
10. Symposium des Kompetenznetzes Leukämien GCP-Fortbildung

Statistik und ethische Aspekte bei klinischen Studien

Prof. Dr. Ulrich Mansmann
IBE, LMU, München
mansmann@ibe.med.uni-muenchen.de

Ethische Belange bei klinischen Studien

- Risiko-Nutzen Abwägung
- Autonomie der Studienindividuen

Nutzen

- Ein komplexes klinisches Konzept muss in ein operationalisierbares meßbares Konstrukt verwandelt werden:
Endpunkte, Strukturen von Nullhypothesen
- Verlängerung des Überlebens
- Ähnliches Überleben in den Therapiegruppen (**Non-Inferiority, Nicht-Unterlegenheit**) aber eine Verbesserung der Lebensqualität um mindestens 10 Scorepunkte (**Substantielle Überlegenheit**)
- Relevanz eines Biomarkers als prädiktiver Faktor: individualisierte Therapieentscheidung führt zu einem verbesserten DFS (disease free survival) in der relevanten Patientengruppe.
- Wie kompliziert kann das Nutzenkonzept werden?

Risiken

- Woher stammen die Informationen:
Investigators Brochure, Rote Liste, Studienregister
- Qualität der vorliegenden Information
- Rote Liste wird oft ignoriert
- Translationale Studien der Phase II:
Erstellen einer Investigators Brochure
- Umgang mit der in der Studie gewonnenen Information zu Risiken:
Oberflächliche explorative Beschreibung

Ethische Belange bei klinischen Studien

- Risiko-Nutzen Abwägung
- Autonomie der Studienindividuen
- **Wissenschaftliche Rechtfertigung**
- **Wissenschaftliche Qualität**
- **Prüfung der wissenschaftlichen Qualität und Biometrie-spezifischer Anforderungen durch die Ethikkommission**

Ethische Belange bei klinischen Studien

- Risiko-Nutzen Abwägung
 - Autonomie der Studienindividuen
 - **Wissenschaftliche Rechtfertigung**
 - **Wissenschaftliche Qualität**
 - **Prüfung der wissenschaftlichen Qualität und Biometrie-spezifischer Anforderungen durch die Ethikkommission**
- Studie als Vertrag zwischen Wissenschaftler und beteiligten Patienten über die Beantwortung einer relevanten Frage.
Respekt vor Studienteilnehmer

Risiko-Nutzen Abwägung

GCP-Leitlinie:

Vor Beginn einer klinischen Prüfung sollten die vorhersehbaren Risiken und Unannehmlichkeiten gegen den zu erwartenden Nutzen für den einzelnen Prüfungsteilnehmer und die Gesellschaft abgewogen werden. Eine klinische Prüfung sollte nur begonnen und fortgesetzt werden, wenn die zu erwartenden Vorteile die Risiken rechtfertigen.

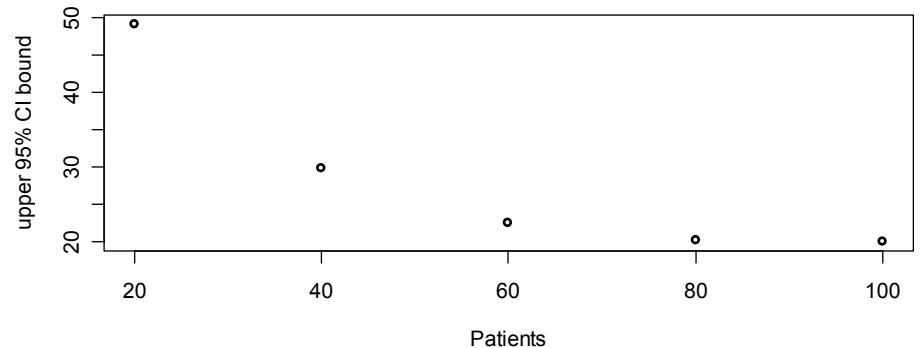
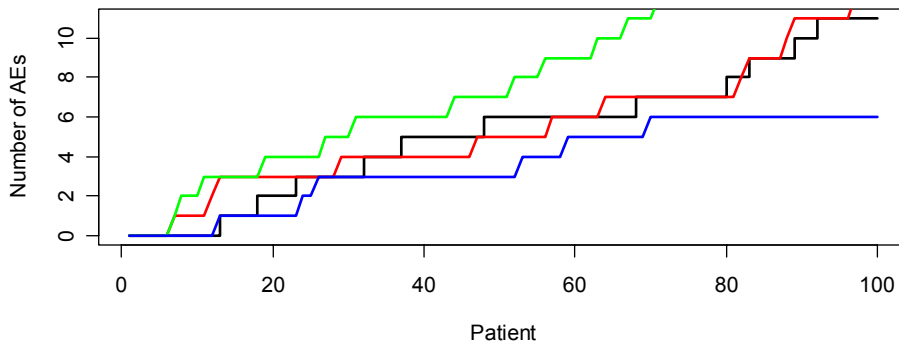
Die Rechte, die Sicherheit und das Wohl der Prüfungsteilnehmer genießen oberste Priorität und haben Vorrang vor den Interessen von Wissenschaft und Gesellschaft.

Risiko-Nutzen Abwägung

- Orientiert sich der Versuchsplan nur am angestrebten Wirkungsnachweis oder sieht er auch Maßnahmen zum Schutz der Patienten vor unerwünschten Nebenwirkungen vor?
 - Können Risiken für die Patienten möglichst früh erkannt werden? (z.B. durch sequentielle Kontrolle der Verträglichkeit) intern
 - Gibt es ein DMSB (Data Monitoring and Safety Board) mit Kompetenz bezüglich eines Studienabbruchs? extern
- Beschaffung und Interpretation von **quantitativer Information** über Risiken, Bearbeiten von Szenarien zur Nutzen/Risiko Bewertung.
- Gibt es klare Abbruchregeln, die durch Nutzen-Risiko-Abwägungen motiviert sind:
für schlechte Datenqualität und unzureichende Compliance,
für geplante und ungeplante Zwischenauswertungen,
für absehbaren Misserfolg, stochastic curtailment?
- Berücksichtigen die Abbruchkriterien im Protokoll Nutzen (futility) und Risiken?
Sehen sie eine sachgerechte Nutzen/Risiko-Abwägung vor?

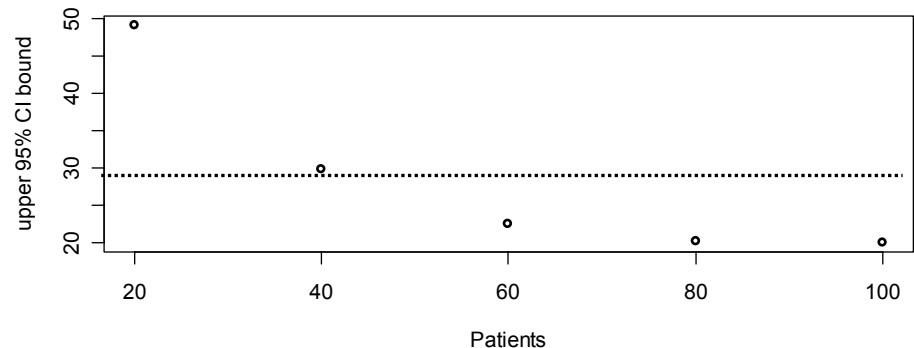
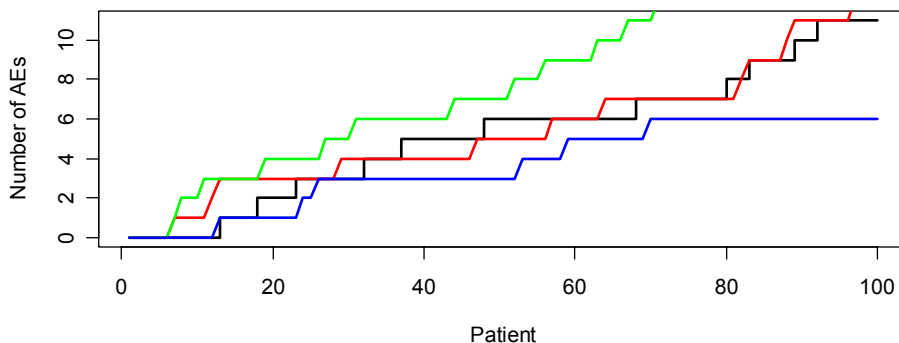
Biometrie spezifische Konsequenzen (I)

- Bewerten der vorliegenden quantitativen Information
Wie sicher / unsicher ist die Information \leftrightarrow Konfidenzintervalle
- *Stop for futility* nach Zwischenauswertung.
- Was ist in der Studie zu erwarten?
Szenarien demonstrieren die Aussagekraft verschiedener Maßnahmen:
Kontrolle der Vergleichsgruppe, Überschuss in der Therapiegruppe?
Formale Tests sind nicht aussichtsreich um Unterschiede zu erkennen.
Durchführen von Benchmarking
Bayesianische Verfahren: **Die W'keit, dass die Toxizitätsrate 10% übersteigt ist 0.8**



Biometrie spezifische Konsequenzen (I)

- Bewerten der vorliegenden quantitativen Information
Wie sicher / unsicher ist die Information \leftrightarrow Konfidenzintervalle
- *Stop for futility* nach Zwischenauswertung.
- Was ist in der Studie zu erwarten?
Szenarien demonstrieren die Aussagekraft verschiedener Maßnahmen:
Kontrolle der Vergleichsgruppe, Überschuss in der Therapiegruppe?
Formale Tests sind nicht aussichtsreich um Unterschiede zu erkennen.
Durchführen von Benchmarking
Bayesianische Verfahren: **Die W'keit, dass die Toxizitätsrate 10% übersteigt ist 0.8**



Regel: Wenn in der Therapiegruppe die obere Schranke der 95% KIs für die Toxizitätsrate über 30% liegt, gilt dies als Signal für nicht-tolerierbare Toxizität (Regel gilt ab einer Gruppengröße von 50 Patienten).

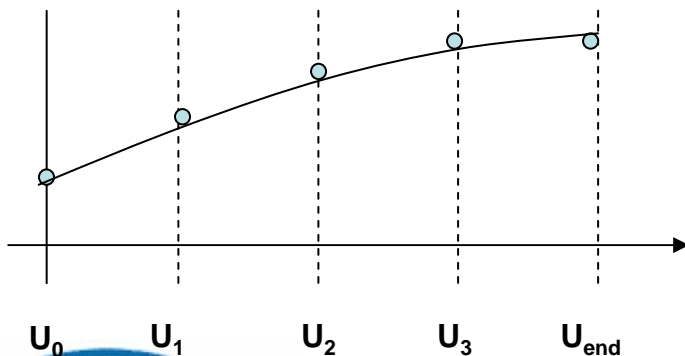
Autonomie von Studienteilnehmern

- Wird in geeigneter und verständlicher Weise über Tatsache und Art der (randomisierten) Therapiezuweisung informiert?
- Kann ein Patient die Studie verlassen, ohne deren Erfolg zu gefährden? Wie kann mit Drop-outs umgegangen werden?
- Was ist eine realistische Drop-out Rate? Konsequenzen für Fallzahl und Design; mit welchen Techniken wird die verloren gegangene Information imputiert? Ist ein Drop-out informativ?
- Eine Studie kann gut geplant sein, aber durch den Verlust von Studienteilnehmern findet ein enormer Powerverlust statt, der ein möglicherweise nicht-signifikantes Studienergebnis nicht mehr angemessen interpretierbar macht.

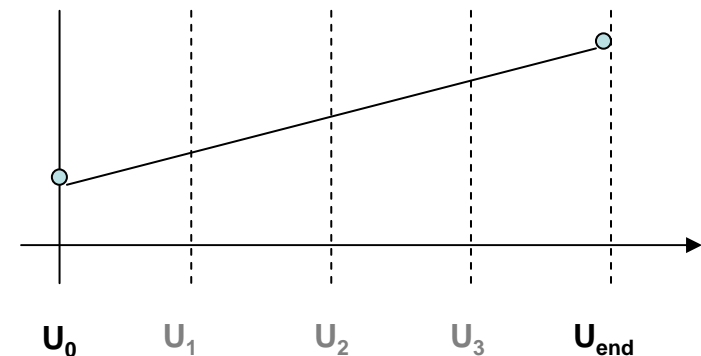
Biometrie spezifische Konsequenzen (II)

- Notwendigkeit eines klaren quantitativen Nutzen-Risiko-Konzeptes und dessen Kommunikation im Aufklärungsgespräch
- Zensierungskonzept bei Überlebenszeiten (Ereignisdaten):
Patient trägt solange zur Studie bei, wie er beobachtbar wird.
Ausscheiden unabhängig vom Krankheitsprozess: Nicht-informative \leftrightarrow **informative Zensur**
- Joint Modelling: Gleichzeitige Analyse von Überleben und Lebensqualität
Biomarker
- Missing at random \leftrightarrow missing not at random
- Verlaufsbeobachtungen versus einfache Vorher-Nachher-Beobachtung

longitudinal



Vorher-Nachher



Wissenschaftliche Rechtfertigung

- Rechtfertigung im Rahmen des vorliegenden Wissens
Rechtfertigung im Rahmen des Entwicklungsprozesses einer Therapie
- Ist die Studie nach aktuellem Erkenntnisstand angebracht und nötig?
Wurde der Kenntnisstand durch eine sachgerechte Synthese der existierenden Evidenz zur untersuchenden Frage (Meta-Analyse *lege artis*) ermittelt?
Wurde bei der Bewertung der Ergebnisse früherer Studien deren methodische Qualität berücksichtigt?
- Ist die Studie nach ihrem Typ richtig in die Entwicklung der Therapie eingeordnet?
Problem: Phase III vor Abschluss der Dosisfindung,
Problem: Ersatz einer pivotalen Zulassungsstudie durch mehrere Kleinstudien
Gibt es Pilot- und Vorstudien, die die Machbarkeit der geplanten Studie belegen.

Phase II: Entwicklung eines klinischen Konzeptes, Prüfen relevanter Aspekte, die im künftigen Konzept zusammengefügt werden.

Phase III: Überprüfung der Wirksamkeit und Sicherheit eines definierten klinischen Therapiekonzeptes.

Biometrie spezifische Konsequenzen (III)

- Zusammenfassung vorhandener Studienergebnisse:
Ist schon ausreichende Evidenz vorhanden? Kann die Frage schon durch eine Meta-Analyse vorhandener Studien beantwortet werden.
Methodik: Systematischer Review
Basiert auf guten Informationen (Reporting)
- Methodische Qualität: CONSORT-Statement (randomisierte Therapiestudie)
Hilft beim Lesen einer Studie,
hilft ebenso bei der Protokollerstellung und beim Schreiben eines Artikels.
A sufficient reporting quality seems achievable with a relatively small additional effort by editors and authors to adhere more strictly to the widely accepted guidelines of the Consolidated Standards on Reporting Trials (CONSORT).
- Reporting von Methoden (Protokoll) und Ergebnissen (Artikel) klinischer Forschung.

CONSORT-Statement

	Item number	Descriptor	Reported on page number
Title and abstract	1	How participants were allocated to interventions (eg, "random allocation", "randomised", or "randomly assigned").	
Introduction			
Background	2	Scientific background and explanation of rationale.	
Methods			
Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	
Objectives	5	Specific objectives and hypotheses.	
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (eg, multiple observations, training of assessors, &c).	
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	
Randomisation			
Sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (eg, blocking, stratification).	
Allocation concealment	9	Method used to implement the random allocation sequence (eg, numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were aware of group assignment. If not, how the success of masking was assessed.	
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses.	
Results			
Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group, report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	
Recruitment	14	Dates defining the periods of recruitment and follow-up.	
Baseline data	15	Baseline demographic and clinical characteristics of each group.	
Numbers analysed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention to treat". State the results in absolute numbers when feasible (eg, 10/20, not 50%).	
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (eg, 95% CI).	
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those prespecified and those exploratory.	
Adverse events	19	All important adverse events or side-effects in each intervention group.	
Discussion			
Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.	
Generalisability	21	Generalisability (external validity) of the trial findings.	
Overall evidence	22	General interpretation of the results in the context of current evidence.	

Checklist of items to include when reporting a randomised trial

Wissenschaftliche Qualität

- Ist die Studie zielführend und kann mit ihr eine klinisch relevante Frage beantwortet werden?
- Schutz der Patienten vor Behandlungen mit der unterlegenen Therapie, Vermeidung falscher Aussagen (Falsch positiv).
- Möglichst frühzeitiges Verfügbarmachen der überlegenen Therapie
- Schutz des Patienten vor Teilnahme an unnötigen Studien

Klinisch relevante Frage

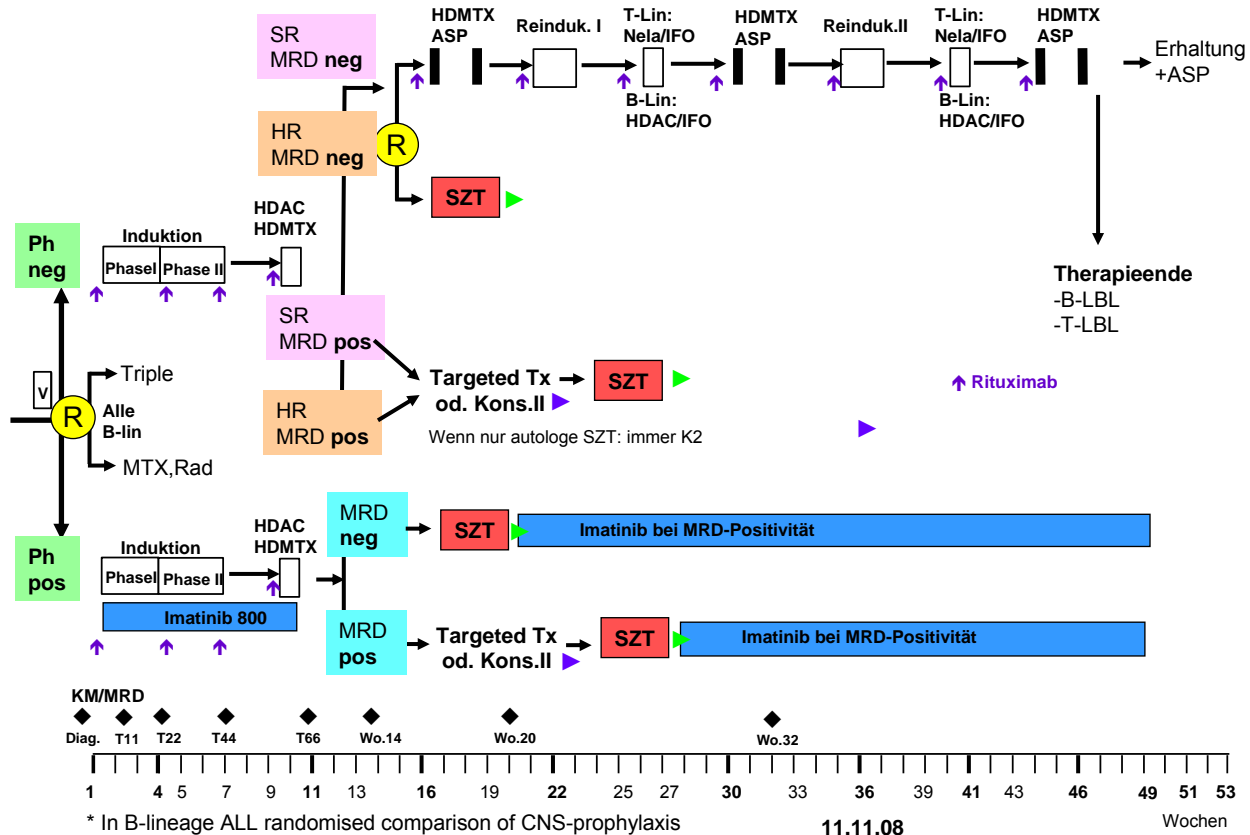
Endpunkte

OS
DFS
LPFS
Lebensqualität
Patient Outcome
Toxizität
Sicherheit
Biologische Outcomes
Marker
Response
...

Fragestellungen

- Es sind eine Reihe von Hypothesen **zielführend** in einer Studie zu prüfen.
- Dies bedarf einer gewissen Systematik, einer **Hypothesenordnung**, wobei sich verschiedenen Hypothesen auf verschiedene Studienpopulationen beziehen können.
- Statistische Verfahren erlauben dann ein methodisch sauberes Vorgehen beim **Testen** der Hypothesen:
 - Hierarchisches Testen, Gate keeping
 - Bonferroni unter Berücksichtigung logischer Randbedingungen
 - Abschlussprinzip
 - etc.

GMALL-Studie 08/2008



Schutz vor einer unterlegenen Therapie

- Entspricht das Protokoll den aktuellen Richtlinien?
- Ist das Studiendesign geeignet, die gestellte Frage zu beantworten?
- Entsprechen die Hypothesen der klinischen Fragestellung?
- Ist das Zielkriterium sachgerecht:
Wirkung oder Wirksamkeit, Begründung von Surrogatendpunkten,
Einbeziehung von Lebensqualität
- Wurden Validität und Reliabilität des Zielkriteriums geprüft?
- Entspricht die Studienpopulation der avisierten med. Indikation?
- Sind die Irrtumswahrscheinlichkeiten richtig gewählt?
- Wird Verzerrungen durch Störgrößen vorgebeugt? (Strat., Adj.)
- Sind Maßnahmen zur Sicherung der Datenqualität vorgesehen?

Biometrie spezifische Konsequenzen (IV)

- Richtlinien für die Biometrie (z.B. ICH E9), Richtlinien für indikations-spezifische Outcome-Messung (Zielkriterien: sachgerecht, valide)
- Auswahl eines geeigneten Studiendesigns: Parallellgruppe, (gruppen-) sequentiell, adaptiv (mit Modifikationsmöglichkeiten während der Studie)
- Entwickeln einer Hypothesenordnung um verschiedene Aspekte der Fragestellung zu formalisieren.
- Sind die Irrtumswahrscheinlichkeiten (alpha, beta, Power) richtig gewählt? Ist eine Fallzahlberechnung erfolgt und somit der Umfang der Studie geklärt, die eine zielführende Antwort auf die Fragestellung erlaubt?
- Verwendung univariater, stratifizierter oder multivariater Analyseverfahren (Strat., Adj.)
- System von Plausibilitätsprüfungen, Zusammenarbeit Biometrie mit Monitoren, Verwendung von eCRFs, ...

Verfügbarmachen der überlegenen Therapie

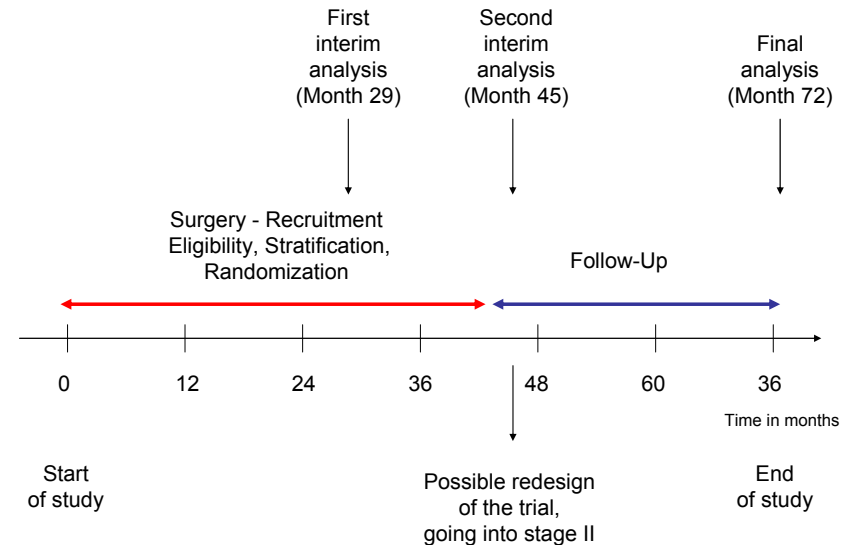
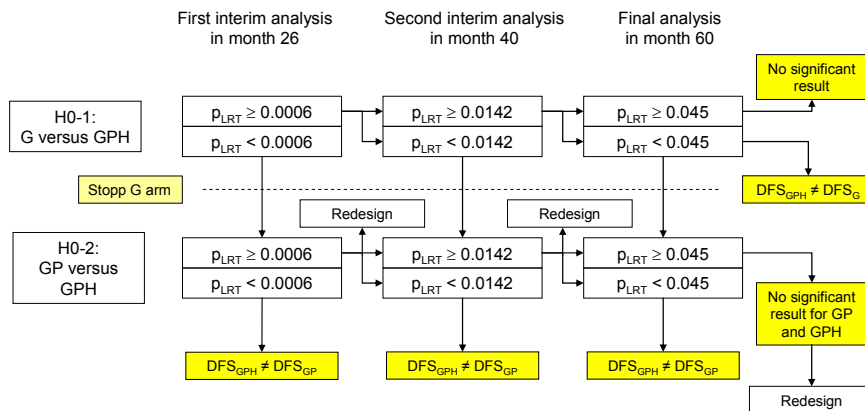
- Ist die Fallzahlplanung sachgerecht und richtig?
- Wird frühes Erkennen der Überlegenheit durch das Studiendesign ermöglicht?
- Gibt es ein unabhängiges Komitee mit Kompetenz bezüglich eines Studienabbruches aufgrund der Ergebnisse einer Zwischenauswertung?
- Sind die Abbruchregeln für eine Zwischenauswertung sinnvoll vorgegeben?

Biometrie spezifische Konsequenzen (V)

- Gruppensequentielle Studien: Plane konservativ, früher Abbruch bei starkem Effekt.

Statistical procedure
for efficacy analysis

3-stage group-sequential trial for two hierarchically
ordered null hypotheses

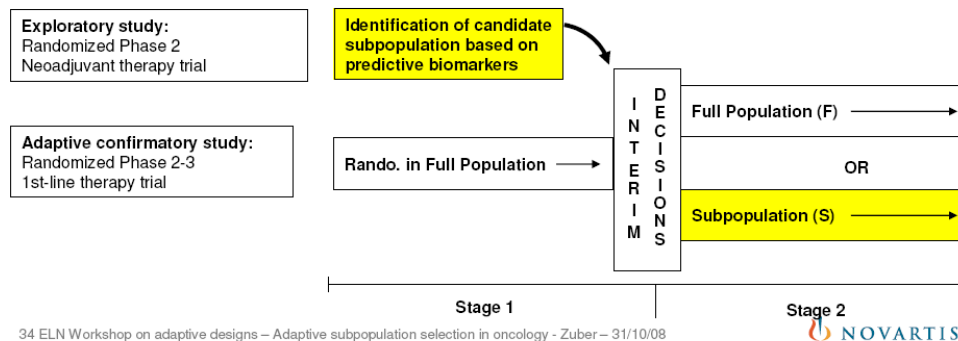


Schutz vor einer unnötigen Studie

- Gibt es zu viele Fragestellungen in der Studie? Multiple Endpunkte ohne klare Prioritäten, multiple Vergleichsgruppen?
- Gibt es eine **echte** Vergleichsgruppe?
- Ist die der Fallzahlplanung zugrunde gelegte relevante Differenz sachgerecht?
 Δ groß \rightarrow Fallzahl klein \rightarrow Gefahr für geringe Power \rightarrow Studie kann keine klare Antwort geben
- Reichen die real verfügbaren Patienten aus, um die Fallzahl im gegebenen Zeitraum zu erreichen?
- Ist das Vorgehen bei „Drop-outs“ oder Protokollverletzern festgelegt?

Biometrie spezifische Konsequenzen (VI)

- Planungsannahmen einer Studie unterliegen immer der Gefahr in der gegebenen Situation unrealistisch zu sein: Mittlerweile gibt es statistische Verfahren, die eine Designmodifikation erlauben und dabei den konfirmatorischen Charakter der Studie nicht beeinträchtigen.
- Es gibt die Möglichkeit in der gleichen Studie eine explorative Phase vor die konfirmatorische Phase zu schalten. Unterschied zur Pilotstudie: Daten der explorativen Phase können in beiden Phasen genutzt werden.



Reporting

**Good reporting is not an optional extra:
it is an essential component of
good research**

“Reporting is not just about telling people what happened in the past. It is a tool for changing the future. Certainly, ‘good’ reporting must be accurate and complete. But this is not sufficient to make a difference. Crucial is the value placed on reporting by the organisation itself.”

Simon Zadek. The impact of public reporting. Oct 2002

<http://www.ethicalcorp.com/content.asp?ContentID=223>

Reporting

519 RCT aus dem Jahr 2000

Wichtige Elemente zu Kernpunkten des Designs wurden nicht berichtet:

- 73% Fallzahlberechnung
- 55% Unklare Definitionen der Endpunkte
- 60% Keine Angaben über Verblindungsmaßnahmen
- 79% Keine Angaben zur Erzeugung der Zufallssequenz für die Randomisierung
- 82% Keine Angaben über die Durchführung der Randomisierung

Chan & Altman (2005) The Lancet,

Typische Erfahrung bei einem Cochrane Review:

Data reporting was poor. 15 trials met the inclusion criteria for this review but only 4 could be included as data were impossible to use in the other 11.

Verfügbarmachen von Protokollen

- Wissenschaftliche Veröffentlichungen sind die wichtigste Form Forschungsergebnisse darzustellen und einer größeren *Community* zugänglich zu machen.
- Qualität und Relevanz der Methoden und Ergebnisse werden durch den veröffentlichten Bericht wahrgenommen und eingeschätzt.
- Der Bericht (Protokoll) ist oft das einzige öffentliche Dokument, das die Bearbeitung der Forschungsfrage belegt.
- Unklare Berichte zu Methoden und Ergebnissen erschweren die objektive Einschätzung der Ergebnisse.
- Schlechtes Berichten von Methoden und Ergebnissen aus der klinischen Forschung impliziert Gefahren für Patienten durch nicht effektive Behandlungsformen.
- Es gibt eine überwältigende Evidenz für das schlechte Berichten von Ergebnissen aus der klinischen Forschung.

Zusammenfassung

- Umsetzung ethischer Prinzipien in der klinischen Forschung haben zu einer enormen statistischen Methodenentwicklung in den vergangenen 20 Jahren geführt.
- Breites statistisches Instrumentarium zur Umsetzung zielführender klinischer Studien mit komplexen Fragestellungen ist vorhanden.
- Dennoch sollten Studien methodisch so einfach wie möglich sein. Abweichungen von dieser Regel benötigen eine gewichtige Begründung: Wichtige Fragen in sehr beschränkten Patientenpopulationen.
- Studienregister zu spezifischen Krankheitsbildern in Studienzentralen sind eine wichtige Informationsquelle für Fragen bei der Planung von Studien. Sie können verwendet werden um Studienszenarien mit dem Studienteam durchzuspielen (Trial Simulation). Dies erleichtert die Wahl zwischen verschiedenen Studiendesigns.
- Protokollentwicklung wird vermehrt auf statistische Simulationen zurückgreifen.
- Die genannten Prinzipien sind für die Begutachtung von Protokollen relevant.

Zusammenfassung

Kapitel 6 der GCP ICH Topic E 6 (R1) Richtlinie für Good Clinical Practice (CPMP/ICH/135/95)

6.	CLINICAL TRIAL PROTOCOL AND PROTOCOL AMENDMENT(S)	36
6.1	GENERAL INFORMATION	36
6.2	BACKGROUND INFORMATION	36
6.3	TRIAL OBJECTIVES AND PURPOSE	37
6.4	TRIAL DESIGN.....	37
6.5	SELECTION AND WITHDRAWAL OF SUBJECTS	37
6.6	TREATMENT OF SUBJECTS.....	38
6.7	ASSESSMENT OF EFFICACY	38
6.8	ASSESSMENT OF SAFETY.....	38
6.9	STATISTICS	38
6.10	DIRECT ACCESS TO SOURCE DATA/DOCUMENTS.....	39
6.11	QUALITY CONTROL AND QUALITY ASSURANCE	39
6.12	ETHICS	39
6.13	DATA HANDLING AND RECORD KEEPING	39
6.14	FINANCING AND INSURANCE.....	39
6.15	PUBLICATION POLICY	39
6.16	SUPPLEMENTS.....	39