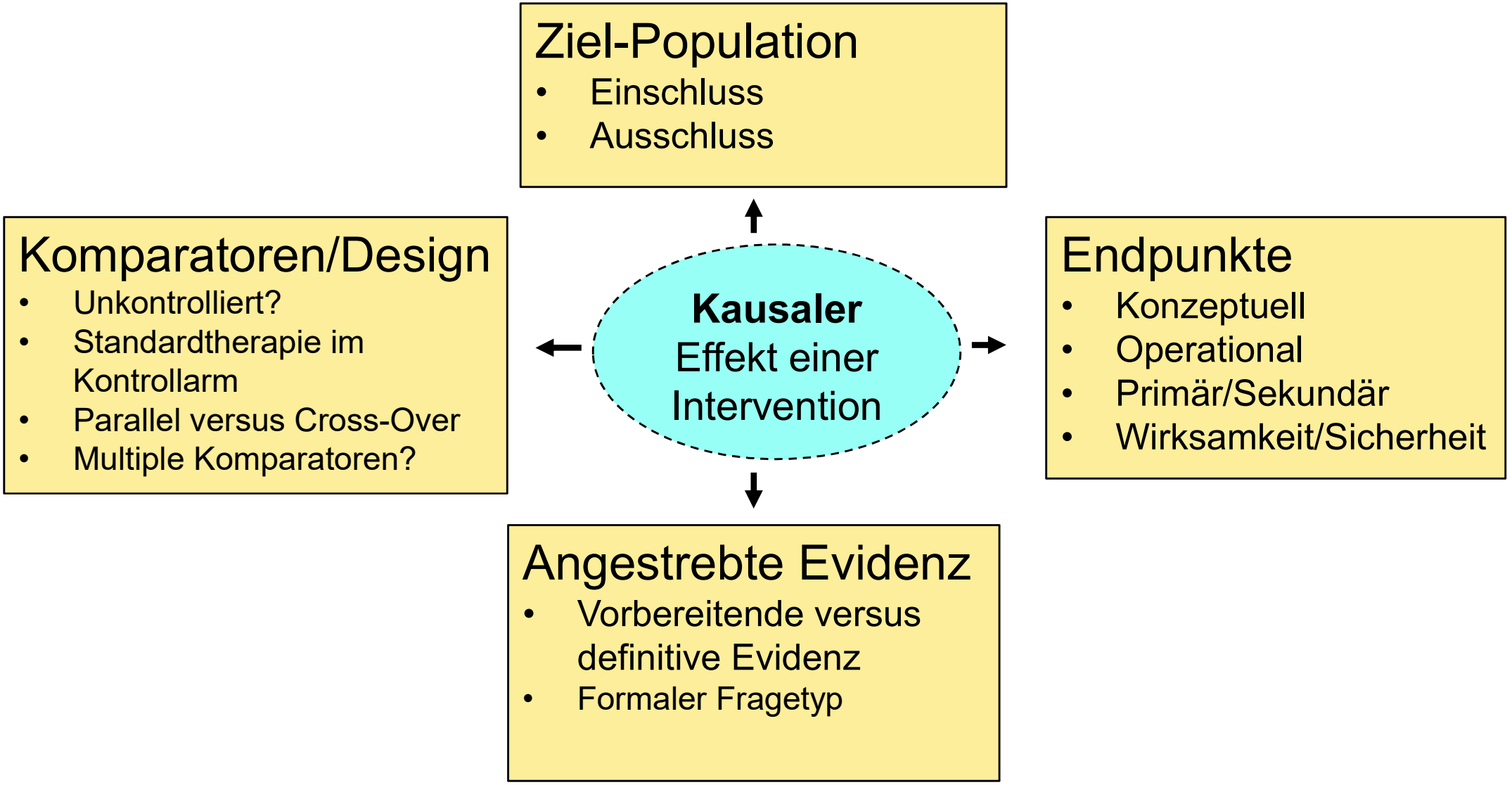
Endpunkte

Dr. Dirk Hasenclever
IMISE

Entfaltung einer Studienidee: Kausaler Effekt einer Intervention



Wahl eines Endpunkts

- **Jeder Studie sollte eine klare Fragestellung zu Grunde liegen.**

- **Zur Spezifikation der Studienfrage gehört**
  - **Klar beschriebener und strikt kontrollierter Eingriff**
  - **Studiendesign geeignet zur Untersuchung der interessierenden Kontraste**
  - **Wohldefinierter Endpunkt**

- **Endpunkt (=Zielkriterium)** ist ein zur Beantwortung der Studienfrage **erhobenes operational erhebbares Kriterium**,
  - welches für jeden Patienten bestimmt wird und
  - an dem der Erfolg des Eingriffs festgemacht wird.

Primäre und sekundäre Endpunkte

- Jede Studie sollte einen **primären Endpunkt** haben:
  - Ergibt sich aus der Frage, die am meisten interessiert
  - Muss durch Studiendesign und Fallzahl adäquat beantwortbar sein

- Daneben können zusätzlich **sekundäre Endpunkte** betrachtet werden:

  z.B.
  - Weitere Aspekte zur Wirksamkeit
  - Sicherheit, Nebenwirkungen, Komplikationen o.ä.

Mehr als einen primären Endpunkt?

- Efficacy und Safety Endpunkte adressieren verschiedene Aspekte. In der Regel keine Adjustierung nötig.
- Manchmal gibt es zwei primäre Endpunkte.
  - Hierarchisch geordnet: Der zweite interessiert nur, wenn der erste signifikant ist. Keine Adjustierung.
  - Gleichberechtigt -> Adjustierung nötig. Z.B:

The Hochberg method will be used to adjust for endpoint multiplicity: If both p-values are below 0.05 we will claim efficacy in both primary endpoints; if otherwise the smallest p-value is below 0.025 we will claim efficacy in the respective primary endpoint. This procedure controls the family-wise error rate (FWER) in the strong sense at a two-sided significance level of 5%. (Hochberg 1988). The Hochberg procedure seems appropriate since we expect non-negatively correlated test-statistics.

Wahl eines Endpunkts: was heißt „besser“ ?

- **Konkurrierende Optimalitätskriterien**

- **Ethisch** wenig invasiv, wenig Zusatzbelastung für Patienten
- **Klinisch** Unterscheide (nur), was für den Patienten einen Unterschied macht
- **Statistisch** möglichst informative Skala:
  binär < ordinal < kontinuierlich
- **Biometrisch** verlässlich messbar: **objektiv und reproduzierbar**
- **Logistisch** lückenlos messbar: **einfach, billig, mit geringem Aufwand**

Krankheitspezifische Endpunkte

- Für viele Krankheiten sind spezifische **Endpunkte bereits etabliert** und ihre **biometrischen Eigenschaften bekannt**.
  (-> EMA Seiten, Literatur)

- In anderen Situationen muss ein **geeigneter Endpunkt erst biometrisch konstruiert und validiert werden**.

Endpunkt - Spezifikation

- **Konzeptueller Endpunkt**
  - Was soll erfasst werden? Klartext!
- **Messverfahren**
  - Welche Datengrundlage
  - Operationale Definition für den einzelnen Patienten
  - **Misst das Messverfahren wirklich, was es messen soll**?
  - Zuverlässig erhebbar? Fehlwerte zu erwarten?
- **Zusammenfassende Statistik** für Endpunkt **in einer Population**
- **Zu schätzendes Differenzmaß**

Bsp: Vergleich roboterassistierte oder konventionelle laparoskopische radikale Prostatektomie

- **Konzeptueller Endpunkt** - Klartext?
  - Stabile (natürliche) Kontinenz nach OP
  - Keine Alltagsbeeinträchtigung durch Inkontinenzproblematik
- **Messverfahren**
  - Datengrundlage: Patiententagebuch, Einlagenverbrauch
  - Operationale Definitionen: Kontinenz dokumentiert falls:
    - Kein Einlagengebrauch über eine Woche / 3 Tage / 1 Tag
    - Nur Sicherheitspads über eine Woche / 3 Tage / 1 Tag
- **Zusammenfassende Statistik** für eine Population
  - Kontinenzrate nach 12 Wochen.
- **Zu schätzendes Differenzmaß**
  - Differenz in Wahrscheinlichkeit

Details sind wichtig

Wie hätten Sie es denn gerne?	
	Total
Keine Pads über drei Tage	**33%**
Keine Pads einmalig	**37%**
Höchstens Sicherheitspad über drei Tage	**45%**
Höchstens Sicherheitspad einmalig	**52%**

Einfache Endpunkte

- Erfolgsraten

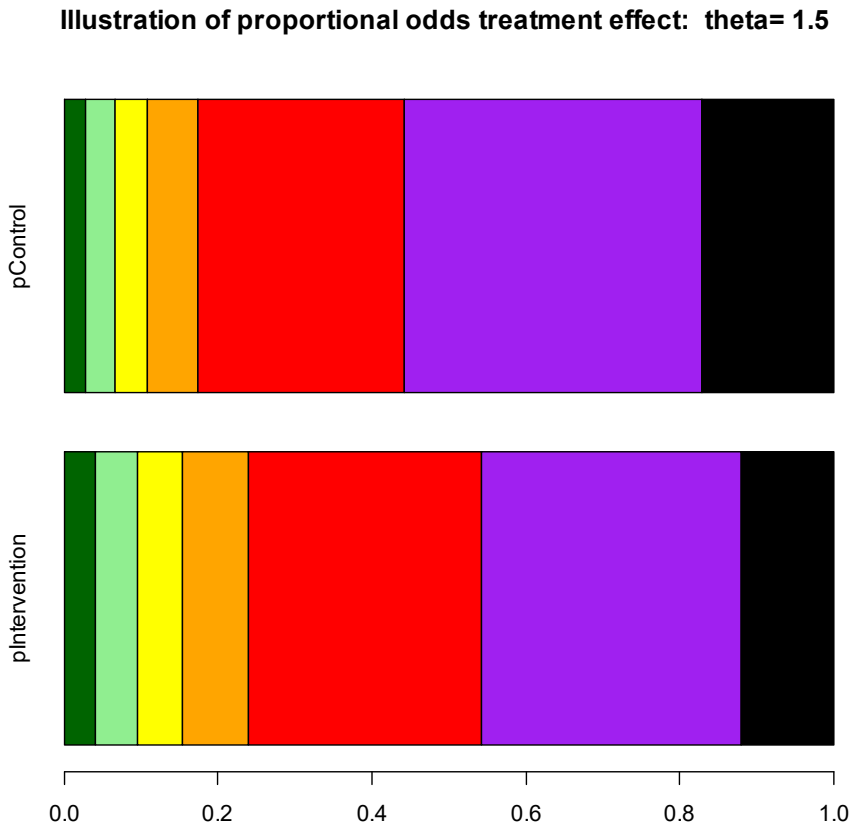
- Reduktion von stetigen Messgrößen
  (Blutdruck, Leberwerte etc.)
  - Welche Differenz ist klinisch relevant?

- Ordinale Scores z.B.:
  - Modifizierte Rankin-Skala
  - Glasgow Coma Score
  - Young Mania Rating Scale

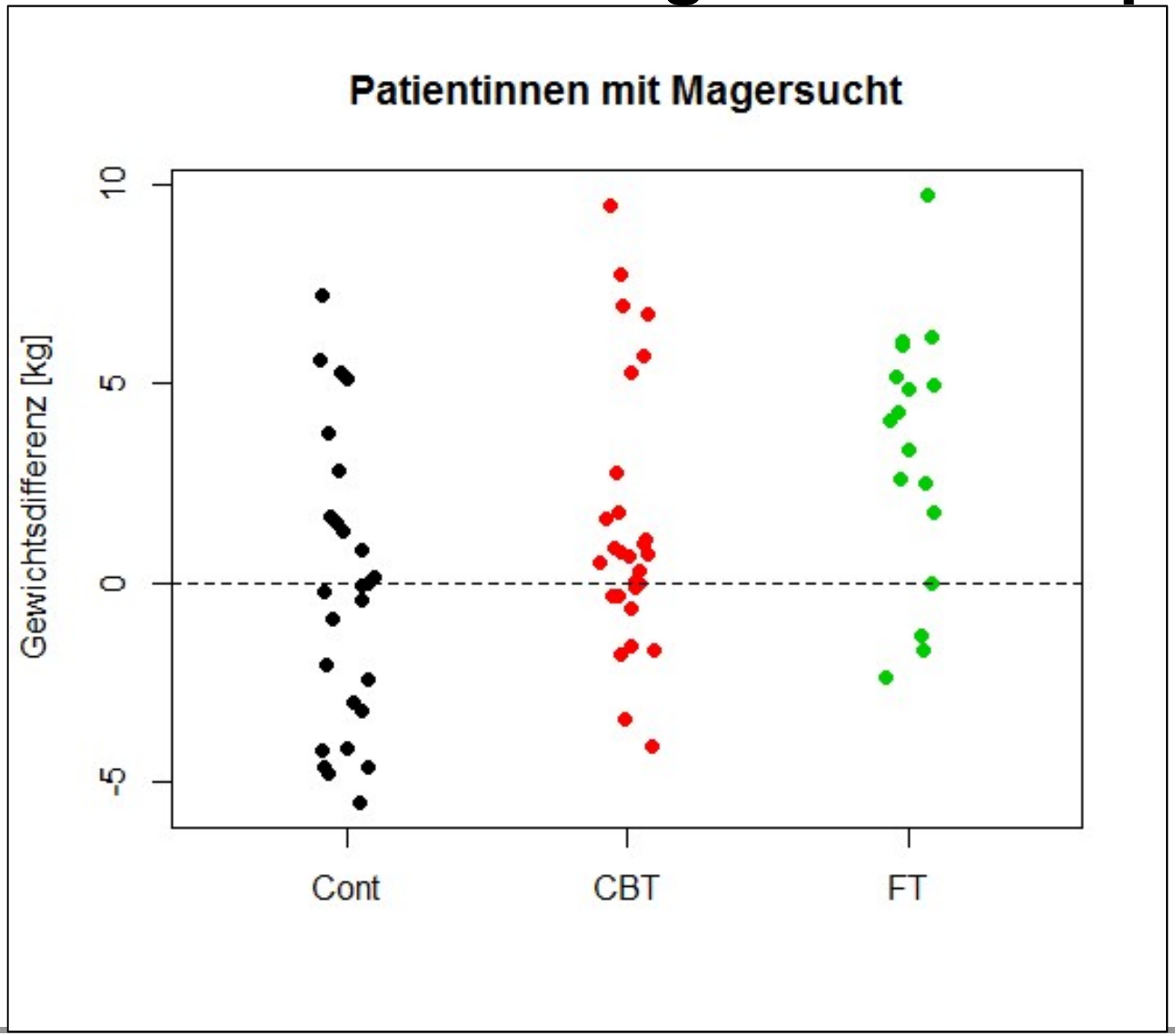
modifizierte Rankin-Skala

- Ausmaß der Behinderung nach einem Schlaganfall
  - 0 - Keine Symptome.
  - 1 - Keine relevante Beeinträchtigung. Kann trotz gewisser Symptome Alltagsaktivitäten verrichten.
  - 2 - Leichte Beeinträchtigung. Kann sich ohne Hilfe versorgen, ist aber im Alltag eingeschränkt.
  - 3 - Mittelschwere Beeinträchtigung. Benötigt Hilfe im Alltag, kann aber ohne Hilfe gehen.
  - 4 - Höhergradige Beeinträchtigung. Benötigt Hilfe bei der Körperpflege, kann nicht ohne Hilfe gehen.
  - 5 - Schwere Behinderung. Bettlägerig, inkontinent, benötigt ständige pflegerische Hilfe.
  - 6 - Tod in Folge des Apoplex.

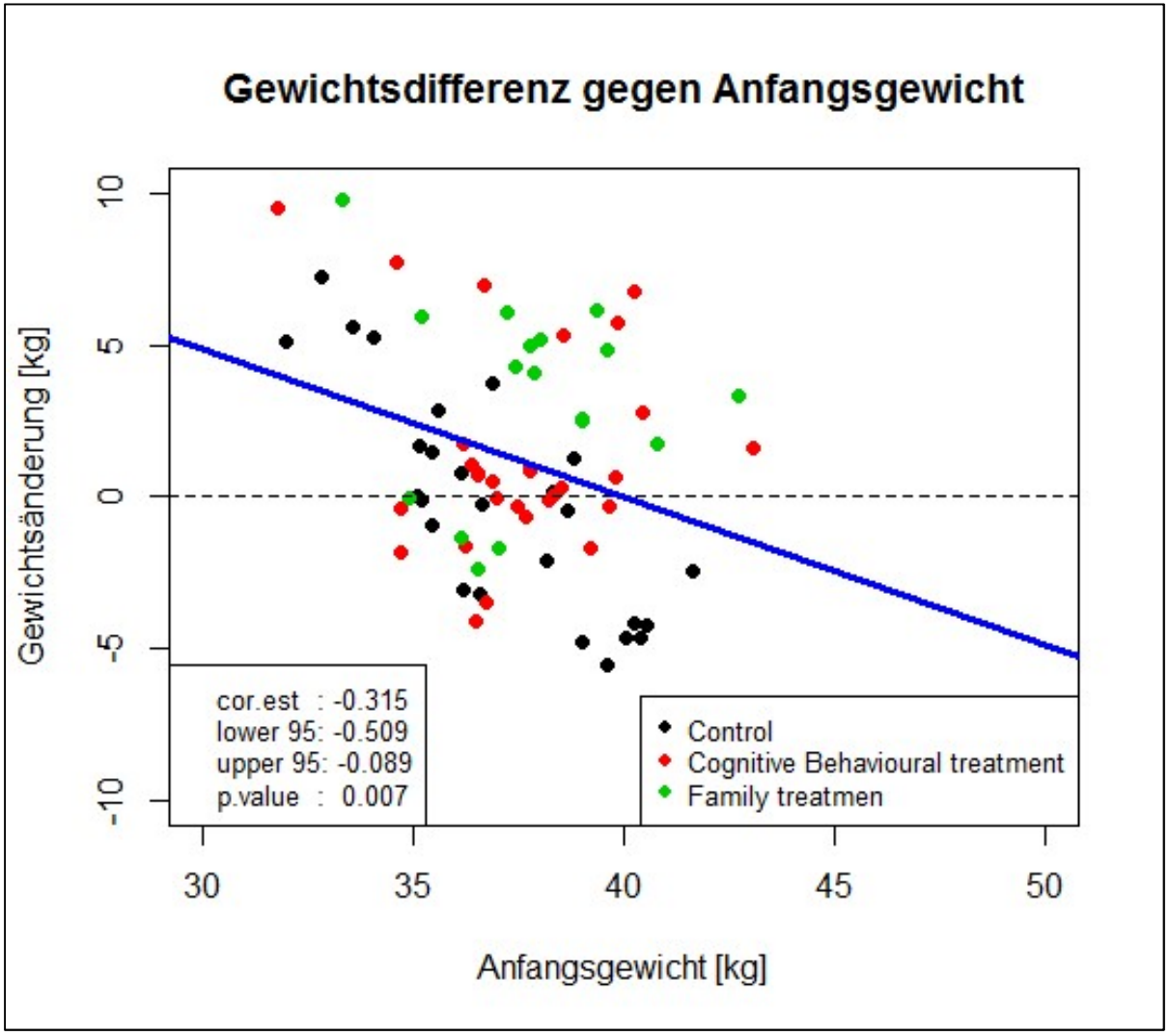
Planungsannahmen für TOMERAS Studie



Vorher-Nachher Daten
Gewichtsveränderung nach Therapieform



Differenzen: Endgewicht-Anfangsgewicht



Qua Konstruktion ist klar, dass eine Korrelation besteht!

“Regression to the mean”

Vergleich mit T-Tests auf Differenzen

```
Welch Two Sample t-test    KONTROLLE gegen CBT

data:  (NACH - VOR)[idx] by THERAPIE[idx]
t = -1.6677, df = 50.971, p-value = 0.1015
alternative hypothesis: true difference in means is not equal to
0
95 percent confidence interval:
  -3.4555870  0.3195395
sample estimates:
  mean in group Cont  mean in group CBT
-0.2041168           1.3639070
```

Bessere Datenbeschreibung und höhere statistische Power mittels Regressionsmodell

Modellansatz

$$Nachher = const + \beta \cdot Vorher + \gamma \cdot Therapieindikator + Fehlerterm$$

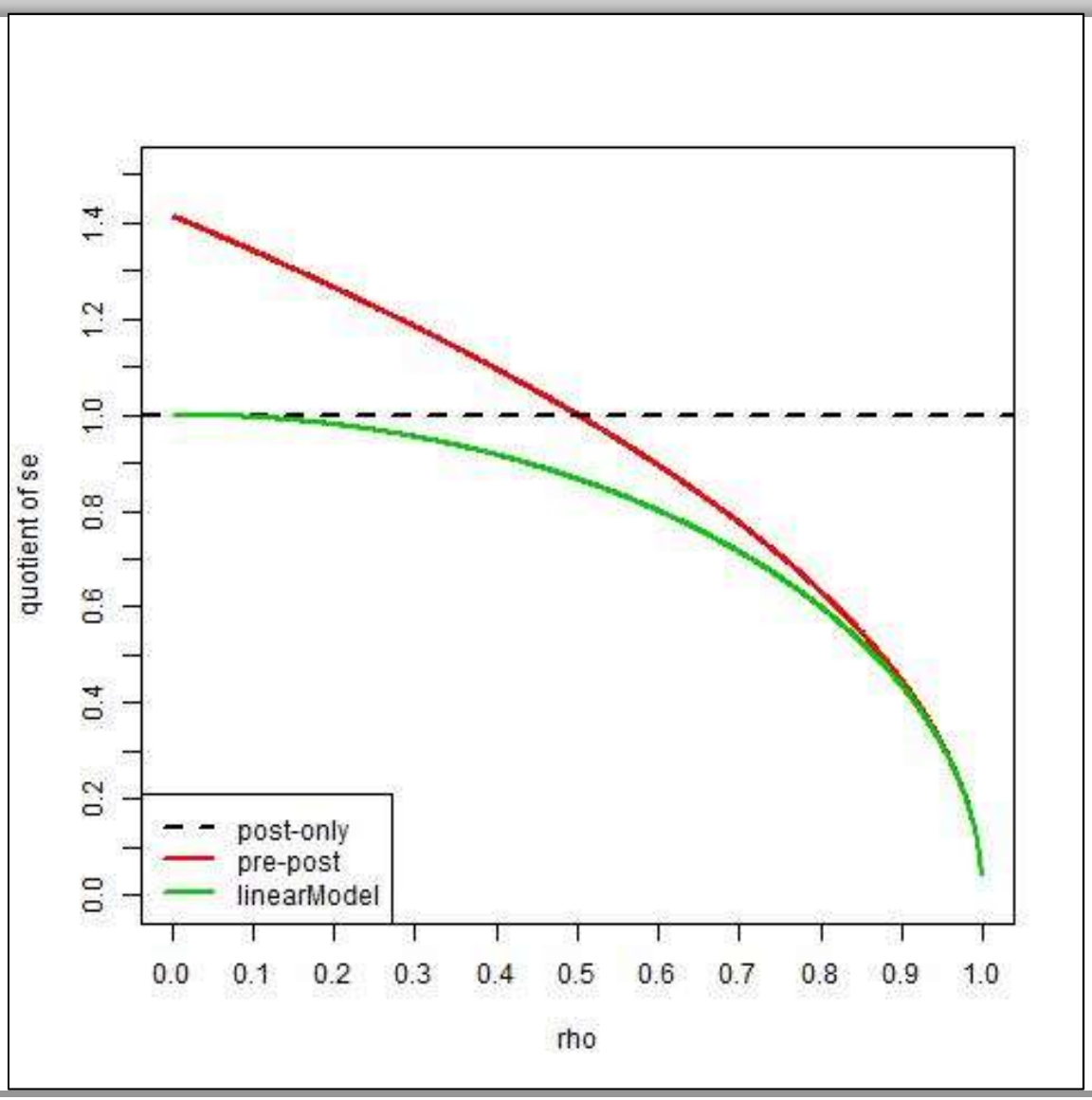
```
> summary(glm(NACH~VOR+THERAPIE,family="gaussian"))
…

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         20.7174    5.9950   3.456   0.000950 ***
  VOR               0.4345     0.1612   2.695   0.008850 **
  THERAPIECBT       1.8584     0.8589   2.164   0.033999 *
  THERAPIEFT        3.9282     0.9948   3.949   0.000189 ***

(Dispersion parameter for gaussian family taken to be 10.01884)

…
```

Analyse NACH?, (NACH-VOR)?
Power always better using linear model



Onkologie (kurativ)

- **Time to event Endpunkte**
  - **Overall Survival (OSV):** Zeit von Studienbeginn bis zum Tod
  - **Progression Free Survival (PFS):** Zeit von Studienbeginn bis zum ersten der folgenden Ereignisse:
    - Progress/Rezidiv
    - Tod
  - **Event Free Survival (PFS):** Zeit von Studienbeginn bis zum ersten der folgenden Ereignisse:
    - Progress/Rezidiv
    - Zweitneoplasie
    - Tod
- **Vorsicht: Viele Varianten**

Wahl eines geeigneten Endpunkts I

- **Beispiel**
  Abhängigkeit des Vergleichs zweier Gruppen vom gewählten Endpunkt

- **Fortgeschrittener Morbus Hodgkin (HD)**
  - Rate kompletter Remissionen: **ca. 75%**
  - Dann im weiteren Verlauf: **25-30 % Rezidive**

- **Frage**
  Hochdosis-Chemotherapie mit autologer Stammzelltransplantation als Konsolidierung einer konventionell erreichten kompletten Remission?

- **Analyse: Matched pairs**
  - Verläufe von 56 Patienten mit Hochdosiskonsolidierung aus europäischen Kliniken gematched mit
  - passenden, konventionell behandelten Patienten aus Daten der Deutschen Hodgkin Studiengruppe

Schmitz N, Hasenclever D, Brosteanu O et al, Blood 86,10 suppl. 1,1995

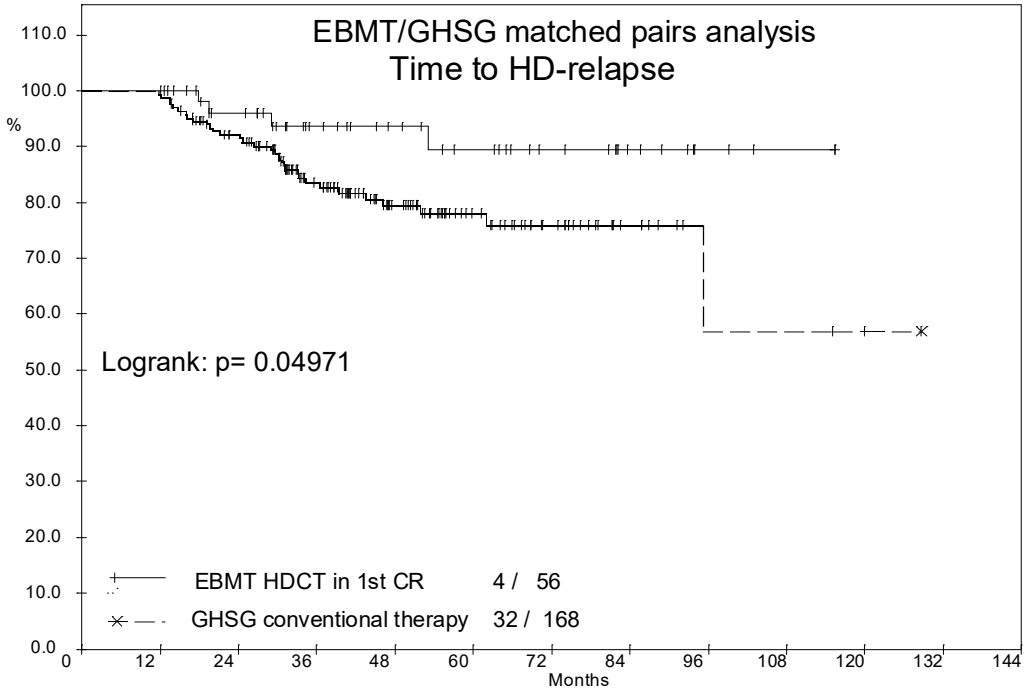
Wahl eines geeigneten Endpunkts II

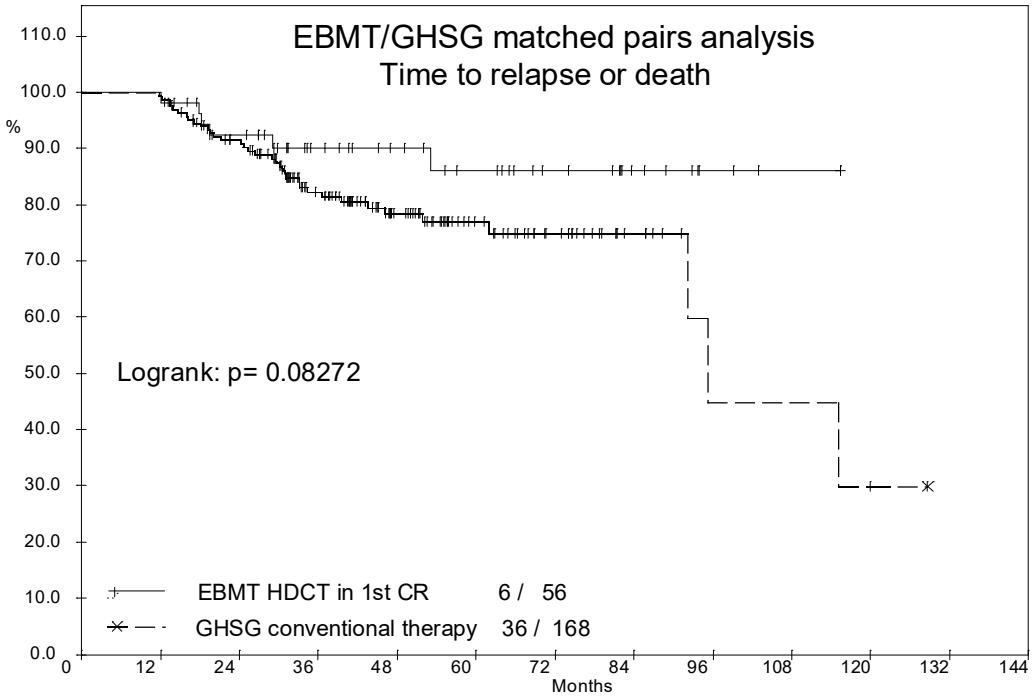
Endpunkt:

- HD-Rezidiv (Tumorkontrolle)

Endpunkt:

- HD-Rezidiv oder
- Tod jeglicher Ursache (insbesondere therapiebedingt)





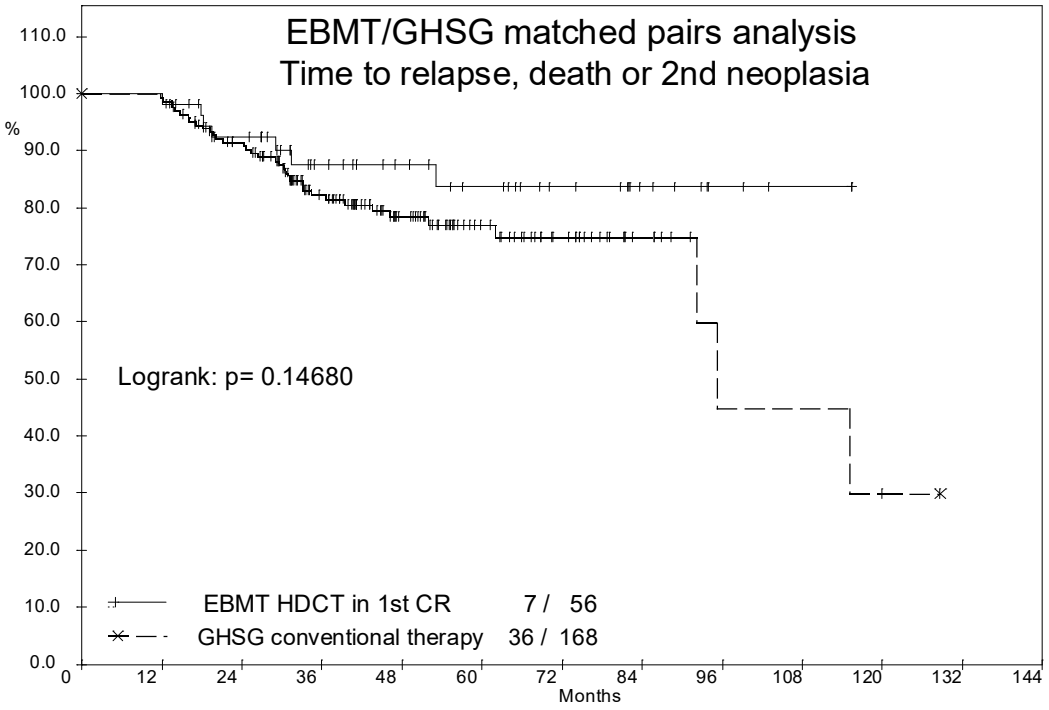
Wahl eines geeigneten Endpunkts III

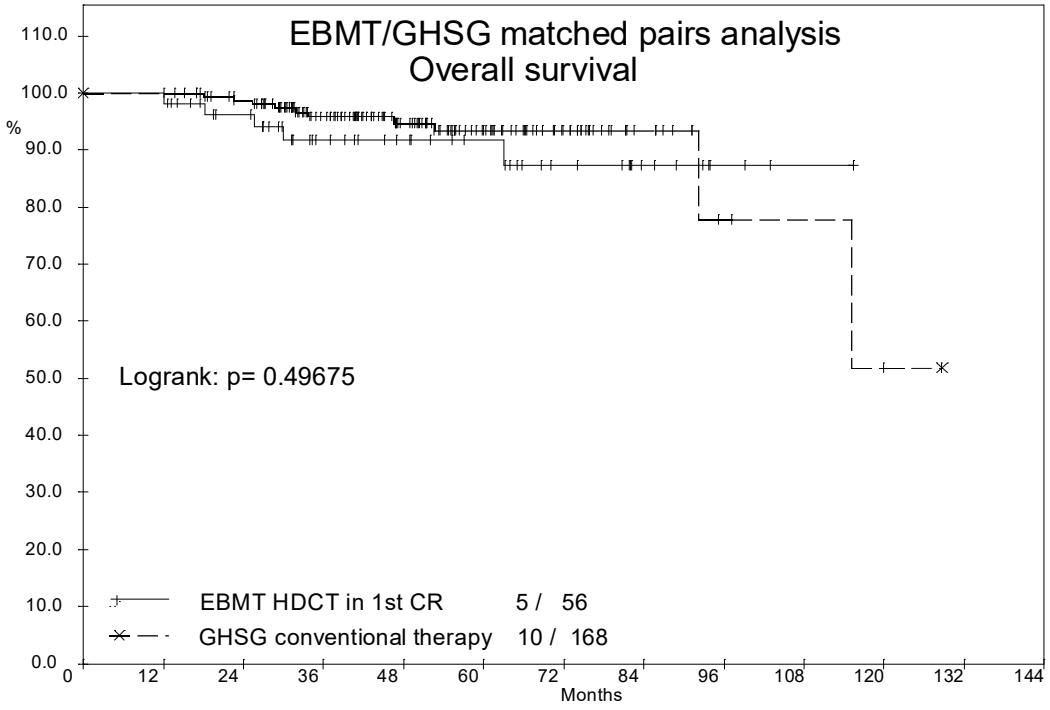
Endpunkt:

- HD-Rezidiv oder
- Tod oder
- Sekundärneoplasie

Endpunkt:

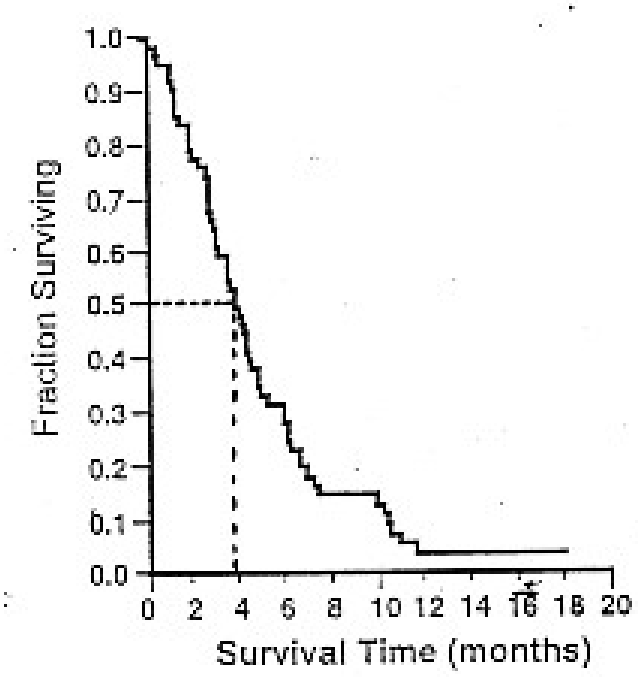
- Tod



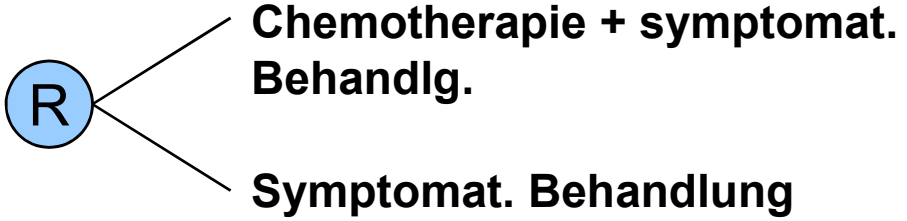


Palliative Therapie I

Beispiel:
Pankreas-Karzinom



Medianes Überleben: 4 –6 Monate



- Mögliche Endpunkte
  - Therapie-Ansprechen (Tumorregression)
  - Verlängerung der Überlebenszeit
  - Lebensqualität
  - Krankenhausfreie Zeit
  - ??

Palliative Therapie II

Möglicher Endpunkt:

Objektive klinisch relevante Symptombesserung

anhaltend über > 1 Monat

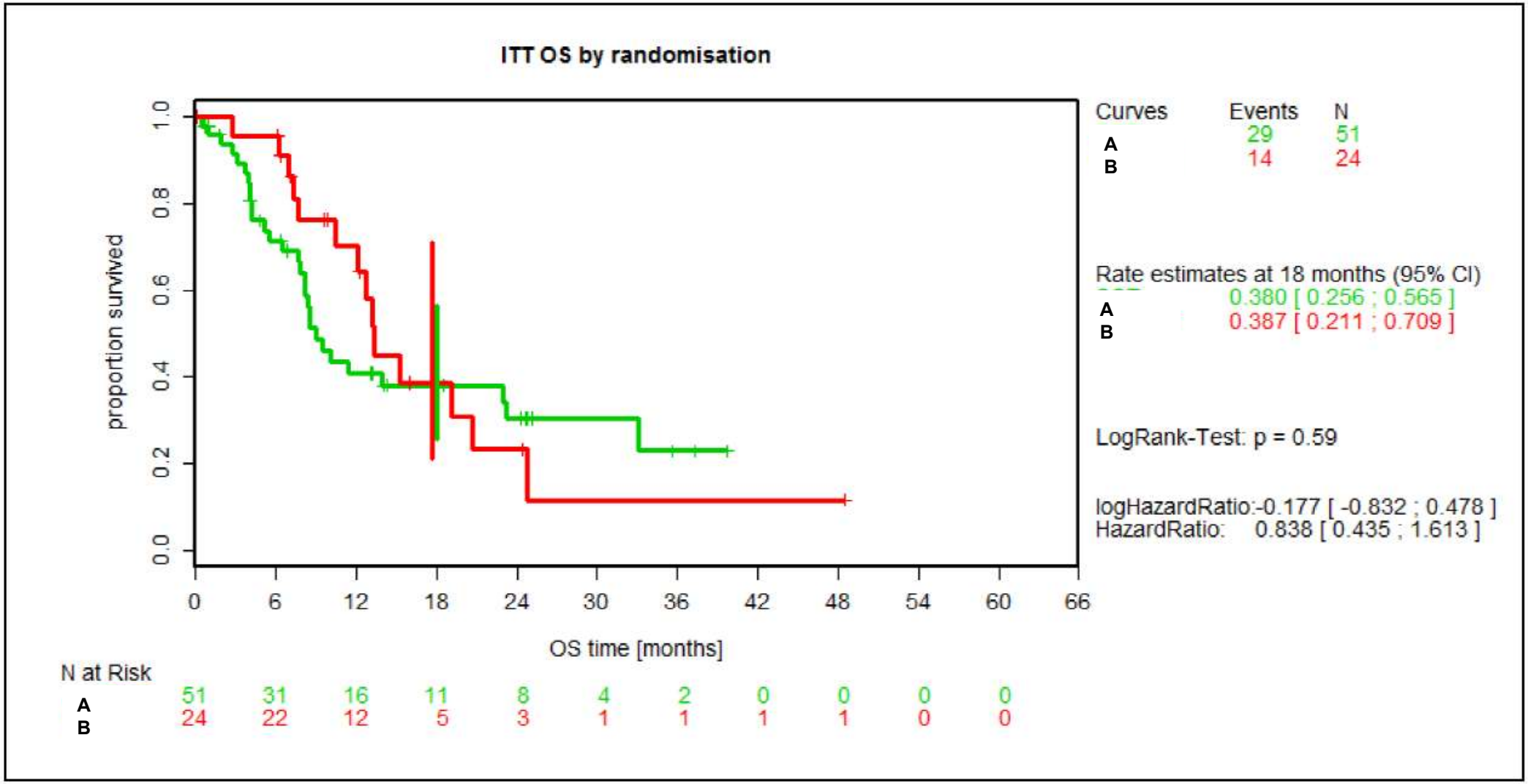
- **Schmerz:** a) ≥ 50% besser auf visueller Analog-Skala
  b) ≤ 50% Schmerzmittelverbrauch
- **Auszehrung:** Gewicht + 7%
- **Allgemeinzustand**: Karnofsky-Index + 20%

Klinischer Gewinn:

Keiner dieser Parameter ist schlechter
und
mindestens einer ist besser !

Burris HA, J Clin Oncol 15:2403-2413,1997

Application of Restricted Mean Survival



Restricted Mean Survival Methodology

Royston and Parmar (2011) have proposed a new 'standard' method of analysis which can be applied when proportional hazards fail and which makes the notion of dependence on the time horizon explicit. Instead of using the log hazard ratio as a measure of difference (which is ill defined without hazard proportionality), they propose to use the difference in **mean survival time within a specified time horizon**.

The restricted mean survival is estimated as the area under the survival curve integrated up to a given time horizon. Confidence intervals for both point estimates and differences in restricted mean survival can be obtained using a recently improved method (Andersen and Perme 2010). Corresponding tests can be used in clinical trials but require specification of a time horizon.

The restricted mean survival is directly interpretable and can be communicated to a patient. E.g. "Your life expectancy over the next two years with treatment A is 11 months; choosing treatment B increases your life expectancy by two months." (Royston & Parmar 2011).

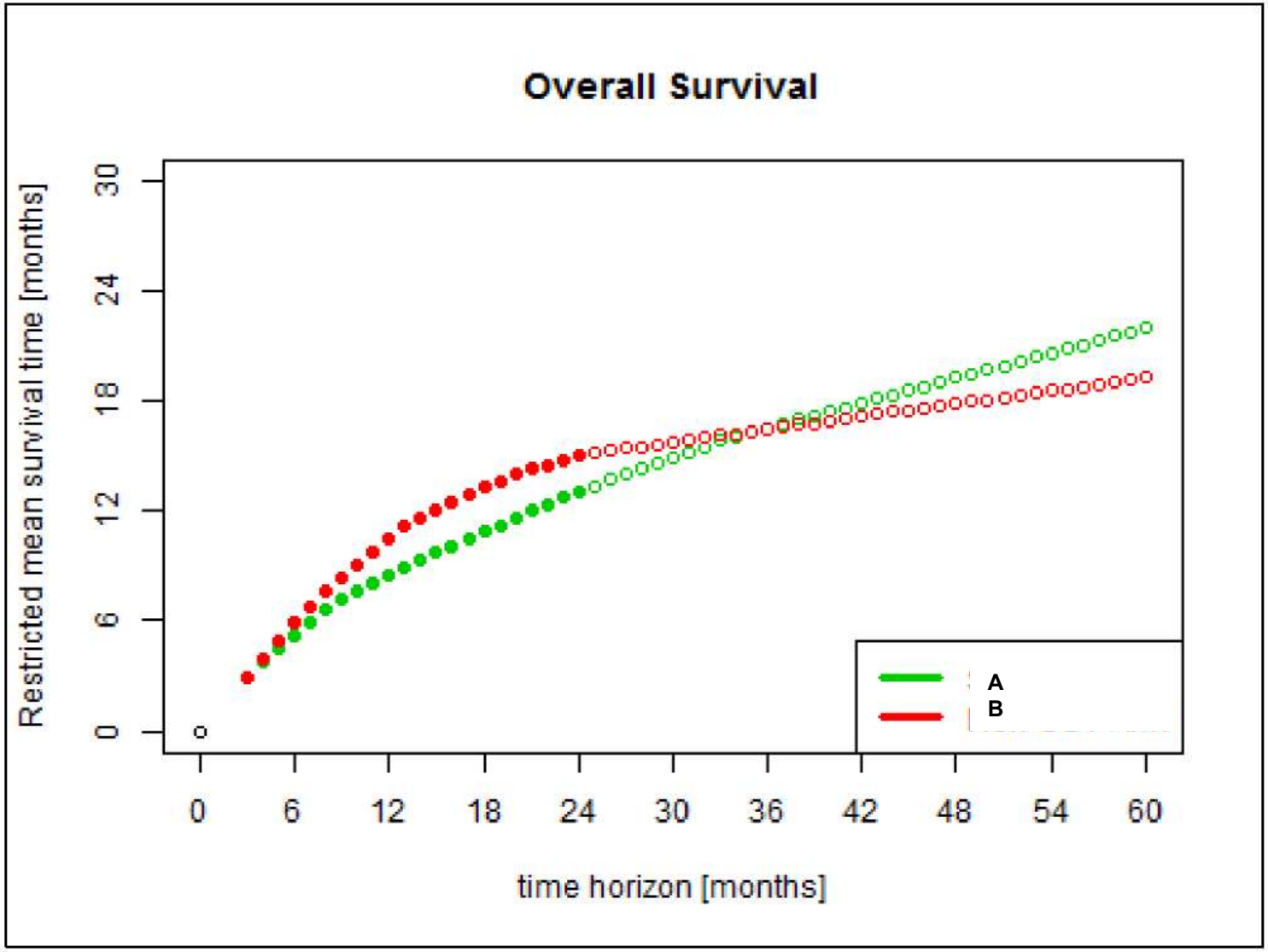
The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt

Patrick Royston*† and Mahesh K. B. Parmar

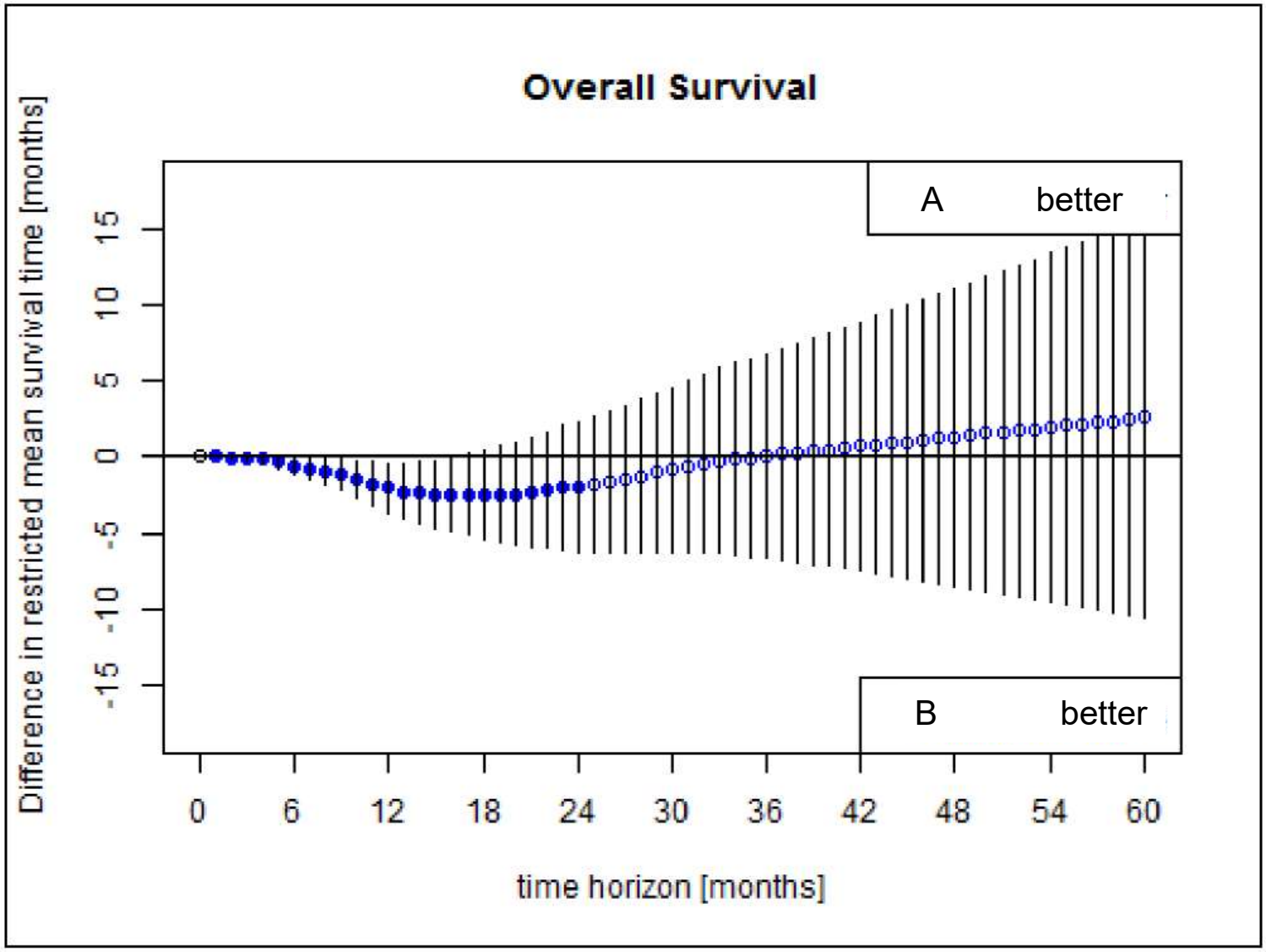
Statist. Med. **2011**, 30 2409–2421

Royston and Parmar *BMC Medical Research Methodology* 2013, **13**:152

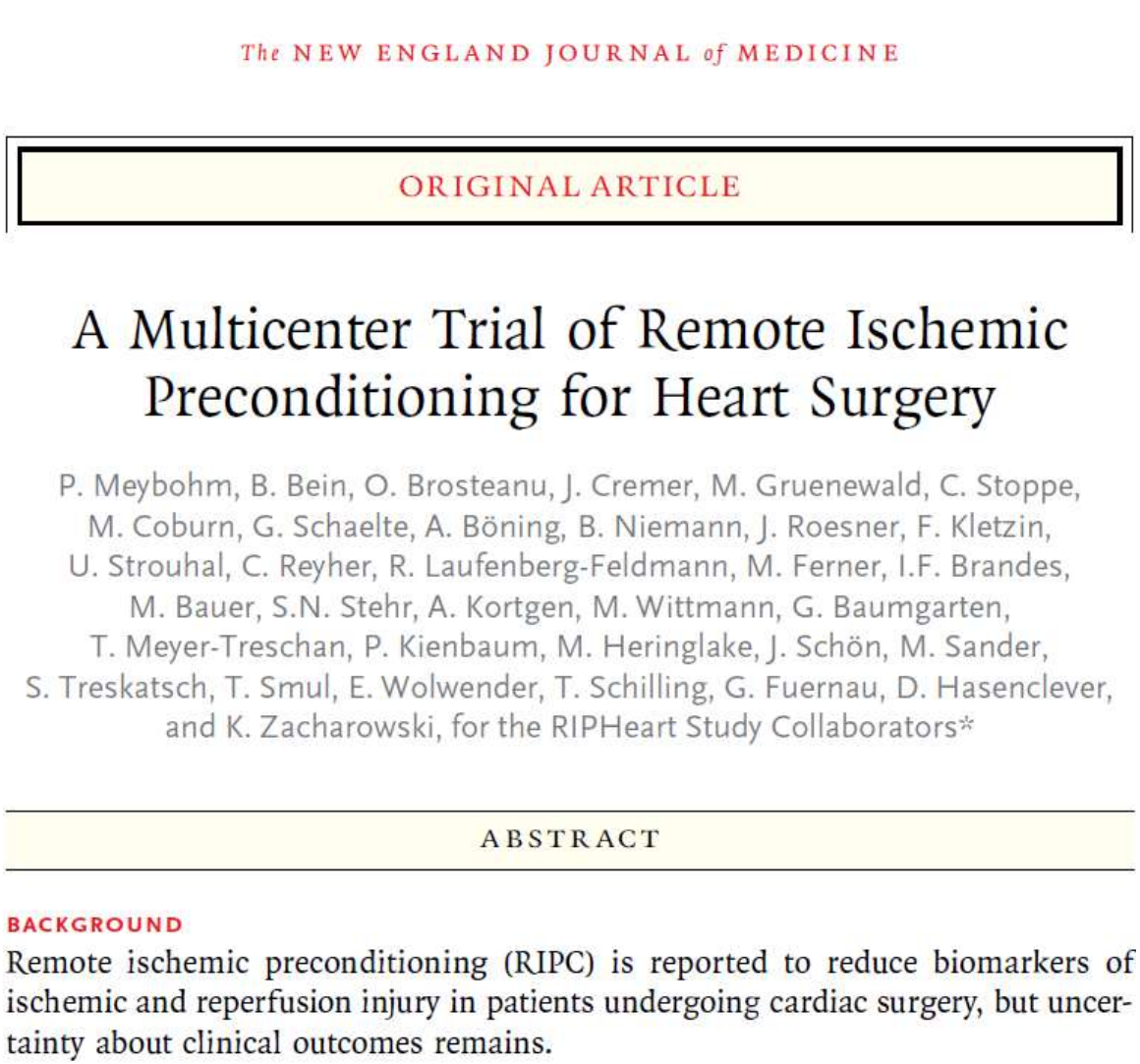
Restricted Mean Survival as a function of the Horizon



Difference in Restricted Mean Survival as a function of the Horizon



Zusammengesetzte Endpunkte I



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Multicenter Trial of Remote Ischemic Preconditioning for Heart Surgery

P. Meybohm, B. Bein, O. Brosteanu, J. Cremer, M. Gruenewald, C. Stoppe, M. Coburn, G. Schaelte, A. Böning, B. Niemann, J. Roesner, F. Kletzin, U. Strouhal, C. Reyher, R. Laufenberg-Feldmann, M. Ferner, I.F. Brandes, M. Bauer, S.N. Stehr, A. Kortgen, M. Wittmann, G. Baumgarten, T. Meyer-Treschan, P. Kienbaum, M. Heringlake, J. Schön, M. Sander, S. Treskatsch, T. Smul, E. Wolwender, T. Schilling, G. Fuernau, D. Hasenclever, and K. Zacharowski, for the RIPHeart Study Collaborators*

ABSTRACT

BACKGROUND

Remote ischemic preconditioning (RIPC) is reported to reduce biomarkers of ischemic and reperfusion injury in patients undergoing cardiac surgery, but uncertainty about clinical outcomes remains.

- **Cardiac surgery with cardiopulmonary bypass** is associated with a predictable incidence of
  - myocardial,
  - neurological and
  - renal dysfunction

due to perioperative ischaemia that accounts for significant morbidity and mortality.

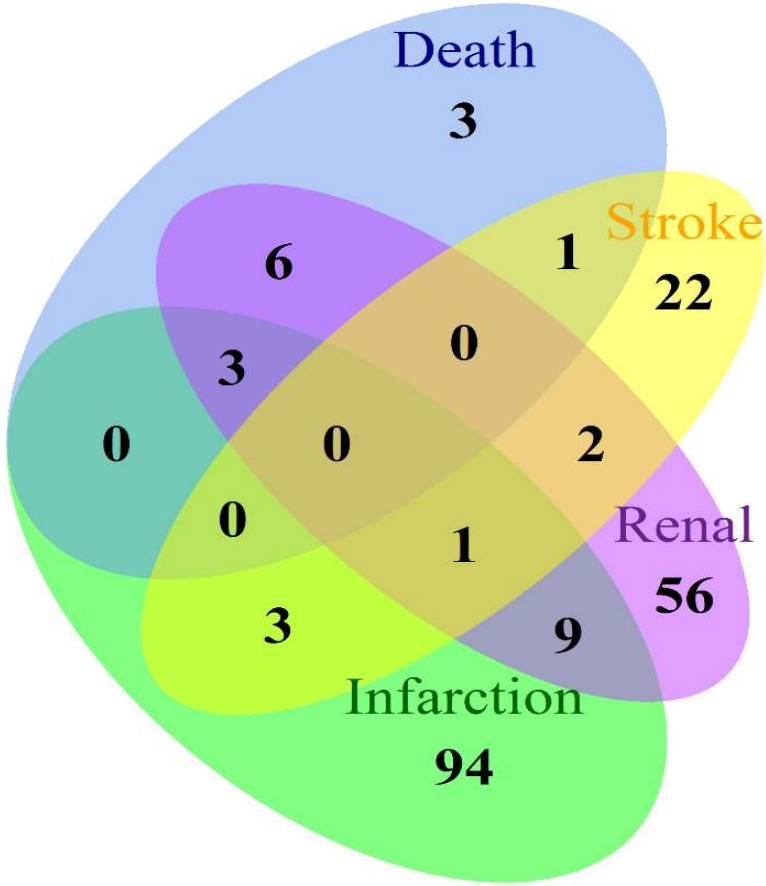
- **Remote ischaemic preconditioning (RIPC)** is a novel, simple, non-invasive and inexpensive intervention by which
  - ischaemia of non-vital tissue (skeletal muscles) protects remote organs (heart, brain and kidney) from a sustained episode of ischaemia.

Primary endpoint

- Endpoint: composite outcome
  - all-cause mortality,
  - non-fatal myocardial infarction
  - any new stroke
  - and/or acute renal failure

within 14 days after surgery compared to control intervention.

- **CAVE**: Composite endpoints can be inadequate
  - If intervention effects components inconsistently.
- Literature in animals and humans suggests effects should be consistently protective in RIPHeart.



Vorsicht mit Übernahme von Definitionen!

Anlehnung an die „Joint Task Force of the European Society of Cardiology/ American College of Cardiology/ American Heart Association for the Redefinition of Myocardial Infarction“

1) Troponin T Anstieg größer als 5-fach des oberen Referenzwertes

UND

2 a) neue pathologische Q-Welle oder **neuer Linksschenkelblock** im EKG

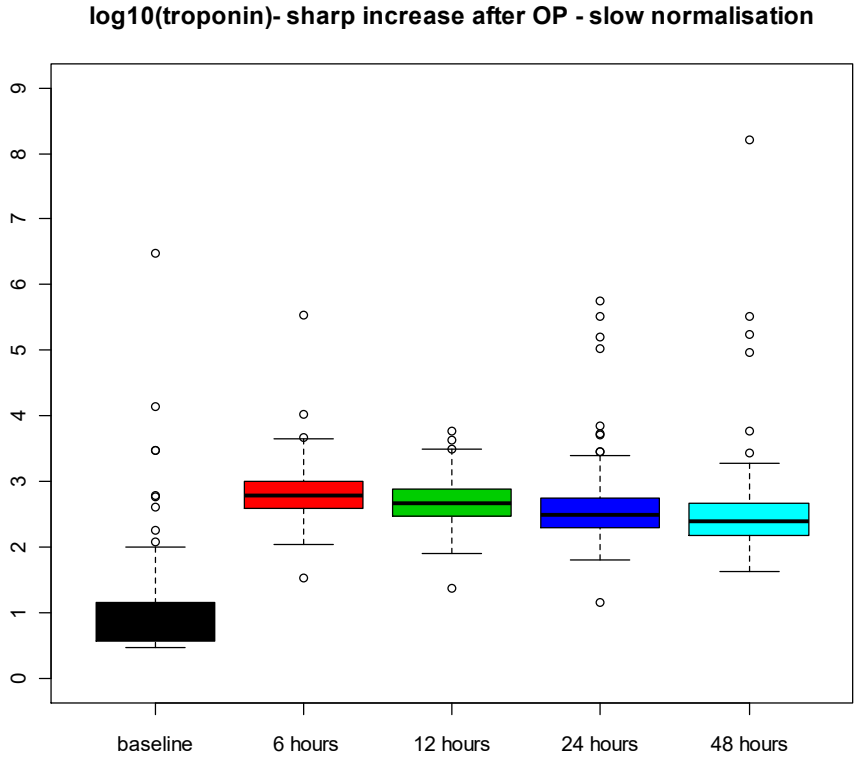
oder

2 b) frischer Ischämienachweis in der Echokardiographie

oder

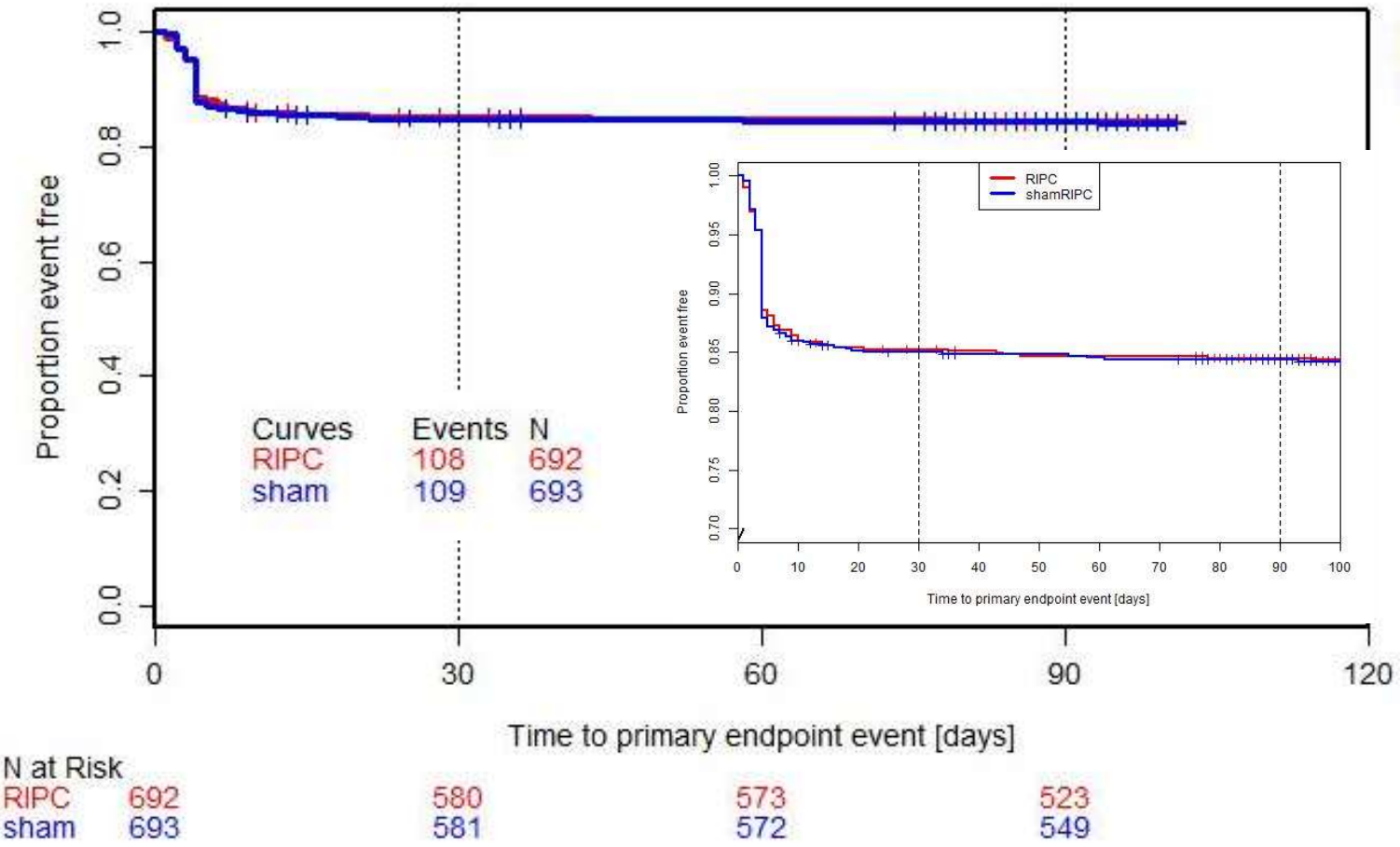
2 c) falls vorhanden, postoperativ Koronarangiographie mit Nachweis von frischem Bypass- bzw. Koronararterienverschluss

EKGs reviewed.
Relevante Inter-reader Varianz !!



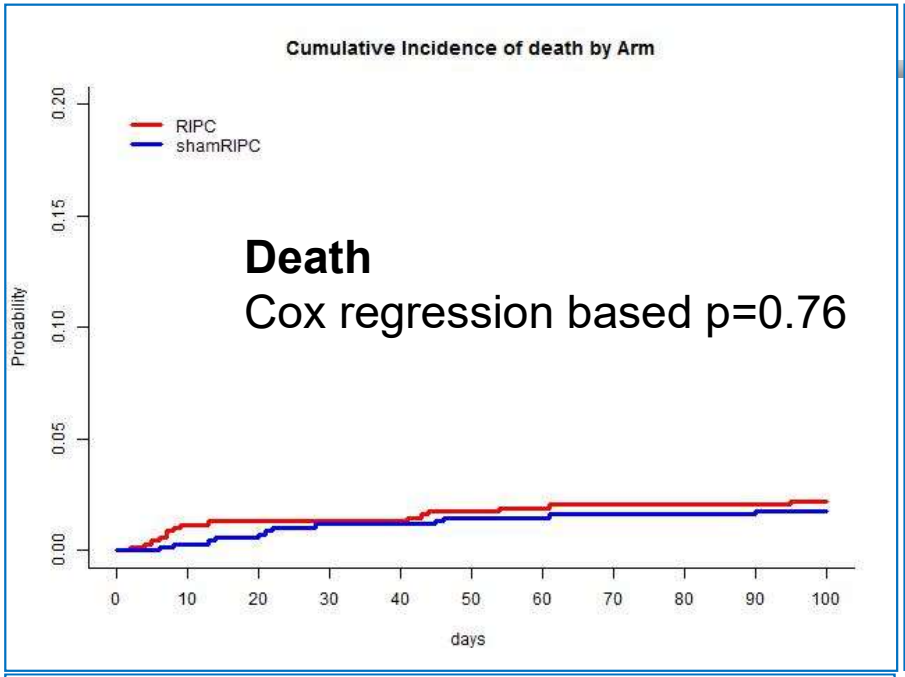
Troponin fast immer erhöht!
Linksschenkelblock KANN transient auftreten.

Time to PE up to d90

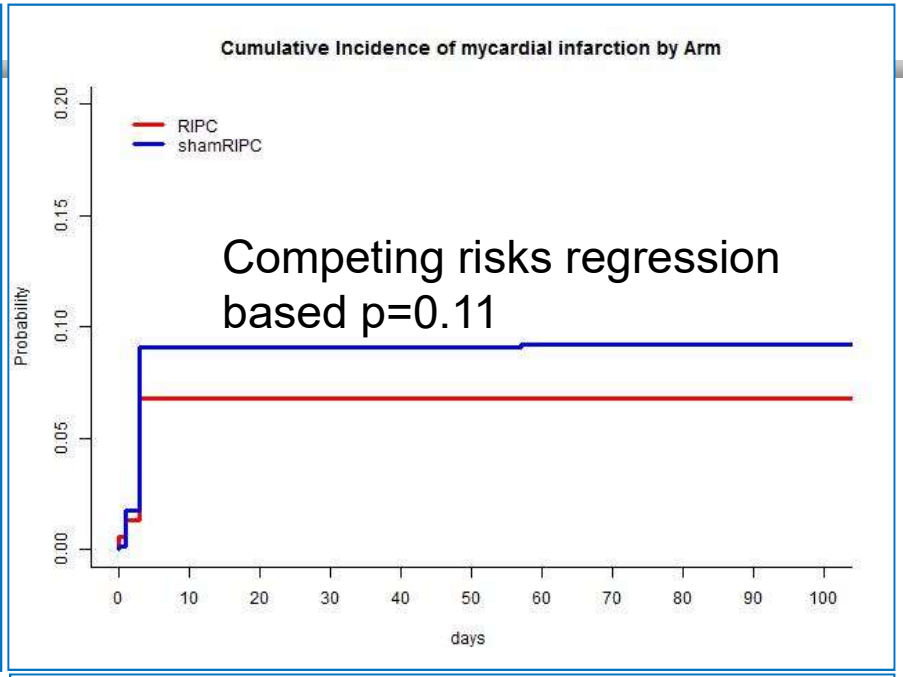


N at Risk				
RIPC	692	580	573	523
sham	693	581	572	549

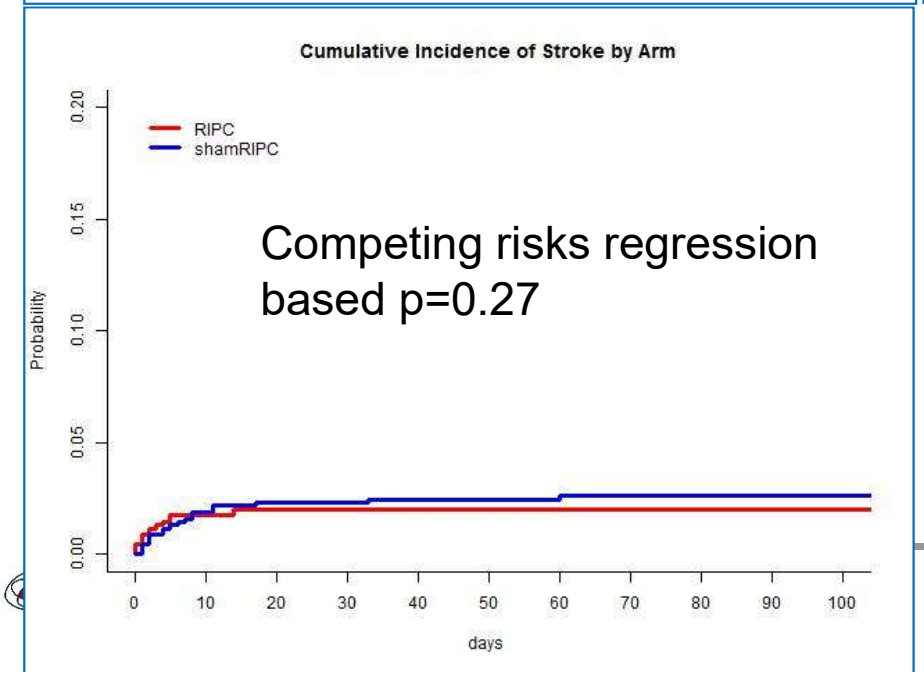
Cumulative incidence up to d90
Time to PE-components



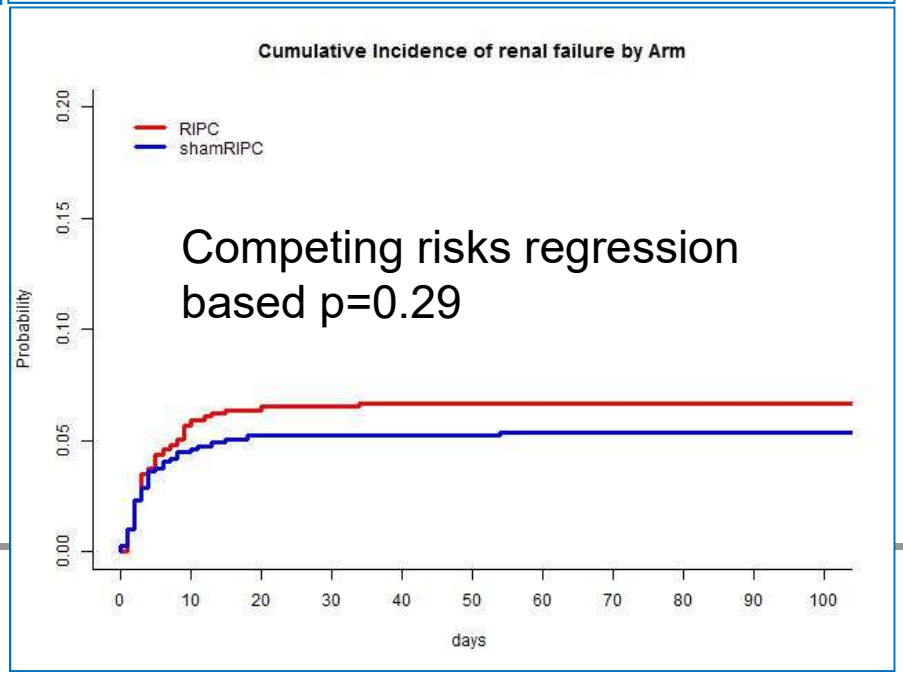
Death
Cox regression based p=0.76



Competing risks regression
based p=0.11



Competing risks regression
based p=0.27



Competing risks regression
based p=0.29

Historischer Vergleich: Induction treatment AML elderly

Table I. Comparison of CR rate and reasons for failure between DAT and SAB.

SAB=same as before

Treatment	Period	Number of patients	CR rate	Induction death	Resistant disease
DAT	1984–90	167	47%	30%	23%
SAB	1990–98	284	61%	15%	24%

Wheatley 2002
Patients ≥ 60 only

SAB=DAT	1990-98	328	62%	16%	22%
MAC	1990-98	656	55%	17%	28%

Goldstone 2001

Auch historische Vergleiche zwischen sukzessiven prospektiven Studien können irreführen!

CR bei AML ist ein verkappter composite endpoint

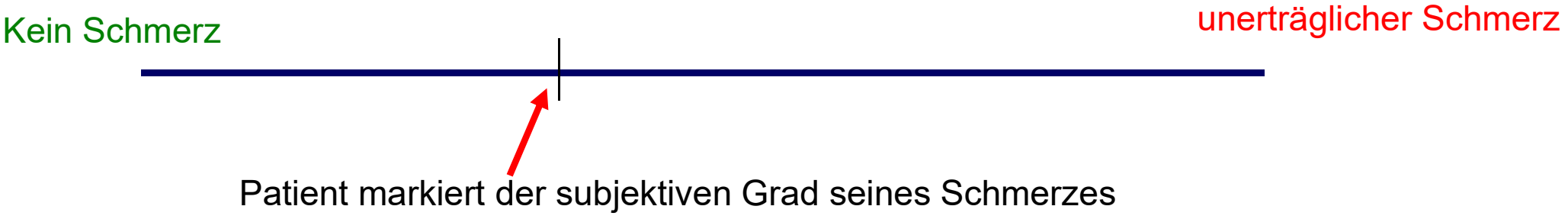
Subjektive Endpunkte: Schmerz

- **Perkutane Ethanolinjektion bei inoperablen Hepatozellulären Karzinom (HCC)**
- Problem: Heftigen Schmerzreaktion durch Austreten von Ethanol aus dem Stichkanal.
- **Vergleiche konventionelle Injektion mit Injektion mittels Spezialnadel, welche zunächst den Stichkanal anästhesiert.**

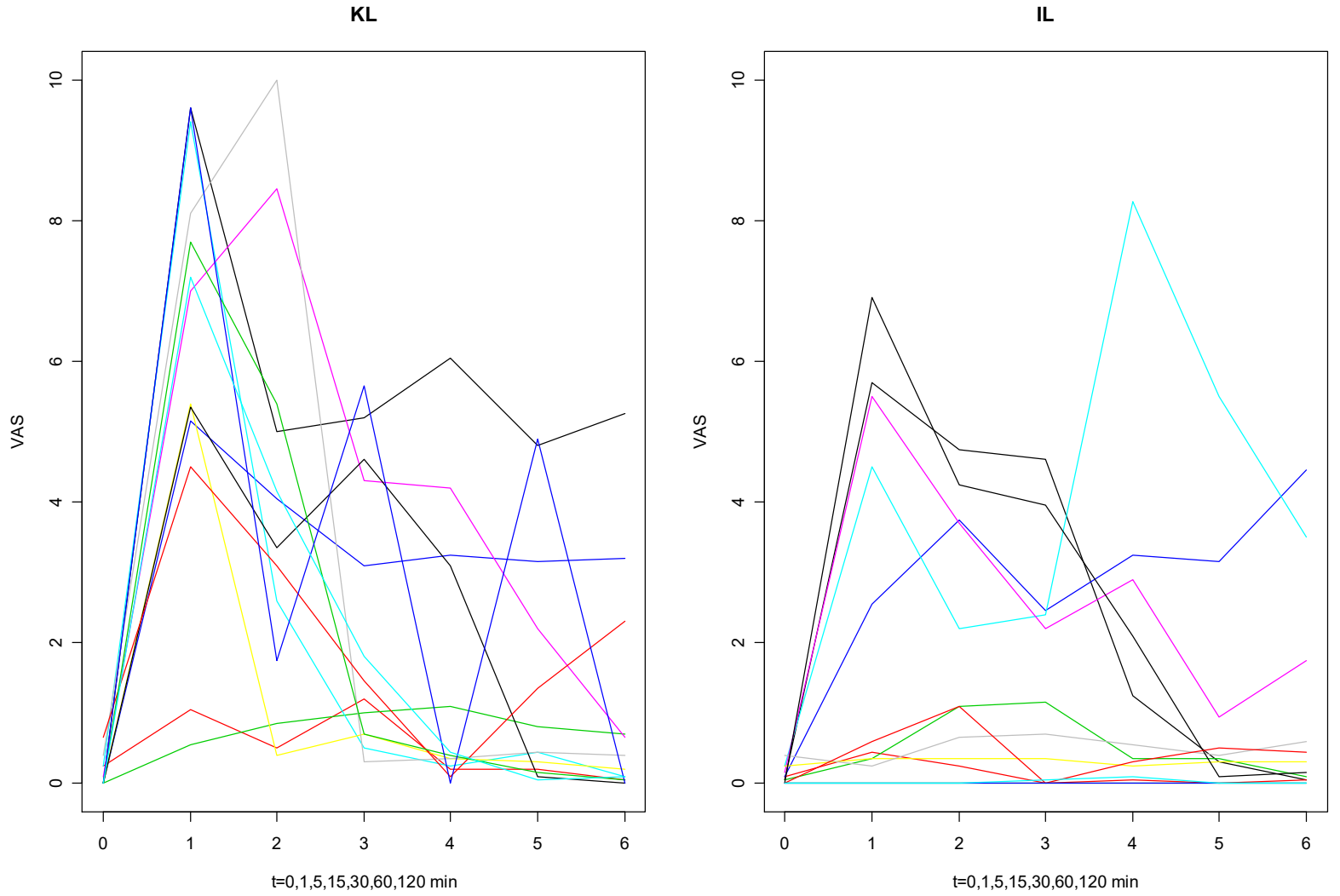
- **Cross-over-Design**.
  - Alle Patienten werden mit beiden Lokalanästhesiemethoden behandelt.
  - Durch das Cross-Over wird angestrebt die interindividuelle Varianz der Patienten bezüglich Schmerzeinschätzung und klinischer Situation als Störgröße auszuschalten.
  - Um etwaige Periodeneffekte auszumitteln, wird die Reihenfolge nach einem Zweierblock-Randomisationsschema festgelegt.

Subjektive Endpunkte: Schmerz II

- Der Studie soll eine aussagekräftige und klinisch leicht interpretierbare Hauptzielgröße zugrunde gelegt werden.
- Dazu definieren wir, wann ein **klare klinische Präferenz für eine der Methoden** vorliegt.
- Idee: Differenz im maximalen subjektiven Schmerzempfinden auf einer Visuellen Analog Skala (VAS):



Schmerzverläufe



Subjektive Endpunkte: Schmerz III

- Wie sind Differenzen auf der VAS Schmerzskala zu interpretieren?
-
- Mehrere Autoren
  - [Wallenstein SL et. al. (1980), Powell CV et al. (2001), Collins SL (1997), DeLoach LJ et al. (1998), Todd KH & Funk JP (1996), Todd KH et al. (1996), Kelly AM (1998)]

  haben durch Befragung von Akutschmerzpatienten die **minimal bedeutsame Differenz einer VAS Schmerzmessung** bestimmt.

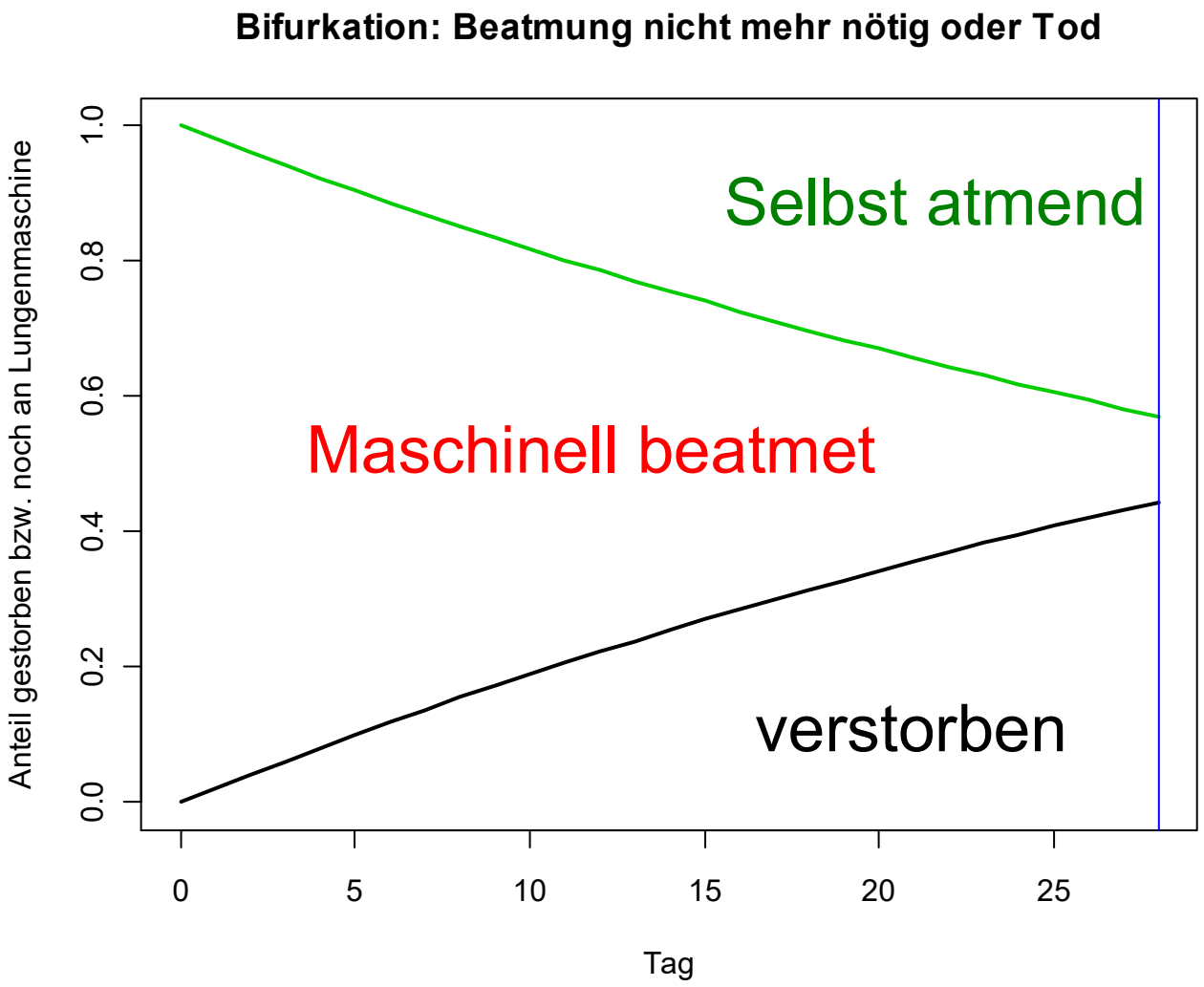
- Die Ergebnisse schwanken zwischen **9 mm und 20 mm**.

Bifurkationsprozess

- Patienten mit mechanischer Beatmung
- Problem: Lungenschädigung
- Frage: Lungenschädigung medikamentös verringern
- Zeithorizont: 28 Tage

- Drei Typen von Verläufen:
  - **Patient stirbt bis Tag 28**
  - **Beatmung wird vor Tag 28 beendet**
  - **Beatmung dauert an Tag 28 noch an (selten)**

Bifurkationsprozess II

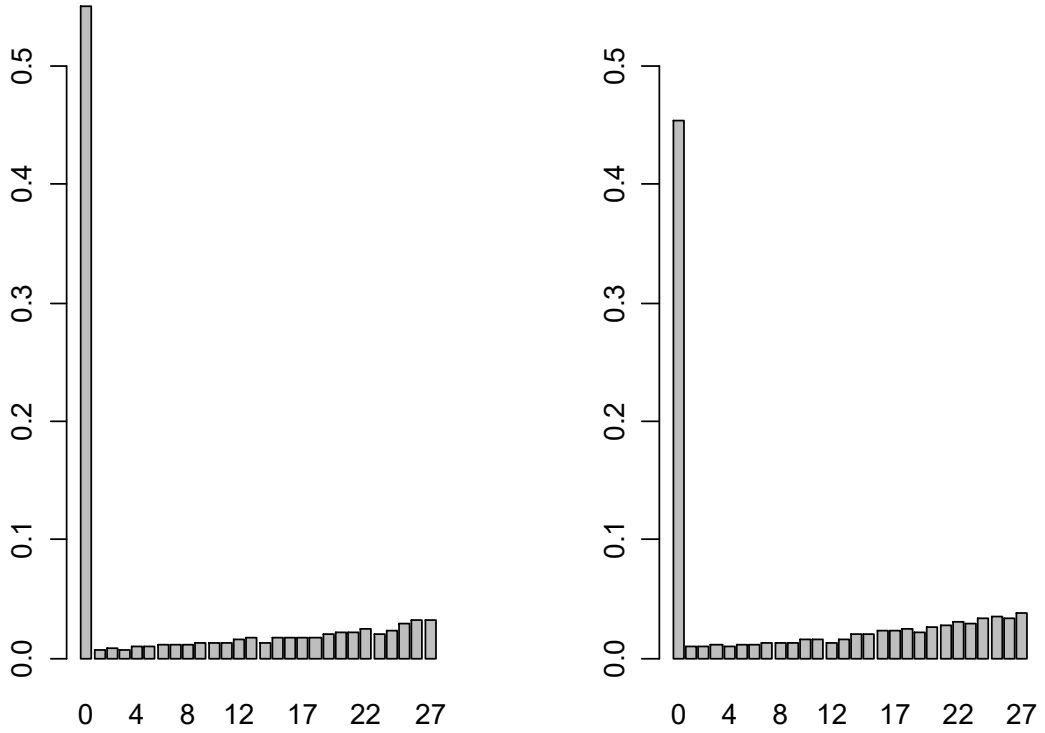


Ventilator free days

- Primary efficacy endpoint:
  - Number of days alive and off ventilator at 28 days after randomisation (VFD)
    - Dauerbeatmung oder Tod = 0 Tage
    - Selbstatmend ab Tag y = 29-y Tage

Ventilator free days II

Planungsszenario für möglichen Unterschied



	A	B
%gestorben	30%	25%
%Selbstatm.	46%	55%
%Dauerbeatm.	24%	20%
Mittelwert	7.7	9.4
Standardabw.	9.8	10.2

Problem: Surrogat Endpunkte

Klinisch relevanter Endpunkt

- tritt **spät** oder **selten** ein
- ist nur **invasiv** oder **teuer** messbar

Lässt sich der **klinisch relevante Endpunkt** ersetzen durch **einen Surrogat-Endpunkt**, der Veränderungen auf einer kürzeren Zeitskala anzeigt ?

Surrogat-Endpunkt

- Ein **Surrogat-Endpunkt** für eine klinische Studie ist
  - ein Laborwert,
  - ein biologischer Marker oder
  - ein histopathologischer oder klinischer Befund,

  der als **Ersatz für den klinisch relevanten Endpunkt** benutzt wird.

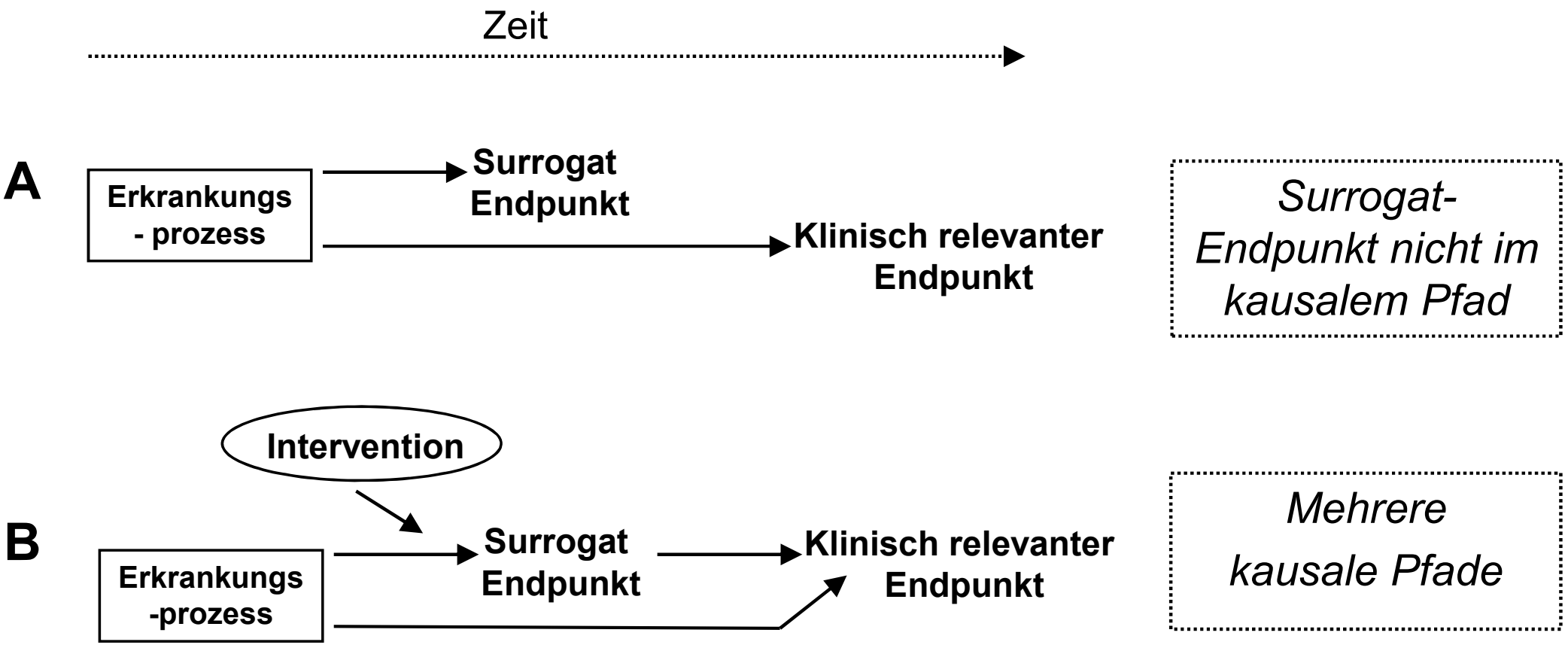
- Surrogat-Endpunkt ist **valide** wenn:
  - **Hohe Korrelation** zwischen Surrogat-Endpunkt und klinischem Endpunkt

    und
  - **Therapieeffekt auf den Surrogatendpunkt** übersetzt sich regelhaft in einen **Therapieeffekt auf den klinisch relevanten Endpunkt**

Korrelation allein ist nicht ausreichend !

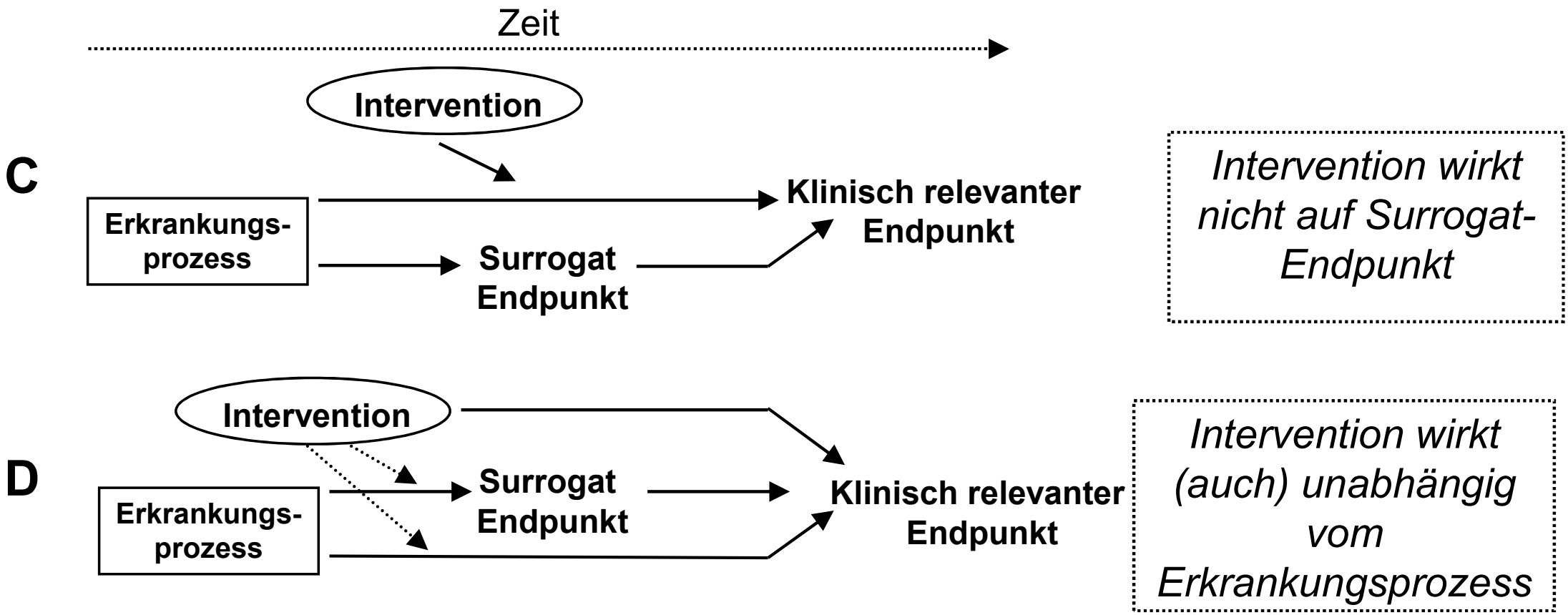
Surrogat-Endpunkt

- **Validität ist nicht gegeben** (trotz Korrelation)



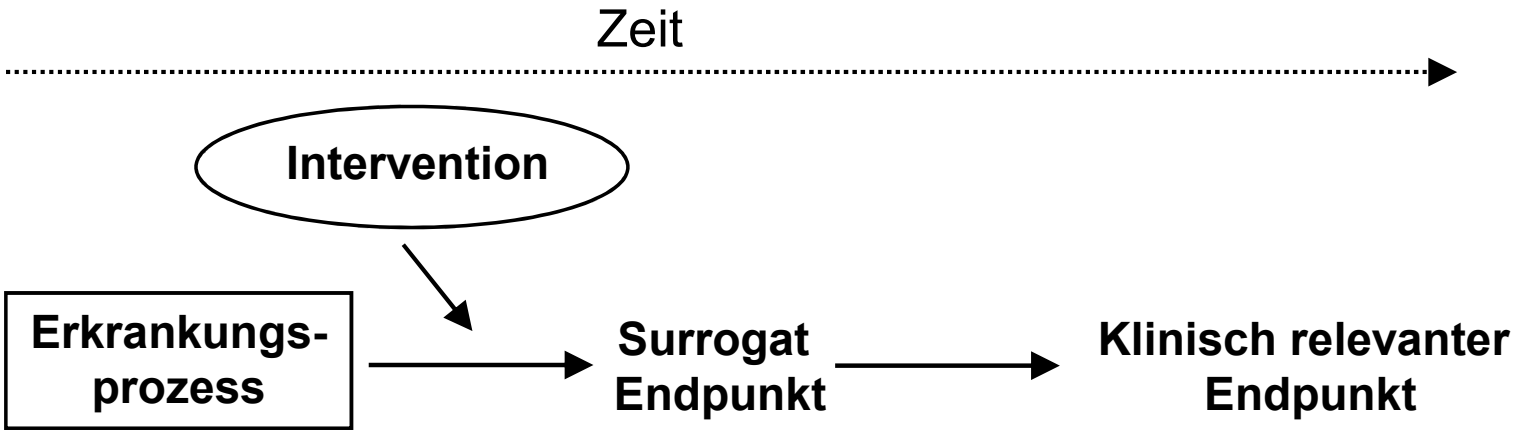
Surrogat-Endpunkt

- **Validität ist nicht gegeben** (trotz Korrelation)



Surrogat-Endpunkt

- **Validität** ist gegeben



- **Vorsicht:**

  auch bei dieser Konstellation kann es zu Irreführungen kommen
  - **Unterschätzung des „wahren“ Therapieeffekts**, z.B. durch „Rauschen“ im Surrogat-Endpunkt
  - **Überschätzung des „wahren“ Effekts**, wenn Effekt auf Surrogat-Endpunkt zu schwach / zu kurz anhaltend ist, um den klinisch relevanten Endpunkt zu beeinflussen

Validierung eines Surrogatendpunkts

- Kann nur **anhand des klinisch relevanten Endpunkts** erfolgen

Notwendig für Validierung:

Statistisch:

- **Daten aus einer randomisierter Studie** (oder aus mehreren randomisierten Studien – Metaanalyse) mit
  - Klinisch relevantem Endpunkt
  - Ausreichender Fallzahl
  - Surrogatendpunkt gemessen
  - Therapieeffekt nachweisbar

Medizinisch-Inhaltlich:

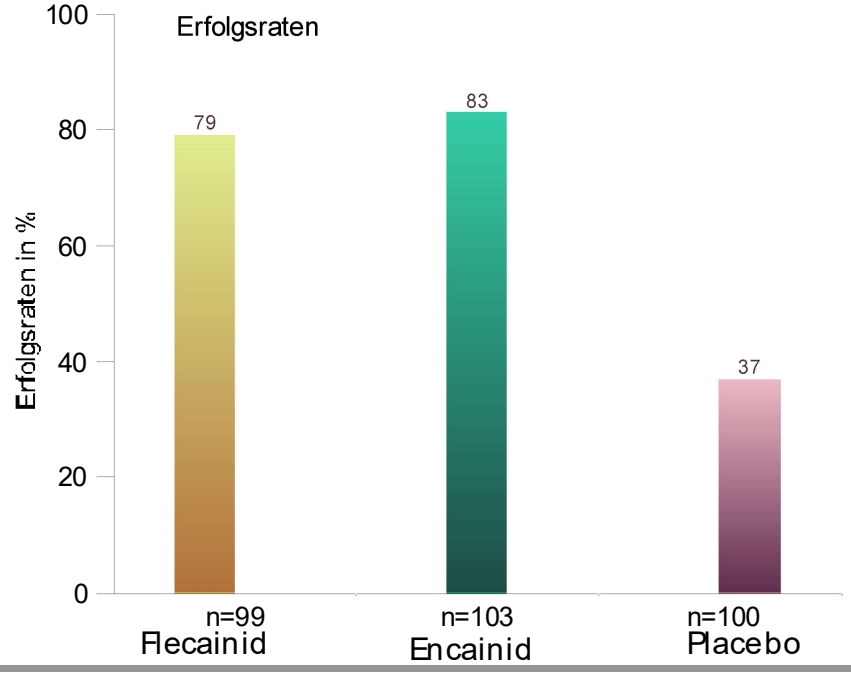
- Einsicht über
  - Kausale Zusammenhänge des Krankheitsprozesses
  - Wirkungsmechanismen der Intervention

Beispiel Surrogatkriterien I

Behandlung von **Arrhythmien nach Myokardinfarkt**

Pilotstudie CAPS:

=> **Können Arrhythmien durch Behandlung mit Encainid oder Flecainid unterdrückt werden ?**



VES: **ventrikuläre Extrasystolen**
VT: **ventrikuläre Tachykardien**

Erfolg: VES ⇓ 70% und VT ⇓ 90%

Am J Cardiol 1988; 61: 501-509

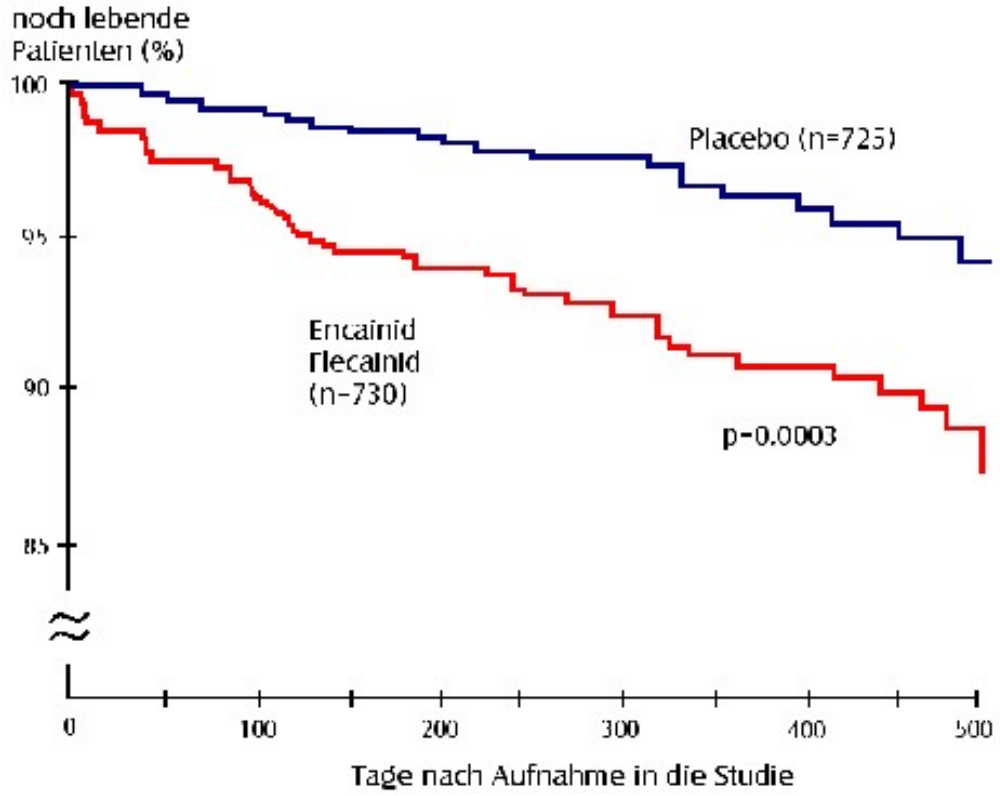
Beispiel Surrogatkriterien II

Behandlung von **Arrhythmien nach Myokardinfarkt**

CAST-Studie:

=> **Verlängert die Therapie mit den Antiarrhythmika**

Überlebenszeit ?



Geplant:
4.400 Patienten in 3 Jahren

=> Studie wurde bei der ersten Zwischenauswertung nach 1,5 Jahren abgebrochen

N Engl J Med 1989; 321: 406-412

Empfehlung zur Studienplanung

- **EMA**

  European Agency for the Evaluation of Medicinal Products

  Europäische Zulassungsbehörde

  http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000085.jsp

  (Stand Mai 2021)

- **Clinical efficacy and safety guidelines**
  - Concept papers / Points to consider / Guidelines:
  - Empfehlungen zu
    - **Studiendesign / Ein- und Ausschlusskriterien / Endpunkte**
  - Für **viele Indikationsgebiete**
    - von Psoriasis über akutes Herzversagen bis zu Schizophrenie

    Vor Studienplanung recherchieren !