

Workshop „Durchführung klinischer Prüfungen nach ICH-GCP“  
(Würzburg, 8. Dezember 2006)



Zentrale für Klinische Studien  
am Universitätsklinikum Würzburg

# **CONsolidated Standards Of Reporting Trials (CONSORT Statement) – Implikationen für klinische Prüfungen**

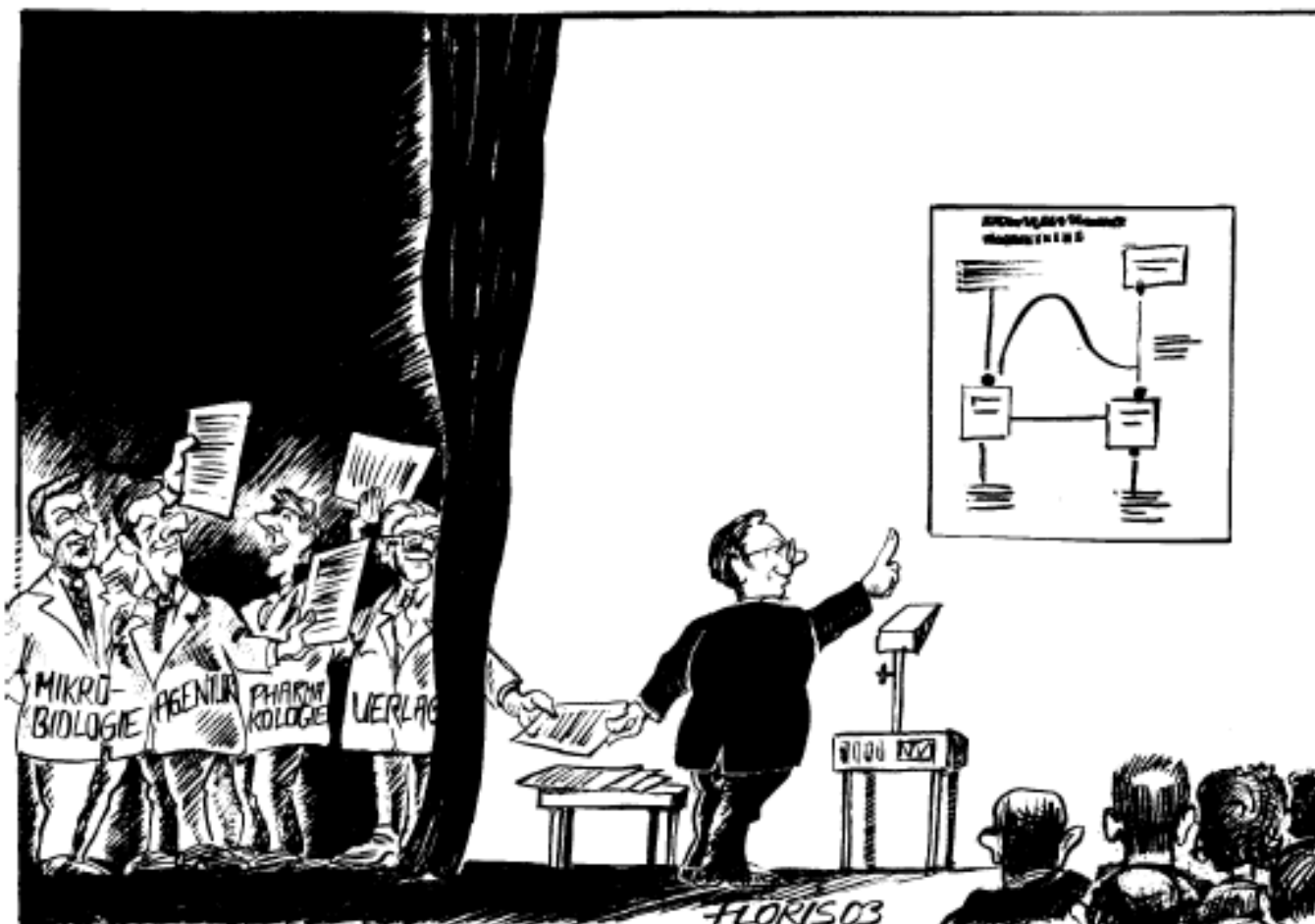
Wilmar Igl, Dipl.-Psych.

Abteilung Biometrie

Zentrale für Klinische Studien (ZKS)  
am Universitätsklinikum Würzburg

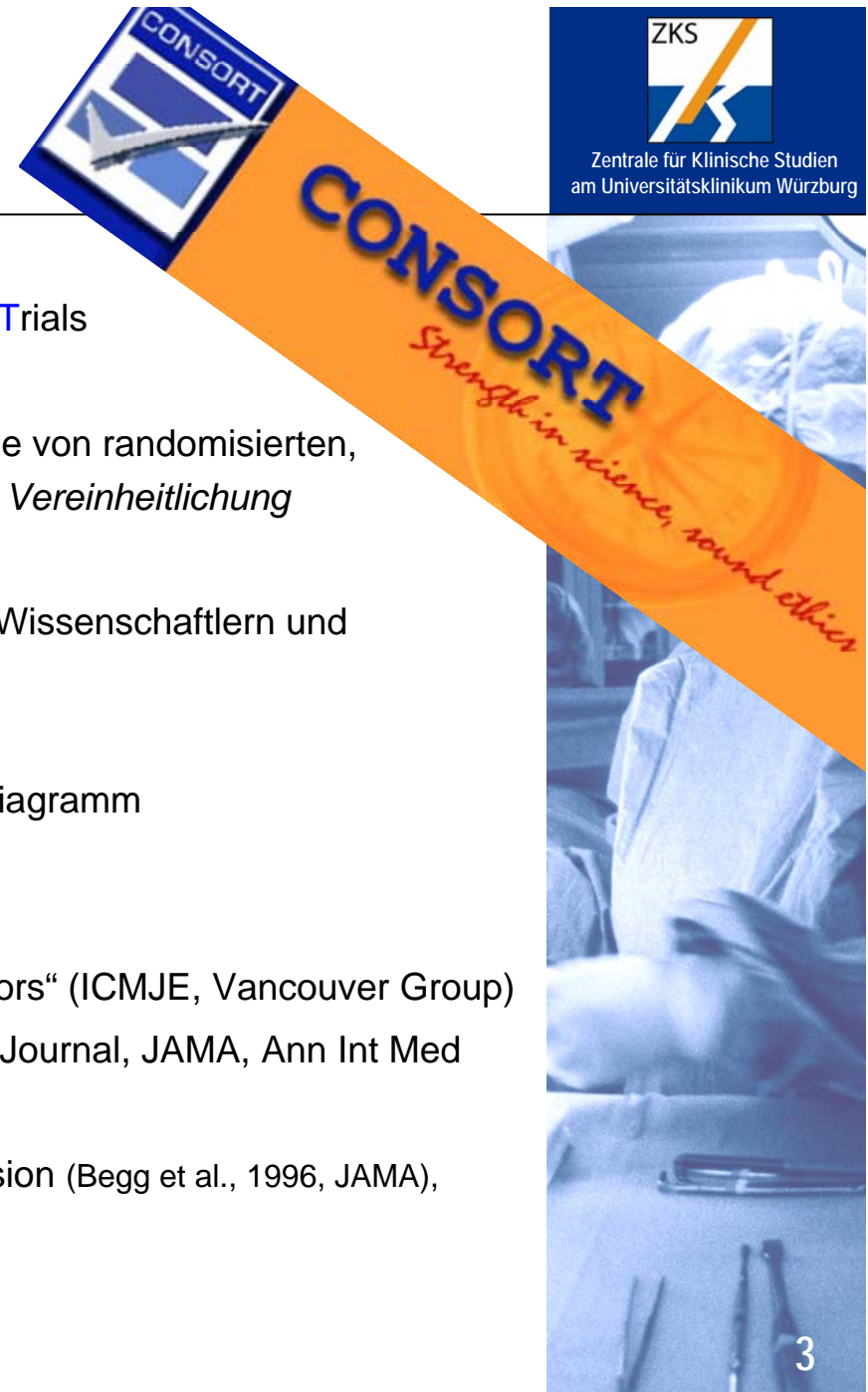


# Transparenz der Forschung?



# CONSORT – Was ist das?

- **CONSORT = CONS**olidated **Standards Of Reporting Trials**
- **Ziel:** Verbesserung der Dokumentation der Ergebnisse von randomisierten, kontrollierten Studien durch höhere *Transparenz* und *Vereinheitlichung*
- **Autoren:** internationale, interdisziplinäre Gruppe von Wissenschaftlern und Herausgebern aus dem biomedizinischen Bereich
- **Inhalt:** Checklisten (5 Bereiche, 22 Items) und Flussdiagramm
- **breite Unterstützung:**
  - „International Committee of Medical Journal Editors“ (ICMJE, Vancouver Group)
  - > 50 Journals, z. B. The Lancet, British Medical Journal, JAMA, Ann Int Med
- **kontinuierliche Weiterentwicklung:** ursprüngliche Version (Begg et al., 1996, JAMA), revidierte Version (Moher et al., 2001, Lancet u. a.)



## Relevanz für klinische Prüfungen

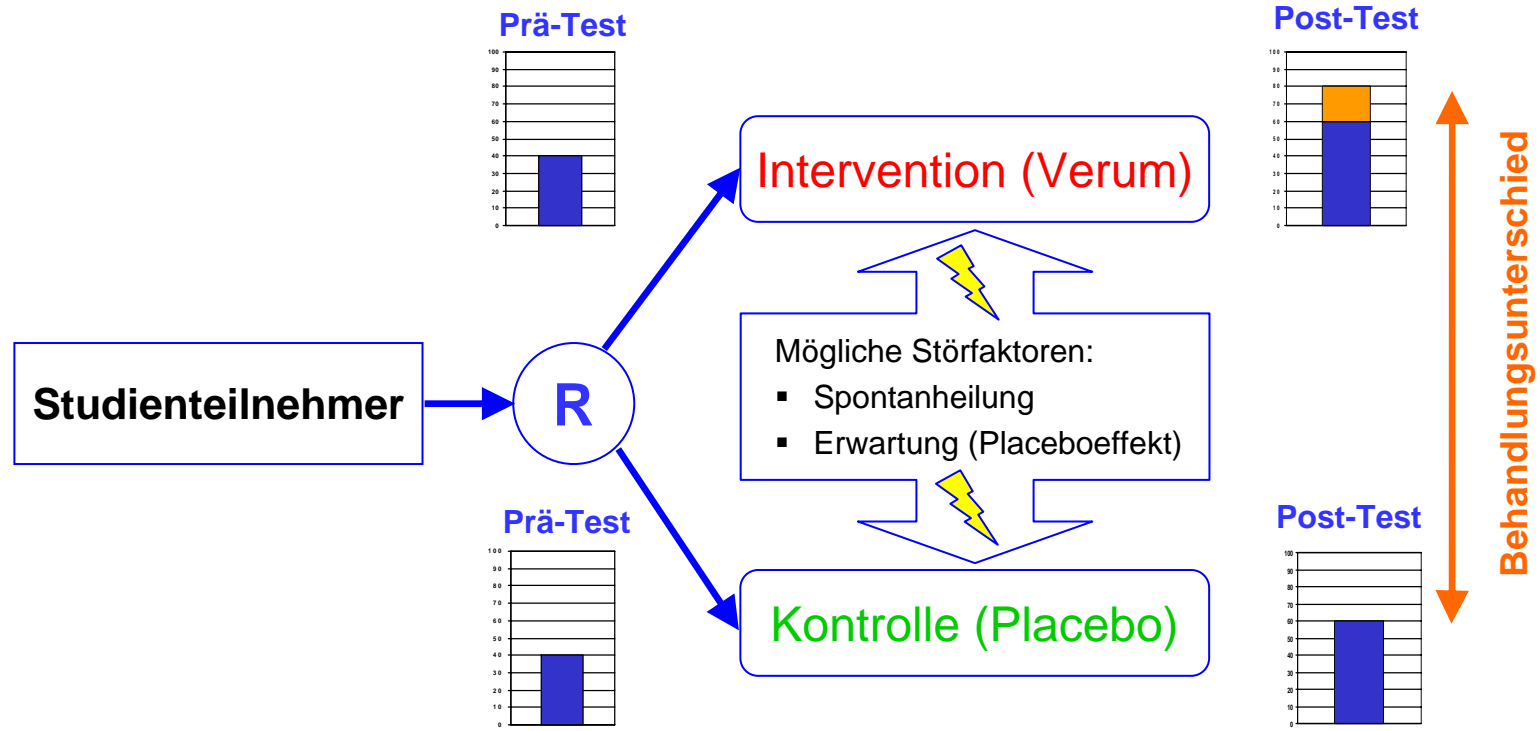
- **Verstehen und Bewerten von Studien wird erschwert** durch unzureichende Dokumentation (z.B. fehlende statistische Kennwerte für Meta-Analysen).
- **Nicht-Berichten** von Informationen hängt mit **Verzerrung („bias“)** der Ergebnisse zusammen (Schulz et al., 1995, JAMA; Moher et al., 1998, Lancet).
- CONSORT ist nicht nur anwendbar zur Dokumentation, sondern auch zur **Planung und Durchführung von Studien** (O'Toole, 1997, BMJ):
  - **Vermeidung methodischer Fehler**, welche die Aussagekraft der Studie beeinträchtigen
  - Erhöhung der Chancen, die **Ergebnisse wirksam publizieren** zu können



# Die randomisierte, kontrollierte Studie (Randomised Controlled Trial, RCT)



# Experimentelle Prüfung durch RCT



R = Randomisierung



# CONSORT Statement

CONsolidated Standards Of Reporting Trials



# Bestandteile des CONSORT-Statements

- Titel und Zusammenfassung
- Einleitung
- **Methoden**
  - Probanden/Patienten
  - Intervention/Behandlung
  - Ziele, Zielkriterien, Fallzahlbestimmung
  - Randomisierung, Verblindung
  - statistische Methoden
- **Ergebnisse**
  - Ein- und Ausschlüsse
  - Untersuchungszeitraum, Messzeitpunkte
  - Stichprobenmerkmale und -größe (vgl. **Flussdiagramm**)
  - Ergebnisse (primär, sekundär) und Schätzmethoden
- **Diskussion**
  - Interpretation
  - Generalisierbarkeit
  - Bewertung der Evidenz





# CONSORT

# Methodik



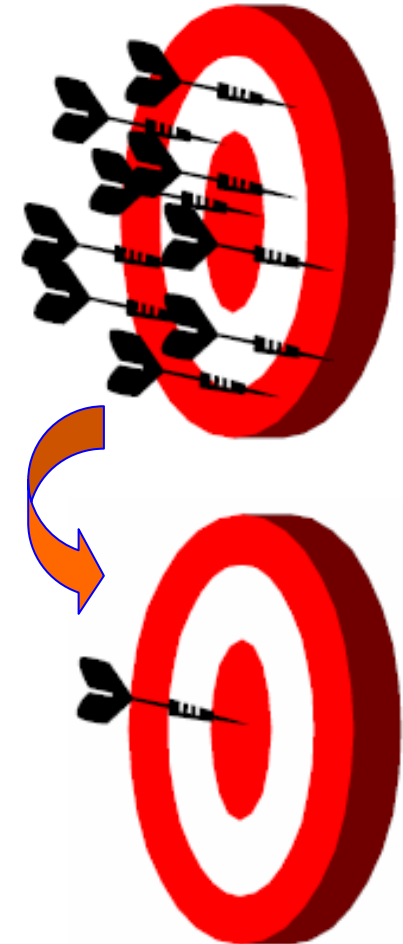
## Fragestellung und Endpunkte (#5, #6)

### Empfehlung:

- eindeutige Formulierung der Fragestellung
- eindeutige Benennung von primären und sekundären Endpunkten
- Festlegung des entscheidenden, primären Messzeitpunkts

### Begründung:

- Die Fragestellung hat zentrale, strukturierende Bedeutung für die gesamte Studie, insbesondere Studiendesign, Endpunkte, Fallzahlplanung und statistische Analyse
- Verwendung von mehreren Endpunkten und Messzeitpunkten führt zu Verzerrungen auf Grund von Multiplizität.



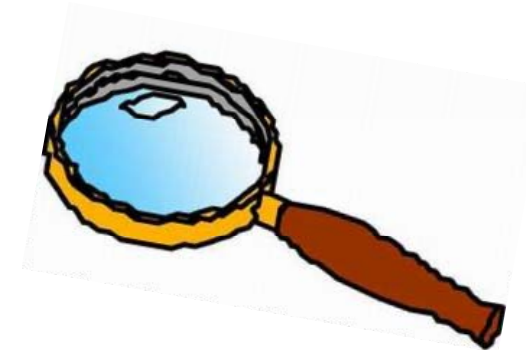
## Fallzahlschätzung (#7)

### Empfehlung:

- Beschreibung des Vorgehens bei der Fallzahlschätzung
- Definition des minimalen klinisch bedeutsamen Unterschieds (MCID) bzw. des erwarteten Effekts
- Berücksichtigung von Studienabbruchern

### Begründung:

- Der Nachweis von Effekten erfordert umso größere Stichproben, je kleiner der Effekt ist.
- Viele Studien mit kleinen Stichprobengrößen schließen irrtümlich auf fehlende Behandlungsunterschiede auf Grund mangelnder Power (Altman & Bland, 1995, BMJ).



## Randomisierung (#8)

### Empfehlung:

- Randomisierung mit geeigneten Methoden (z.B. Tabellen, Software)
- ausreichende Angaben zur Randomisierungsmethode
- Benennung der Einschränkungen (z. B. Blockbildung, Stratifizierung)



### Begründung:

- Studien mit fehlenden Angaben zur Randomisierung neigen zu Verzerrungen (Schulz et al., 1995, JAMA).
- In weniger als 50% aller Studien wird die Randomisierungsmethode angemessen erläutert (Schulz et al., 1994, JAMA).
- „Zufällig“ wird häufig nicht in seiner exakten, statistischen Bedeutung gebraucht (z. B. Zuweisung nach Geburtsdatum, Anreisedatum,...).



## Concealed Allocation, Verblindung (#9, #10, #11)

### Empfehlung:

- Trennung von Erstellung und Zuordnung der Randomisierungscode
- externe Randomisierung
- Begründung von fehlender Verblindung
- Kontrolle und Angaben über den Erfolg der Verblindung

### Begründung:

- Studien mit fehlenden Angaben zur verblindeten Zuordnung (concealed allocation) neigen zu Verzerrungen (Schulz et al., 1995, JAMA).
- Fehlende Verblindung von Behandler, Beurteiler oder Auswerter führt zu progressiven Ergebnissen (Noseworthy et al., 1994, Neurology).
- Bei bis zu 38% der untersuchten Studien fehlen Angaben zur Verblindung (Cheng et al., 2000, Pediatr Pulmonol).



# CONSORT

# Ergebnisse



## Flussdiagramm der Patienten (#13)

### Empfehlungen:

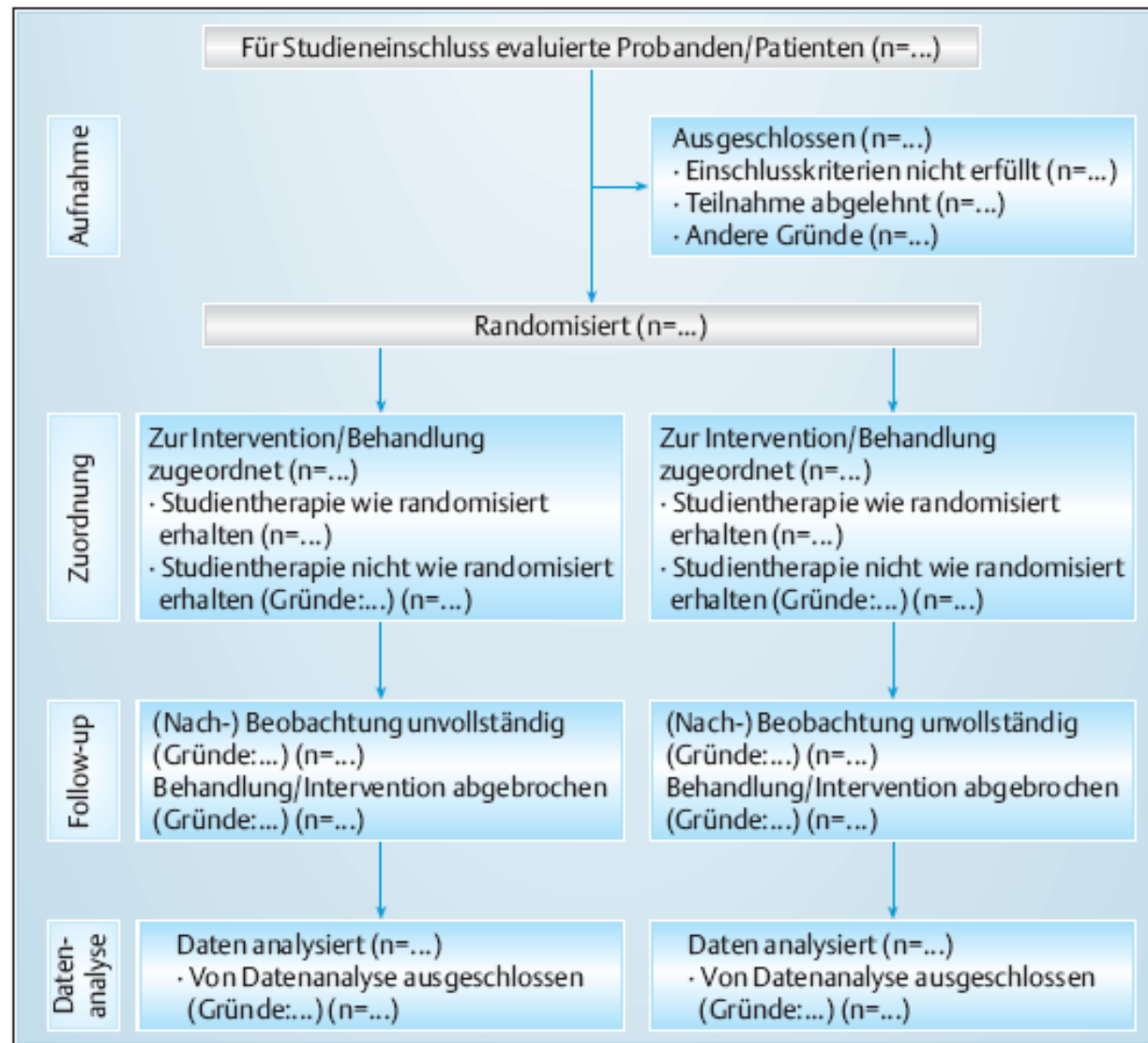
- Angaben der Patientenzahlen in den Behandlungsgruppen über alle Messzeitpunkte der Studie
- Durchführung von *Per-protocol-* und *Intention-to-treat*-Analysen zur Abschätzung des maximalen und minimalen Effekts
- Beschreibung von Abweichungen vom Studienprotokoll

### Begründung:

- Einschränkungen der Generalisierbarkeit und der Wirkung der Intervention durch selektiven *Drop-out* (*Loss-to-follow-up*) möglich (z. B. durch negative Nebenwirkungen)
- Angaben zu eligiblen Patienten und *Loss-to-follow-up* fehlen in bis zu 20% der geprüften Studien (Egger et al., 2001, JAMA).

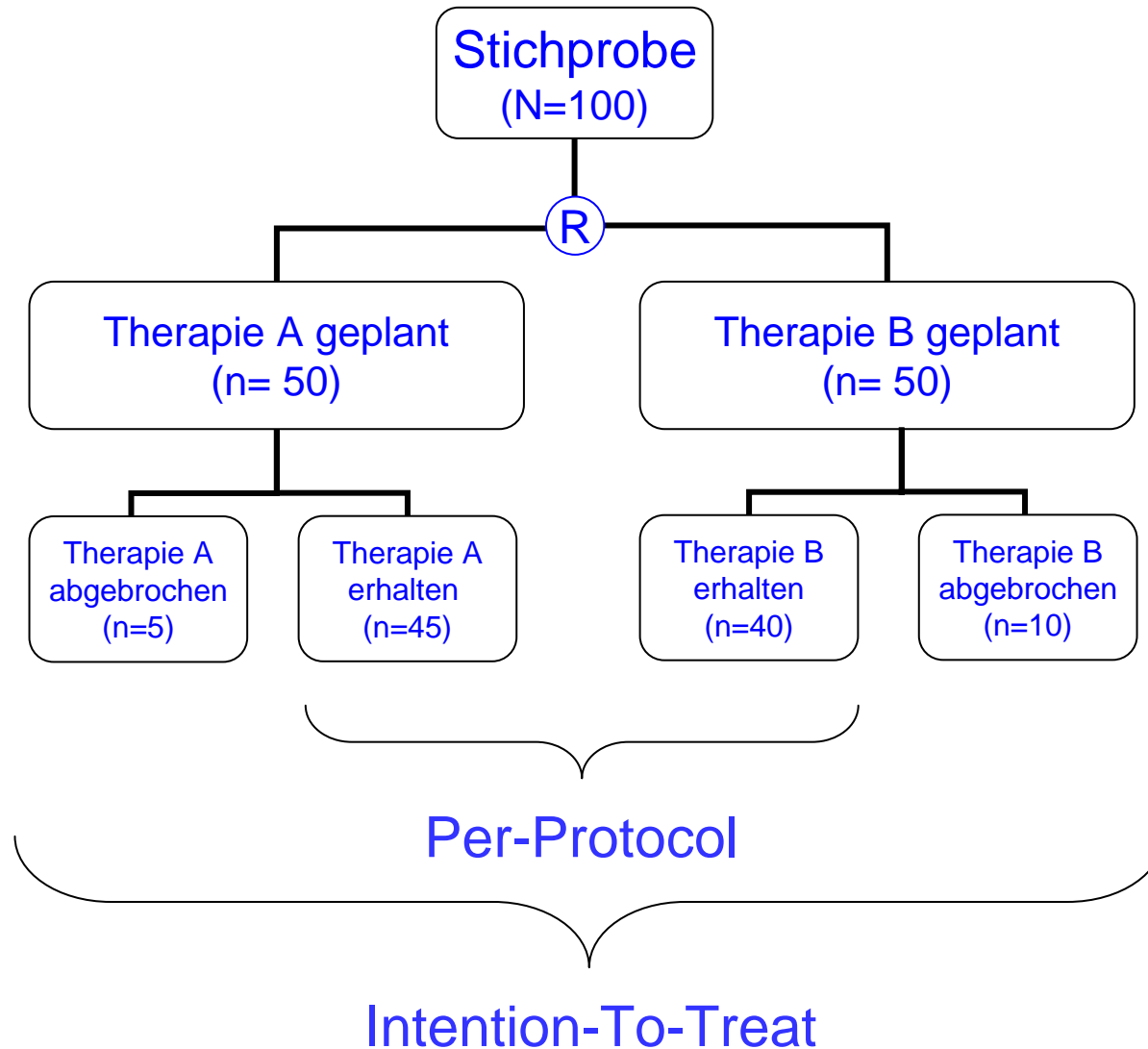


# Flussdiagramm (Vorlage)





## Beispiel: Studie mit Abbrechern



## Stichprobenmerkmale (#16)

### Empfehlungen:

- (tabellarische) Angaben zu Geschlecht, Alter, Diagnosen u. a. getrennt nach Behandlungsgruppen
- Angaben zur Variabilität kontinuierlicher Variablen (SD, nicht SE/CI !)
- Beurteilung von Baselineunterschieden (Signifikanztests nicht angemessen, klinische Bedeutsamkeit wichtiger).

### Begründung:

- Einschlusskriterien lassen nur begrenzte Aussagen über die tatsächlich untersuchte Patientengruppe zu.

(vgl. Chappell et al., 1999, Lancet)



**Table 6. Item 15: Example of Reporting of Baseline Demographic and Clinical Characteristics of Trial Groups†**

Characteristic	Vitamin Group (n = 141)	Placebo Group (n = 142)
Mean age $\pm$ SD, y	28.9 $\pm$ 6.4	29.8 $\pm$ 5.6
Smokers, n (%)	22 (15.6)	14 (9.9)
Mean body mass index $\pm$ SD, kg/m <sup>2</sup>	25.3 $\pm$ 6.0	25.6 $\pm$ 5.6
Mean blood pressure $\pm$ SD, mm Hg		
Systolic	112 $\pm$ 15	110 $\pm$ 12
Diastolic	67 $\pm$ 11	68 $\pm$ 10
Parity, n (%)		
0	91 (65)	87 (61)
1	39 (28)	42 (30)
2	9 (6)	8 (6)
>2	2 (1)	5 (4)
Coexisting disease, n (%)		
Essential hypertension	10 (7)	7 (5)
Lupus or antiphospholipid syndrome	4 (3)	1 (1)
Diabetes	2 (1)	3 (2)

† Adapted from part of Table 1 of reference 111.

## Statistische Methoden und Ergebnisse (#12, #17)

### Empfehlungen:

- (tabellarische) Darstellung der Ergebnisse (primäre, sekundäre Endpunkte)
- Angabe von (standardisierten) Effektstärken (z. B. Cohen's d, Odds Ratio, Hazard Ratio u. a.)
- Verwendung von 95%-Konfidenzintervallen als Maß für statistische Bedeutsamkeit und den wahrscheinlichen wahren Effekt
- Angabe von exakten  $p$ -Werten (z. B.  $p = 0.021$  statt  $p < 0.05$ )
- Berücksichtigung von Kovariaten bei Baselineunterschieden der Gruppen z.B. Kovarianzanalyse (ANCOVA), Regressionsanalysen

### Begründung:

- Unkorrekte Anwendung statistischer Methoden in bis zu 63% der untersuchten Studien (Gotsche, 1989, Control Clin Trials)
- Ergebnisse von statistischen Tests ( $p$ -Werte) sind unzureichend, da sie wesentlich von der Stichprobengröße abhängen.
- Baselineunterschiede verzerren Behandlungseffekte (z.B. Regression zur Mitte).



## Beispiel: Ergebnisdarstellung

End Point	Etanercept Group (t1 / t2) (n=30)	Placebo Group (t1 / t2) (n=30)	Difference (95%-CI)	p- Value
Unit of Measurement	%	%	%	
<b>Primary Outcome</b>				
Psoriatic Arthritis Response criteria	26 / 87	7 / 23	63 (44-83)	< 0.001
<b>Secondary Outcome</b>				
Proportions of patients meeting ACR criteria				
ACR20	22 / 73	4 / 13	60 (40-80)	< 0.001
ACR50	15 / 50	1 / 3	47 (28-66)	<0.001
ACR70	4 / 13	0 / 0	13 (1-26)	0.04

(vgl. Mease et al., 2000, Lancet)



## Exkurs: Prüfung der Wirksamkeit vs. Sicherheit (I)

- Bei klinischen Prüfungen ist nicht nur die **Wirksamkeit** (*effectiveness*), sondern auch die **Sicherheit** (*safety*) der Therapie wichtig.
- **Unterschiedliche statistische Vorgehensweisen:**
  - Prüfung auf Überlegenheit (der Wirksamkeit)
  - Prüfung auf Nicht-Unterlegenheit bzw. Äquivalenz (der Sicherheit)
- **Mögliche Ansätze** zur Prüfung der Sicherheit (Prüfung auf Nicht-Unterlegenheit bzw. Äquivalenz):
  1. **Deskriptive Prüfung:**

Beispiel: „*Es kann gleiche Sicherheit angenommen werden, wenn der Unterschied der Komplikationsrate zwischen den Behandlungen weniger als 10% beträgt.*“
  2. **Inferenzielle Prüfung:**

Beispiel: „*Es kann gleiche Sicherheit angenommen werden, wenn der wahre Unterschied der Komplikationsrate mit einer Wahrscheinlichkeit von 95% im Intervall [-10%;+10%] liegt.*“



# Exkurs: Prüfung der Wirksamkeit vs. Sicherheit (II)

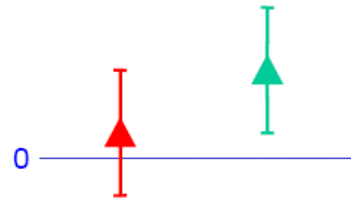
 = Wirkungsunterschied mit 95%-KI

rot = Nullhypothese beibehalten

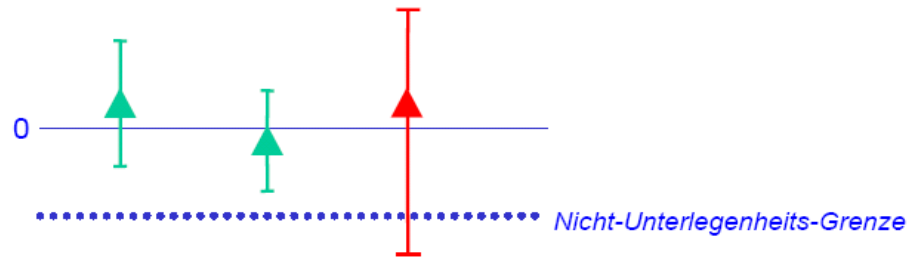
grün = Alternativhypothese angenommen

## Studienziel

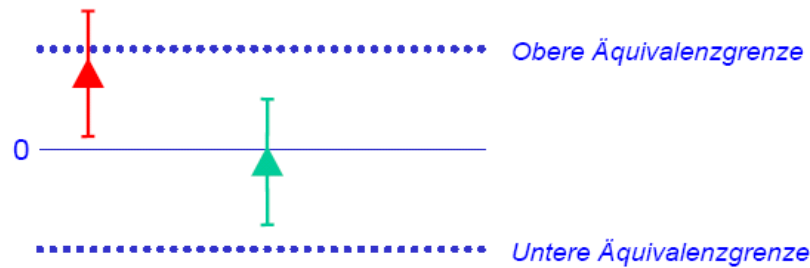
Überlegenheit:



Nicht-Unterlegenheit:



Äquivalenz:



nach Meinert, 2005



# CONSORT

## Diskussion



## Interpretation (#20)

### Empfehlung:

- Strukturierung der Diskussion
  - Zusammenfassung
  - Erklärungen für die gefundenen Ergebnisse
  - Einordnung in bisherige Forschung
  - Einschränkungen der Aussagekraft (!)
  - Auswirkungen auf Forschung und Praxis
- Unterscheidung zwischen statistischer Signifikanz und klinischer Bedeutsamkeit
- Diskussion der Genauigkeit der Ergebnisse („Ist bei fehlender statistischer Signifikanz möglicherweise ein klinischer bedeutsamer Effekt im Konfidenzintervall enthalten?“)



### Begründung:

- Ergebnisse werden nicht ausreichend kritisch bewertet (Horton, 1995, BMJ).



# Zusammenfassung



## CONSORT in a nutshell ...

- klare Herleitung und Definition der **Fragestellung**
- Festlegung der **primären und sekundären Endpunkte**
- Untersuchung **ausreichender Fallzahlen** (Fallzahlschätzung)
- **Kontrolle von Störfaktoren** (Randomisierung, Verblindung)
- Auswertung mit **angemessenen (statistischen) Methoden**  
(Intention-to-treat, Berücksichtigung von Baselineunterschieden)
- Darstellung der Ergebnisse mit **aussagekräftigen Statistiken**  
(Effektstärken, Konfidenzintervalle)
- **(selbst)kritische Diskussion** der Bedeutung der Ergebnisse



## Literatur

- Moher, D., Schulz, K. F., Altman, D. G. für die CONSORT Gruppe (2004). Das CONSORT Statement: Überarbeitete Empfehlungen zur Qualitätsverbesserung von Reports randomisierter Studien im Parallel-Design. *Deutsche Medizinische Wochenschrift*, 129, T16-T20.
- (!) Moher, D., Schulz, K. F., Altman, D. G., for the CONSORT Group (2001). The CONSORT statement: Revised recommendations for improving the quality of parallel-group randomised trials. *The Lancet*, 357, 1191-1194.
- (!) Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gøtzsche, P. C., Lang, T., for the CONSORT Group (2001). The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. *Annals of Internal Medicine*, 134, 663-694.
- Begg, C. B., Cho, M. K., Eastwood, S., Horton, R., Moher, D., Olkin, I., Rennie, D., Schulz, K. F., Simel, D. L., Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *Journal of the American Medical Association*. 276, 637-639.
- Campbell, M. K., Elbourne, D. R., Altman, D. G. (2004). CONSORT Statement: Extension to cluster randomised trials. *British Medical Journal*, 328, 702-708.



***Vielen Dank!***

**Kontakt: [igl.w@zkws.de](mailto:igl.w@zkws.de)**

