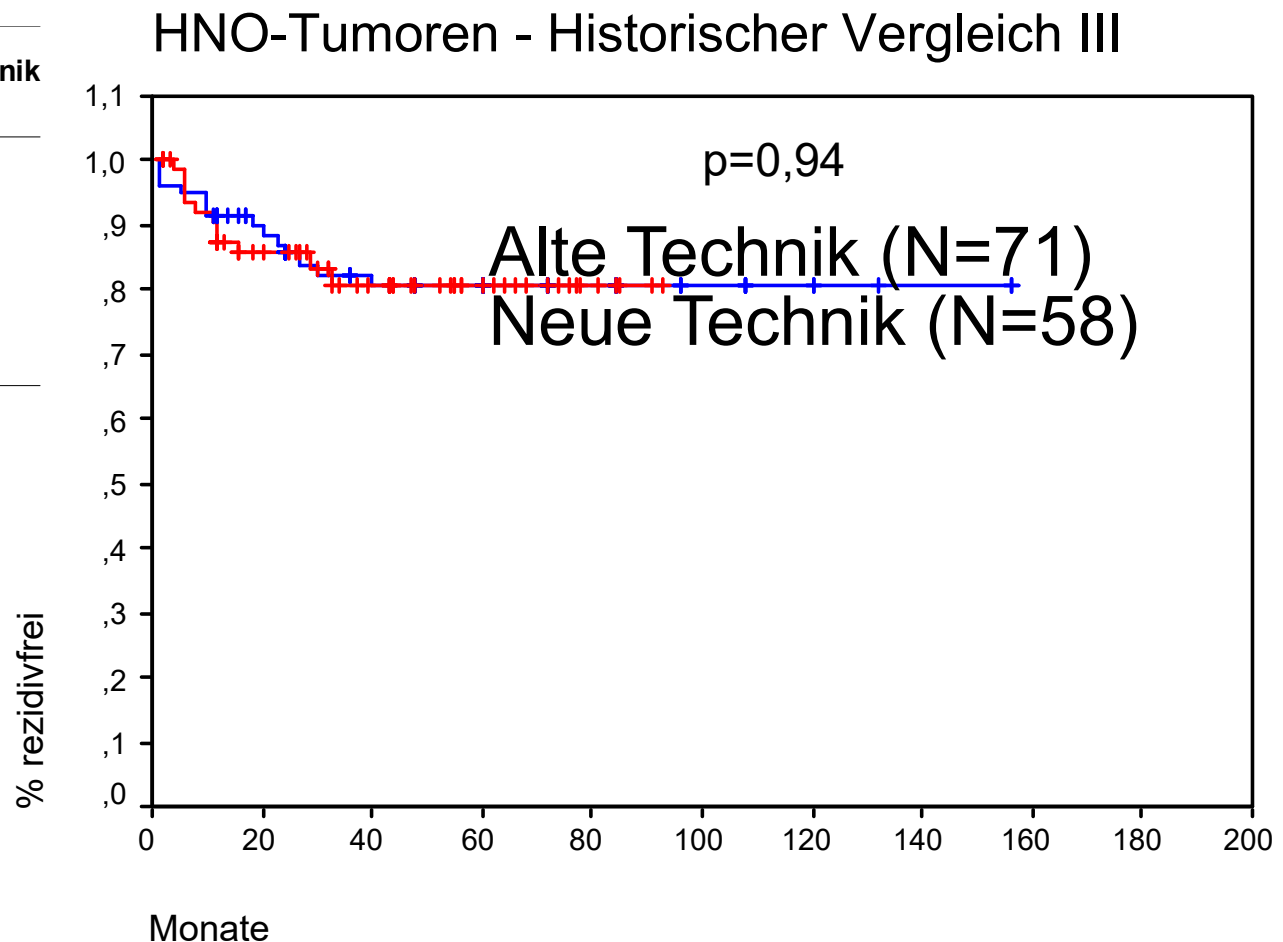
Informative Zensur – durch Zusammenhang zwischen Informationsfluss und Outcome



CAVE bei retrospektiven Historischen Vergleichen

Lokalisation des Tumors	alte Technik N=119	neue Technik N=89
Lippe	1	27
Augenlid	47	4
Sonstige	71	58

Nach Weglassen von Untergruppen mit drastischer Struktur-Ungleichheit



Verzerrung im Prozess der Veröffentlichung

- **Publikations Bias:** Selektive Publikation signifikanter Ergebnisse
- **Verzögerungs Bias:** Schnellere Publikation signifikanter Ergebnisse
- **Sprachen Bias:** Sprachwahl je nach Ergebnis
- **Selektiver Bericht:** Selektiver Bericht von signifikanten Endpunkten

Kann das Zufall sein?

Dioxin verhindert die Zeugung von Jungen

(Yahoo! Schlagzeilen: Montag 29. Mai 2000 10.00h)

Eine hohe Dioxin-Belastung verringert die Chance für Männer, einen Jungen zu zeugen. Darauf weist das Wissenschaftsmagazin „MorgenWelt“ mit Blick auf eine italienische Studie hin. Die Forscher hatten die Langzeitfolgen des Dioxin-Unfalls von Seveso im Jahre 1976 untersucht....

Insgesamt hatten die betroffenen Männer seit dem Unfall mit unbelasteten Frauen 88 Jungen und 103 Mädchen gezeugt. Normalerweise liegt die Geburtenrate dagegen bei 106 Jungen zu 100 Mädchen...

Kann das Zufall sein? II

Dioxin verhindert die Zeugung von Jungen

(Yahoo! Schlagzeilen: Montag 29. Mai 2000 10.00h)

Erwartet: $p_0 = 106/206 = 51,5\%$

Beobachtet: $p = 88/191 = 46,1\%$

```
> binom.test(88,191, p = 106/206)
```

Exact binomial test

data: 88 and 191

number of successes = 88, number of trials = 191, **p-value = 0.1477**

alternative hypothesis: true probability of success is not equal to 0.5145631

95 percent confidence interval:

0.3885281 0.5341757

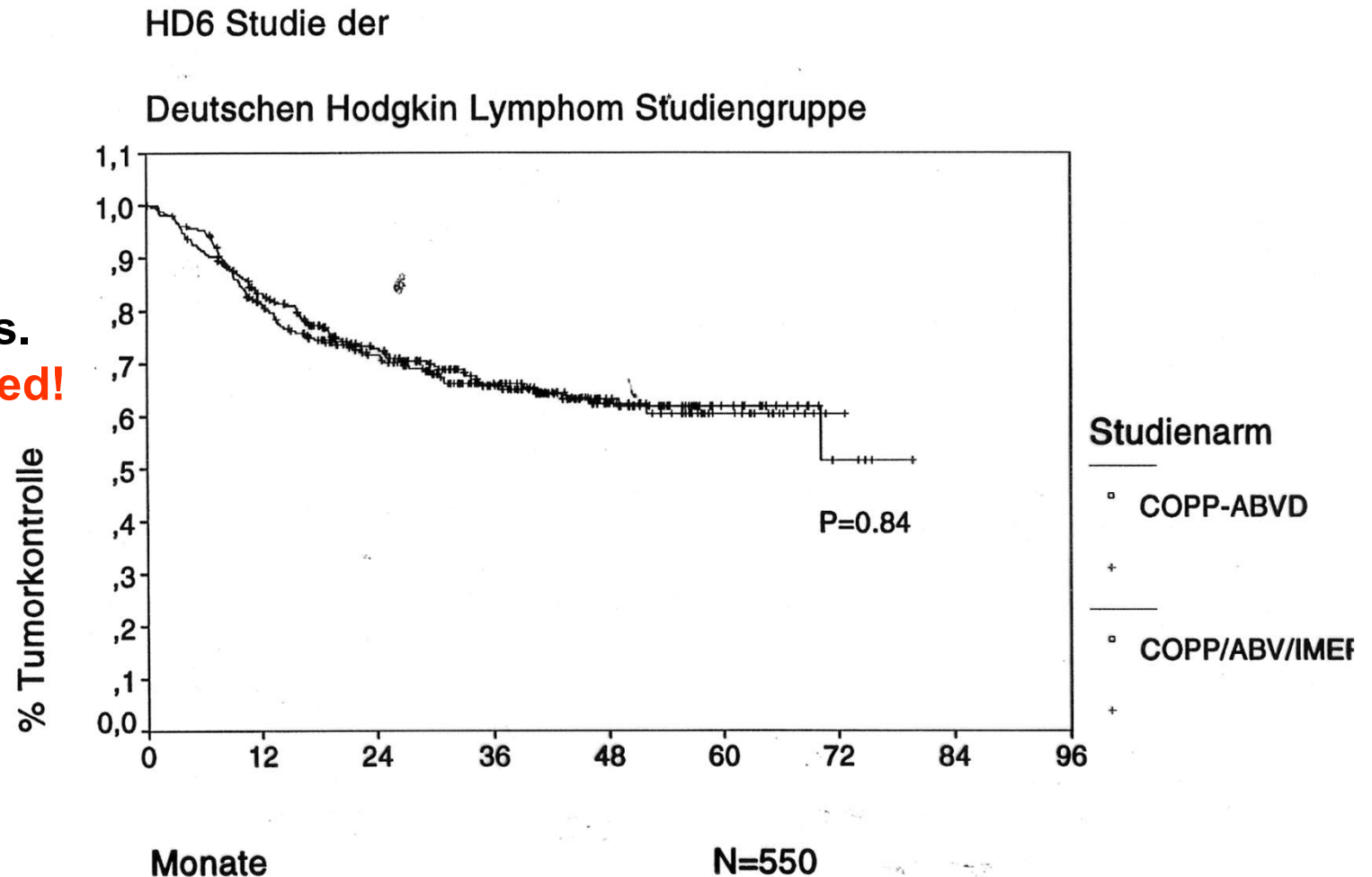
sample estimates:

probability of success: 0.460733

Fazit: **Beobachtete Abweichung ist durchaus mit einem Zufallsbefund verträglich.**

Subgruppenanalyse HD6

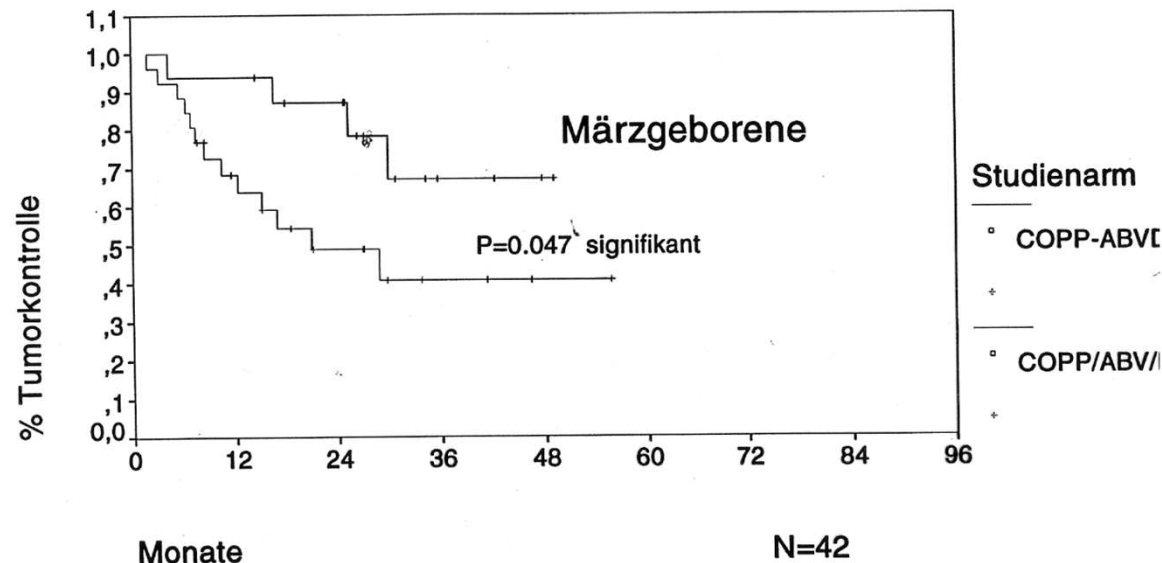
Randomisierter
Chemotherapie
Vergleich in
fortgeschrittenen
Stadien
eines
Hodgkin Lymphoms.
Keinerlei Unterschied!



Subgruppenanalyse HD6

HD6 Studie der

Deutschen Hodgkin Lymphom Studiengruppe



Subgruppenanalyse nach Geburtsmonat

CAVE: Multiples Testen

Test des Therapieeffekts $p=0.84$

Subgruppenanalyse nach Geburtsmonat:

Januar	$p=0.87$
Februar	$p=0.56$
März	$p=0.047$ formal signifikant
April	$p=0.95$
Mai	$p=0.28$
Juni	$p=0.21$
Juli	$p=0.76$
August	$p=0.075$ Trend in entgegengesetzte Richtung..
September	$p=0.94$
Oktober	$p=0.91$
November	$p=0.77$
Dezember	$p=0.22$

Valide Daten

- **Schlechte Datenqualität:**

Falsche, gefälschte, verfälschte, unvollständige Daten

- kontrolliert durch **Professionelle Qualitätssicherung,
Datenmanagement und -monitoring**

- **Systematische Verzerrungen**

- Nicht valide Messmethode

- Selektive Auswahl der
Beobachtungen

kontrolliert durch **Biometrie**

Validierte Messmethoden

Biasdiagnostik & Versuchplaung

- **Zufallsbefunde**

kontrolliert durch

Statistik

Versuchsplanung: Herstellen der Vergleichbarkeit!

Zeitpunkt	Heterogenitätsquellen	Ziel	
bei Eintritt in die Studie	Patientcharakteristika (bekannte und latente Faktoren, die den Therapieerfolg beeinflussen)	Strukturgleichheit	Randomisieren Störgrößen Ausschalten/ mitmessen
während der Durchführung der Studie	Behandlung (Strategien bei Unverträglichkeit der Therapie, Zeitschema, supportive Maßnahmen, Pflege)	Beobachtungs- und Behandlungsgleich- heit	Standardisieren Qualitäts- kontrolle
	Beobachtung (Zeitschema, Beurteilungskriterien)		Valide Mess- größen Verblinden Unabhängiger Review
nach Beendigung der Studie	Auswertung (Zählweise nicht protokollgerecht behandelter Patienten)	Auswertungsgleich- heit	Vorab- Analyseplan

Hinweise

■ **Vorlesungsseite:** Folien und Skripte

<https://www.imise.uni-leipzig.de/Lehre/Semester/2020-21/MI-Biometrie-2>

Zugang: Teilnehmer / Voodoo

Wer es genauer wissen will - Begleitlektüre:

* Michael Lavine: Introduction to statistical thought

<https://people.math.umass.edu/~lavine/Book/book.html>

* C. M. Grinstead and J. L. Snell : Introduction to Probability.

http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html

* Lothar Sachs & Jürgen Heddreich: Angewandte Statistik 13. Auflage Springer

* Wolfgang Kohn: Statistik, Springer

R Project for Statistical Computing:

<http://www.r-project.org/>

R-Studio: Entwicklungsumgebung für R

<https://rstudio.com/>

Bis zur nächsten Übung

- Auf der Vorlesungseite zum 'runterladen:
- Bitte lesen! bzw. bearbeiten!
- <https://www.rwi-essen.de/unstatistik/>
 - Einen Beitrag zum Referieren in der Übung vorbereiten
 - Jahresarchiv: Nehmen Sie einen Beitrag aus dem Jahr 20XX, wenn XX Ihr Geburtsmonat ist. (Sonst wählen)
- Übungsblatt 1: Besprechung 2020-11-10