

Studiendesign und Statistik: Interpretation publizierter klinischer Daten

Dr. Antje Jahn

Institut für Medizinische Biometrie, Epidemiologie und Informatik
Universitätsmedizin Mainz

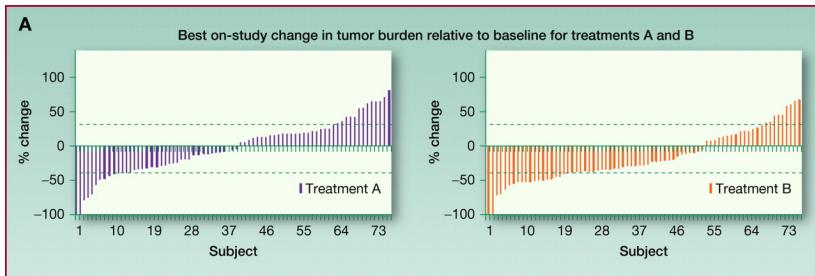
Hämatologie im Wandel, 28/02/2015

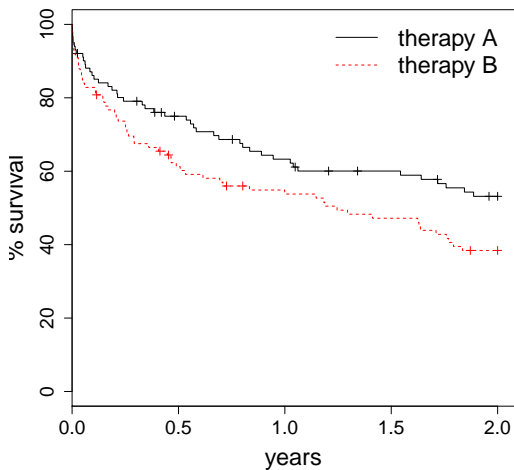


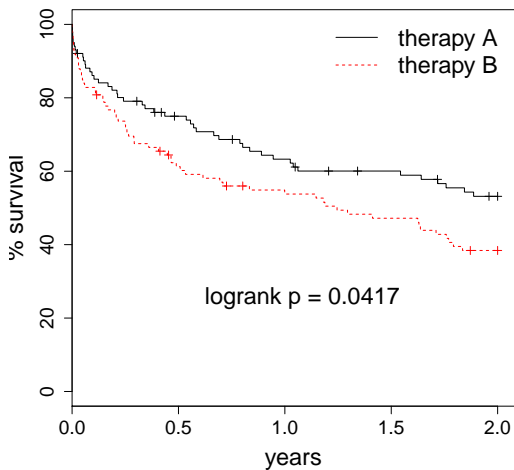
UNIVERSITÄTS**medizin.**

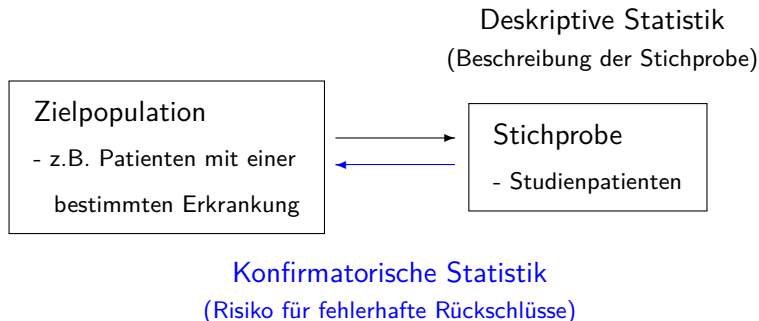
MAINZ

- Statistische Signifikanz
- Multiples Testen und Zwischenauswertungen
- (Cox-)Regressionen und Hazard Ratio
- Konfidenzintervalle
- Auswertungspopulationen









p-Wert:

Die Wahrscheinlichkeit, dass ein Gruppenunterschied von mind. der Größe wie in der Studie beobachtet rein aus Zufall, d.h. ohne Therapieeffekt, auftritt

⇒ Kleine p-Werte sprechen eher für einen Therapieeffekt

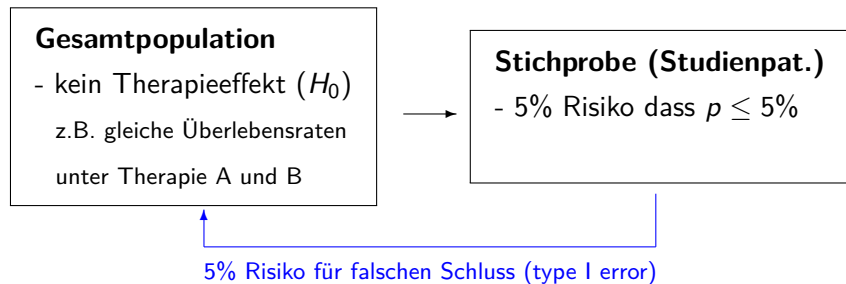
Gesamtpopulation

- kein Therapieeffekt (H_0)
z.B. gleiche Überlebensraten
unter Therapie A und B



Stichprobe (Studienpat.)

- 5% Risiko dass $p \leq 5\%$



- Das **Signifikanzniveau α** ist das maximal tolerable Risiko für einen Fehler 1. Art und wird zu Studienbeginn festgelegt.
- Ein statistischer Test liefert ein Entscheidungskriterium, welches dies sicherstellt.

ICH-E9: „Conventionally the probability of type I error is set at 5% or less... „

CAVE

Nicht-Signifikanz ($p > \alpha$)

≠

Äquivalenznachweis

Signifikanz ($p \leq \alpha$)

≠

klinische Relevanz

- Je größer die Fallzahl, desto kleiner der p-Wert

Variable	Therapie A	Therapie B	p-Wert
Patienten mit Progression (Studie 1)	20/50 (40%)	30/50 (60%)	0.046
Patienten mit Progression (Studie 2)	10/25 (40%)	15/25 (60%)	0.157

- Ein p-Wert liefert nur eine Aussage über die statistische Signifikanz eines Ergebnisses und keine Aussage über die klinische Relevanz

- Statistische Tests berechnen ein Entscheidungskriterium (p-Wert), ob ein Therapieeffekt als nachgewiesen betrachtet werden kann - so dass der Fehler 1. Art kontrolliert wird
- Es gibt verschiedene statistische Tests, d.h. Methoden einen p-Wert zu berechnen
- Welcher Test geeignet ist, hängt vom Studiendesign, der Fragestellung, der Datenstruktur und der Null- und Alternativhypothese ab.



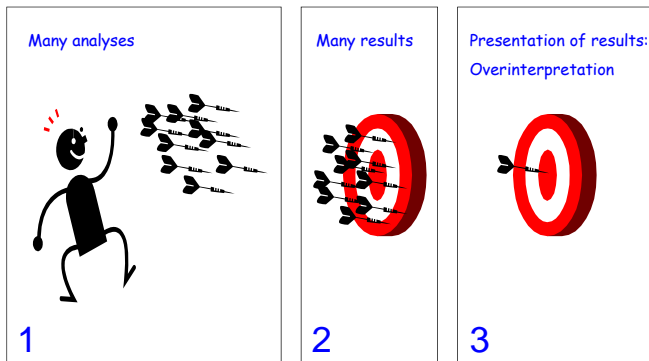
„Müsli macht Männer“ (Süddeutsche, 2008)

„Mehr Söhne durch Müsli?“ (Stern, 2008)

„Schlemmen steigert Chance auf Jungen“ (Spiegel, 2008)



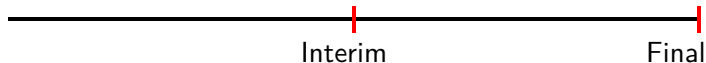
Problem of multiple Testing



In jeder Studie sind viele Endpunkte von Interesse. Wie kann das Risiko eines Fehlers 1.Art dennoch kontrolliert werden?

- Auswahl eines Primärziels, z.B. Überlebensraten
- α -Adjustierung: Anpassung des Signifikanzniveaus für jeden einzelnen Test, so dass insgesamt das Fehlerrisiko $\leq 5\%$

- 5-Jahres-Studie zum Vergleich der Überlebensraten zweier Therapien A und B mit Zwischenauswertung nach 2.5 Jahren



⇒ α -Adjustierung erforderlich!

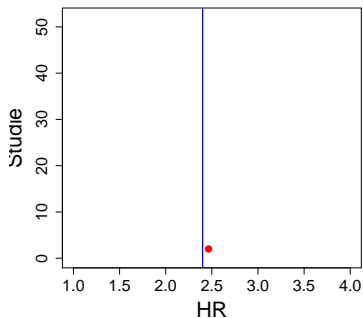
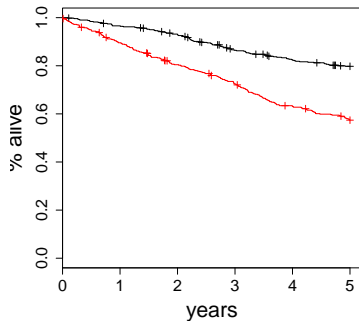
- Studienprotokoll: „... the significance level for the final analysis was set at 4.8% after adjustment by the O'Brien and Fleming method.“

„ICH-E9: Conventionally the probability of type I error is set at 5% or less as dictated by any adjustments made necessary for multiplicity considerations,,

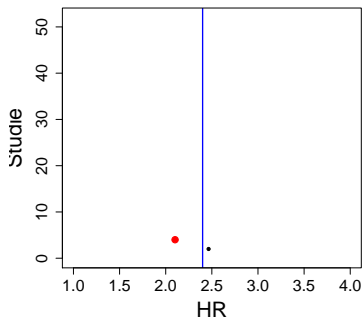
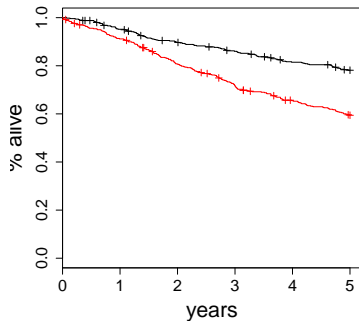
- Regressionsanalyse bei Ereigniszeitdaten: Cox-Regression
- Berücksichtigung von Zensierungen
- Schätzer für den Therapieeffekt: Hazard Ratio
 - relative Veränderung im hazard (\approx unmittelbares Ereignisrisiko eines bisher ereignisfreien Patienten) unter Therapie A im Vergleich zu Therapie B
 - Hazard Ratio $\left\{ \begin{array}{l} > 1, \text{ erhöhtes Risiko unter Therapie A} \\ = 1, \text{ gleiches Risiko} \\ < 1, \text{ reduziertes Risiko unter Therapie A} \end{array} \right.$

- Der HR wird aus einer **Stichprobe** der Gesamtpopulation geschätzt und kann daher vom wahren HR abweichen

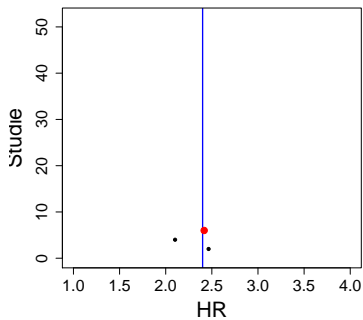
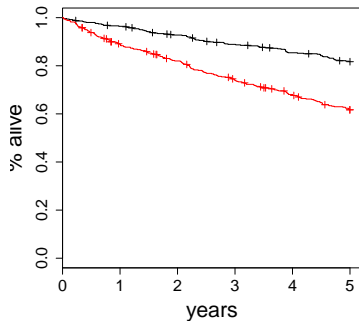
Kaplan Meier plot



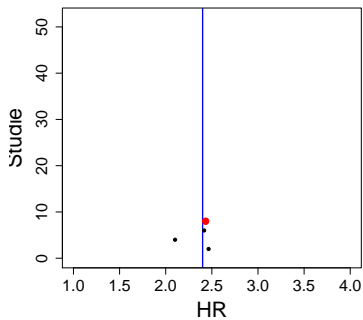
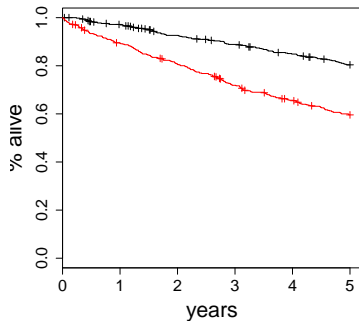
Kaplan Meier plot



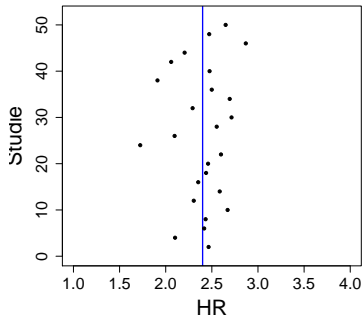
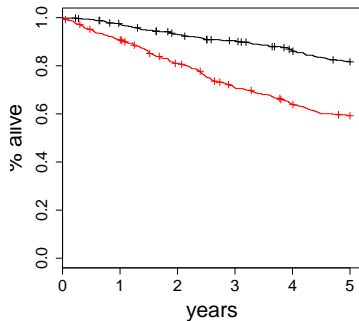
Kaplan Meier plot



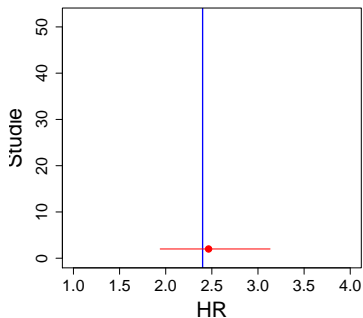
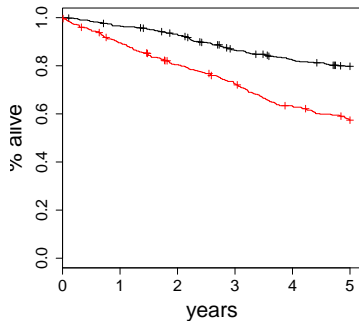
Kaplan Meier plot



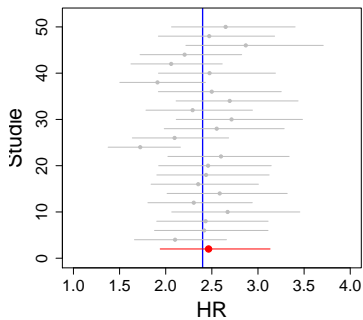
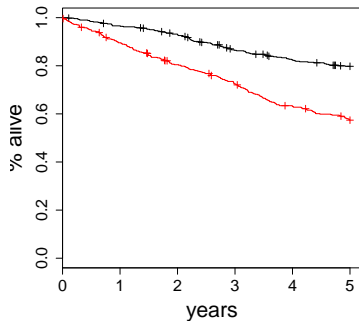
Kaplan Meier plot



Kaplan Meier plot



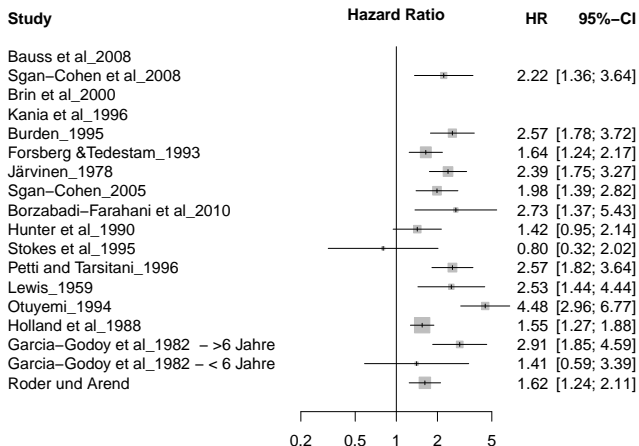
Kaplan Meier plot



Konfidenzintervall (CI)

- Man legt das Konfidenzniveau $1 - \alpha$ fest, i.d.R. $1 - \alpha = 95\%$
- Das 95%-CI für einen Parameter ist ein Intervall um den Schätzer aus der Stichprobe, der in 95% der Fälle den wahren Parameter enthält.
- Man nimmt an, dass der wahre Parameter im CI liegt, da man damit in 95% aller Fälle richtig liegt.

Meta-Analysen - Forest Plot



In einer randomisierten Studie bei Männern nach Herzinfarkt zum Vergleich eines Beta-Blockers mit Placebo waren in beiden Gruppen ca. 32% der Patienten non-compliant.

- Sollte man diese Patienten von der Auswertung ausschließen?

- Intention to Treat (ITT) Population:
Alle randomisierten/eingeschlossenen Patienten
Primärpopulation in Überlegenheitsstudien
- Per-Protocol Population:
Alle Patienten ohne schwerwiegende Protokollverstöße
Primärpopulation in Äquivalenzstudien
- Safety-Evaluable Population:
Alle Patienten, die mindestens einmal Therapie erhielten
Primärpopulation für Sicherheitsparameter

„No clever analysis can rescue a poorly conducted trial,,

„The greatest challenge is in the interpretation of results,,

DeMets D.L.: Statistical issues in interpreting clinical trials. Journal of Internal Medicine, 2004; 255: 529-53