
Interdisziplinäre Sitzung:
Medizinische Epidemiologie, Informatik und Biometrie
im Netzwerk der pädiatrischen Onkologen

51. Jahrestagung der GMDS
Leipzig, 11. September 2006

Erfahrung mit flexibler Gestaltung
von klinischen Studien in der
pädiatrischen Onkologie und Hämatologie

Dr. Andreas Faldum
Angela Emser

Institut für Medizinische Biometrie,
Epidemiologie and Informatik
Johannes Gutenberg-Universität Mainz

faldum@imbei.uni-mainz.de
emser@imbei.uni-mainz.de

Biometrische Betreuung von Studien der GPOH

- HIT-2000
- HIT-GBM
- HIT-Endo
- HIT-REZ
- HIT-LGG
- SIOP-LGG
- NB2004
- MET
- SAA
- LCH

Flexible Gestaltung klinischer Studien

- Warum brauchen wir mehr Flexibilität?

Um auf neue Erkenntnisse sowie Änderungen im Studienumfeld reagieren zu können.

- Intern: Ergebnisse der ZA, Internationalisierung
- Extern: Ergebnisse anderer Studien, AMG

- Wann sollte man ein flexibles Design wählen?

Immer wenn eine Studie lange dauert!

- Insbesondere bei Überlebenszeitstudien
- Insbesondere bei Studien für krebskranke Kinder

- Ethische wie ökonomische Notwendigkeit
- Vorsicht: Wahrung der wissenschaftlichen Integrität

Helmut Schäfer and Hans-Helge Müller

Statistics in Medicine

2001; 20: 3741-3751

Modification of the **sample size** and
the **schedule** of interim analyses
in **survival trials**
based on **data inspections**

HIT – SIOP PNET 4

- Kinder mit einem **Medulloblastom** ohne MTS, 4-21 Jahre alt
- Studienleitung: **Dr. S. Rutkowski, Dr. F. Deinlein** – Würzburg
- Randomisierte, offene **Überlebenszeitstudie** (EFS)
- Konventionelle vs. hyperfraktionierte **Radiotherapie**
- Start: **2001**, Rekrutierung: **5 Jahre**, Nachbeobachtung: **2 Jahre**
- Gruppensequentielles Design: **2 Zwischenanalysen** (2 und 4 J.)
- **Amendment: Adaptive Design** nach **Schäfer** und **Müller**
- 2003: **Fusion von SIOP PNET 4** (Österreich, Belgien, Frankreich, Italien, Niederlande, Skandinavien, Spanien, Schweiz, UK) **mit HIT 2000**
- Wunsch: **Änderung der Studiendauer**

Vorbehalte gegen die Methode von Schäfer und Müller

- Zu wenig Erfahrung mit dieser neuen Methode
- Bisher: Änderung des Studiendesigns ohne adaptive Methoden
- DMC als neutrales Gremium soll Entscheidung treffen
- Approximation der Verteilung der Logrank Teststatistiken durch eine multivariate Normalverteilung: Große Fallzahl
- Keine unverzerrten Schätzer und Konfidenzintervalle
 - G. Wassmer / W. Brannath et. al.: Hazard Ratio
 - A. Emser, A. Faldum: Kumulative Überlebensraten
- HIT – SIOP PNET IV: Grundlage der Designmodifikation = a priori berechnete Rekrutierungszahlen
- Aber: Keine Kontrolle des Gesamtniveaus und der Power!

HIT – GBM D

- Kinder mit einem malignen Gliom, 3-18 Jahre alt
- Studienleitung: PD H. Kramm – Halle
- Randomisierte, offene Überlebenszeitstudie (OS)
- Induktion: Simultane Radiochemotherapie \pm MTX
- Start: 2003, Rekrutierung: 2,5 Jahre, Nachbeobachtung: 3 Jahre
- Studiendesign: Fester Stichprobenumfang / keine ZA
- Wechsel d. Studienleitung: Prof. Wolff 2005 / Dr. Wagner 2006
- Problem: Fehlende Zeit für ein neues Protokoll
- Wunsch: Verlängerung der Studie um 1 Jahr
- Entscheidung nach ungeplanter ZA (Amend.: Schäfer & Müller)
- Entscheidungskriterium: Bedingte Power in mehreren Szenarien

Neues nominelles Signifikanzniveau

Nominelles Niveau für die **Endanalyse aller Daten** mit dem Logrank Test in **Studien ohne weitere Zwischenanalysen**:

$$\alpha_{\text{neu,+}} =$$

$$\Phi \left(\frac{\left(\sqrt{k_1 - k_0} - \sqrt{k_2 - k_0} \right) \cdot \sqrt{k_0} \cdot \Phi^{-1} \left(\frac{p_0}{2} \right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1} \left(\alpha_{\text{alt,+}} \right)}{\sqrt{k_1 - k_0} \cdot \sqrt{k_2}} \right)$$

Neues nominelles Signifikanzniveau

Nominelles Niveau für die Endanalyse aller Daten mit dem Logrank Test in Studien ohne weitere Zwischenanalysen:

$$\alpha_{\text{neu,+}} = \Phi \left(\frac{\left(\sqrt{k_1 - k_0} - \sqrt{k_2 - k_0} \right) \cdot \sqrt{k_0} \cdot \Phi^{-1} \left(\frac{p_0}{2} \right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1} \left(\alpha_{\text{alt,+}} \right)}{\sqrt{k_1 - k_0} \cdot \sqrt{k_2}} \right)$$

neues nominelles Signifikanzniveau (einseitig)

Neues nominelles Signifikanzniveau

Nominelles Niveau für die Endanalyse aller Daten mit dem Logrank Test in Studien ohne weitere Zwischenanalysen:

$$\alpha_{\text{neu,+}} = \Phi \left(\frac{\left(\sqrt{k_1 - k_0} - \sqrt{k_2 - k_0} \right) \cdot \sqrt{k_0} \cdot \Phi^{-1} \left(\frac{p_0}{2} \right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1} \left(\alpha_{\text{alt,+}} \right)}{\sqrt{k_1 - k_0} \cdot \sqrt{k_2}} \right)$$

altes nominelles Signifikanzniveau (einseitig)

Neues nominelles Signifikanzniveau

Nominelles Niveau für die Endanalyse aller Daten mit dem Logrank Test in Studien ohne weitere Zwischenanalysen:

$$\alpha_{\text{neu,+}} = \Phi \left(\frac{\left(\sqrt{k_1 - k_0} - \sqrt{k_2 - k_0} \right) \cdot \sqrt{k_0} \cdot \Phi^{-1} \left(\frac{p_0}{2} \right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1} \left(\alpha_{\text{alt,+}} \right)}{\sqrt{k_1 - k_0} \cdot \sqrt{k_2}} \right)$$

p-Wert des Logrank Test der Zwischenanalyse (zweiseitig)

Neues nominelles Signifikanzniveau

Nominelles Niveau für die Endanalyse aller Daten mit dem Logrank Test in Studien ohne weitere Zwischenanalysen:

$$\alpha_{\text{neu,+}} = \Phi \left(\frac{(\sqrt{k_1} - k_0 - \sqrt{k_2} - k_0) \cdot \sqrt{k_0} \cdot \Phi^{-1}\left(\frac{p_0}{2}\right) + \sqrt{k_2} - k_0 \cdot \sqrt{k_1} \cdot \Phi^{-1}(\alpha_{\text{alt,+}})}{\sqrt{k_1} - k_0 \cdot \sqrt{k_2}} \right)$$

Anzahl der Ereignisse bis zur Zwischenanalyse

Neues nominelles Signifikanzniveau

Nominelles Niveau für die Endanalyse aller Daten mit dem Logrank Test in Studien ohne weitere Zwischenanalysen:

$$\alpha_{\text{neu,+}} = \Phi \left(\frac{(\sqrt{k_1 + k_0} - \sqrt{k_2 - k_0}) \cdot \sqrt{k_0} \cdot \Phi^{-1}\left(\frac{p_0}{2}\right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1}(\alpha_{\text{alt,+}})}{\sqrt{k_1 + k_0} \cdot \sqrt{k_2}} \right)$$

Anzahl der Ereignisse bei der ursprünglich geplanten Endanalyse

Neues nominelles Signifikanzniveau

Nominelles Niveau für die Endanalyse aller Daten mit dem Logrank Test in Studien ohne weitere Zwischenanalysen:

$$\alpha_{\text{neu,+}} = \Phi \left(\frac{\left(\sqrt{k_1 - k_0} - \sqrt{k_2 - k_0} \right) \cdot \sqrt{k_0} \cdot \Phi^{-1} \left(\frac{p_0}{2} \right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1} \left(\alpha_{\text{alt,+}} \right)}{\sqrt{k_1 - k_0} \cdot \sqrt{k_2}} \right)$$

Anzahl der Ereignisse bei der neu geplanten Endanalyse

Neues nominelles Signifikanzniveau

Nominelles Niveau für die Endanalyse aller Daten mit dem Logrank Test in Studien ohne weitere Zwischenanalysen:

$$\alpha_{\text{neu,+}} =$$

$$\Phi \left(\frac{\left(\sqrt{k_1 - k_0} - \sqrt{k_2 - k_0} \right) \cdot \sqrt{k_0} \cdot \Phi^{-1} \left(\frac{p_0}{2} \right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1} \left(\alpha_{\text{alt,+}} \right)}{\sqrt{k_1 - k_0} \cdot \sqrt{k_2}} \right)$$

Voraussetzung für die Methode nach Schäfer und Müller:

Kenntnis über die **Anzahl der Ereignisse bei ursprünglich** sowie **neu geplanter Endanalyse**

Neues nominelles Signifikanzniveau

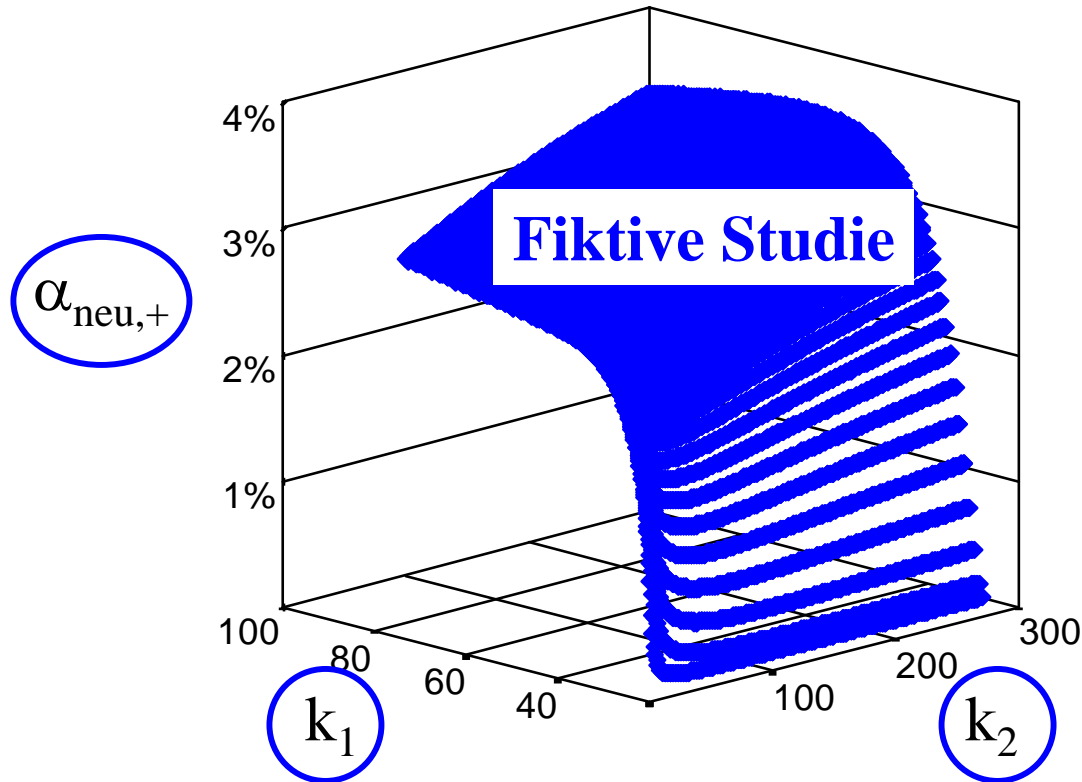
Nominelles Niveau für die Endanalyse aller Daten mit dem Logrank Test in Studien ohne weitere Zwischenanalysen:

$$\alpha_{\text{neu,+}} =$$

$$\Phi \left(\frac{\left(\sqrt{k_1 - k_0} - \sqrt{k_2 - k_0} \right) \cdot \sqrt{k_0} \cdot \Phi^{-1} \left(\frac{p_0}{2} \right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1} \left(\alpha_{\text{alt,+}} \right)}{\sqrt{k_1 - k_0} \cdot \sqrt{k_2}} \right)$$

Aber: **Dauer der GBM D-Studie** richtet sich nach der **Kalenderzeit**
⇒ In der Zwischenanalyse sind die **Anzahl der Ereignisse bei ursprünglich** sowie **neu geplanter Endanalyse unbekannt**

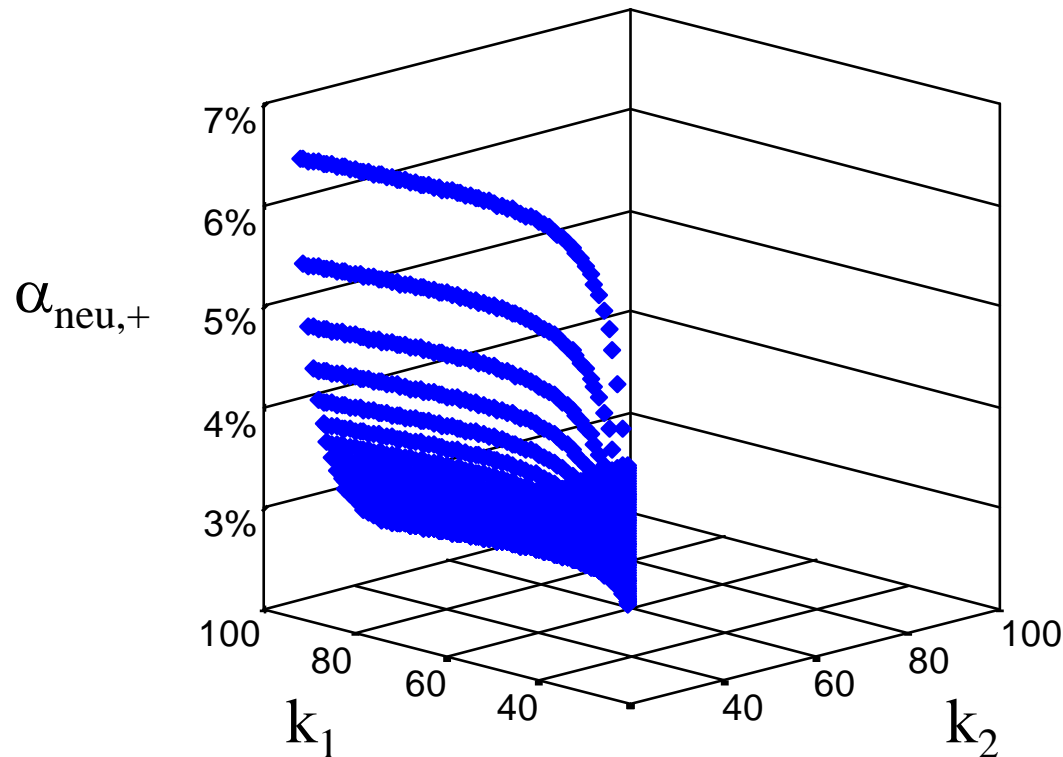
Neues nominelles Niveau in Abhängigkeit der Anzahl der Ereignisse bei ursprünglich und neu geplanter Endanalyse



Verlängerung
der Studie:

$$k_2 > k_1$$

Neues nominelles Niveau in Abhängigkeit der Anzahl der Ereignisse bei ursprünglich und neu geplanter Endanalyse



Verkürzung

der Studie:

$$k_2 < k_1$$

Fazit

- Bei **kalenderzeitabhängiger** Planung einer Überlebenszeitstudie:
 - **Verkürzung nicht möglich:**
 - **Anzahl d. Ereignisse bei ursprünglicher EA ist unbekannt**
 - **Spannweite des neuen nominellen Niveaus ist groß**
 - **Verlängerung nur möglich, bei**
 - **genauer Festlegung des Ablaufs der Designänderung**
 - **strenger Vertraulichkeit der Studiendaten bis zur EA**
 - **a priori sowie posteriori Analyse** des Einflusses möglicher Ereigniszahlen auf das neue nominelle Niveau
- **Insgesamt: Je flexibler** ein verwendetes Design ist, **desto sorgfältiger** muss die Planung der Studie erfolgen!!!

Dr. Andreas Faldum
Angela Emser

Institut für Medizinische Biometrie,
Epidemiologie and Informatik
Johannes Gutenberg-Universität Mainz

faldum@imbei.uni-mainz.de
emser@imbei.uni-mainz.de

Inhalt

- Zusammenarbeit zwischen GPOH und IMBEI
- Flexible Gestaltung von Überlebenszeitstudien
- Beispiele
- Besonderheiten bei Studienplanung nach Kalenderzeit

Beginn der Zusammenarbeit von GPOH und IMBEI

- Gesellschaft für Pädiatrische Onkologie und Hämatologie: Therapieoptimierungsstudien für krebskranke Kinder
- Vor 25 Jahren: Gründung des DKKR am IMBEI
- IMBEI: zentrale Anlaufstelle für biometrische und informatische Fragen
 - Partielle Beratung bei statistischen Problemen und Dokumentationsfragen
 - Seit vielen Jahren Mitarbeit in Studienkommissionen
- Unbefriedigende Situation:
 - Nur sporadische Kontakte
 - Oft nur Troubleshooting

Intensivierung der Zusammenarbeit

- Seit 2001: Förderung
 - durch die Deutsche Kinderkrebsstiftung der DLFH
 - einer zentralen biometrischen Betreuung am IMBEI
 - für 5 Hirntumorstudien der GPOH
- Seit 2002: Angebot einer intensiven biometrischen Betreuung durch das IMBEI an alle Kinderkrebsstudien
- Seit 2004: Intensivierte Kooperation mit weiteren Studien
- Zentrale biometrische Betreuung:
 - Projektleitung: A. Faldum
 - Wissenschaftliche Mitarbeiter: A. Emser, H. Götte
 - Dokumentarin (1/2 Stelle): S. Berkefeld

Zentrale biometrische Betreuung von Studien der GPOH

- HIT-2000

- HIT-GBM

- HIT-REZ

- HIT-REZ

- HIT/SIOP-LGG

Deutsche Kinderkrebsstiftung

- NB2004

- MF7

- A

„Neue“ Kooperationen

- DSMB

Intensivierung der Zusammenarbeit

- | | | |
|--|---|----------------------------------|
| • HIT-2000 | | |
| Dr. S. Rutkowski, Dr. F. Deinlein - Würzburg | | |
| • HIT-G | | Prof. F. Berthold, |
| PD H. Kramm - Halle | | Dr. T. Simon, Dr. B. Hero - Köln |
| • ... | | Prof. P. Bucsky – Lübeck |
| PD H. Müller – Oldenburg | | Prof. H. Gadner - Wien |
| HIT-REZ | | Dr. M. Führer - München |
| Prof. U. Bode, Dr. G. Fleischhack - Bonn | | |
| • HIT-LGG | ⇒ | Internationalisierung |
| Dr. A. Gnekow - Augsburg | | |

Biometrische Begleitung der klinischen Studien

Vereinbarung mit Prof. J. Kühl (Hirntumorstudien, 2000)

- Mitwirkung bei der Formulierung der **Studienziele**
- Mitwirkung bei **Haupt- und Nebenfragestellungen**
- Formulierung der **Hypothesen**
- Auswahl der **statistischen Tests**
- Wahl geeigneter **Verfahren für Zwischenanalysen**
- **Power / Fallzahlberechnung**
- Formulierung von **Abbruchkriterien**
- Verfassen der biometrischen Teile des **Studienprotokolls**

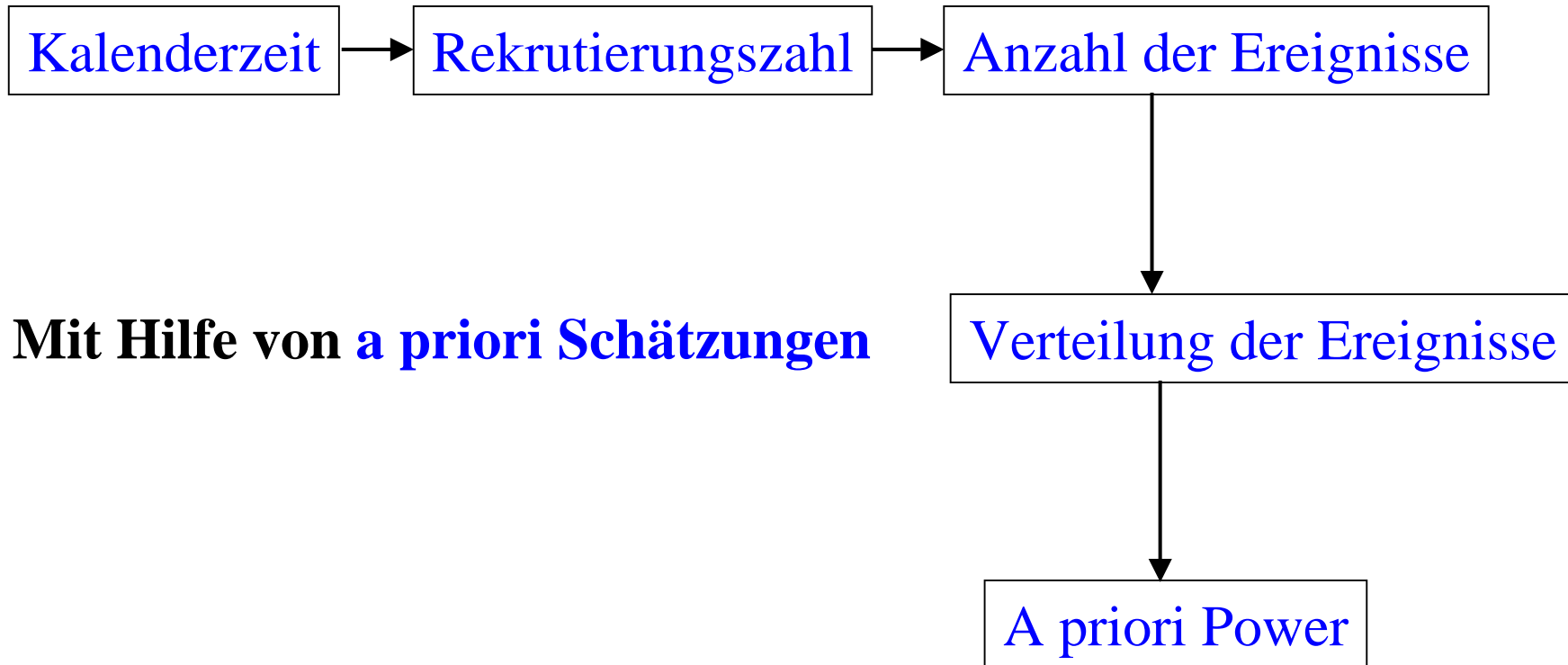
Biometrische Begleitung der klinischen Studien

- Erstellen von **Randomisierungslisten**
- Konzeption der **Auswertungsdatei**
- Mitarbeit bei der **Qualitätssicherung**
- Überwachung der **Abbruchkriterien** bei **Sicherheitsanalysen**
- Durchführen von **Zwischenanalysen**
- Teilnahme an **Kommissionssitzungen / DMC**
- **Endauswertung**
- Mitarbeit beim Erstellen des **Abschlussberichts**
- **Publikationen**
- ...

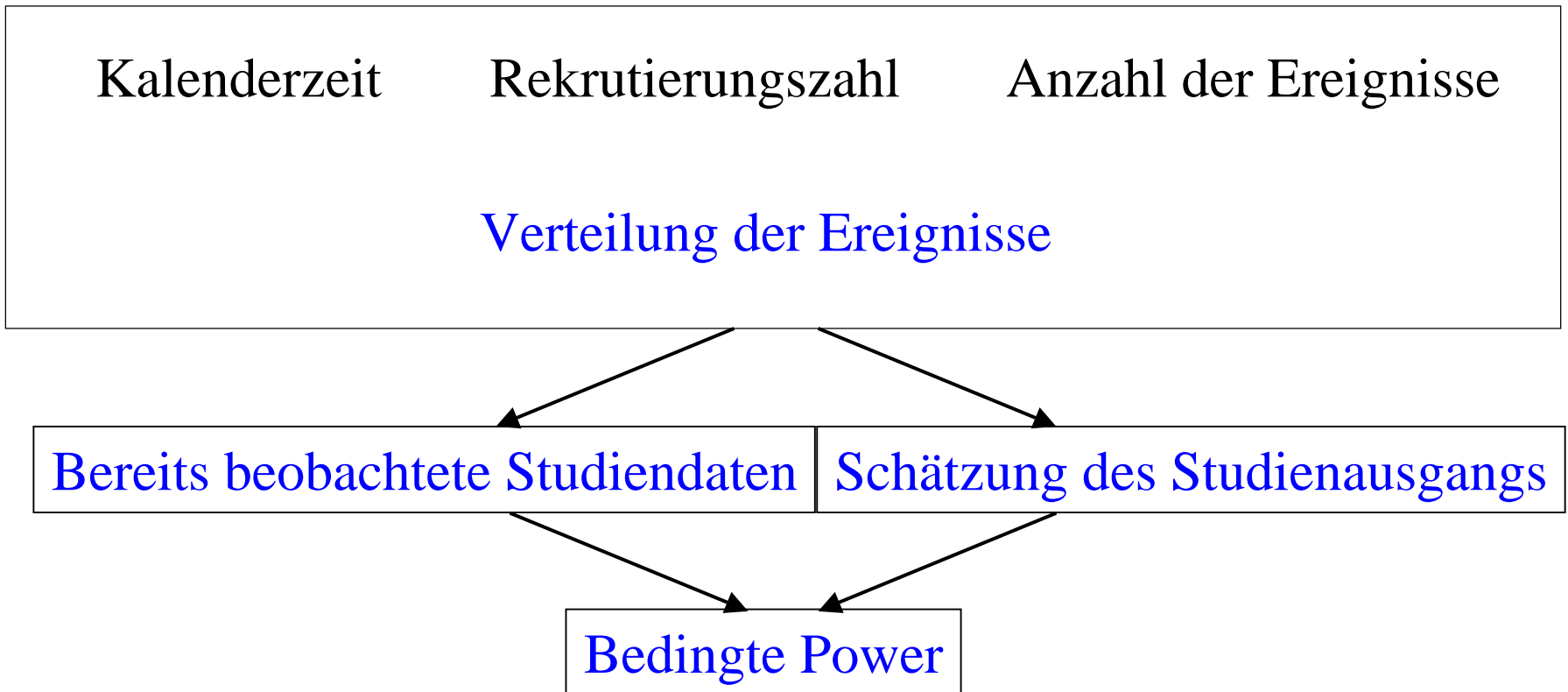
Idee der Methode von Schäfer und Müller

- Problem: Patienten ohne Ereignis verbleiben nach einer Zwischenanalyse in der Studie → **abhängige Stufen** → adaptive Standardverfahren sind nicht anwendbar
- Bestimme das **neue nominelle Signifikanzniveau** mit Hilfe der “Conditional Rejection Error Probability”:
Wahrscheinlichkeit für einen Fehler 1. Art im geplanten Studiendesign bedingt auf die beobachteten Daten
- **Änderungen** des Studiendesigns sind **erlaubt**, solange sich die **Conditional Rejection Error Probability** nicht ändert
- Das a priori festgelegte **Signifikanzniveau** wird **eingehalten**
- Die **Power der weiteren Studie** kann bestimmt werden

A priori Berechnung der Power in Überlebenszeitstudien



Berechnung der Power in einer laufenden Überlebenszeitstudien



Neues nominelles Signifikanzniveau

- Berechnung des nominellen Signifikanzniveaus für die Endanalyse aller Daten mit dem Logrank Test

$$\alpha_{\text{neu,+}} = \frac{\left(\left(\sqrt{k_1 - k_0} - \sqrt{k_2 - k_0} \right) \cdot \sqrt{k_0} \cdot \Phi^{-1} \left(\frac{p_0}{2} \right) + \sqrt{k_2 - k_0} \cdot \sqrt{k_1} \cdot \Phi^{-1}(\alpha_{\text{alt,+}}) \right)}{\sqrt{k_1 - k_0} \cdot \sqrt{k_2}}$$

Verteilungsfunktion und Inverse der Standardnormalverteilung

Möglichkeiten der a priori Planung des Studienendes

Kalenderzeit

Rekrutierungszahl

Anzahl der Ereignisse

Möglichkeiten der a priori Planung des Studienendes

Kalenderzeit

Rekrutierungszahl

Anzahl der Ereignisse

Bessere Planbarkeit der:

- Organisation
- Finanzierung
- Nachfolgestudie

Möglichkeiten der a priori Planung des Studienendes

Kalenderzeit

Rekrutierungszahl

Anzahl der Ereignisse

Einfache Erfassung

Möglichkeiten der a priori Planung des Studienendes

Kalenderzeit

Rekrutierungszahl

Anzahl der Ereignisse

Enger Zusammenhang zur Power

Möglichkeiten der a priori Planung des Studienendes

Kalenderzeit

Rekrutierungszahl

Anzahl der Ereignisse

Voraussetzung für die
Anwendung der Methode
nach Schäfer und Müller

Möglichkeiten der a priori Planung des Studienendes

Kalenderzeit

Rekrutierungszahl

Anzahl der Ereignisse

Zeitplan der **HIT-GBM** Studie

Neues nominelles Signifikanzniveau

- Berechnung des nominellen Signifikanzniveaus für die Endanalyse aller Daten mit dem Logrank Test
- Formel nur für **Studien ohne weitere Zwischenanalysen** gültig:
 - Einfacher
 - **Realistisch**: Entscheidung gegen Ende einer Studie akut
 - Auch mit weiteren Zwischenanalysen berechenbar
- **Bekannt sind**: Altes nominelles Signifikanzniveau, p-Wert des Logrank Tests sowie **Anzahl der Ereignisse bis zur ZA**
- **Bekannt sein müssen** aber auch: die **Anzahl der Ereignisse bei ursprünglich sowie neu geplanter Endanalyse**
- **Aber**: **Dauer der GBM D-Studie** richtet sich nach **Kalenderzeit**

Fiktive Studie

- Randomisierte Überlebenszeitstudie
- Standardbehandlung A vs. neue Behandlung B
- Primäre Analyse mit dem Logrank Test
- Studiendesign: fester Stichprobenumfang / keine ZA
- Signifikanzniveau: 5% (zweiseitig)
- Power = 80% ~ Rekrutierung: 5 J., Nachbeobachtung: 2 J.
- Probleme: 4,5 Jahre nach Studienstart:
 - Rekrutierung ist langsamer als geplant
 - Externe Studie: kleinerer Unterschied zwischen A und B
- Wunsch: Verlängerung der Studie um 2 Jahre
- Methode nach Schäfer und Müller nach ungeplanter ZA

Bedingte Power und neues nominelles Niveau

Fiktives Ergebnis der **Zwischenanalyse**:

- Altes nominelles Signifikanzniveau (einseitig): 0,025
- P-Wert des Logrank Tests: 0,2084
- Anzahl der Ereignisse: 27

Schätzung nach				
Bisherigen Daten				
Studienprotokoll				
Externe Studie				

Bedingte Power und neues nominelles Niveau

Fiktives Ergebnis der Zwischenanalyse:

- Altes nominelles Signifikanzniveau (einseitig): 0,025
- P-Wert des Logrank Tests: 0,2084
- Anzahl der Ereignisse: 27

Schätzung nach	k_1	k_2		
Bisherigen Daten	57	95		
Studienprotokoll	71	122		
Externe Studie	66	112		

Bedingte Power und neues nominelles Niveau

Fiktives Ergebnis der Zwischenanalyse:

- Altes nominelles Signifikanzniveau (einseitig): 0,025
- P-Wert des Logrank Tests: 0,2084
- Anzahl der Ereignisse: 27

Schätzung nach	k_1	k_2	Bedingte Power	
Bisherigen Daten	57	95	56%	
Studienprotokoll	71	122	90%	
Externe Studie	66	112	48%	

Bedingte Power und neues nominelles Niveau

Fiktives Ergebnis der Zwischenanalyse:

- Altes nominelles Signifikanzniveau (einseitig): 0,025
- P-Wert des Logrank Tests: 0,2084
- Anzahl der Ereignisse: 27

Schätzung nach	k_1	k_2	Bedingte Power	Neues nominelles Niveau
Bisherigen Daten	57	95	56%	0,0258
Studienprotokoll	71	122	90%	0,0275
Externe Studie	66	112	48%	0,0270

Bedingte Power und neues nominelles Niveau

Fiktives Ergebnis der Zwischenanalyse:

- Altes nominelles Signifikanzniveau (einseitig): 0,025
- P-Wert des Logrank Tests: 0,2084
- Anzahl der Ereignisse: 27

Kalkulation des neuen nominellen Niveaus $\alpha_{\text{neu,+}}$ für:

- Anzahl der Ereignisse bei der ursprünglich geplanten Endanalyse: $k_1 = 28, \dots, 100$
- Anzahl der Ereignisse bei der neu geplanten Endanalyse: $k_2 = 28, \dots, 300$

Fazit

- Bei kalenderzeitabhängiger Planung einer Überlebenszeitstudie:
 - Verkürzung nicht möglich (k_1 bleibt unbekannt)
 - Verlängerung nur möglich, bei
 - genauer Festlegung des Ablaufs der Designänderung
 - strenger Vertraulichkeit der Studiendaten bis zur EA
- Gefahr bei adaptiven Plänen: schlechte a priori Planung, Manipulation, Verletzung der wissenschaftlichen Integrität
- Aber: adaptive Designs ermöglichen notwendige Änderungen der Studie unter Kontrolle des Signifikanzniveaus
- Insgesamt: je flexibler ein verwendetes Design ist, desto sorgfältiger muss die Planung der Studie und Formulierung des Studienprotokolls erfolgen!!!