

Multivariate Verfahren

Einheit 4: Mehrebenenanalyse - Lineare gemischte Modelle (3)

Wintersemester 2025 | Prof. Dr. Stephan Goerigk



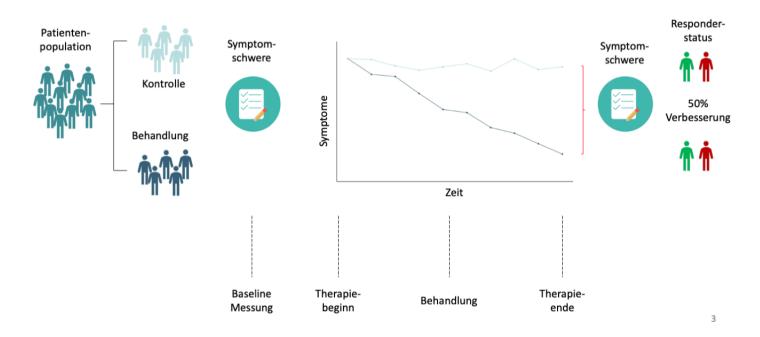
Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

- Klinische Trials sind systematische Studien, die dazu dienen, die Sicherheit und Wirksamkeit klinischer Interventionen wie Medikamente, psychotherapeutische Behandlungen oder neue Therapieverfahren zu testen.
- Ziele klinischer Trials:
 - Sicherheit und Wirksamkeit nachweisen
 - Vergleich mit bestehenden Standardbehandlungen
 - Neue Interventionen validieren



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Beispiel: Das randomisiert kontrollierte Trial (RCT)





Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Phasen Klinischer Studien

Phase	Ziel	Teilnehmer	Fokus
Phase I (Erste Studien am Menschen)	Prüfung der Sicherheit und Festlegung einer Dosierung (bei Medikamenten)	Kleine Gruppe gesunder Freiwilliger (oder Patienten bei riskanten Therapien)	Toxizität, Nebenwirkungen, Dosisfindung
Phase II (Erste Wirksamkeitsprüfung)	Ermittlung der Wirksamkeit und weitergehende Sicherheitsprüfung	Größere Gruppe von Patienten, die an der zu behandelnden Krankheit leiden	Erste Nachweise für die Wirksamkeit, Untersuchung von Nebenwirkungen
Phase III (Bestätigung der Wirksamkeit und Vergleich)	Vergleich mit bestehenden Standardbehandlungen und Verifikation der Langzeitwirkung und -sicherheit	Große Patientengruppen (hundert bis tausende)	Bestätigung der Ergebnisse aus Phase II, häufig Randomisierte Kontrollierte Studien (RCTs)
Phase IV (Überwachung nach Markteinführung [Post-Marketing-Studien])	Langzeitüberwachung der Sicherheit, Wirksamkeit und Nebenwirkungen nach der Markteinführung	Noch breitere Anwendungspopulation im klinischen Alltag	Erkennung seltener Nebenwirkungen, Langzeitergebnisse und Vergleich mit neuen Entwicklungen

Merke: Klinische Studien sind der Goldstandard zur Sicherstellung der Wirksamkeit und Sicherheit neuer Behandlungsansätze.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Spezifische Aspekte für Psychotherapie-Studien

- Phasenmodelle können angepasst werden (z. B. Pilotstudien statt Phase I).
- Besondere Herausforderungen: Langfristige Überprüfung der Wirksamkeit, Placeboeffekte, Durchführung randomisierter Studien.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Design	Definition_Beschreibung	Vorteile	Beispiele
Randomisierte Kontrollierte Studien (RCTs)	Teilnehmer werden zufällig in Interventions- und Kontrollgruppen aufgeteilt.	Hohe interne Validität durch Randomisierung, Minimierung von Bias.	Vergleich einer neuen kognitiven Verhaltenstherapie (CBT) mit einer Standardtherapie.
Nicht-Randomisierte Studien	Kein Zufallsmechanismus bei der Gruppenzuweisung.	Praktisch, wenn Randomisierung nicht möglich ist; geringere Kosten.	Praktische Anwendungen, wenn Randomisierung nicht möglich ist.
Cross-Over-Designs	Teilnehmer erhalten abwechselnd verschiedene Behandlungen.	Jeder Teilnehmer fungiert als eigene Kontrolle.	Vergleich von zwei Psychotherapieansätzen bei derselben Personengruppe zu unterschiedlichen Zeitpunkten.
Beobachtungsstudien	Beobachtungen von Teilnehmern ohne experimentelle Eingriffe (z.B., Kohortenstudien, Fall-Kontroll-Studien).	Einsatz für Hypothesenbildung, Langzeitrisiken.	Kohortenstudien zur Langzeitverfolgung, Fall-Kontroll-Studien.

Merke: Die Wahl des Designs hängt von den Studienzielen, ethischen Überlegungen und praktischen Herausforderungen ab.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Randomisierte Kontrollierte Studien (RCTs)

Randomisierung

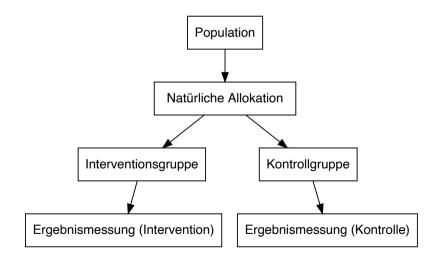
Interventionsgruppe

Kontrollgruppe

Ergebnismessung (Intervention)

Ergebnismessung (Kontrolle)

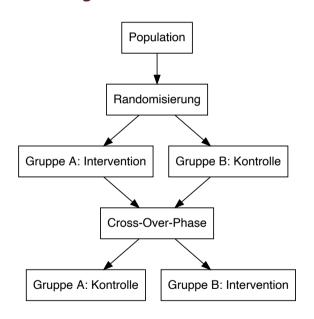
Nicht-Randomisierte Studien (Kohortenstudien)



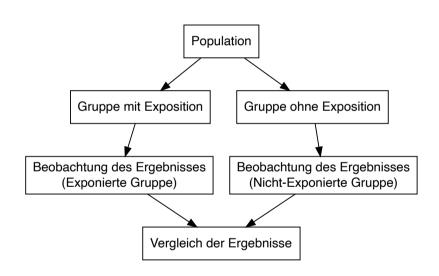


Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Cross-Over-Designs



Beobachtungsstudien





Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Studiendesign für Phase-III RCT (Methodenteil) - CONSORT Checkliste

Abschnitt/Thema	Punkt Nr.	Checkliste Punkt
Studiendesign	3a	Beschreibung des Studiendesigns (z.B. parallel, faktoriell) einschließlich des Zuweisungsverhältnisses
	3b	Wichtige Änderungen der Methoden nach Beginn der Studie (z.B. Einschlusskriterien), mit Begründungen
Teilnehmer	4a	Einschlusskriterien für Teilnehmer
	4b	Kontexte und Orte, an denen die Daten erhoben wurden
Interventionen	5	Die Interventionen für jede Gruppe mit ausreichenden Details zur Replikation, einschließlich wie und wann sie verabreicht wurden
Outcomes	6a	Vollständig definierte, vorab festgelegte primäre und sekundäre Endpunkte, einschließlich wie und wann sie gemessen wurden
	6b	Änderungen der Endpunkte nach Beginn der Studie, mit Begründungen
Stichprobengröße	7a	Wie die Stichprobengröße bestimmt wurde
	7b	Falls zutreffend, Erklärung von Zwischenanalysen und Abbruchkriterien

https://www.equator-network.org/reporting-guidelines/consort/



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Studiendesign für Phase-III RCT (Methodenteil) - CONSORT Checkliste

Abschnitt/Thema	Punkt Nr.	Checkliste Punkt
Randomisierung: Sequenzgenerierung	8a	Methode zur Generierung der zufälligen Zuweisungssequenz
	8b	Art der Randomisierung; Details zu Einschränkungen (wie Blockbildung und Blockgröße)
Randomisierung: Zuweisungsmechanismus	9	Mechanismus zur Implementierung der zufälligen Zuweisungssequenz (wie nummerierte Behälter), einschließlich Maßnahmen zur Geheimhaltung der Sequenz
Randomisierung: Durchführung	10	Wer die zufällige Zuweisungssequenz erstellt hat, wer die Teilnehmer eingeschrieben hat und wer die Teilnehmer zugewiesen hat
Verblindung	11a	Falls durchgeführt, wer nach der Zuweisung zur Intervention verblindet war (z.B. Teilnehmer, Pflegekräfte, Ergebnisprüfer) und wie
	11b	Falls relevant, Beschreibung der Ähnlichkeit der Interventionen
Statistische Methoden	12a	Statistische Methoden zum Vergleich von Gruppen für primäre und sekundäre Endpunkte
	12b	Methoden für zusätzliche Analysen, wie Subgruppenanalysen und angepasste Analysen

https://www.equator-network.org/reporting-guidelines/consort/



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Beispieltext Studiendesign für Phase-III RCT (Methodenteil)

The [...] trial was a double-blind, placebo-controlled randomized clinical trial that enrolled patients with [...]. Randomization was performed using [...]. The study protocol was described elsewhere [Referenz] and performed with no significant changes. The trial was conducted at [...] study centers (Name der Institute) from [...] to [...]. The study was registered on [...], reported per the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline for nonpharmacological treatments, and approved by the [Name der Ethikkommission]. Participants signed informed consent forms per the Declaration of Helsinki guidelines.

Die [...]-Studie war eine doppelblinde, placebokontrollierte, randomisierte klinische Studie, die Patienten mit [...] einschloss. Die Randomisierung wurde mithilfe von [...] durchgeführt. Das Studienprotokoll wurde an anderer Stelle beschrieben [Referenz] und ohne signifikante Änderungen durchgeführt. Die Studie wurde in [...] Studienzentren (Name der Institute) von [...] bis [...] durchgeführt. Die Studie wurde am [...] registriert, gemäß den CONSORT-Richtlinien für die Berichterstattung über nicht-pharmakologische Behandlungen berichtet und von der [Name der Ethikkommission] genehmigt. Die Teilnehmer unterzeichneten Einverständniserklärungen gemäß den Richtlinien der Deklaration von Helsinki.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Endpunkte und Outcomes in Klinischen Studien

- Endpunkte sind messbare Ereignisse oder Ergebnisse, um den Erfolg oder Misserfolg einer Intervention zu bewerten.
- Sie geben an, worauf eine Studie abzielt, und dienen als Grundlage zur Beurteilung der Wirksamkeit und Sicherheit.
- Endpunkte werden i.d.R. hierarchisiert (z.B. Primäre vs. Sekundäre Endpunkte)



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Primäre Endpunkte

- Definition: Der primäre Endpunkt ist das Hauptziel der Studie und stellt die wichtigste Messgröße zur Bewertung der Wirksamkeit dar.
- Beispiele:
 - In einer Psychotherapiestudie: Reduktion von Depressionssymptomen gemessen durch eine validierte Skala (z. B. Beck-Depressions-Inventar).
 - o In einer Medikamentenstudie: Überlebensrate, Symptomfreiheit nach einer bestimmten Zeit.
- ightarrow Bedeutung: Signifikanz des primären Endpunkts entscheidet über den Erfolg oder Misserfolg der Studie.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Sekundäre Endpunkte

- Definition: Sekundäre Endpunkte sind zusätzliche Messgrößen, die zur umfassenderen Bewertung der Intervention herangezogen werden.
- Beispiele (bezogen auf Depression als primären Endpunkt):
 - Verbesserung der Lebensqualität.
 - Reduktion von Angstzuständen bei einer Therapie.
 - Nebenwirkungen oder Langzeiteffekte.
- ightarrow Bedeutung: Sekundäre Endpunkte liefern ergänzende Informationen und helfen, weitere Aspekte der Wirksamkeit und Sicherheit zu verstehen.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Primäre Endpunkte vs. sekundäre Endpunkte (Studienbeispiel - Methodenteil)

Praktische Bedeutung für Psychotherapie-Studien:

- Ein klar definierter primärer Endpunkt hilft, den Fokus der Studie zu behalten und die Hauptfrage zu beantworten.
- Sekundäre Endpunkte bieten wertvolle zusätzliche Einsichten und unterstützen die Interpretation der Ergebnisse.

Primäre Endpunkte	Sekundäre Endpunkte	
Hauptziel der Studie	Ergänzende, zusätzliche Ziele	
Wesentliche Grundlage für Studienerfolg	Hilfreich zur umfassenden Bewertung	
Zentrale Messgröße für die Wirksamkeit	Weitere Messgrößen zur Unterstützung	



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Primäre Endpunkte vs. sekundäre Endpunkte

ightarrow Signifikanz des primären Endpunkts entscheidet über den Erfolg oder Misserfolg der Studie.

Zentral für die statistische Auswertung:

- Stichprobenumfangsplanung (Poweranalyse) wird i.d.R. am erwarteten Effekt im primären Endpunkt orientiert
- Signifikanzniveau gilt für Hypothesentest des primären Endpunkts (konfirmatorisches Ergebnis)
- ullet (mehrere) sekundäre Endpunkte o lpha-Fehler Kumulierung (erhöhte Typ-I-Fehlerrate)
 - \circ Problem: Stichprobengröße (N) durch Effekt im primären Endpunkt festgelegt \to Typ-I-Fehlerkorrektur würde bewirken, dass sekundäre Endpunkte nur schwer signifikant werden
 - \circ Lösung in der Praxis: Hypothesentests für sekundäre Endpunkte werden nicht für lpha-Fehler korrigiert
 - Vorsicht bei der Interpretation: Ergebnisse multipler, nicht-kontrollierter Tests sind explorative Ergebnisse (nicht hypothesenpr
 üfend, sondern hypothesengenerierend)



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Beispieltext Endpunkte (Methodenteil)

Assessments were performed by trained psychiatrists and psychologists blinded for patients' condition. Participants were assessed at baseline and then [...] weeks, [...] weeks, [...] weeks (primary endpoint), and [...] weeks (secondary endpoint) after treatment onset. Adverse effects were recorded at [...] weeks after treatment onset. The primary outcome was the change in [...] scores until [...] weeks. Secondary outcomes included clinical response (defined as 50% or greater improvement) and remission (defined as a ... score less than ...) as well as changes in the [...] scales.

Die Messungen wurden von geschulten Psychiatern und Psychologen durchgeführt, die für die Gruppenzugehörigkeit der Patienten verblindet waren. Die Teilnehmer wurden zu Beginn (Baseline) sowie [...] Wochen, [...] Wochen, [...] Wochen (primärer Endpunkt) und [...] Wochen (sekundärer Endpunkt) nach Beginn der Behandlung beurteilt. Nebenwirkungen wurden [...] Wochen nach Beginn der Behandlung erfasst. Der primäre Endpunkt war die Veränderung der [...]-Werte bis [...] Wochen. Sekundäre Endpunkte umfassten das klinische Ansprechen (definiert als eine Verbesserung von 50% oder mehr) und die Remission (definiert als ein [...]-Wert unter ...) sowie Veränderungen auf den [...]-Skalen.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Itention-to-treat vs. Per-protocol Analysen

Intention-to-Treat Analyse (ITT)

- Definition: Alle Teilnehmer, die einer Behandlungsgruppe zu Beginn der Studie zufällig zugewiesen wurden, werden in die Analyse einbezogen, unabhängig davon, ob sie die zugewiesene Behandlung vollständig durchgeführt haben oder nicht.
- Vorteile:
 - o Erhalt der Randomisierung: Minimiert Bias und bewahrt die Vergleichbarkeit der Gruppen.
 - Realitätsnähe: Gibt ein realistisches Bild der Wirksamkeit einer Intervention in der Praxis.
- Beispiel: Ein Teilnehmer, der eine Therapie vorzeitig abbricht, wird weiterhin der ursprünglichen Gruppe zugerechnet.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Itention-to-treat vs. Per-protocol Analysen

Per-Protocol Analyse (PP)

- Definition: Nur die Teilnehmer, die die Behandlung genau nach Protokoll durchgeführt haben, werden in die Analyse aufgenommen.
- Vorteile:
 - o Reine Wirksamkeitsprüfung: Bietet eine Bewertung der optimalen Wirkung der Intervention unter idealen Bedingungen.
- Nachteil:
 - Gefahr von Bias: Durch den Ausschluss von Teilnehmern, die das Protokoll nicht befolgen, kann die Vergleichbarkeit zwischen den Gruppen beeinträchtigt werden.
- Beispiel: Teilnehmer, die eine Therapie abbrechen, werden nicht in die Analyse einbezogen.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Itention-to-treat vs. Per-protocol Analysen

Praktische Relevanz:

- Die ITT-Analyse ist der Goldstandard für die Bewertung der Wirksamkeit in klinischen Studien, da sie die tatsächliche klinische Praxis besser widerspiegelt.
- Per-Protocol-Analysen bieten zusätzliche Informationen zur maximalen Wirksamkeit der Behandlung.
- Die Wahl zwischen ITT- und Per-Protocol-Analysen beeinflusst die Interpretation der Studienergebnisse und sollte klar definiert und begründet sein.

ITT-Analyse	Per-Protocol-Analyse	
Berücksichtigt alle Teilnehmer	Nur Teilnehmer, die das Protokoll einhalten	
Erhalt der Randomisierung	Risiko von Verzerrungen durch Ausfälle	
Realitätsnahe Ergebnisse	Ergebnisse unter idealen Bedingungen	



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Monozentrische vs. Multizentrische Studien

Monozentrische Studien:

- Definition: Studien, die an einem einzigen Standort (z.B. Klinik oder Forschungseinrichtung) durchgeführt werden.
- Vorteile:
 - o Einheitliche Bedingungen: Kontrollierte und konsistente Durchführung.
 - Kostengünstiger: Oft einfacher und kostengünstiger durchführbar.
 - o Leichtere Koordination: Weniger organisatorischer Aufwand.
- Nachteile:
 - o Geringere Generalisierbarkeit: Ergebnisse sind möglicherweise nicht auf andere Standorte oder Populationen übertragbar.
 - o Begrenzte Teilnehmerzahl: Weniger Teilnehmer im Vergleich zu multizentrischen Studien.
- Beispiele: Pilotstudien, kleinere klinische Studien zur Überprüfung erster Hypothesen.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Monozentrische vs. Multizentrische Studien

Multizentrische Studien:

- Definition: Studien, die an mehreren Standorten gleichzeitig durchgeführt werden.
- Vorteile:
 - Höhere Generalisierbarkeit: Ergebnisse sind repräsentativer und übertragbarer auf verschiedene Populationen und Standorte.
 - o Größere Teilnehmerzahl: Mehr Teilnehmer erhöhen die statistische Aussagekraft.
 - o Breitere Vergleichbarkeit: Variabilität zwischen Standorten ermöglicht umfassendere Beurteilungen.
- Nachteile:
 - Komplexere Koordination: Höherer organisatorischer Aufwand und Kommunikationsbedarf.
 - Höhere Kosten: Meist teurer und ressourcenintensiver.
- Beispiele: Große klinische Studien zu neuen Medikamenten, nationale oder internationale Studien.
- ightarrow Zentrumseffekte sollten (statistisch) berücksichtigt werden



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Monozentrische vs. Multizentrische Studien

Multizentrische Studien - statistische Kontrolle von Zentrumseffekten im LMM (2 Optionen)

1 - Kontrolle mittels Kovariate (Fixed Effekt):

```
lmer(y ~ time * group + center + (time|subject))
```

- leichter zu schätzen (nur 1 Modellkoeffizient), aber komplexere Hierarchie nicht abbildbar
- 2 Kontrolle mittels Random Effekt:

```
lmer(y ~ time * group + (time|center/subject))
```

• komplexer zu schätzen (k Varianzen für pro Zentrum; kann in kleinen Stichproben zu Konvergierungsproblemen führen), aber flexibler



Poweranalyse

Zur Erinnerung - Warum Poweranalyse?

- Bestimmung der benötigten Stichprobengröße: Um sicherzustellen, dass die Studie ausreichend Teilnehmer hat, um statistisch signifikante Effekte zu erkennen (β -Fehler vermeiden).
- Optimierung von Ressourcen: Ressourcen wie Zeit und Kosten effizient einsetzen, indem die Stichprobengröße so gewählt wird, dass unnötig große Studien vermieden werden.

Immer nötige Zutaten (neben anderen - je nach Modell):

- Signifikanzniveau (lpha)
- gewünschte Power (1β) , i.d.R. mindestens 80%
- Effektschätzung
- Stichprobengröße (oft gesucht)



Poweranalyse

Poweranalyse für das LMM - Herausforderungen

- Anzahl von Messungen muss auf mehreren Ebenen berücksichtigt werden (Messungen, Personen, Zentren...)
- Für eine genaue Schätzung müssen viele Komponenten bekannt sein
 - o Für fixed Effekte sind i.d.R. Schätzungen/Vorstellungen verfügbar (Effektstärke)
 - Für Varianzkomponenten (random Effekte) i.d.R. unbekannt.
 - Größere Vorstudien hilfreich, unterscheiden sich jedoch oft in Details
 - Kleinere Pilotstudien hilfreich, jedoch oft zu klein für sinnvolle Schätzung
- Einführen von Kovariaten kann die optimalen Stichprobengrößen auf jedem Level ändern



Poweranalyse

Simulationsbasierte Poweranalyse

- Für genaue Powerschätzungen müssen Daten (inklusive fixed und random Effekte simuliert werden)
- Schätzer für fixed und random Effekte müssen angegeben werden
- Logik:
 - Daten erstellen, die in etwa so aussehen, wie in unserer Studie
 - In diesen Daten LMM berechnen und den Effekt, der uns interessiert auf Signifikanz prüfen
 - Vorgang für stabile Ergebnisse oft wiederholen (i.d.R. 1000 oder 10000 mal)
 - Power (%) = Anteil der Durchgänge, in denen Test signifikant war
- Implementierung simulationsbasierte Poweranalysen im R-Paket simr



Poweranalyse

Simulationsbasierte Poweranalyse

Datensatzstruktur erstellen (Anzahl Personen, Gruppen, Zeitpunkte):

Blick in den simulierten Datensatz (erste 10 Zeilen):



Poweranalyse

Simulationsbasierte Poweranalyse

Wir wollen in der Studie folgendes Modell nutzen:

```
lmer(y \sim group * time + (1|class/id), data = data)
```

Fixed Effekte und Varianzkomponenten müssen für die genaue Simulation angegeben/geschätzt werden:

```
## Intercept und slopes für group, time und treatment:time
fixed <- c(5, 0, 1, 0.6)

## Random intercepts für Teilnehmer gruppiert in Klassen

rand <- list(0.5, 0.1)

## Residualvarianz

res <- 1</pre>
```



Poweranalyse

Simulationsbasierte Poweranalyse

Basierend auf den Parametern erstellen wir ein "künstliches" Regressionsmodell:

```
model <- makeLmer(y ~ group*time + (1|class/id), fixef=fixed, VarCorr=rand, sigma=res, data=data)</pre>
model
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ group * time + (1 | class/id)
   Data: data
REML criterion at convergence: 478.8519
Random effects:
                     Std.Dev.
 Groups
        Name
 id:class (Intercept) 0.7071
 class
       (Intercept) 0.3162
 Residual
                     1.0000
Number of obs: 150, groups: id:class, 50; class, 5
Fixed Effects:
      (Intercept)
                  grouptherapy
                                                  time grouptherapy:time
              5.0
                                 0.0
                                                   1.0
                                                                       0.6
```

Poweranalyse

Simulating: |

Simulationsbasierte Poweranalyse

Wir analysieren die erreichte Power für den Interaktionseffekt:

```
powerSim(model, nsim=100, test = fixed("grouptherapy:time", method = "t"))
```



Poweranalyse

Klassische Poweranalyse

- Aufgrund der Schwierigkeiten bzgl. Schätzung der Varianzkomponenten wird in der Praxis oft eine klassische Poweranalyse herangezogen
- Statt eines LMMs wird eine konventioneller Test (ohne random Effekte) geprüft
- Diese Schätzung ist konsevativer, da Berücksichtigung von random Effekten die Power tendenziell eher erhöht
- Beispiel:
 - Im LMM: Unterschied zwischen den durchschnittlichen Steigungen von 2 Gruppen (t-verteilt)
 - Konventionell: Unterschied zwischen 2 Durchschnittswerten (t-verteilt)
- Diese Analysen können auf Basis herkömmlicher Effektstärkeschätzungen berechnet werden (fixed Effekte)
- Implementierung klassischer Poweranalysen im R-Paket pwr



Poweranalyse

Klassische Poweranalyse - Beispiel 2 Gruppen RCT

Two-sample t test power calculation

n = 63.76561
 d = 0.5
 sig.level = 0.05
 power = 0.8
alternative = two.sided

NOTE: n is number in *each* group

- Beispiel: Vergleich der durchschnittlichen Veränderungen von 2 Gruppen
- Annahme eines mittleren Effekts (d=0.5, nach Cohen)
- $\alpha = .05$
- $1-\beta=.80$ (wird i.d.R. als Untergrenze definiert)
- ullet Es werden N=128 Personen benötigt (n=64 pro Gruppe)



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Exkurs: Kategoriale Outcomes für Therapieansprechen

Response (50% Verbesserung):

- Definition: Eine Verbesserung um 50% oder mehr im Vergleich zum Ausgangswert eines Messinstruments (z.B. Skalen zur Symptomerfassung).
- Bedeutung:
 - Wird häufig als Kriterium zur Beurteilung des klinischen Ansprechens einer Therapie verwendet.
 - Ein Patient, der eine 50%ige Verbesserung zeigt, wird als "Responder" bezeichnet.
- Beispiel: Bei der Behandlung von Depressionen bedeutet dies eine Reduktion der Symptome um mindestens 50% gemäß einer Skala wie dem Hamilton-Depressions-Skala (HAMD).



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Exkurs: Kategoriale Outcomes für Therapieansprechen

Remission:

- Definition:
 - o Ein Zustand, bei dem die Symptome einer Erkrankung auf ein normales oder nahezu normales Niveau reduziert werden
 - Häufig definiert durch einen festgelegten Schwellenwert auf einer Symptombewertungsskala.
- Bedeutung:
 - Remission bedeutet, dass die Krankheitssymptome minimal oder nicht mehr nachweisbar sind und der Patient sich weitgehend normal fühlt.
 - Sie stellt ein anspruchsvolleres Ziel als eine bloße Verbesserung dar.
- Beispiel: Ein Patient gilt als in Remission, wenn der HAMD-Score unter einen bestimmten Wert fällt, der als symptomfrei oder symptomarm definiert wird (z.B. HAMD ≤ 7).



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Exkurs: Kategoriale Outcomes für Therapieansprechen

Reliable change index (RCI)

- Statistisches Maß, das verwendet wird, um die Signifikanz einer Veränderung auf individueller Ebene zu beurteilen.
- Zeigt an, ob die Veränderung in den Testergebnissen einer Person zwischen zwei Zeitpunkten größer ist, als es durch Messfehler (eines Tests) erklärbar wäre

$$RCI = \frac{Post\text{-}Test\text{-}Score - Pre\text{-}Test\text{-}Score}{Standard Error of Difference (SE_{diff})}$$

$$ext{SE}_{ ext{diff}} = \sqrt{2 \cdot (ext{Standard Error of Measurement})^2}$$

ightarrow Der Standardfehler der Messung basiert auf der Reliabilität des Tests und seiner Standardabweichung.



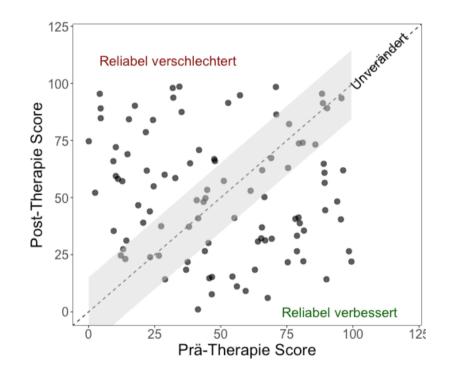
Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Exkurs: Kategoriale Outcomes für Therapieansprechen

Reliable change index (RCI)

Interpretation des RCI:

- Positiver Wert (> 1.96): Die Veränderung ist signifikant und kann als echte Verbesserung interpretiert werden
 - bei einem z-Wert von 1.96 entspricht dies einer Signifikanz auf dem 5%-Niveau
- Negativer Wert (< -1.96): Die Veränderung ist signifikant und zeigt eine Verschlechterung an.
- Wert im Bereich von -1.96 bis 1.96: Keine verlässliche Veränderung; die Veränderung könnte durch Messfehler bedingt sein.





Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Exkurs: Kategoriale Outcomes für Therapieansprechen

Reliable change index (RCI)

Beispiel:

Ein Patient erhält zu Beginn einer Therapie (Pre-Test) einen Score von 20 auf einer Depressionsskala und am Ende (Post-Test) einen Score von 10. Mit einem berechneten $SE_{\rm diff}$ von 4 beträgt der RCI:

$$RCI = \frac{10 - 20}{4} = -2.5$$

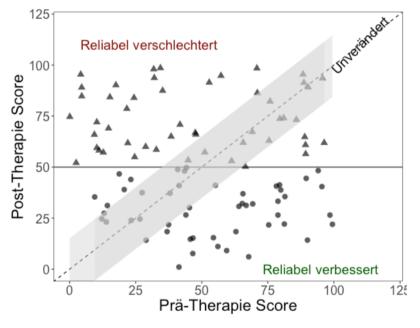
Da der Wert kleiner als -1.96 ist, zeigt er eine verlässliche Verbesserung.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Exkurs: Kategoriale Outcomes für Therapieansprechen - Reliable and clinically significant change

- Oft wird der RCI mit einem Remissionskriterium kombiniert
- Verbesserungen sind dann vom Messfehler befreit und klinisch bedeutsam
- Hier wurde ein Remissionskrierium von Score < 50 gewählt



• in Remission A nicht in Remission



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Sensitivitätsanalysen

- Definition: Eine Sensitivitätsanalyse bewertet, wie empfindlich die Ergebnisse einer klinischen Studie gegenüber Änderungen in den Annahmen, Methoden oder Parametern des Modells sind.
- Ziel: Sicherstellung der Robustheit und Zuverlässigkeit der Studienergebnisse durch die Analyse, wie sich Änderungen auf die Hauptschlussfolgerungen auswirken.

Typische Anwendungsfälle in Klinischen Studien

- Umgang mit fehlenden Daten: Analyse der Auswirkungen verschiedener Annahmen über fehlende Werte (z.B. vollständig zufälliges Fehlen, imputierte Werte).
- Abweichungen vom Protokoll: Berücksichtigung von Teilnehmern, die sich nicht an das Studienprotokoll gehalten haben (z.B. Intention-to-Treat-Analyse vs. Per-Protocol-Analyse).
- Änderung von Modellparametern: Bewertung, wie sich Änderungen in den Parametern oder Variablen auf die Ergebnisse auswirken.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Beispieltext Endpunkte (Methodenteil)

The sample size was estimated for a power of 80% and a 2-tailed a-level of 5% for a 2-group comparison of change over time in the [primary outcome]. Our study was powered to detect a between-group difference of d=0.5, which represents a moderate effect size (Cohen, 1992). Therefore, a targeted sample of 128 patients (50 per group) was obtained. Data were analyzed in the intention-to-treat sample. Analyses were performed in R in its most recent version. Results were significant at a P value less than .05. For all continuous outcomes, we calculated 2-level linear mixed-effects regression models (LMM), assuming a linear relationship over time with 5 (up to week 6) repeated measurements per patient, respectively. The model included time, group and their interaction as fixed effects. Measurements were assumed to be nested within patients. Effect sizes were calculated as Cohen's d. As secondary outcomes, response (at least 50 improvement from baseline on the [primary outcome]) and remission ([primary outcome] scores below XX) were computed for each group.

Die Stichprobengröße wurde für eine Power von 80% und ein zweiseitiges α-Niveau von 5% für einen Vergleich der Veränderung über die Zeit in der [primären Zielgröße] zwischen zwei Gruppen geschätzt. Unsere Studie war darauf ausgelegt, einen Unterschied zwischen den Gruppen mit d=0,5 zu erkennen, was eine moderate Effektgröße darstellt (Cohen, 1992). Daher wurde eine angestrebte Stichprobe von 128 Patienten (50 pro Gruppe) erreicht. Die Daten wurden in der Intention-to-Treat-Stichprobe analysiert. Die Analysen wurden in der aktuellsten Version von R durchgeführt. Die Ergebnisse wurden als signifikant betrachtet, wenn der P-Wert kleiner als 0.05 war. Für alle kontinuierlichen Endpunkte berechneten wir zweistufige lineare gemischte Regressionsmodelle (LMM), wobei wir von einer linearen Beziehung über die Zeit mit 5 (bis zur Woche 6) Messwiederholungen pro Patient ausgingen. Das Modell umfasste Zeit, Gruppe und deren Interaktion als feste Effekte. Die Messungen wurden als in den Patienten verschachtelt betrachtet. Effektgrößen wurden als Cohen's d berechnet. Als sekundäre Endpunkte wurden das Ansprechen (mindestens 50% Verbesserung gegenüber dem Ausgangswert auf der [primären Zielgröße]) und die Remission ([primäre Zielgröße]-Werte unter XX) für jede Gruppe berechnet.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Studiendesign für Phase-III RCT (Ergebnisteil) - CONSORT Checkliste

Abschnitt/Thema	Punkt Nr.	Checkliste Punkt
Teilnehmerfluss (Diagramm wird empfohlen)	13a	Für jede Gruppe die Anzahl der zufällig zugewiesenen Teilnehmer, die die vorgesehene Behandlung erhalten haben und die für das primäre Ergebnis analysiert wurden
	13b	Für jede Gruppe Verluste und Ausschlüsse nach der Randomisierung, zusammen mit Begründungen
Rekrutierung	14a	Zeitpunkte, die die Rekrutierungs- und Nachbeobachtungszeiträume definieren
	14b	Gründe, warum die Studie beendet oder gestoppt wurde
Baseline-Daten	15	Eine Tabelle mit den demografischen und klinischen Ausgangsdaten für jede Gruppe
Analysierte Zahlen	16	Für jede Gruppe die Anzahl der in jeder Analyse eingeschlossenen Teilnehmer (Nenner) und ob die Analyse nach den ursprünglichen Zuweisungsgruppen durchgeführt wurde

https://www.equator-network.org/reporting-guidelines/consort/



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Studiendesign für Phase-III RCT (Ergebnisteil) - CONSORT Checkliste

Abschnitt/Thema	Punkt Nr.	Checkliste Punkt	
Outcomes und Schätzungen	17a	Für jedes primäre und sekundäre Ergebnis die Ergebnisse für jede Gruppe und die geschätzte Effektgröße sowie deren Präzision (z.B. 95%-Konfidenzintervall)	
	17b	Für binäre Outcomes wird die Darstellung sowohl absoluter als auch relativer Effektgrößen empfohlen	
Zusätzliche Analysen	18	Ergebnisse anderer durchgeführter Analysen, einschließlich Subgruppenanalysen und angepasster Analysen, unter Unterscheidung zwischen vordefinierten und explorativen Analysen	
Schädigungen	19	Alle wichtigen Schädigungen oder unbeabsichtigte Effekte in jeder Gruppe (für spezifische Leitlinien siehe CONSORT für Schädigungen)	

https://www.equator-network.org/reporting-guidelines/consort/



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Beispielszenario: Vergleich von Kognitiver Verhaltenstherapie (KVT) vs. Kognitive Verhaltenstherapie + Muttraining

Die Studie soll die Wirksamkeit von Kognitiver Verhaltenstherapie (KVT) vs. KVT + Muttraining bei Patienten mit generalisierter Angststörung vergleichen. Es handelt sich um eine doppelblinde, randomisierte klinische Studie, in der 128 Patienten zufällig einer der beiden Therapieformen zugewiesen werden. Die Behandlungen werden über 6 Wochen hinweg in wöchentlichen Sitzungen durchgeführt.

Ziel der Studie: Primärer Endpunkt ist die Veränderung der Angstwerte, gemessen anhand der Hamilton-Angst-Skala (HAM-A), 6 Wochen nach Therapiebeginn. Sekundäre Endpunkte umfassen die Response, Remission (definiert als HAM-A-Wert unter 7) sowie die Verbesserung der Lebensqualität, gemessen durch das SF-36-Formular.

Ablauf: Die Teilnehmer werden von geschulten Therapeuten behandelt, die für die spezifische Therapieform trainiert sind. Die Bewertung der Ergebnisse erfolgt durch verblindete Psychologen. Nebenwirkungen oder unerwünschte Ereignisse werden dokumentiert.

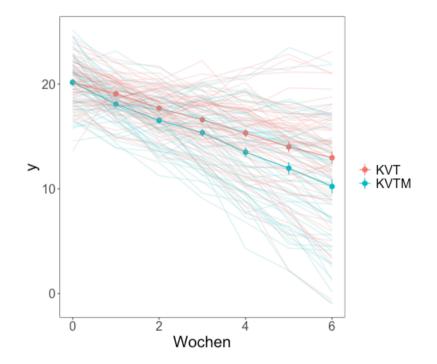
Hypothesen: Es wird erwartet, dass beide Therapieformen zu einer signifikanten Reduktion der Angst führen, jedoch mit unterschiedlichen Schwerpunkten in der Symptomreduktion und Lebensqualitätsverbesserung.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

```
ggplot(data, aes(x = time, y = y, colour = treatment)) +
  stat_summary(aes(group = subject), geom = "line") +
  labs(colour = "", x = "Wochen", y = "HAM-A") +
  theme(text = element_text(size = 20), legend.position = "right")
```

- Darstellung der Daten in Linegraph (Deskriptivstatistik)
- I.d.R. bekommt primärer Endpunkt einen eigenen Graphen.





Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

```
mod = lmer(y ~ time * treatment + (time|subject), data = data)
summary(mod)
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ time * treatment + (time | subject)
  Data: data
REML criterion at convergence: 3359.8
Scaled residuals:
    Min
            10 Median
                            30
                                  Max
-3.2231 -0.5451 0.0004 0.5428 3.1874
Random effects:
Groups Name
                     Variance Std.Dev. Corr
 subject (Intercept) 3.9859 1.9965
                    1.0146 1.0073
                                     -0.59
                     0.9966 0.9983
 Residual
Number of obs: 896, groups: subject, 128
Fixed effects:
                  Estimate Std. Error t value
(Intercept)
                   20.1978
                              0.2636 76.610
                   -1.2172
time
                              0.1281 -9.502
treatmentKVTM
                   -0.2454
                              0.3729 -0.658
time:treatmentKVTM -0.3928
                             0.1812 -2.168
Correlation of Fixed Effects:
           (Intr) time trKVTM
           -0.598
tretmntKVTM -0.707 0.423
tm:trtmKVTM 0.423 -0.707 -0.598
```



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Veränderung innerhalb beider Gruppen und deren Differenz darstellen:

```
library(emmeans)
em = emmeans::emtrends(mod, pairwise ~ treatment, var = "time")
em
$emtrends
 treatment time.trend
                        SE df lower.CL upper.CL
               -1.22 0.128 126
                                 -1.47
 KVT
                                         -0.964
 KVTM
               -1.61 0.128 126
                               -1.86 -1.357
Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
$contrasts
 contrast
           estimate
                       SE df t.ratio p.value
 KVT - KVTM
              0.393 0.181 126 2.168 0.0320
Degrees-of-freedom method: kenward-roger
```



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

```
em
$emtrends
```

treatment time.trend SE df lower.CL upper.CL KVT -1.22 0.128 126 -1.47 -0.964 KVTM -1.61 0.128 126 -1.86 -1.357

Degrees-of-freedom method: kenward-roger Confidence level used: 0.95

```
$contrasts
contrast estimate SE df t.ratio p.value
KVT - KVTM 0.393 0.181 126 2.168 0.0320
```

Degrees-of-freedom method: kenward-roger

Beispieltext im Ergebnisteil:

Es zeigten sich signifikante Verbesserungen im HAM-A für die KVT Gruppe $(\beta_{KVT}=-7.32,KI_{95\%}-8.82~{\rm bis}~-5.78)~{\it und die KVTM Gruppe} \\ (\beta_{KVTM}=-9.66,KI_{95\%}-11.16~{\rm bis}~-8.14).~{\it Es wurde ein signifikanter} \\ {\it Zeit x Gruppe Interaktionseffekte mit stärken Verbesserungen für die KVTM Gruppe gefunden } (\beta_{Diff}=2.36,t_{126}=2.17,p=.032).$

Vorsicht: Steigungen müssen mit 6 multipliziert werden, um Veränderung bis zum Endpunkt (Woche 6) der Studie zu berechnen.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Berechnung der Effektsärke

Beispieltext im Ergebnisteil:

Die Effektstärke für den Zwischengruppeneffekt betrug ein Cohen's d von 0.69 $(KI_{95\%}=0.06~{
m bis}~1.32)$. Dies entspricht einem mittleren Effekt.



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

Berechnen der Response und Remissionsraten:

```
data_w = pivot_wider(data, id_cols = c(subject, treatment), names_from = time, names_prefix = "week", values_from = y,
head(data w, 10)
# A tibble: 10 × 9
  subject treatment week0 week1 week2 week3 week4 week5 week6
    <int> <fct>
                  65 KVT
                   20.1 21.1 17.9 18.0 17.7 14.9 13.5
                        18.5 15.0 15.5 12.2 11.9 10.4
       66 KVT
                   18.1
       67 KVT
                   22.2
                        20.4 20.1 19.3 19.3 17.8 15.6
       68 KVT
                   17.2 18.0 16.2 15.2 14.7 13.7 13.4
       69 KVT
                        21.1 21.0 18.6 20.2 15.8 18.1
                   18.9
       70 KVT
                   19.8
                        20.1 20.0 18.3 18.0 16.3 15.8
 7
       71 KVT
                   21.1 19.8
                             17.7 14.2 11.9
                                              8.22 5.93
 8
       72 KVT
                   21.0 17.3 15.3 12.4 7.81 5.60 5.04
 9
       73 KVT
                   20.4 16.4 16.2 14.5 14.4 12.4 11.8
       74 KVT
                   16.4 19.0 17.0 17.9 18.4 18.0 18.8
10
```



Statistische Analyse von Wirksamkeitsstudien (Efficacy trials)

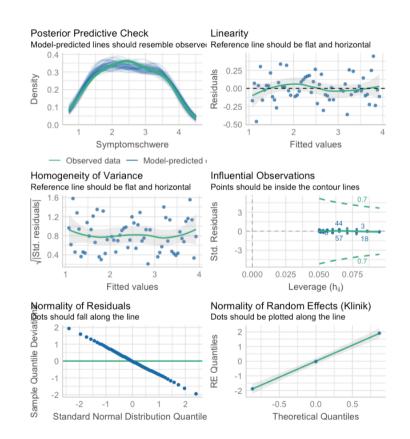
Berechnen der Response und Remissionsraten:

```
data w = pivot wider(data, id cols = c(subject, treatment), names from = time, names prefix = "week", values from = y,
data w$response = data w$week6 < (data w$week0 * 0.5)
data_w$remission = data_w$week6 < 7</pre>
table(data_w$response, data_w$treatment)
                                                                            prop.table(table(data_w$response, data_w$treatment), 2)
      KVT KVTM
                                                                                              KVTM
 FALSE 48
                                                                             FALSE 0.750000 0.515625
 TRUE 16
                                                                             TRUE 0.250000 0.484375
table(data w$remission, data w$treatment)
                                                                            prop.table(table(data_w$remission, data_w$treatment), 2)
      KVT KVTM
                                                                                              KVTM
                                                                             FALSE 0.859375 0.718750
           18
                                                                             TRUE 0.140625 0.281250
```



Modellvoraussetzungen

```
model = lmer(Symptomschwere ~ Therapiestunden + (1|Klin-
library(performance)
check_model(model)
```





Troubleshooting

- Warning: "model failed to converge"
- Keine Modellkonvergierung (vermutliche zu viele Parameter bei zu kleiner Stichprobe)
- Ergebnis darf nicht interpretiert werden Schätzwerte könnten falsch sein
- Lösung 1: Modell vereinfachen, Variablen transformieren (zentrieren, standardisieren)
- Lösung 2: Anderen Optimizer verwenden (mehr Iterationen bei maximum likelihood Schätzung erzwingen)

```
# anderen Optimizer testen bei Konvergierungsproblemen

lmer(y ~ x + (1|group), data = data, control = lmerControl(optimizer = "Nelder_Mead"))
```



Troubleshooting

- Warning: "singular fit"
- Grund 1: Mindestens ein Random Effekt ist 0 (oder beinahe
 0)
- Grund 2: Eine Korrelation zwischen Random Effekten ist -1 oder 1 (oder beinahe)
- Nicht zwingend problematisch, aber Hinweis, dass Modell vereinfacht werden könnte
- Kann auch potentiell auf Multikolliniarität hinweisen (hohe Korrelation zwischen Prädiktoren)

```
model = lmer(Symptomschwere ~ Therapiestunden + (Therapiestunden|Klinik), data = data)
boundary (singular) fit: see help('isSingular')
 summary(model)
Linear mixed model fit by REML ['lmerMod']
Formula: Symptomschwere ~ Therapiestunden + (Therapiestunden | Klinik)
  Data: data
REML criterion at convergence: -9.8
Scaled residuals:
   Min 1Q Median
-2.3469 -0.5817 -0.1171 0.5837 2.5651
Random effects:
                        Variance Std.Dev. Corr
 Klinik (Intercept) 3.182983 1.78409
         Therapiestunden 0.001463 0.03825 1.00
 Residual
                        0.033578 0.18324
Number of obs: 60, groups: Klinik, 3
               Estimate Std. Error t value
(Intercept)
              3.8132 1.0376 3.675
Therapiestunden -0.4511 0.0465 -9.701
Correlation of Fixed Effects:
           (Intr)
Therapstndn 0.367
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

Take-aways



- Lineare gemischte Modelle (LMMs) sind ein wichtiges statistisches Instrument zur Analyse hierarchischer und wiederholter Messdaten, wie sie in klinischen Studien häufig auftreten.
- In klinischen Studien sind primäre und sekundäre Endpunkte entscheidend für die Bewertung der Wirksamkeit und Sicherheit einer Intervention. Primäre Endpunkte bestimmen den Studienerfolg, während sekundäre Endpunkte ergänzende Informationen liefern.
- Poweranalysen für LMMs erfordern Simulation auf Basis vieler Axiome, oft wird daher eine konventionelle Poweranalyse durchgeführt.
- Diskrete Endpunkte wie Response, Remission oder RCI können ebenfalls zur Bewertung von Therapieerfolg herangezogen werden.
- Die CONSORT Checkliste ist ein Katalog zur Qualitätskontrolle beim Berichten von Methoden und Ergebnissen klinischer Studien.