



UniversitätsKlinikum Heidelberg

Bewertung klinischer Relevanz – zwischen 0.2 und ärztlichem Urteil

Meinhard Kieser

Institut für Medizinische Biometrie und Informatik

Universität Heidelberg



Inhalt

- **Klinische Relevanz – was ist das und wo liegt das Problem?**
- **Welche Methoden zur Bewertung der klinischen Relevanz gibt es?**
- **Wann ist ein Effekt klinisch relevant?**
- **Diskussion**



Beispiel: Fingolimod bei Multipler Sklerose

The NEW ENGLAND JOURNAL *of* MEDICINE

ESTABLISHED IN 1812

FEBRUARY 4, 2010

VOL. 362 NO. 5

A Placebo-Controlled Trial of Oral Fingolimod in Relapsing Multiple Sclerosis

Ludwig Kappos, M.D., Ernst-Wilhelm Radue, M.D., Paul O'Connor, M.D., Chris Polman, M.D.,
Reinhard Hohlfeld, M.D., Peter Calabresi, M.D., Krzysztof Selmaj, M.D., Catherine Agoropoulou, Ph.D.,
Malgorzata Leyk, Ph.D., Lixin Zhang-Auberson, M.D., Ph.D., and Pascale Burtin, M.D., Ph.D.,
for the FREEDOMS Study Group*

- **Ergebnis Expanded Disability Status Scale (EDSS) nach 24 Monaten**

Placebo: 0.13 ± 0.94

1.25 mg Fingolimod: -0.03 ± 0.88

Bewertung der klinischen Relevanz – Methode „Eine Regel für alles“

The screenshot shows a Mozilla Firefox browser window with the address bar containing the URL: <http://www.google.de/search?hl=de&client=firefox-a&hs=sww&rls=org.mozilla%3Ade%3Aofficial&q=effect+size+calculator&oq=effect+size+calculator&a>. The search results page displays the Google logo, a search bar with the text "effect size calculator", and a search button. Below the search bar, it indicates "Suche" and "Ungefähr 7.730.000 Ergebnisse (0,16 Sekunden)".

The search results are categorized into several sections:

- Alles**:
 - [Effect Size Calculator](http://www.uccs.edu/~faculty/lbecker/) - [Diese Seite übersetzen](#)
Effect Size Calculators. Calculate Cohen's d and the effect-size correlation, rYI, using --. means and standard deviations. independent groups t test values and df ...
 - [Effect Size Calculator - Cognitive Flexibility Lab](http://cognitiveflexibility.org/efficientsize/) - [Diese Seite übersetzen](#)
18 Jun 2008 – This **calculator** evaluates the **effect size** between two means (i.e., Cohen's d; Cohen, 1988), which is the difference between means divided by ...
 - [Effect Size Calculator : The Campbell Collaboration - Framed With ...](http://www.campbellcollaboration.org/.../effect_size_i...) - [Diese Seite übersetzen](#)
The Campbell Collaboration. C2 Home · Training · Research · Knowledge translation · Evidence-based decisions and practice · Links · C2 Policy doc ...
- Heidelberg**: Standort ändern
- Das Web**:
 - [Effect size - Wikipedia, the free encyclopedia](http://en.wikipedia.org/wiki/Effect_size) - [Diese Seite übersetzen](#)
Another example, where effect sizes may be distorted is in a multiple trial experiment where the **effect size calculation** is based on the averaged or aggregated ...
 - [The Meta-Analysis Calculator](http://www.lyonsmorris.com/ma1/index.cfm) - [Diese Seite übersetzen](#)
The Meta Analysis **Calculator**: Data Entry ... Chi Square with One df to **Effect Size**. One way F statistic to **Effect Size**, One-tailed Probability (pvalue) to **Effect Size** ...



Bewertung der klinischen Relevanz – Methode „Eine Regel für alles“ (2)

Calculate d and r using means and standard deviations

Calculate the value of Cohen's d and the effect-size correlation, $r_{Y\lambda}$, using the means and standard deviations of two groups (treatment and control).

$$\text{Cohen's } d = M_1 - M_2 / \sigma_{\text{pooled}}$$

$$\text{where } \sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$$

Compute

$$r_{Y\lambda} = d / \sqrt{(d^2 + 4)}$$

Note: d and $r_{Y\lambda}$ are positive if the mean difference is in the predicted direction.

Group 1	Group 2
M_1 <input type="text"/>	M_2 <input type="text"/>
SD_1 <input type="text"/>	SD_2 <input type="text"/>
<input type="button" value="Reset"/>	
Cohen's d	effect-size r
<input type="text"/>	<input type="text"/>



Bewertung der klinischen Relevanz – Methode „Eine Regel für alles“ (2)

Calculate d and r using means and standard deviations

Calculate the value of Cohen's d and the effect-size correlation, $r_{Y\lambda}$, using the means and standard deviations of two groups (treatment and control).

Cohen's $d = M_1 - M_2 / \sigma_{\text{pooled}}$

where $\sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$

Compute

$r_{Y\lambda} = d / \sqrt{(d^2 + 4)}$

Note: d and $r_{Y\lambda}$ are positive if the mean difference is in the predicted direction.

Group 1	Group 2
M_1 <input type="text" value="0.13"/>	M_2 <input type="text" value="-0.03"/>
SD_1 <input type="text" value="0.94"/>	SD_2 <input type="text" value="0.88"/>
<input type="button" value="Reset"/>	
Cohen's d	effect-size r
<input type="text"/>	<input type="text"/>



Bewertung der klinischen Relevanz – Methode „Eine Regel für alles“ (2)

Calculate d and r using means and standard deviations

Calculate the value of Cohen's d and the effect-size correlation, $r_{Y\lambda}$, using the means and standard deviations of two groups (treatment and control).

$$\text{Cohen's } d = M_1 - M_2 / \sigma_{\text{pooled}}$$

$$\text{where } \sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$$

Compute

$$r_{Y\lambda} = d / \sqrt{(d^2 + 4)}$$

Note: d and $r_{Y\lambda}$ are positive if the mean difference is in the predicted direction.

Group 1	Group 2
M_1 <input type="text" value="0.13"/>	M_2 <input type="text" value="-0.03"/>
SD_1 <input type="text" value="0.94"/>	SD_2 <input type="text" value="0.88"/>
<input type="button" value="Reset"/>	
Cohen's d	effect-size r
<input type="text" value="0.175"/>	<input type="text" value="0.087"/>



Bewertung der klinischen Relevanz – Methode „Eine Regel für alles“ (3)

Effect Size and Clinical/Practical Significance

...

Cohen originally proposed the following interpretation:

Cohen's <i>d</i>	Interpretation
0.2	Small
0.5	Medium
0.8	Large

Quelle: http://www.umdj.edu/idsweb/shared/effect_size.htm

→ Effekt nicht klinisch relevant?



Bewertung der klinischen Relevanz – Methode „Es gibt keine Regel“

- „Anhänger der Evidenzbasierten Medizin (EBM) versuchen, aus **angeblich allgemeingültigen Erkenntnissen** Aussagen über die Therapie im Einzelfall abzuleiten.“
- „Eine patientenzentrierte Versorgung würde eben **genau nicht vom Wissen über allgemeine Zusammenhänge ausgehen**, die dann auf den Einzelfall zurechtgestutzt werden, sondern sie würde radikal beim Patienten ansetzen.“ (Walach, 2010)



Klinische Relevanz – was ist das und wo liegt das Problem?

- „*Statistical significance* refers to whether or not the value of a statistical test exceeds some prespecified level.
Clinical significance refers to the medical importance of a finding. *The two often agree but not always.*“ (Friedman, 1998)
- „Für die Beurteilung der klinischen Relevanz von Studienergebnissen gibt es noch *kein breit akzeptiertes Vorgehen.*“ (IQWiG, 2011)



Klinische Relevanz – was ist das und wo liegt das Problem?

- „In broad usage, the „practical clinical significance“ answers the question, *how effective is the intervention or treatment*, or how much change does the treatment cause? In terms of testing clinical treatments, practical significance optimally yields *quantified information about the importance of a finding, using metrics such as effect size...*“ (Wikipedia, 1998)
- „The extent to which therapy moves *someone* outside the range of the dysfunctional population or within the range of the functional population“ (Jacobson & Truax, 1991)



Klinisch relevant – für wen?

- **Patientenpopulation**

„Mittlere Verbesserung des EDSS-Scores unter Fingolimod um 0.03 Punkte, mittlere Verschlechterung unter Placebo um 0.13 Punkte.“

- **individueller Patient**

„Progression der Einschränkung liegt vor bei Verschlechterung des EDSS-Scores um 0.5 Punkte (Ausgangswert ≥ 5.5) bzw. 1 Punkt (Ausgangswert < 5.5).“

Welche Methoden und Kriterien werden zur Bewertung der klinischen Relevanz verwendet?



Methoden zur Bewertung der klinischen Relevanz

- **Signifikanztest auf relevante Überlegenheit (Victor, 1987)**
 - „Effekt $> \delta$?”
 - „... testing should no longer be done against the null effect but against a non-relevant improvement ...“
(Heidrich et al., 1995)
 - Vorteil: Signifikanz impliziert Relevanz
 - Nachteile:
 - erhebliche Erhöhung der notwendigen Fallzahl
 - Aussage für Patientenpopulation, nicht für individuellen Patienten



Methoden zur Bewertung der klinischen Relevanz (2)

- **Beurteilung anhand des beobachteten Effektes**
 - *„Establishing a clinically relevant benefit ... is accomplished by considering the point estimates of the difference between the test product and placebo...”* (CHMP, 2005)
 - abgeschwächte Variante des Tests auf relevante Überlegenheit ($\alpha = 0.5$ statt $\alpha = 0.05$)
 - Fallzahlberechnung für klinische Studien ✓
(Kieser & Hauschke, 2005)
 - Vorteil: „akzeptable“ Erhöhung der notwendigen Fallzahl
 - Nachteil: Aussage für Patientenpopulation, nicht für individuellen Patienten



Methoden zur Bewertung der klinischen Relevanz (3)

- **Responderanalysen**

- Beispiel MS: „*The key secondary endpoint was ... confirmed disability progression, defined as an increase of one point in the EDSS score (or half a point if the baseline EDSS score was equal to 5.5).*” (Kappos et al., 2010)
- Anteil der Patienten, die klinisch bedeutsames Ergebnis erzielen
- Fallzahlberechnung für klinische Studien ✓
(Kieser et al., 2004)
- in zahlreichen Guidelines empfohlen (z.B. EMA-Guidelines Multiple Sklerose, 2006; Alzheimer Demenz, 2008; Parkinson, 2008; Epilepsie, 2010; Depression, 2011; ...)



Methoden zur Bewertung der klinischen Relevanz – Responderanalysen

- **Vorteile**

- einfaches Maß für Effekt in der **Patientenpopulation**: $p_2 - p_1$
- basiert auf Erfolgskriterium für **individuellen Patienten**: $X \geq c$

- **Nachteile**

- Informationsverlust durch Dichotomisierung (Uryniak et al., 2011; Fedorov et al., 2009)
- Beliebigkeit des Cutpoints; Beispiel EDSS bei MS:
 - 0.5 / 1.0 Punkte bei Ausgangswert $> 5.5 / \leq 5.5$ (CHMP, 2006)
 - 0.5 / 1.0 Punkte bei Ausgangswert $\geq 5.5 / < 5.5$ (Kappos et al., 2010)
 - 1.5 Punkte bei Ausgangswert 0 (Giovannoni et al., 2010)
 - ...



Methoden zur Bewertung der klinischen Relevanz – alternativer Vorschlag

- **relativer Effekt $P(X_2 > X_1)$**
 - “Wahrscheinlichkeit, dass ein Patient, der mit Therapie 2 behandelt wird, ein besseres Ergebnis erzielt als ein Patient, der mit Therapie 1 behandelt wird.”
 - Interpretation: Wahrscheinlichkeit, dass der Arzt mit Wahl von Therapie 2 den Hippokratischen Eid erfüllt:
“*To help or, at least, do no harm*”. (Bordley, 2009)
- **Vorteile:** wie bei Responderanalyse ...
- **ohne die Nachteile** der Responderanalyse!

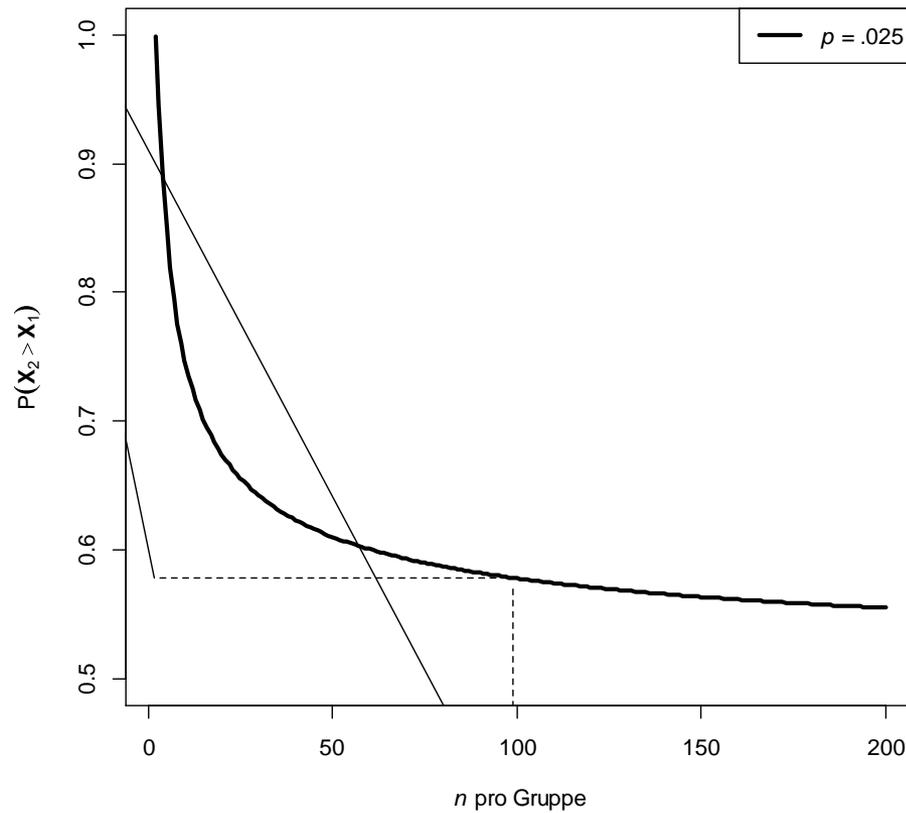


Methoden zur Bewertung der klinischen Relevanz – relativer Effekt

- **allgemeines Konzept mit engen Beziehungen zu „alten Bekannten“**
 - normalverteilte Endpunkte: $P(X_2 > X_1) = \Phi\left(\frac{1}{\sqrt{2}} \cdot \frac{\Delta}{\sigma}\right)$
 - binäre Endpunkte: $P(X_2 > X_1) = 0.5 + 0.5 \cdot (p_2 - p_1)$
 - Überlebenszeiten: $P(X_2 > X_1) = HR / (1 + HR)$
- **Schätzung und Konfidenzintervalle ✓**
(Brunner & Munzel, 2000; Newcombe, 2006)
- **Fallzahlplanung für klinische Studien ✓**
(Kieser et al., 2011)
- **günstige statistische Eigenschaften**

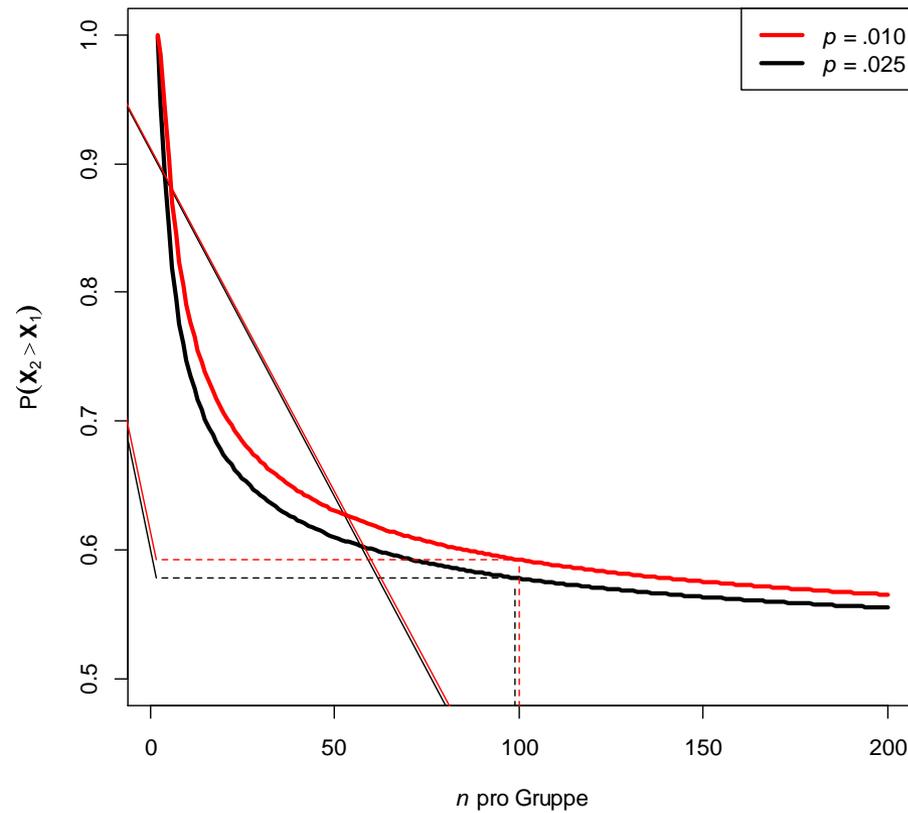


Zusammenhang zwischen p-Wert und relativem Effekt – Signifikanz und Relevanz



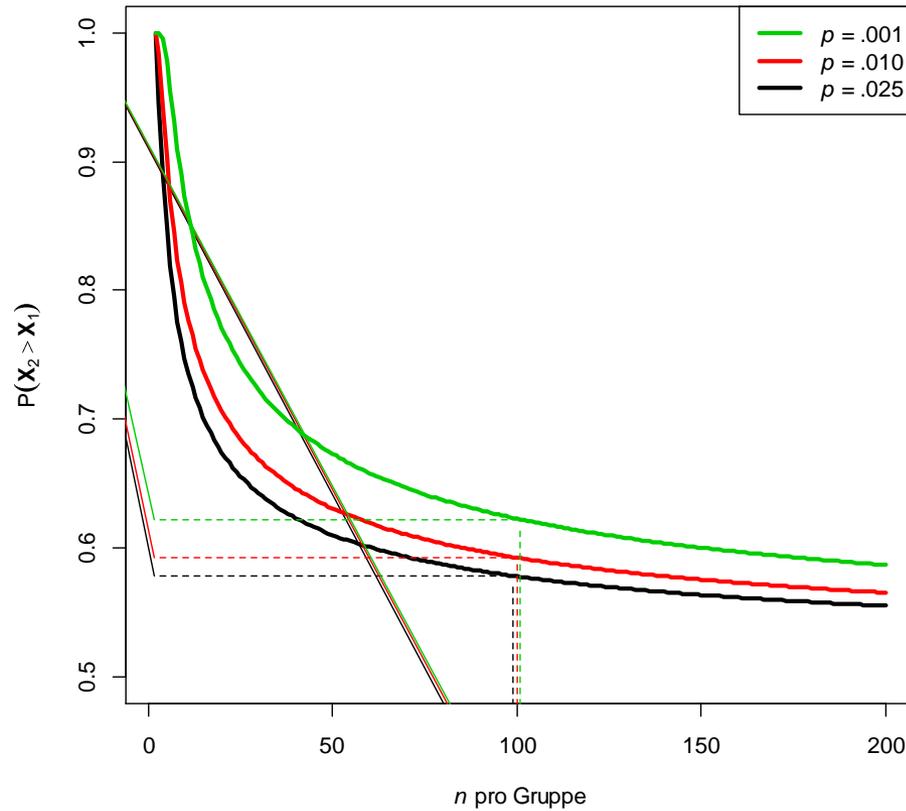


Zusammenhang zwischen p-Wert und relativem Effekt – Signifikanz und Relevanz





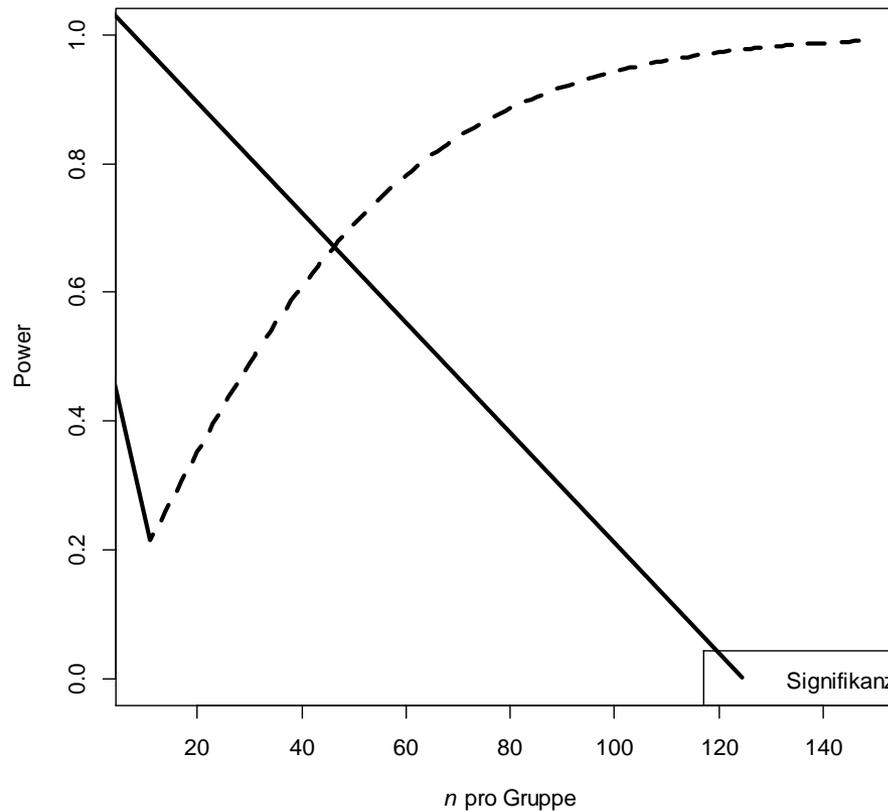
Zusammenhang zwischen p-Wert und relativem Effekt – Signifikanz und Relevanz





Powervergleich Responderanalyse versus relativer Effekt

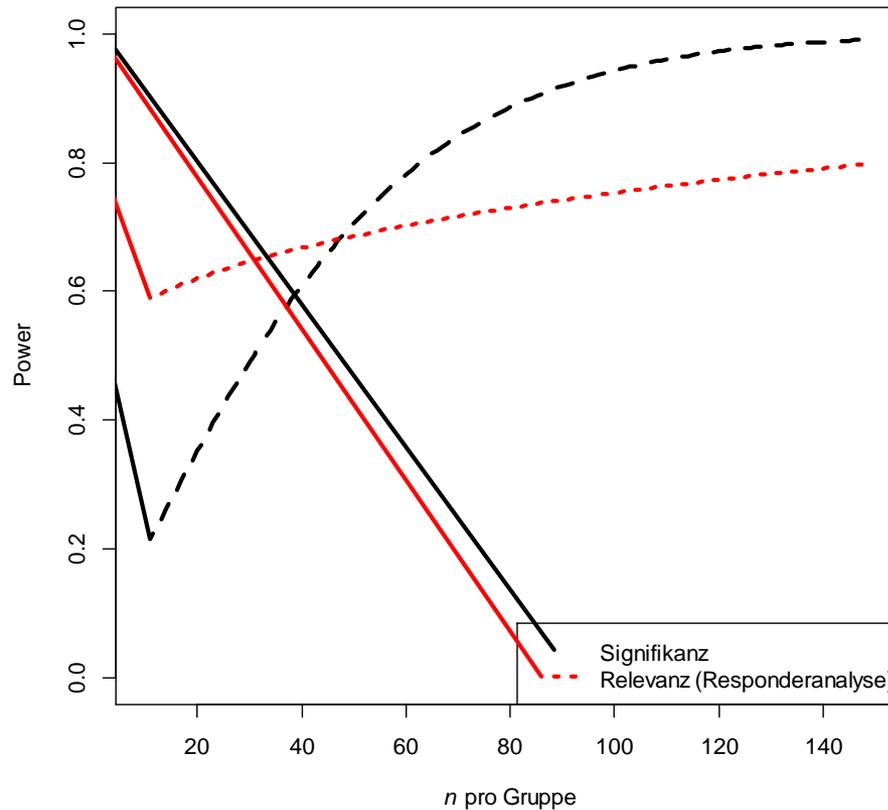
Beispiel: Power für $\Delta / \sigma = 0.5$, d.h. $P(X_2 > X_1) = 0.638$





Powervergleich Responderanalyse versus relativer Effekt

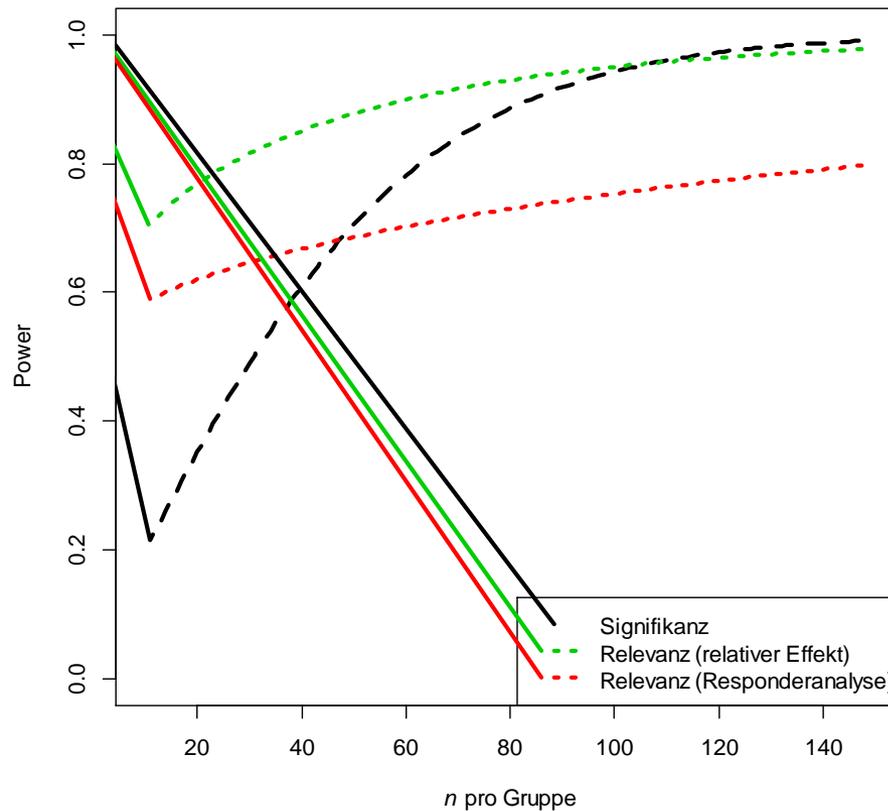
Beispiel: Power für $\Delta / \sigma = 0.5$, d.h. $P(X_2 > X_1) = 0.638$





Powervergleich Responderanalyse versus relativer Effekt

Beispiel: Power für $\Delta / \sigma = 0.5$, d.h. $P(X_2 > X_1) = 0.638$





Wann ist ein Effekt klinisch relevant?

- **Wahl der Relevanzschwelle** abhängig von
 - Schwere der Erkrankung
 - Prävalenz der Erkrankung
 - anderen Nutzen und Risiken der Behandlung
 - ...



Wann ist ein Effekt klinisch relevant?

- **Wahl der Relevanzschwelle** abhängig von
 - Schwere der Erkrankung
 - Prävalenz der Erkrankung
 - anderen Nutzen und Risiken der Behandlung
 - ...
- **Beispiel:** „kleiner“ Effekt evtl. klinisch relevant, wenn
 - schwerwiegende Erkrankung
 - Behandlung nicht mit Risiken verbunden



Wann ist ein Effekt klinisch relevant? (2)

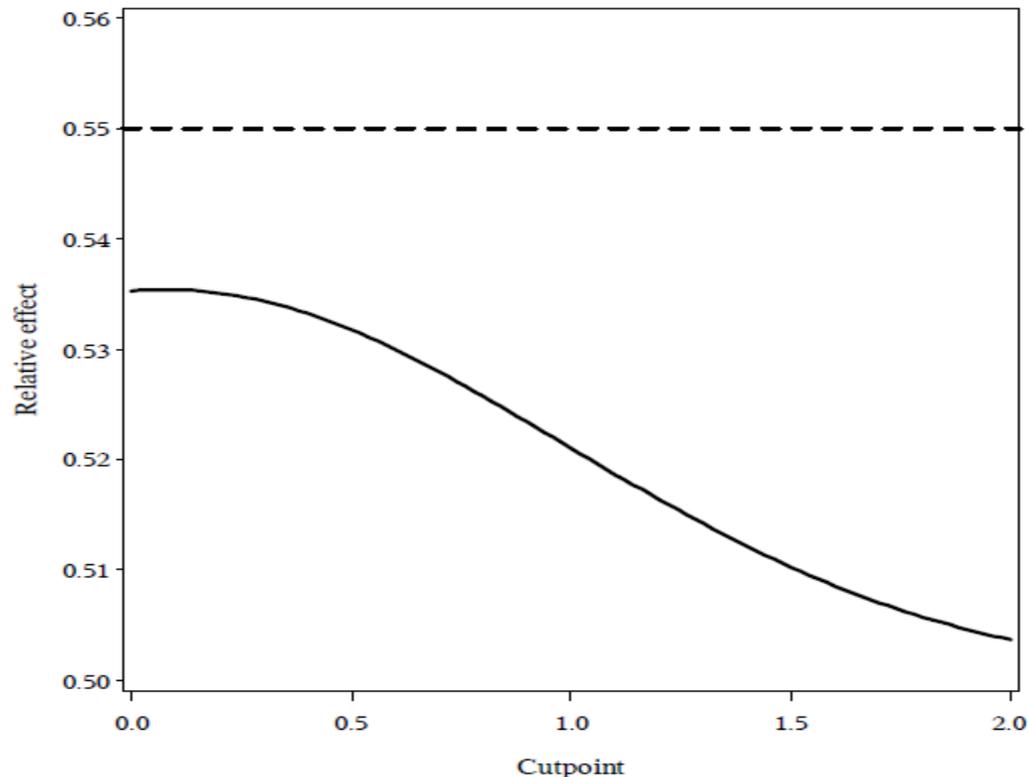
- **Wahl der Relevanzschwelle** abhängig von
 - Schwere der Erkrankung
 - Prävalenz der Erkrankung
 - anderen Nutzen und Risiken der Behandlung
 - ...
- **Beispiel:** „großer“ Effekt evtl. klinisch irrelevant, wenn
 - leichte und vorübergehende Befindlichkeitsstörung
 - Behandlung mit Risiken verbunden



Wann ist ein Effekt klinisch relevant? (3)

- **Wahl der Relevanzschwelle** auch abhängig von **Skalenniveau und Definition des Endpunktes**

Beispiel: relativer Effekt für verschiedene Dichotomisierungen





Wann ist ein Effekt klinisch relevant? (4)

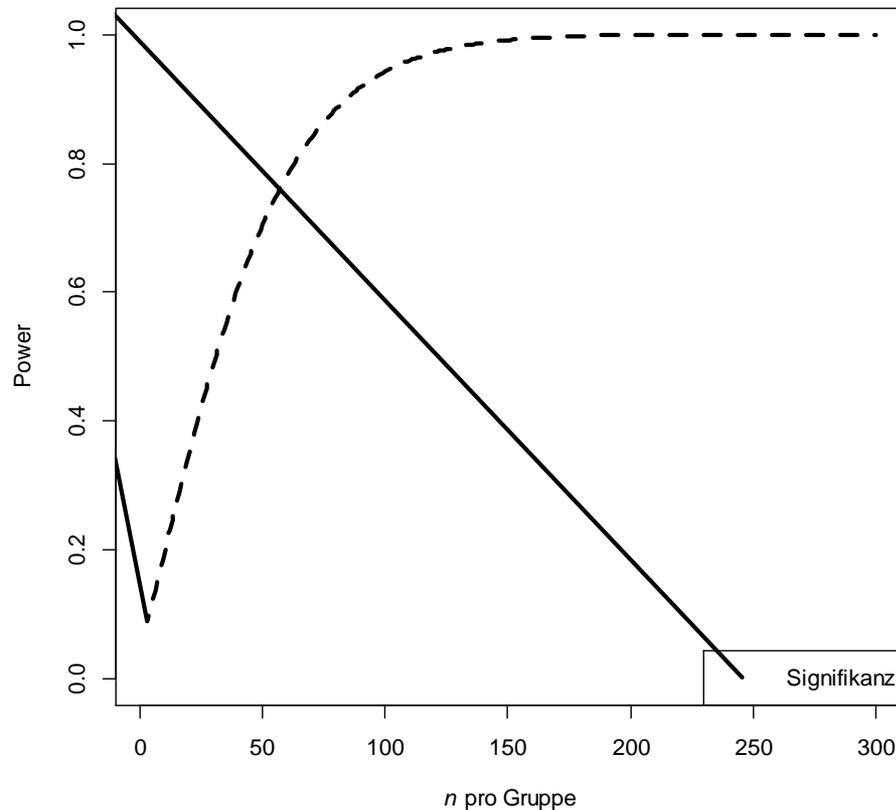
- **beobachteter Effekt oder Konfidenzintervall über Relevanzschwelle?**
 - *„... by considering the point estimates of the difference...”* (CHMP, 2005)
 - *„Liegt das zum beobachteten Effekt korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle...”* (IQWiG, 2011)



Wann ist ein Effekt klinisch relevant? (5)

- beobachteter Effekt oder Konfidenzintervall?

Beispiel: Power für $\Delta / \sigma = 0.5$, d.h. $P(X_2 > X_1) = 0.638$

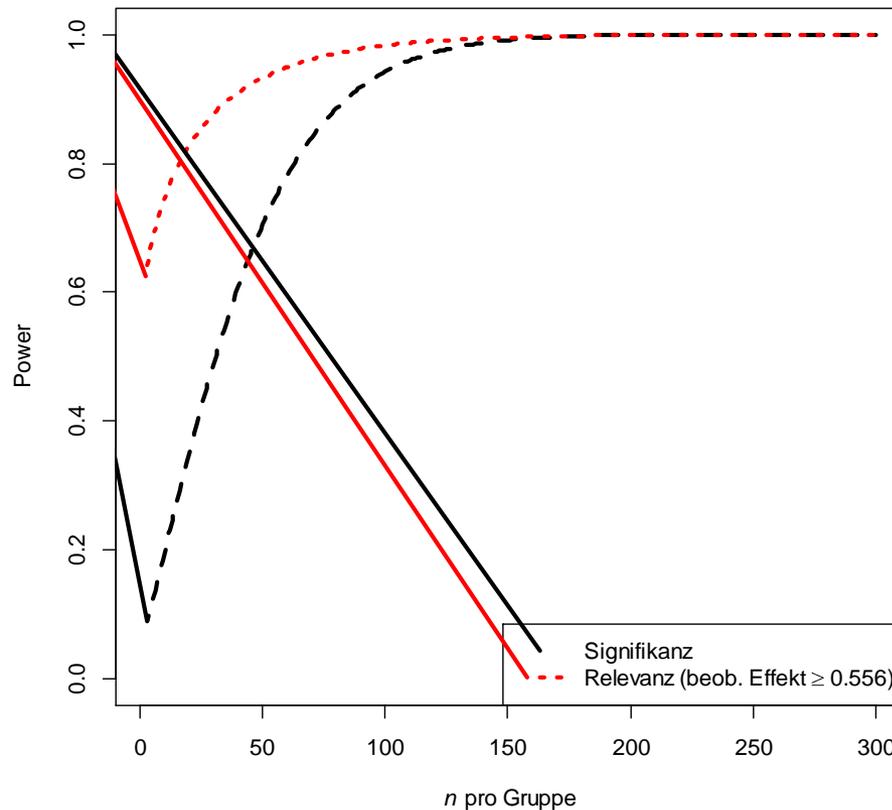




Wann ist ein Effekt klinisch relevant? (5)

- beobachteter Effekt oder Konfidenzintervall?

Beispiel: Power für $\Delta / \sigma = 0.5$, d.h. $P(X_2 > X_1) = 0.638$

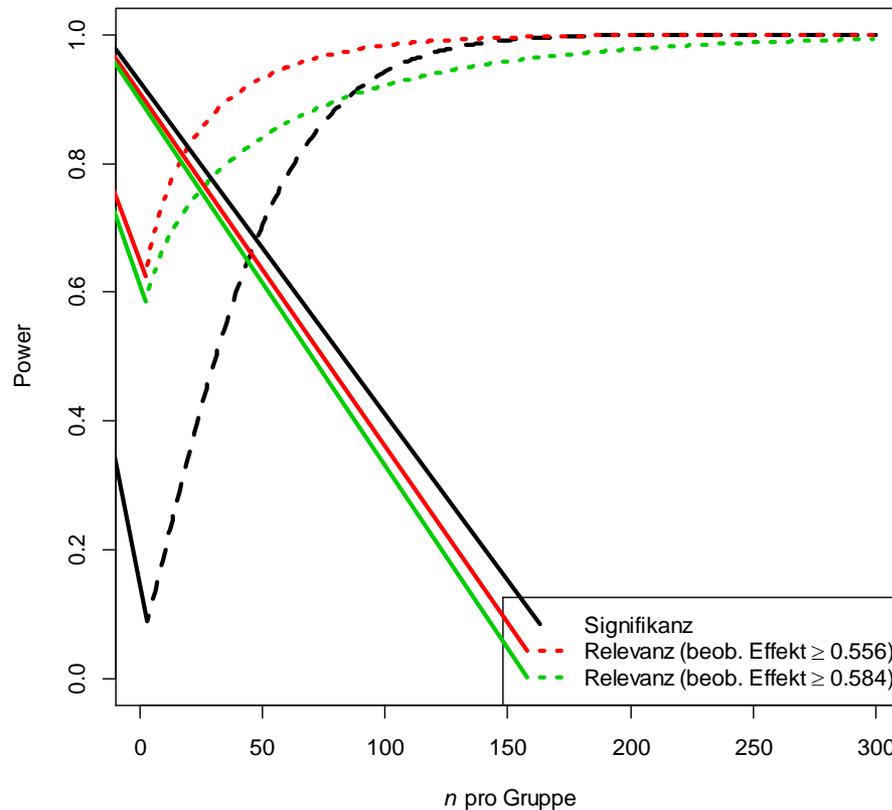




Wann ist ein Effekt klinisch relevant? (5)

- beobachteter Effekt oder Konfidenzintervall?

Beispiel: Power für $\Delta / \sigma = 0.5$, d.h. $P(X_2 > X_1) = 0.638$

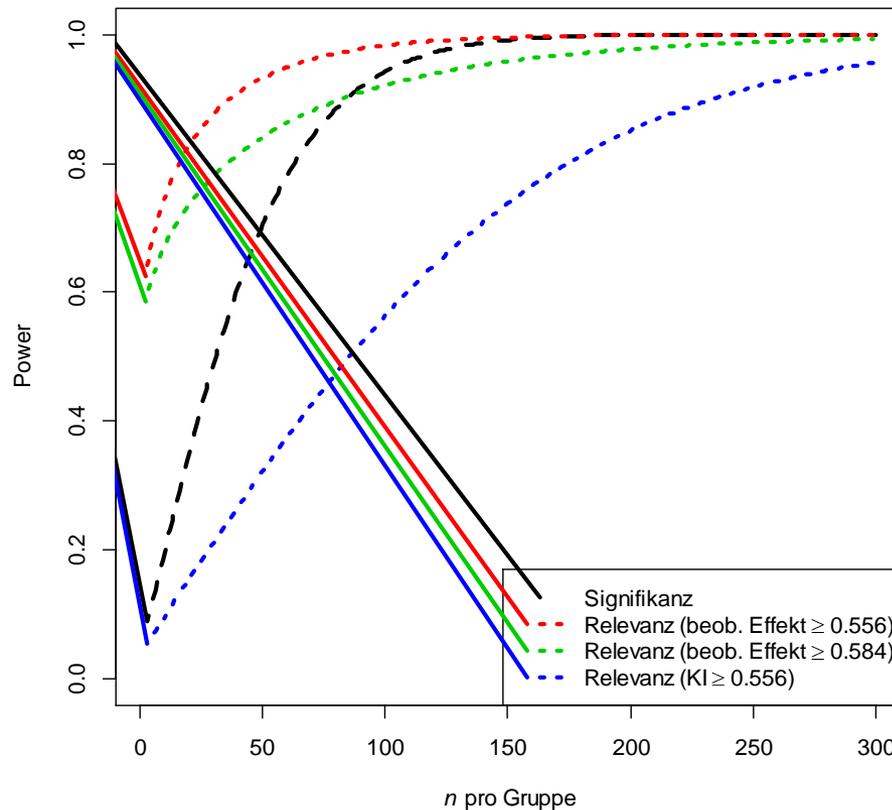




Wann ist ein Effekt klinisch relevant? (5)

- beobachteter Effekt oder Konfidenzintervall?

Beispiel: Power für $\Delta / \sigma = 0.5$, d.h. $P(X_2 > X_1) = 0.638$





Wann ist ein Effekt klinisch relevant? (6)

- **beobachteter Effekt oder Konfidenzintervall über Relevanzschwelle?**
 - **Abwägung zwischen**
 - **gewählter Relevanzschwelle**
 - **Grad der notwendigen statistischen Absicherung**
 - **realisierbarer Fallzahl**



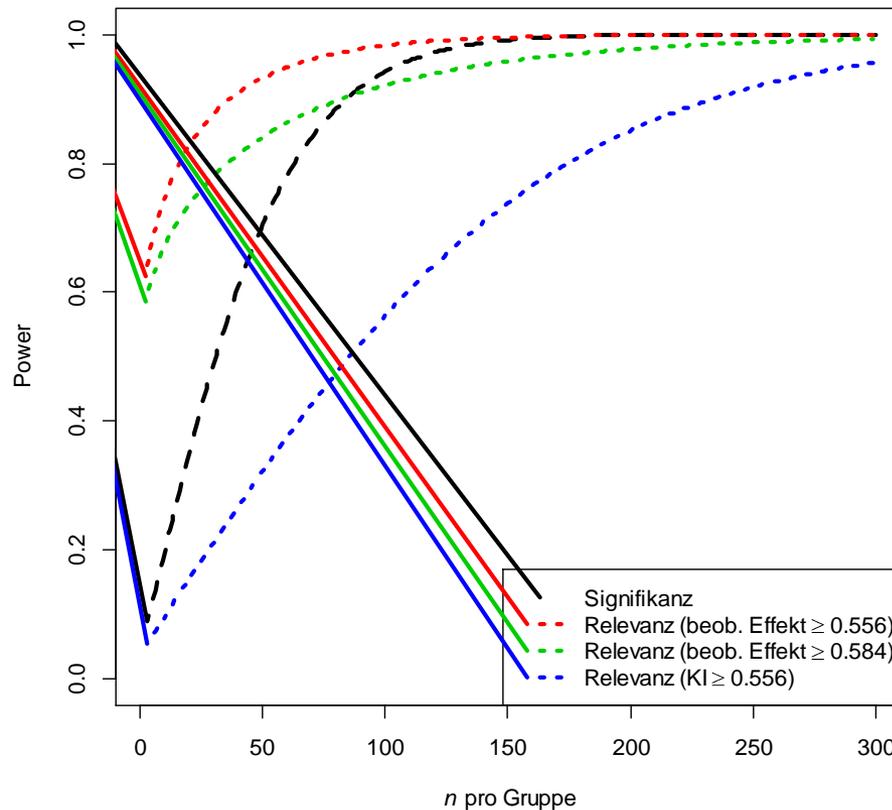
Bedeutung der Prä-Spezifizierung

- **Design und Fallzahl klinischer Studien** sind auf jeweilige Ziele ausgerichtet
 - **Spezialfall Beurteilung der klinischen Relevanz:** Ergebnis hängt ab von
 - tatsächlichem Effekt
 - Methode
 - Relevanzschwelle
 - beobachteter Effekt oder Konfidenzintervall
- **Notwendigkeit der Festlegung vor Studiendurchführung**



... oder Berücksichtigung bei der Wahl der Bewertungskriterien

Beispiel: Power für $\Delta / \sigma = 0.5$, d.h. $P(X_2 > X_1) = 0.638$

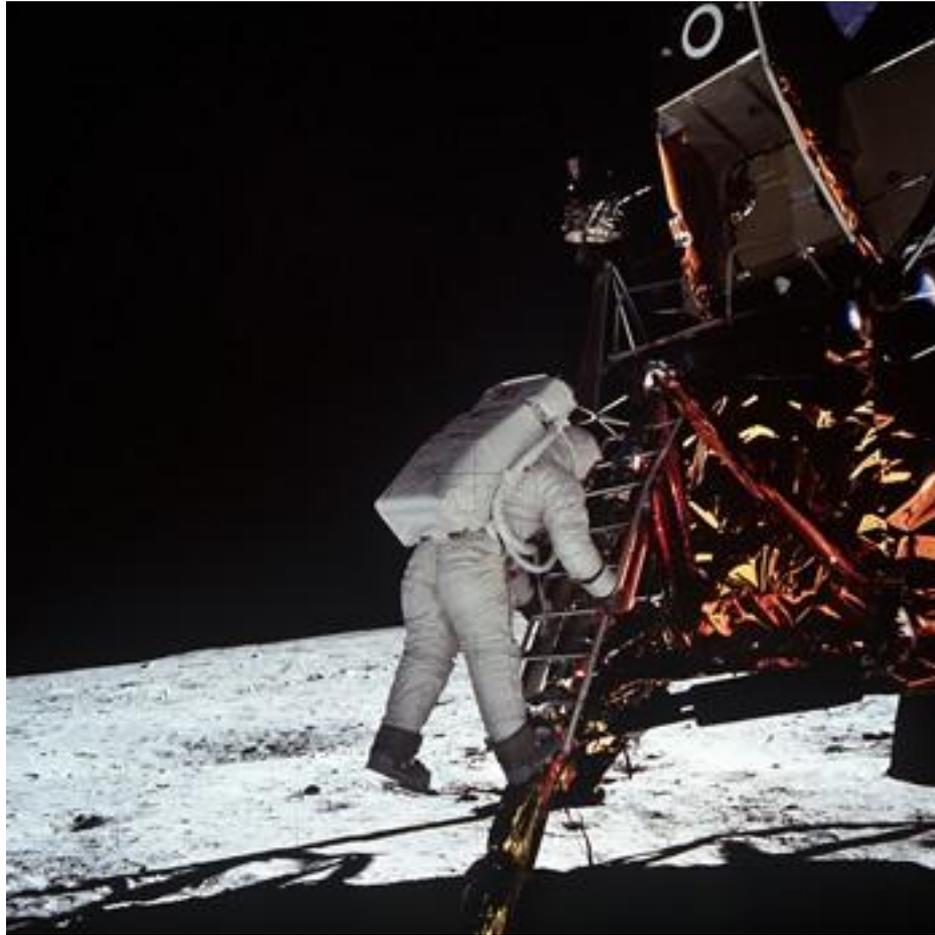




Was war das *relevantere* Ergebnis der NASA-Mondfahrten ...



Was war das *relevantere* Ergebnis der NASA-Mondfahrten ...





... oder das?





Diskussion

- **Regeln oder Willkür? Prinzipien!**



Diskussion

- **Regeln oder Willkür? Prinzipien!**
 - **Methode**
 - Interpretation für indiv. Patienten und Population möglich
- **Responderanalyse (mit versch. Cutpoints!) oder relativer Effekt**



Diskussion

- **Regeln oder Willkür? Prinzipien!**
- **Methode**
 - Interpretation für indiv. Patienten und Population möglich
- **Responderanalyse (mit versch. Cutpoints!) oder relativer Effekt**
- **Wahl der Relevanzschwelle:** abhängig von
 - Indikation, anderen Nutzen und Risiken, ...
- **kontextbezogene Festlegung**



Diskussion

- **Regeln oder Willkür? Prinzipien!**
- **Methode**
 - Interpretation für indiv. Patienten und Population möglich
- **Responderanalyse (mit versch. Cutpoints!) oder relativer Effekt**
- **Wahl der Relevanzschwelle:** abhängig von
 - Indikation, anderen Nutzen und Risiken, ...
- **kontextbezogene Festlegung**
- **beobachteter Effekt oder Konfidenzintervall:** abhängig von
 - gewählter Relevanzschwelle, Realisierbarkeit, Grad der notwendigen statistischen Absicherung
- **sinnvolle Balance zwischen Notwendigem und Machbarem**