

Studiendesign und- biometrie für die Praxis: Grundlagen und Anwendung für Publikationen

Dr. Antje Jahn

Institut für Medizinische Biometrie, Epidemiologie und Informatik
Universitätsmedizin Mainz

Hämatologie im Wandel, 02/03/2013

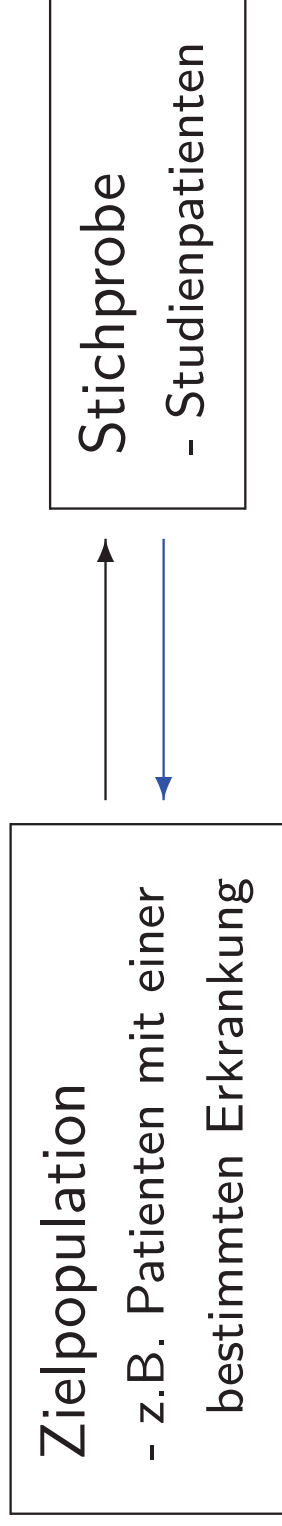


UNIVERSITÄTS**medizin.**
MAINZ

- Statistische Tests und p-Werte
- Multiples Testen und Zwischenbewertungen
- Konfidenzintervalle
- Auswertungspopulationen
- Regressionsanalysen

Was ist Biometrie?

Deskriptive Statistik
(Beschreibung der Stichprobe)



Konfirmatorische Statistik
(Risiko für fehlerhafte Rückschlüsse)

Studienergebnis (Beispiel)

Variable	Therapie A N=50	Therapie B N=50	p-Wert
Patienten mit Progression	20 (40%)	30 (60%)	0.046

Die ersten Spalten sind deskriptive Ergebnisse, der p-Wert ist ein konfirmatorisches Ergebnis

Statistische Tests

Illustratives Beispiel:

- Überprüft werden soll, ob Otitis Media (OM) öfter das Ohr betrifft, auf dem das Kind normalerweise schläft
- Bei 10 Kindern mit OM war das betroffene Ohr tatsächlich das, auf dem sie meist schlafen
- Wenn Schlafverhalten das OM-Risiko nicht beeinflusst, könnte man solch ein Ergebnis auch beobachten - allein aus Zufall
- Dies hätte eine Wahrscheinlichkeit von $(1/2)^{10} < 0.001$ ("p-Wert")
- Aufgrund der minimalen Wahrscheinlichkeit nehmen wir an, dass der beobachtete Zusammenhang zwischen Schlafverhalten und OM **kein** Zufallseffekt war, sondern tatsächlich besteht.

Das ist das Prinzip aller statistischen Tests!

Statistische Tests - der p-Wert

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

Statistische Tests - der p-Wert

Variable	Therapie A N=50	Therapie B N=50	p-Wert
Patienten mit Progression	20 (40%)	30 (60%)	0.046

Statistische Tests - Studienhypothesen

H_0 = Nullhypothese

= Die Annahme über den Therapieeffekt, die widerlegt werden soll

H_1 = Alternativhypothese

= Die Annahme über den Therapieeffekt, die nachgewiesen werden soll

Beispiel:

H_0 = Das Progressionsrisiko ist unter beiden Therapien gleich

H_1 = Die Therapien unterscheiden sich im Progressionsrisiko

Statistische Tests - Fehler 1. und 2. Art

"Wahrheit"
(Gesamtpopulation)

	H_0	H_1
Fazit aus der Studie (Stichprobe)	H_0 ✓	H_1 type II error
	type I error	✓

Beispiel:

H_0 = Das Progressionsrisiko ist unter beiden Therapien gleich

H_1 = Die Therapien unterscheiden sich im Progressionsrisiko

Statistische Tests - Signifikanzniveau

- 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... „

Statistische Tests - Testentscheidung

$p \leq \alpha$	$p > \alpha$
Die Nullhypothese wird abgelehnt, die Alternativhypothese (das Studienziel) ist nachgewiesen	Die Nullhypothese wird beibehalten, die Alternativhypothese ist nicht nachgewiesen
Das Ergebnis ist statistisch signifikant	

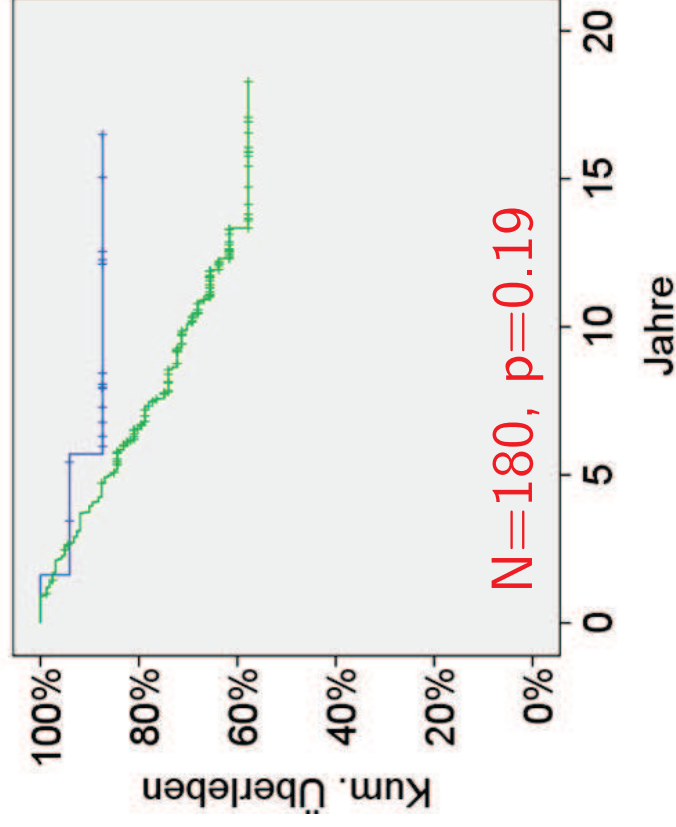
Statistische Tests - Testentscheidung

CAVE

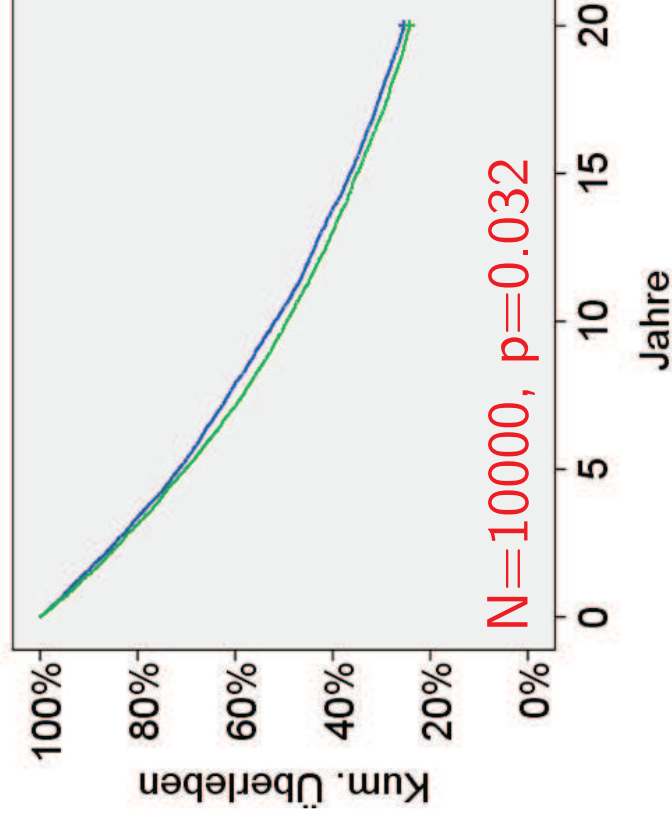
Nicht-Signifikanz ($p > \alpha$) \neq Beweis der Nullhypothese (z.B. keine Äquivalenz folgerbar)	Signifikanz ($p \leq \alpha$) \neq klinische Relevanz
--	--

Signifikanz \neq Klinische Relevanz

N0-Brustkrebs: Hochrisiko vs.
Niedrigrisiko (St.Gallen-Score)



simulierte Daten



Signifikanz \neq 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

- 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.

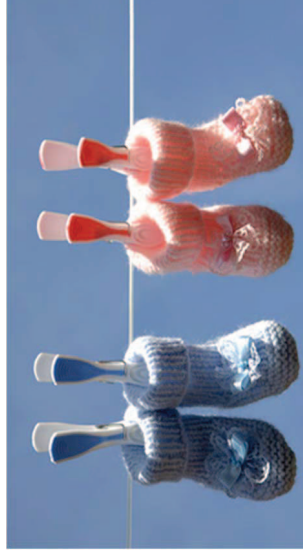
Multiple Testen



„Müsli macht Männer“ (Süddeutsche, 2008)

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


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



Multiples Testen

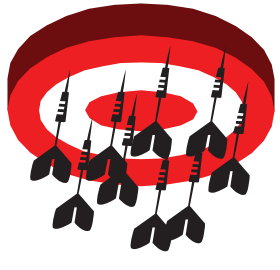
Problem of multiple Testing

Many analyses



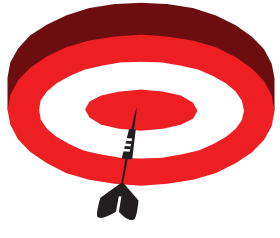
1

Many results



2

Presentation of results:
Overinterpretation



3

Multiples Testen

Angenommen, die Studie umfasst eine weitere Therapiegruppe C

Variable	Gruppenvergleich	local type 1 error rate
Progression	A vs B	5%
	A vs C	5%
	B vs C	5%

overall type 1 error rate $> 5\%$

Lösungen:

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

Multiple Testen: Zwischenauswertungen

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



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

Multiples Testen: Zwischenauswertungen



local type 1 error rate

Interim: A vs B | 5%

Final: A vs B | 5%

overall type 1 error rate $> 5\%$

⇒ α -Adjustierung erforderlich!

Multiples Testen

„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,”

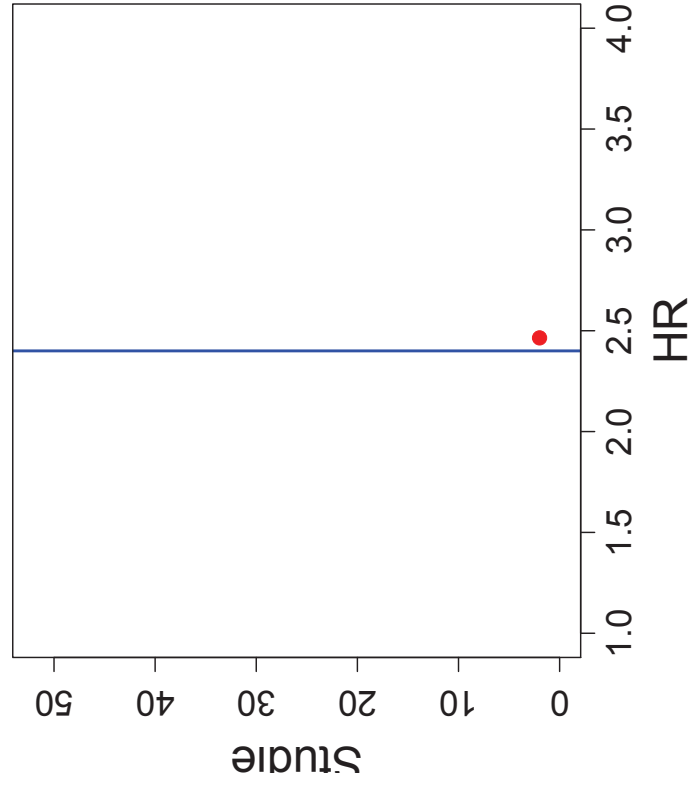
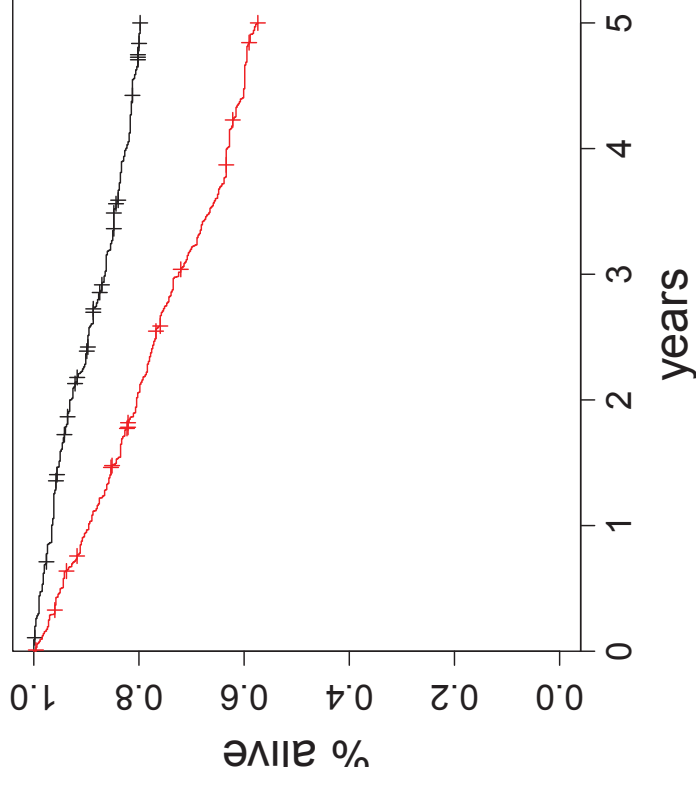
Schätzer und Konfidenzintervalle

Ziel: Schätzung eines Populationsparameters

- Beispiel: Hazard Ratio (HR)
- Der HR wird aus einer **Stichprobe** der Gesamtpopulation geschätzt und kann daher vom wahren HR abweichen

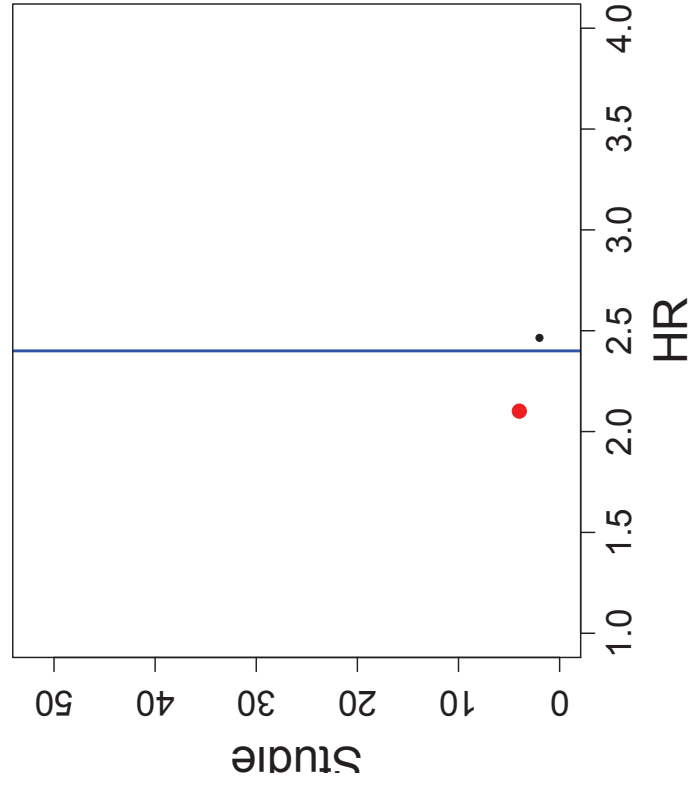
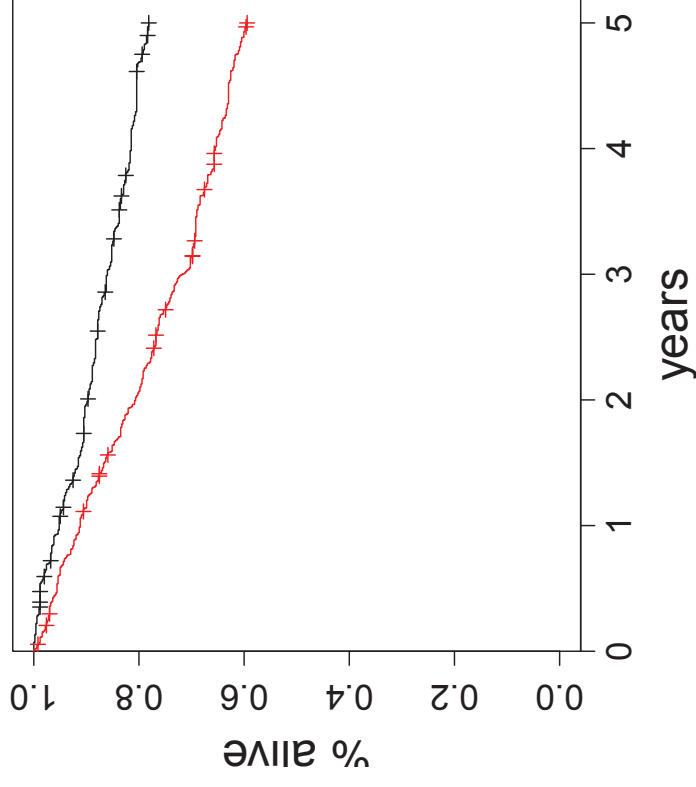
Schätzer und Konfidenzintervalle

Kaplan Meier plot



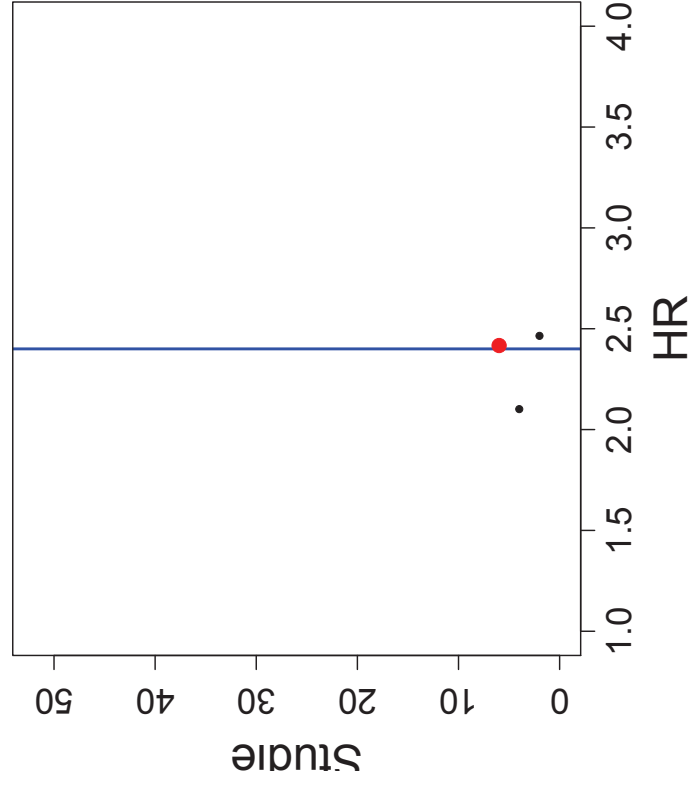
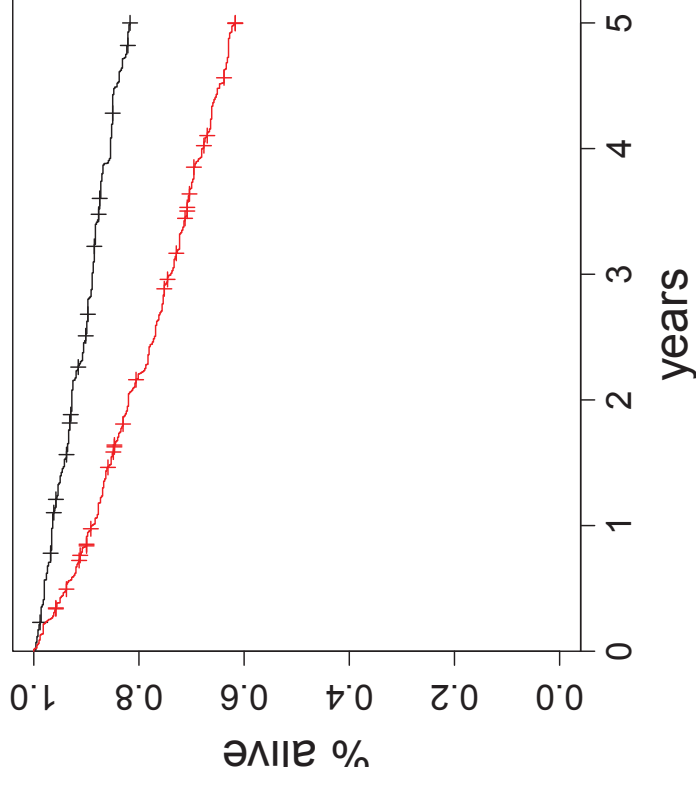
Schätzer und Konfidenzintervalle

Kaplan Meier plot



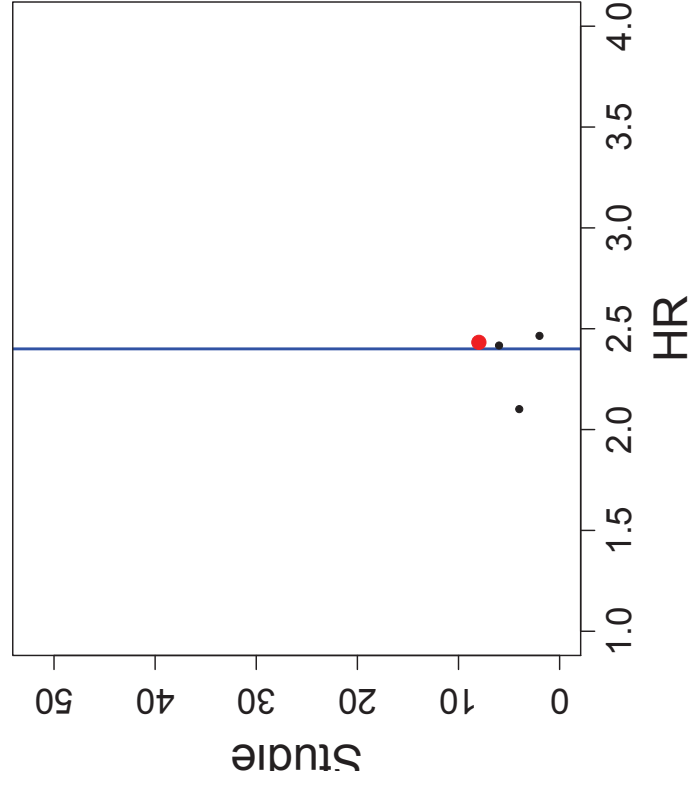
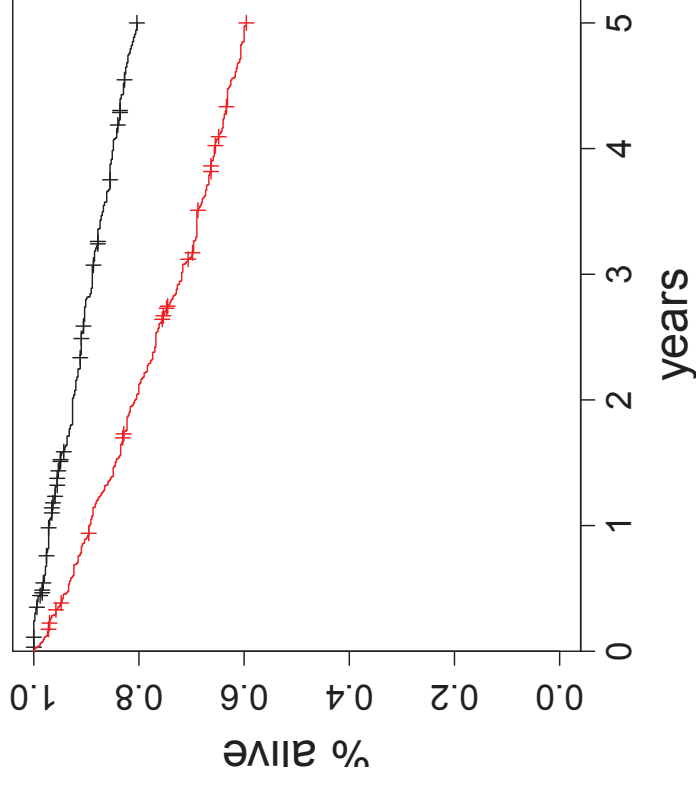
Schätzer und Konfidenzintervalle

Kaplan Meier plot



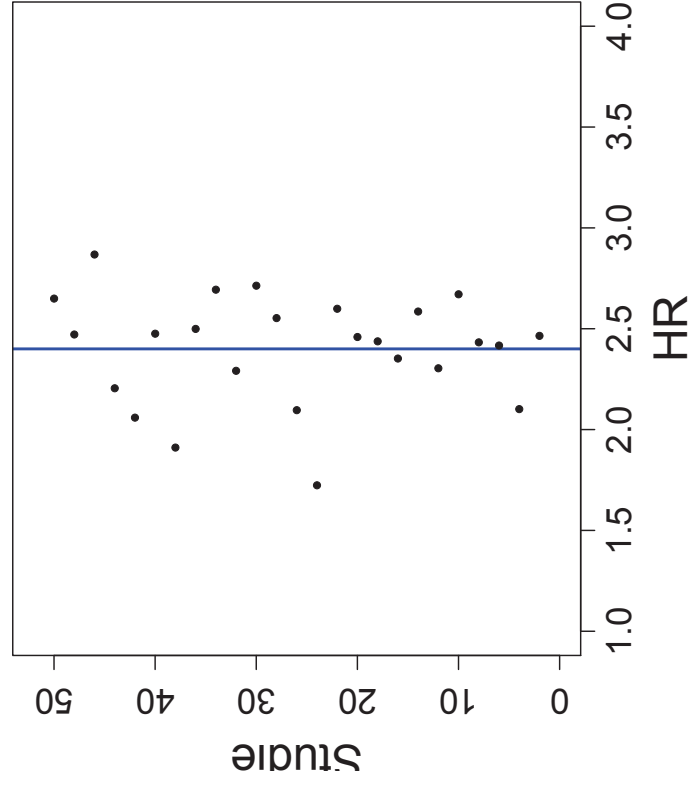
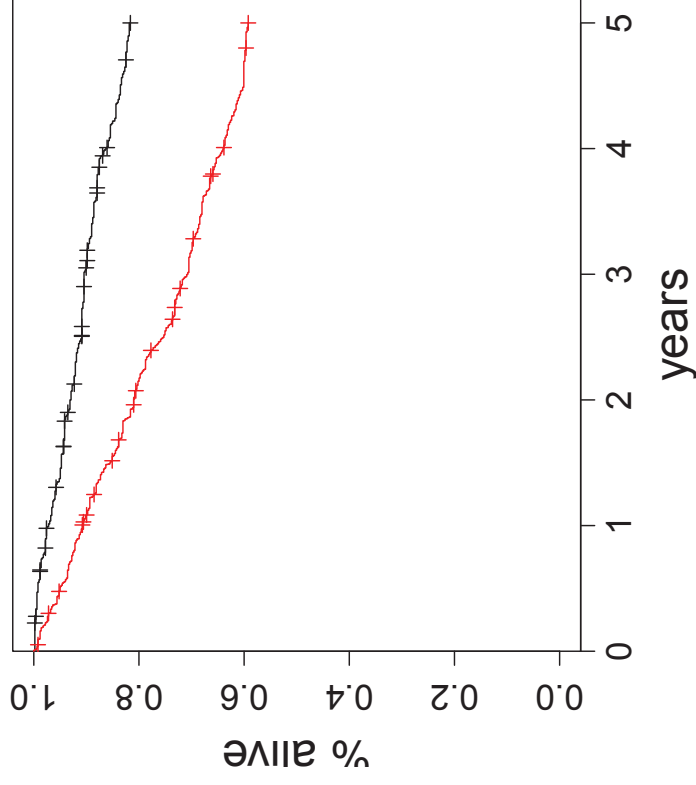
Schätzer und Konfidenzintervalle

Kaplan Meier plot



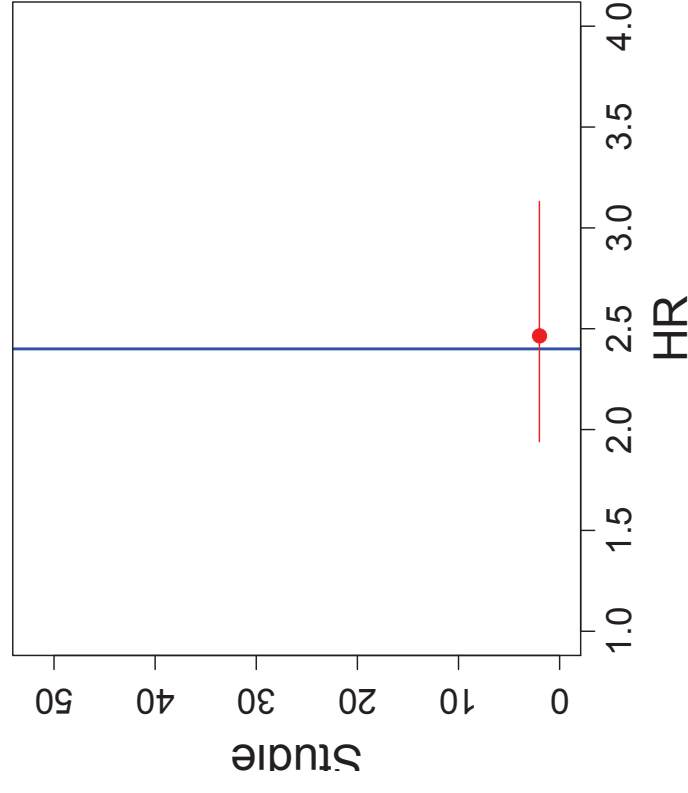
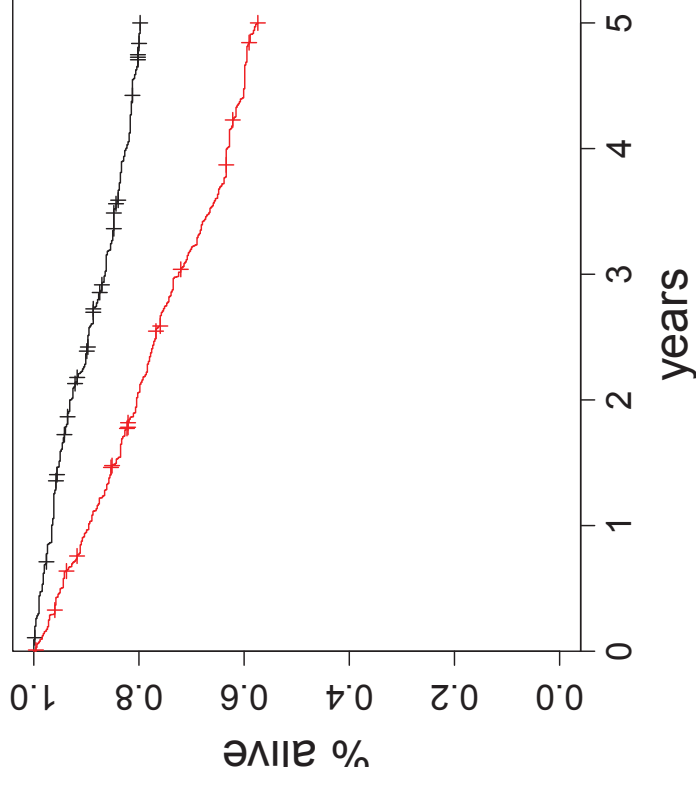
Schätzer und Konfidenzintervalle

Kaplan Meier plot



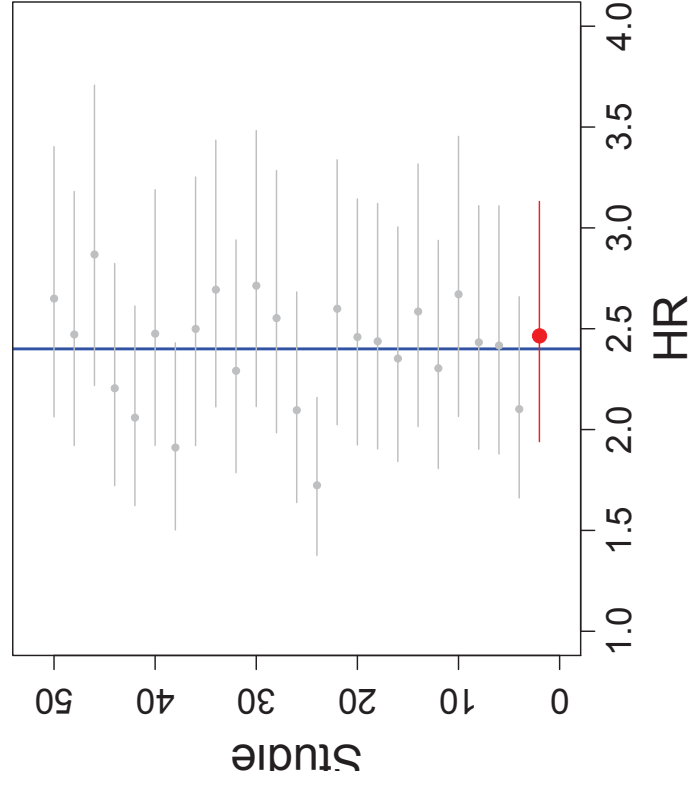
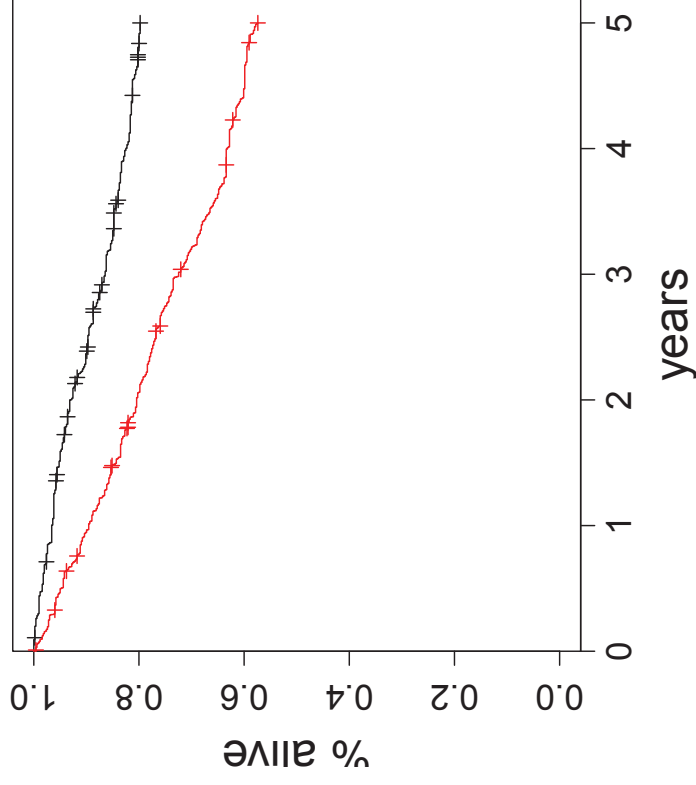
Schätzer und Konfidenzintervalle

Kaplan Meier plot



Schätzer und Konfidenzintervalle

Kaplan Meier plot



Schätzer und Konfidenzintervalle

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.

Auswertungspopulationen

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?

Auswertungspopulationen

- 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

Take Home Message

„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