

Statistische Methoden in der Brustkrebsforschung

Topics

- 17:00 - 17:05 **Begrüßung**
- 17:05 - 17:15 **Endpunkte und konfirmatorische Analysen**
Dr. Natalie Filmann
- 17:15 - 17:25 **Supportive Analysen**
Vanessa Schaser
- 17:25 - 17:35 **Auswertung binärer Endpunkte**
Dr. Natalie Filmann
- 17:35 - 17:45 **Auswertung von Überlebenszeitdaten**
Vanessa Schaser
- 17:45 - 18:00 **Q&A Session**
Dr. Natalie Filmann, Vanessa Schaser

Endpunkte und konfirmatorische Analysen



Dr. Natalie Filmann
GBG Forschungs GmbH

Primäre Endpunkte vs. Sekundäre Endpunkte

Der Primäre Endpunkt entscheidet über den Erfolg einer Studie

- Wichtigstes Ergebnis, dass den Nutzen einer Behandlung zeigen soll
 - Beispiele: Heilung, längeres Überleben, weniger Rückfälle
- Falls möglich: messbar/objektiv erhebbar, relevant für das Therapiegebiet und aussagekräftig
- Fallzahlberechnungen basieren auf dem primären Endpunkt
- Zulassung hauptsächlich basierend auf Ergebnis des primären Endpunkts:
 - **Primärer Endpunkt fällt negativ aus -> Studie ist negativ, auch wenn sekundäre Endpunkte positiv sind!**
- Oft nur ein primärer Endpunkt definiert, Co-primäre/multiple primäre Endpunkte aber möglich

Sekundäre Endpunkte unterstützen das Ergebnis einer Studie

- Beispiele: zusätzliche Effektivitätsvariablen, Nebenwirkungsprofil, Lebensqualität
- Studie wird nicht darauf gepowert → Vorteile in diesem Endpunkt eventuell nicht nachweisbar
- Präspezifizierte Hierarchie im Testen dieser Endpunkte möglich

Explorative Endpunkte oft Hypothesen generierend

- Noch weniger streng als sekundär, z.B. Analysen von neuen Biomarkern

Der Primäre Endpunkt entscheidet über den Erfolg einer Studie

Aufstellen einer relevanten Fragestellung (Hypothese)

Inkl. Wahl des primären Endpunkts und des statistischen Verfahrens zur Beantwortung der Fragestellung



Festlegung der Studienpopulation



Bestimmung der Fallzahl



...



Statistische Entscheidung



Interpretation

→ **Primäre Analyse**

Erinnerung: P-Werte

Definition: Die Wahrscheinlichkeit zufällig einen Unterschied in der zugrundeliegenden Population zu beobachten, wenn in Wirklichkeit für die Gesamtpopulation kein Unterschied besteht

→ Entscheidung über Erfolg basierend auf p-Werten (und Signifikanzniveau α)

$p \leq \alpha$: statistisch signifikant

$p > \alpha$: nicht statistisch signifikant

Fehlinterpretation nicht statistisch signifikantes Ergebnis:

Es konnte kein statistisch signifikanter Unterschied zwischen den Gruppen nachgewiesen werden.

- P-Werte liefern keine Information zur Effektgröße und der klinischen Relevanz eines Ergebnisses
- In p-Werten spiegelt sich keine Unsicherheit des Ergebnisses wider
 - Zusätzliche Berichterstattung des Effektschätzers wie bspw. Odds Ratio, Hazard Ratio mit dem zugehörigen Konfidenzintervall (→ mehr dazu später)

P-Werte allein berichten – nicht empfohlen

Fragestellung: Welche dieser fiktiven Studien zeigt einen Anstieg in der pCR-Rate?

Studie	P-Wert
Innovative 007	0.02
JustMissedEffect 098	0.055
ScheinEffect 123	0.041
Spekulation 992	0.039

Unklarheiten aus der Tabelle

Wie groß war der Effekt?

Welche Unsicherheit beinhaltet das Ergebnis?

In welche Richtung geht der Effekt?

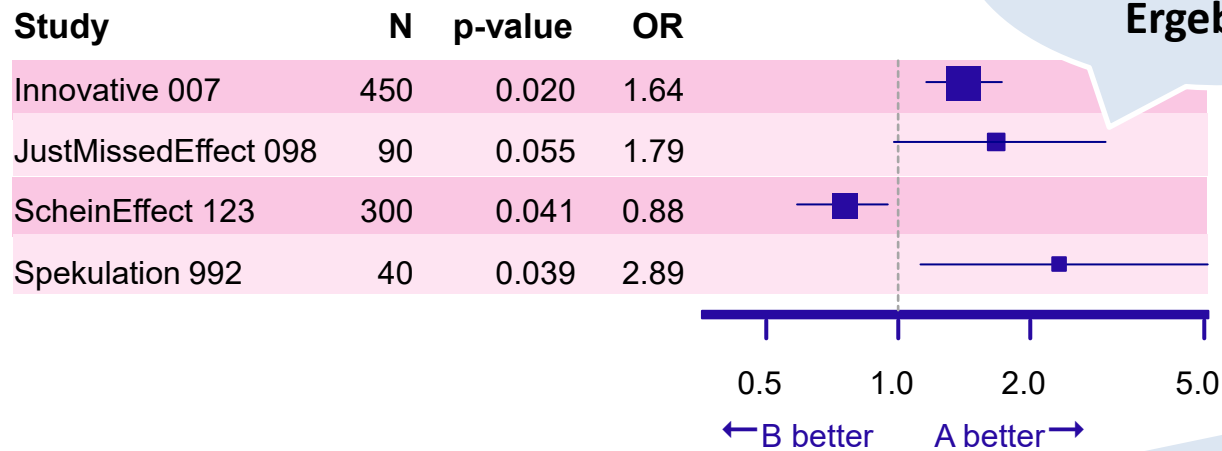
Wie groß war die Studienpopulation?

P-Werte allein berichten – nicht empfohlen (2)

Wie groß war die Studienpopulation?

Wie groß war der Effekt?

Welche Unsicherheit beinhaltet das Ergebnis?



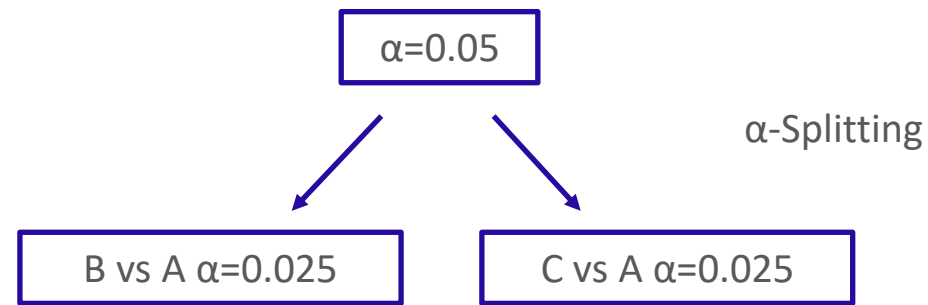
In welche Richtung geht der Effekt?

Multiple Endpunkte in klinischen Studien

Warum mehrere Endpunkte?

- Komplexe Krankheitsbilder erfordern mehrere Bewertungskriterien (z.B. Überleben, Lebensqualität, Sicherheit)
- Unterschiedliche regulatorische/interventionelle Ziele (FDA/EMA)
 - FDA bevorzugt oft OS
- Problem: Multiplizität erhöht Fehlerrate (Typ-I-Fehler) → Verfahren zur Kontrolle
 - Bonferroni(-Holm), Hochberg, Fixed Sequence Testing, Gatekeeping, Fallback-Verfahren (alpha-recycling)
- Regulatorische Anforderungen (EMA/FDA): Klare Spezifikation des Verfahrens im Studienprotokoll

Fiktives Beispiel, Bonferroni
3-Armige Studie, B vs A, C vs A



Supportive Analysen



Vanessa Schaser
GBG Forschungs GmbH



Univariate vs. Multivariate Analyse

Univariate Analyse

- Der Effekt **einer** Variable auf einen Endpunkt wird betrachtet
 - z.B. Effekt der Therapie auf die pathologische Komplettremission (pCR)

Multivariate Analyse

- Endpunkt wird analysiert unter Berücksichtigung **mehrerer** Variablen gleichzeitig
 - z.B. Effekt der Behandlung, der Tumorgröße, des Nodalstatus und des Subtypes auf das Erreichen einer pCR



Stratifikation in Studiendesign und Analyse

Stratifizierte Randomisierung:

- Teilnehmer werden vor der Zufallszuteilung in Untergruppen (Strata) nach wichtigen prognostischen Merkmalen (z.B. Nodalstatus, Vortherapien) aufgeteilt.
- Zweck: Sicherstellung der gleichmäßigen Verteilung dieser Merkmale in den Behandlungsarmen, um Verzerrungen zu minimieren.

Stratifikation in der Analyse:

- Umgang der Stratifikationsvariablen bei Protokollerstellung/Erstellung des Analyseplans definieren
- Spezielle Testverfahren (Mantel Haenszel, stratifizierter Log-Rank) oder Berücksichtigung der Strata als Variable im Modell



Subgruppenanalysen

Ziel: Überprüfung ob sich Behandlungseffekt in einzelnen Patientenpopulationen unterscheidet

- Mögliche SG: BL-Charakteristika (z.B. Alter, Geschlecht), Therapie (z.B. Dosis,...) oder Krankheitscharakteristika (z.B. Biologischer Subtyp,...)
- Wichtige Subgruppen können auch im α -Splitting aufgenommen werden

Herausforderungen:

- Relevante Subgruppen sollten präspezifiziert werden (Cherry-Picking vermeiden)
- Subgruppen können zu klein werden, um aussagekräftige Schätzer zu berechnen
- Erschwerte Interpretation aufgrund von gegensätzlichen Ergebnissen
- Multiples Testen
- Alle Subgruppenkategorien berichten, um vollumfängliches Bild zu erhalten
- Oft wenig Power, Durchführung von Meta-Analysen sinnvoll für Einschätzung eines Subgruppeneffekts
- Interaktions p-Werte können Einordnung/Interpretation unterstützen

Subgruppenanalysen (2)

Multiples Testen

- Bei Testung mehrere Subgruppen: potenziell signifikantes Ergebnis ist wahrscheinlich

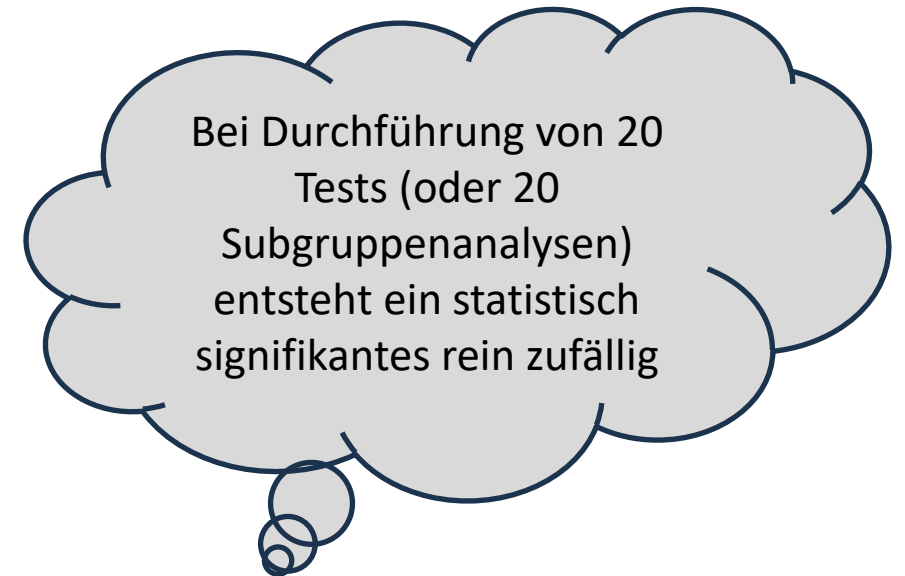
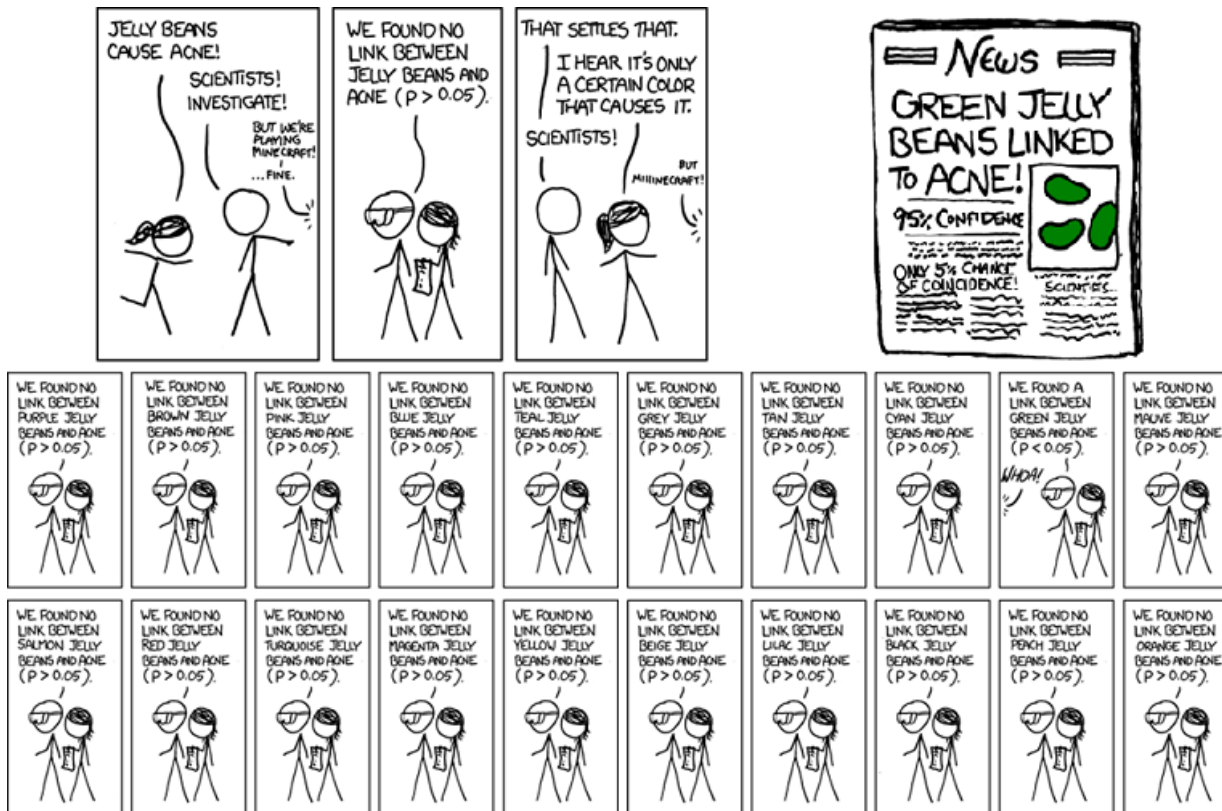


Figure from: [xkcd: Significant](#)

Auswertung binärer Endpunkte



Dr. Natalie Filmann
GBG Forschungs GmbH

Analyse des binären Endpunkts

- **Klassisches Beispiel:** Therapieansprechen, Vergleich zwischen zwei Behandlungsgruppen
- **Berichtet werden:** Raten mit Konfidenzintervallen, p-Werte
- **Testverfahren:** Chi-Quadrat-Test, exakter Fisher-Test, Mantel-Haenszel-Test (→ zur Berücksichtigung von Stratifikationsfaktoren)

	Arm A	Arm B
pCR	77 (38.5%)	54 (26.9%)
no pCR	123	147

→ 77/200 Patienten mit pCR in Arm A (38.5%, 95%-KI 31.7-45.6),
54/201 Patienten mit pCR in Arm B (27%, 95%-KI 20.9-33.6)

→ $p=0.0174$ (Chi-Quadrat-Test)

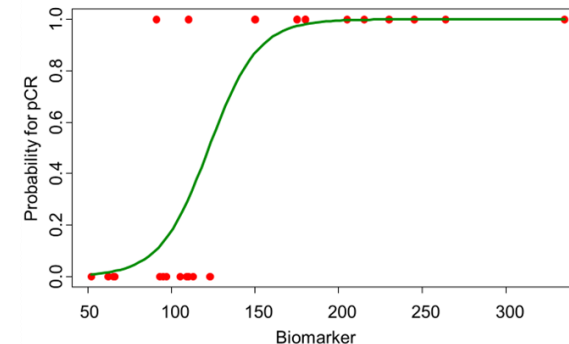
Fiktives Beispiel

Logistische Regression: Für binäre Zielgrößen

- **Logistische Regression:** Speziell für Zielvariablen mit nur zwei Kategorien (binär/dichotom)
- **Modelliert:** Die *Wahrscheinlichkeit* des Eintretens eines der beiden möglichen Ereignisse

→ (Fiktives) Beispiel: Analyse des Ansprechens auf eine Therapie:

„Wie stark beeinflusst ein bestimmter Biomarker die Wahrscheinlichkeit des Eintretens einer pCR?“



- **Testet:** Hat eine bestimmte Einflussgröße einen Effekt auf die Wahrscheinlichkeit des Eintretens eines der beiden Ereignisse?
- **Effektstärke:** Die Stärke des Einflusses wird durch das Odds Ratio (OR) quantifiziert
- **Multivariate Analysen:** Mehrere Einflussgrößen gleichzeitig können in ein Modell eingeschlossen werden

Was ist ein Odds Ratio?

- **Zentrales Maß:** Das Odds Ratio (OR) ist das zentrale Maß für die Effektstärke in der logistischen Regression.
- **Odds (Chance oder Wettverhältnis):** Verhältnis der 'Versuche', bei denen ein Ereignis eintritt, zu denen, bei denen es *nicht* eintritt (→ Wahrscheinlichkeit/Gegenwahrscheinlichkeit).
 - Beispiel: 30 von 100 zeigen Ereignis, 70 nicht = Odds von 30 zu 70 ≈ 0.43 .
 - Wichtig: Wahrscheinlichkeiten (0-1) und Odds (0-unendlich) sind nicht identisch!
- **Odds Ratio Definition:** Das Verhältnis der Odds, dass ein Ereignis in einer Gruppe A (z.B. Expositionsgruppe) eintritt, zu den Odds, dass dasselbe Ereignis in einer Gruppe B (z.B. Kontrollgruppe) eintritt.

	Arm A	Arm B
pCR	77 (38.5%)	54 (26.9%)
no pCR	123	147

- Odds für Arm A und B sind $77/123 \approx 0.63$ und $54/147 \approx 0.37$
- Odds Ratio ist $0.63/0.37 \approx 1.7$



Interpretation von OR-Werten

- **OR quantifiziert:** Um welchen Faktor sich die Chancen für ein Ereignis ändern, wenn eine unabhängige Variable um eine Einheit zunimmt oder von einer Referenzkategorie zu einer anderen gewechselt wird.
- **OR > 1:** Erhöhte Chance/Risiko für das Ereignis.
- **OR < 1:** Reduzierte Chance/Risiko; protektiver Faktor.
- **OR = 1:** Kein Unterschied in den Chancen; der Faktor hat keinen Einfluss.
- **Beispiele (fiktiv):**
 - **Alter (pro Jahr):** OR = 1.074. Für jedes zusätzliche Lebensjahr steigt die Chance, den Endpunkt zu erreichen, um das 1.074-fache (7.4% Erhöhung).
 - **Alter (>50 Jahre vs. <50 Jahre):** OR = 4.37. Die Chance den Endpunkt zu erreichen ist bei Patienten über 50 um den Faktor 4.37 größer.

Fiktives Beispiel: Logistische Regression

Variable	univariate		
	OR	95%-CI	p-value
Arm			
Control	Ref.		
Experimental	2.95	1.71-5.10	<0.001
Breast cancer subtype			
HER2-/HR+	Ref.		<0.001
TNBC	4.70	3.00-7.34	<0.001
HER2+	4.39	3.10-6.21	<0.001
Age, years			
<40	Ref.		
≥40	0.42	0.14-1.30	0.131
cN			
cN0	Ref.		
cN+	0.57	0.41-0.80	0.001
Grading			
G1-2	Ref.		
G3	1.54	1.14-2.07	0.004

- **Referenzgruppe**

- **Bedeutung des 95%-Konfidenzintervalls (KI):**

- Mit 95%iger Wahrscheinlichkeit liegt der wahre OR-Wert in dem 95% KI.



- **Statistische Signifikanz:**

- Ein OR ist signifikant, wenn das 95%-KI den Wert **1.0 nicht einschließt**. - Analog p-Wert ≤ 5%
- Liegt 1.0 im Intervall, ist der Effekt statistisch nicht signifikant.
- OR=0.42 (95%-KI: 0.14-1.30) ist nicht signifikant (1.0 liegt im KI).
- OR=2.95 (95%-KI: 1.71-5.10) ist signifikant (1.0 liegt nicht im KI).



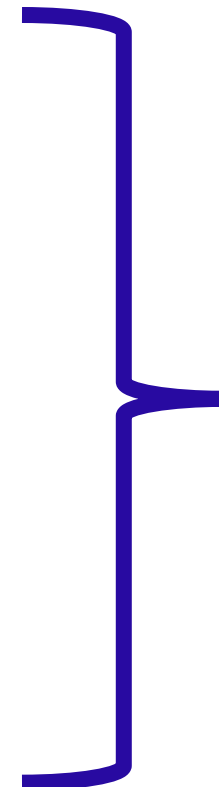
Fiktives Beispiel: Logistische Regression (2)

Variable	univariate		
	OR	95%-CI	p-value
Arm			
Control	Ref.		
Experimental	2.95	1.71-5.10	<0.001
Breast cancer subtype			
HER2-/HR+	Ref.		<0.001
TNBC	4.70	3.00-7.34	<0.001
HER2+	4.39	3.10-6.21	<0.001

 Hat der Subtyp generell Einfluss auf pCR?
 Vergleiche TNBC vs HER2-/HR+ (=Referenz) und
 HER2+ vs HER2-/HR+ (=Referenz)

Fiktives Beispiel: Uni vs. Multivariate Logistische Regression

Variable	univariate		
	OR	95%-CI	p-value
Arm			
Control	<i>Ref.</i>		
Experimental	2.95	1.71-5.10	<0.001
Breast cancer subtype			
HER2-/HR+	<i>Ref.</i>		<0.001
TNBC	4.70	3.00-7.34	<0.001
HER2+	4.39	3.10-6.21	<0.001
Age, years			
<40	<i>Ref.</i>		
≥40	0.42	0.14-1.30	0.131
cN			
cN0	<i>Ref.</i>		
cN+	0.57	0.41-0.80	0.001
Grading			
G1-2	<i>Ref.</i>		
G3	1.54	1.14-2.07	0.004



5 univariate log.
Regressionsmodelle

Fiktives Beispiel: Uni vs. Multivariate Logistische Regression (2)

Variable	univariate			multivariate		
	OR	95%-CI	p-value	OR	95%-CI	p-value
Arm						
Control	Ref.			Ref.		
Experimental	2.95	1.71-5.10	<0.001	2.48	1.35-4.55	0.003
Breast cancer subtype						
HER2-/HR+	Ref.		<0.001	Ref.		<0.001
TNBC	4.70	3.00-7.34	<0.001	4.32	2.85-6.56	<0.001
HER2+	4.39	3.10-6.21	<0.001	3.53	2.45-5.11	<0.001
Age, years						
<40	Ref.			Ref.		
≥40	0.42	0.14-1.30	0.131	0.59	0.17-2.05	0.401
cN						
cN0	Ref.			Ref.		
cN+	0.57	0.41-0.80	0.001	0.75	0.52-1.06	0.112
Grading						
G1-2	Ref.			Ref.		
G3	1.54	1.14-2.07	0.004	1.27	0.96-1.70	0.095

▪ Multivariate logistische Regression:

- Ein gemeinsames Modell für die genannten Einflussgrößen
- Odds Ratios sind adjustiert in Bezug auf die weiteren Variablen

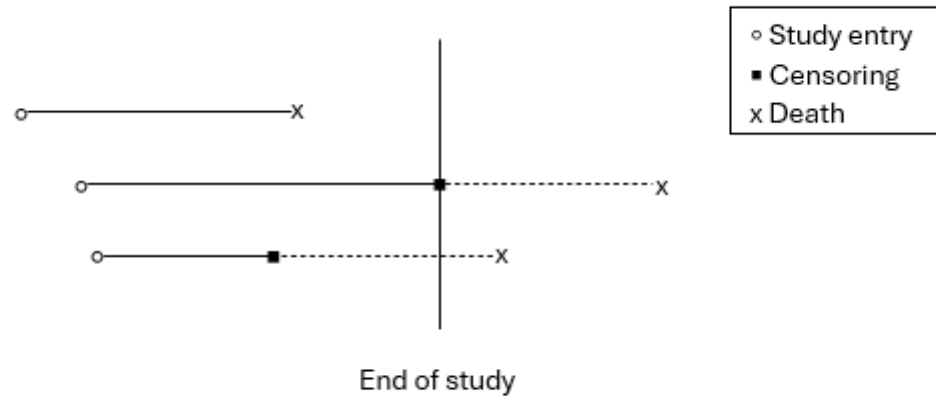
Auswertung von Überlebenszeitdaten



Vanessa Schaser
GBG Forschungs GmbH

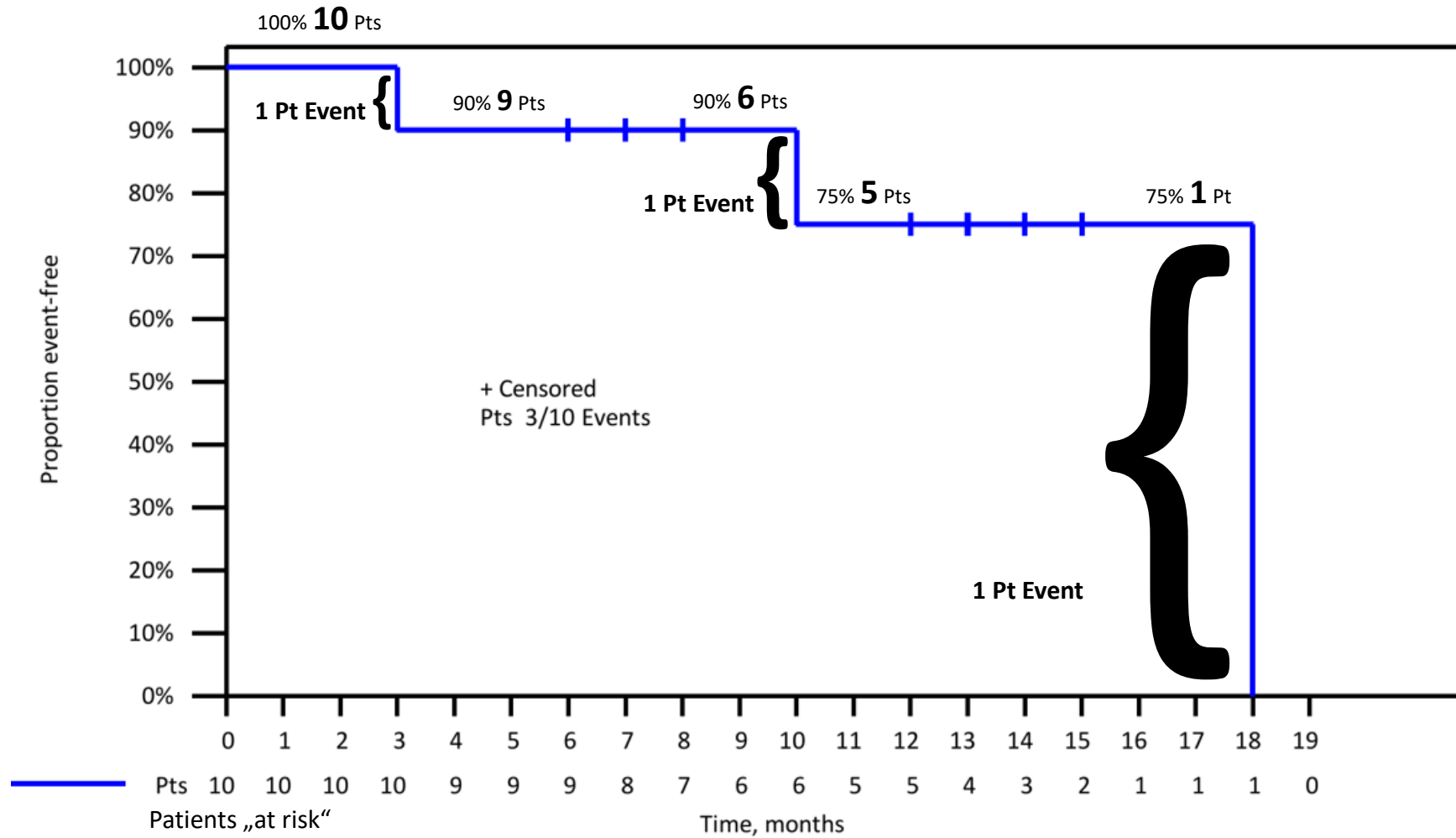
Überlebenszeitanalyse

Von Interesse: Zeit bis zum Auftreten eines Ereignisses, bspw. Tod, Rückfall, ...



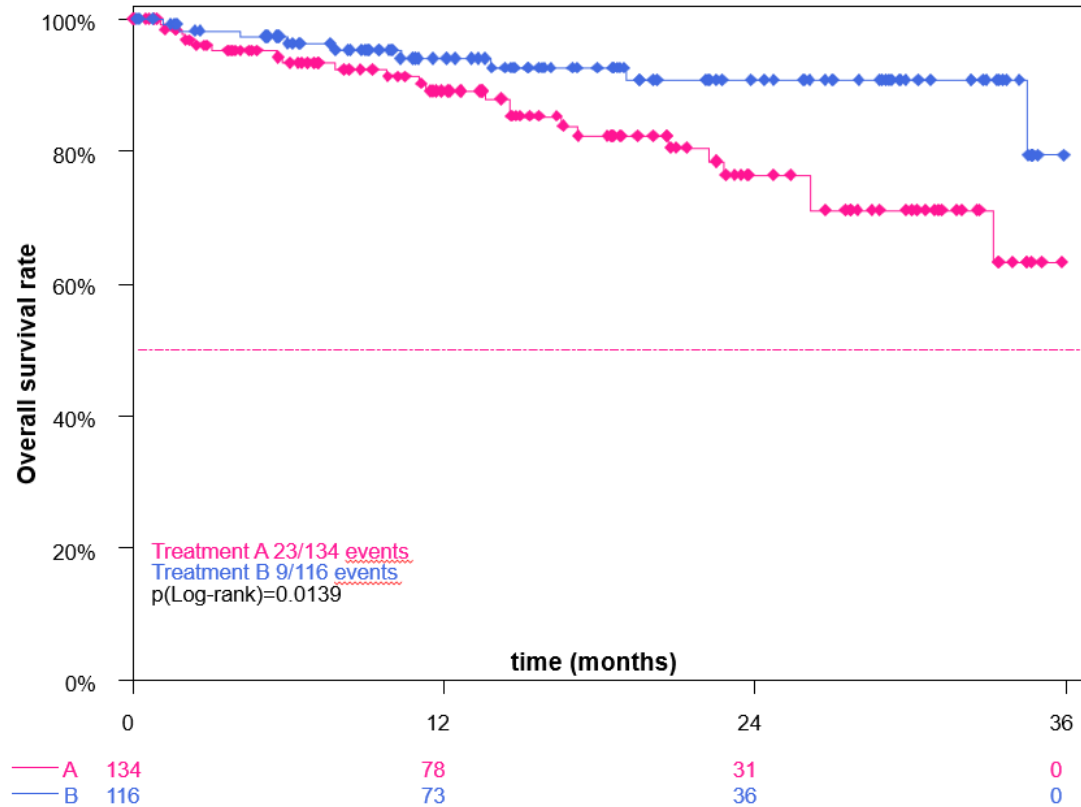
- Ereignis muss nicht zwingend für die gesamte Studienpopulation beobachtet werden (z.B. Patientin ist lost to follow-up, Drop-out, Ende der Studie, ...) – das zu beobachtende Ereignis wird als zensiert betrachtet
- Standardmethoden können nicht angewandt werden

Kaplan-Meier Kurve



Kaplan-Meier Kurve (2)

Beispiel



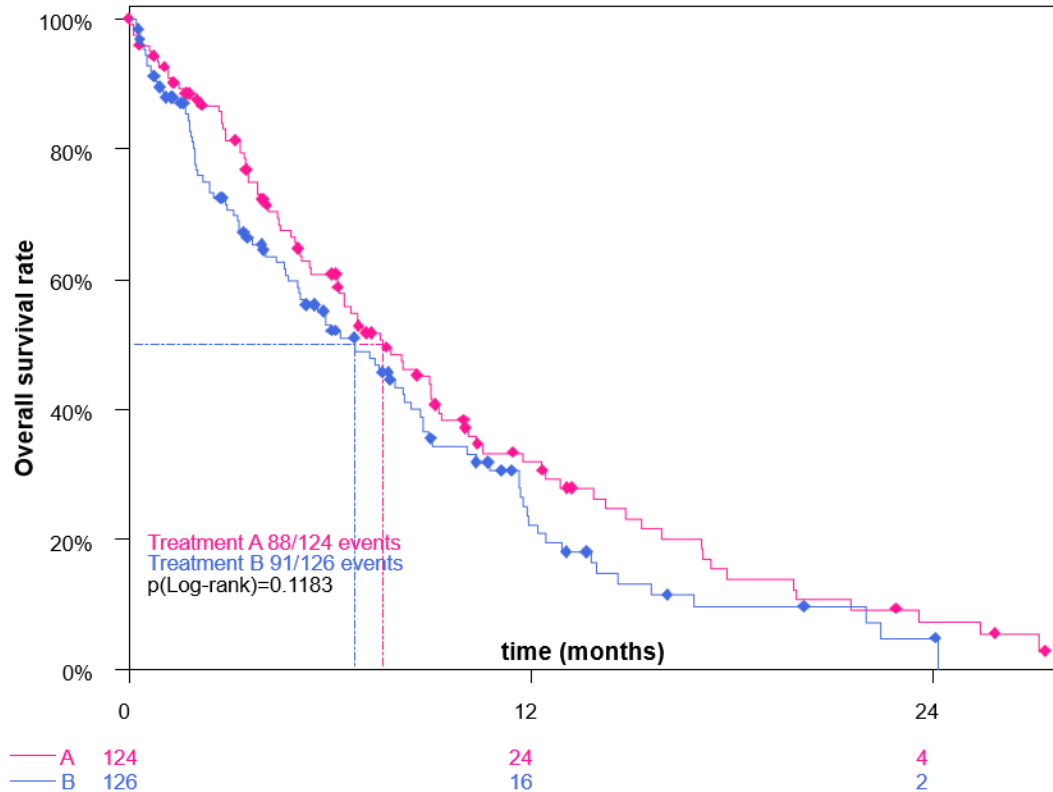
Plot based on simulated data

Interpretation?

- Median wurde in keinem der Arme erreicht
- Weniger OS-Ereignisse im Treatment B Arm
- Wahrscheinlichkeit bis zum 24. Monat kein OS-Ereignis entwickelt zu haben ist ca. 90% im Treatment B Arm und 75% im Treatment A Arm

Kaplan-Meier Kurve (3)

Beispiel



Plot based on simulated data

Interpretation?

- Medianes Überleben nach ca. 7 Monaten für Treatment B und nach ca. 8 Monaten für Treatment A erreicht
- Weniger OS- Ereignisse im Treatment Arm A



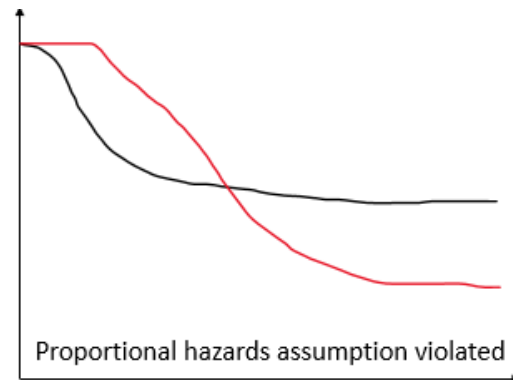
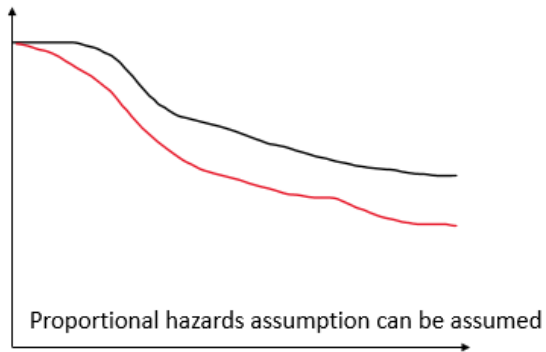
Cox Regression

- Von Interesse: wie groß ist der beobachtete Unterschied zwischen den Armen
- Cox proportional hazards model wird verwendet um diesen Effekt durch **Hazard Ratio** zu schätzen
- Ergebnisse können hinsichtlich der Effektstärke interpretiert werden
 - **HR > 1: erhöhtes Risiko für Eintreten eines Ereignisses**
 - **HR < 1: reduziertes Risiko für Eintreten eines Ereignisses**
 - **HR = 1: kein Effekt**
- Einfluss einer oder mehrerer Variablen auf Zielgröße kann betrachtet werden

Log-Rank vs. Cox-Regression

Gemeinsamkeiten

- Annahme der Proportionalen Hazards

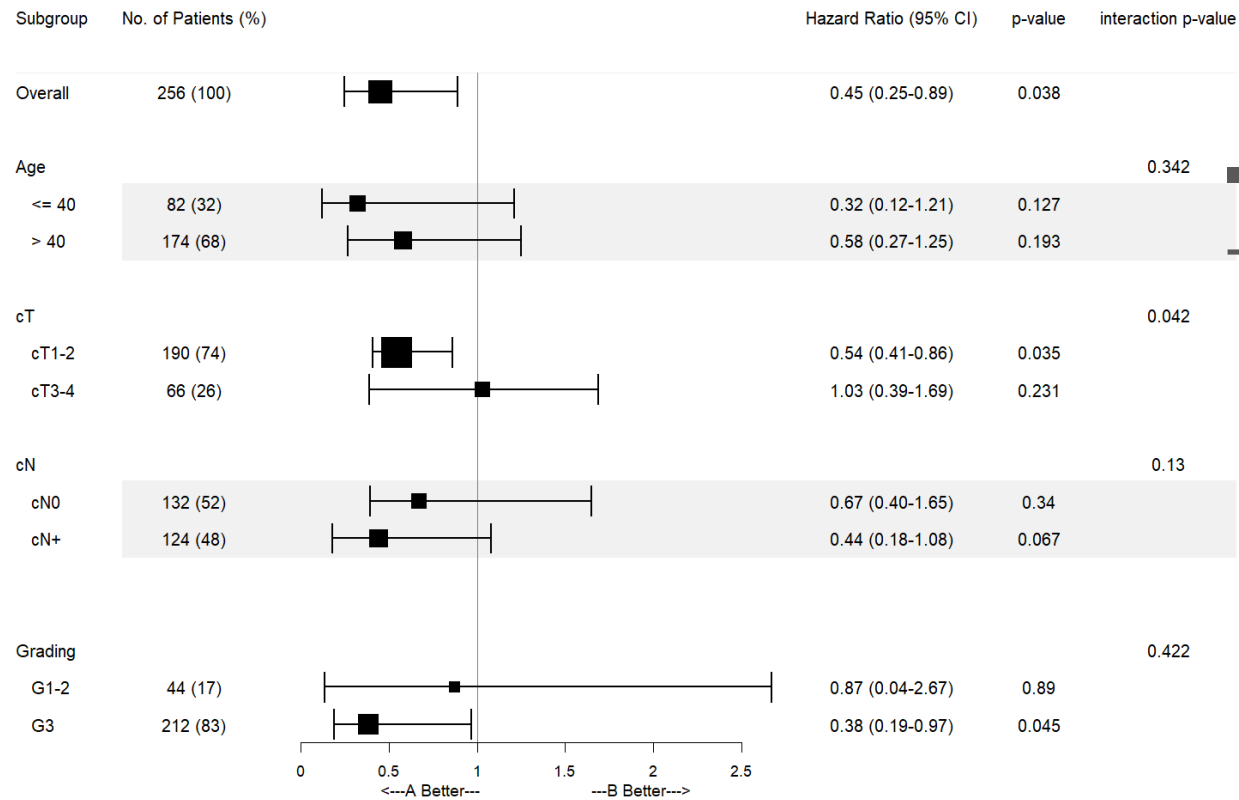


Unterschiede

Log-rank	Cox-regression
Nur p-Wert als Ergebnis	Effekt (HR) mit 95% CI und p-Wert für diesen Effekt
Kovariaten können nicht eingeschlossen werden	Kovariaten können eingeschlossen werden
Vergleich von Überlebenszeitkurven	Testet ob bestimmte Variable(n) Effekt auf Überleben haben

Cox Regression (2)

Beispiel SG Analyse



- Für jede SG-Kategorie wird Schätzer aus einem separaten Modell berechnet (d.h. insgesamt 8 Tests für HR und 4 Tests für Interaktion)
- Interaktions p-Wert für cT signifikant
→ Potenziell unterschiedlicher Effekt für Patientinnen mit cT1-2 und cT3-4

Fictive example – for illustrative purposes only

Q&A Session



Vielen Dank!

Die Slides erhalten Sie im Nachgang
über unsere GBG Library!