IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Institute for Quality and Efficiency in Health Care

IQWiG im Dialog 2010

Bewertung der klinischen Relevanz bei der Nutzenbewertung

Ein formalisiertes Vorgehen zur Bewertung der Relevanz von Gruppenunterschieden auf Skalen

Stefan Lange, Thomas Kaiser, Yvonne Beatrice-Schüler, Guido Skipka, Volker Vervölgyi, Beate Wieseler

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

"Klinische Studien werden mit dem Ziel durchgeführt, Ergebnisse zu liefern, die in der Praxis zu einer Verbesserung der Patientenversorgung führen. Allgemein wird jedoch beklagt, dass die Charakterisierung von Studienergebnissen als 'signifikant' oder 'nicht signifikant' keine Aussagen über ihre klinische Relevanz erlaubt."

Windeler J, Conradt C. Wie können "Signifikanz" und "Relevanz" verbunden werden?. Med Klin 1999; 94: 652-5.

"The scientific basis for recommending donepezil, rivastigmine, or galantamine as preferred treatment for patients with Alzheimer's disease is questionable because minimal benefits were measured on rating scales and the methodological quality of the available trials was poor. … The gains of 1.5-3.9 points in cognitive function, as measured with the Alzheimer's disease assessment scale, fall below the 4 points that a panel of experts from the US Food and Drug Administration proposed as the minimum of a clinically important effect."

Kaduszkiewicz H, Zimmermann T, Beck-Bornholdt H-P, van den Bussche H. Cholinesterase inhibitors for patients with Alzheimer's disease: systematic review of randomised clinical trials. BMJ 2005; 331: 321-7.

"Das Institut soll Nutzenbewertungen erarbeiten, die eine Aussage über den Beitrag neuer Arzneimittel zur Verbesserung der medizinischen Behandlung von Patienten beinhalten. Hierzu soll das Institut auch erarbeiten, für welche Patientengruppen ein neues Arzneimittel eine maßgebliche Verbesserung des Behandlungserfolgs erwarten lässt mit dem Ziel, dass diese Patienten das neue Arzneimittel erhalten sollen. Diese Abgrenzung soll auch Inhalt der Bewertungen des Instituts sein."

BT-Drucksache 15/1525, S. 88, Begründung zum GMG

"We have developed an approach to elucidating the significance of changes in score in quality of life instruments by comparing them to global ratings of change. Using this approach we have established a plausible range within which the minimal clinically important difference (MCID) falls. … This information will be useful in interpreting questionnaire scores, both in individuals and in groups of patients participating in controlled trials, and in the planning of new trials."

Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989; 10: 407-15.

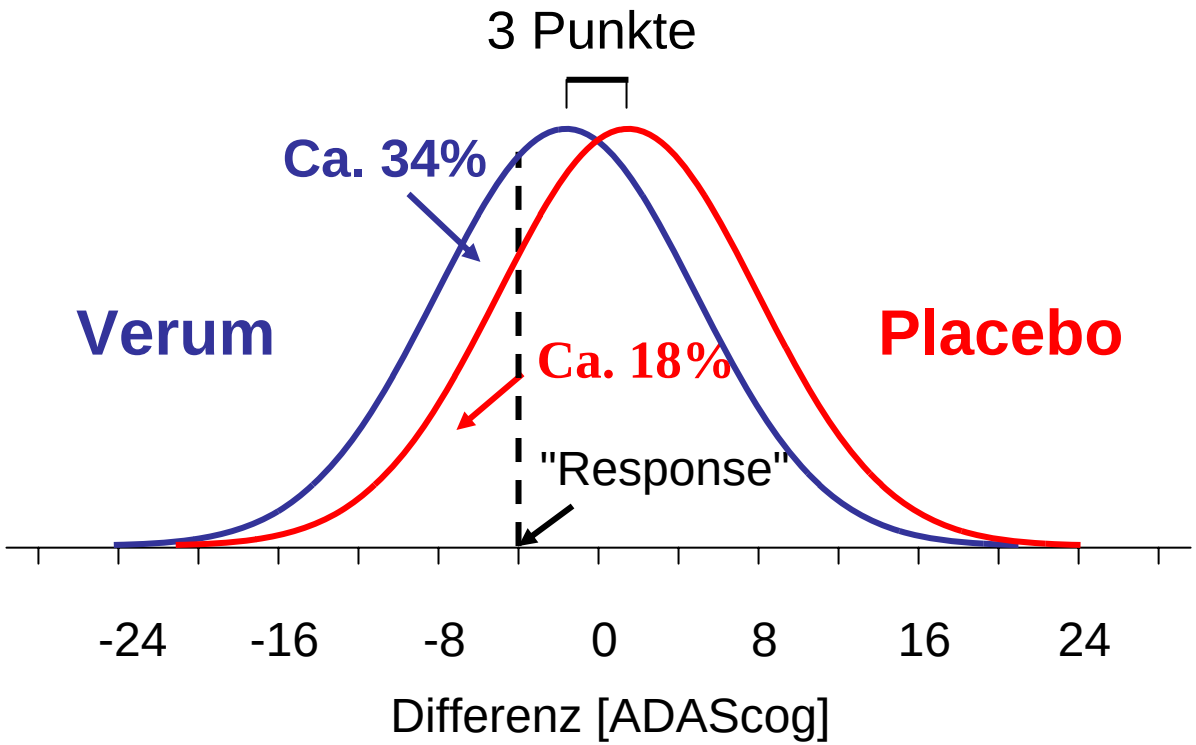
"Clinicians and investigators tend to assume that if the mean difference between a treatment and a control is appreciably less than the smallest change that is important, then the treatment has a trivial effect. This may not be so."

"Even if the mean difference between a treatment and a control is appreciably less than the smallest change that is important, treatment may have an important impact on many patients"

Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. BMJ 1998; 316: 690-3.

?

M(C)ID = geeignetes Responsekriterium?



Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Draft Guidance. 2006

" … extremely complex concepts such as quality of life, which is widely understood to be a multidomain concept … "

"The amount and kind of evidence that the FDA expects to support a labeling claim measured by a PRO instrument is the same as that required for any other labeling claim."

Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Draft Guidance. 2006

"Smallest difference that is considered clinically important; this can be a specified difference (the minimum important difference (MID)) or, in some cases, any detectable difference. The MID is used as a benchmark to interpret mean score differences between treatment arms in a clinical trial."

"Responder definition - Change in score that would be clear evidence that an individual patient experienced a treatment benefit. Can be based on experience with the measure using a distribution-based approach, a clinical or nonclinical anchor, an empirical rule, or a combination of approaches."

Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Draft Guidance. 2006

"If the MID is truly to be the smallest effect considered meaningful, however, it would be logical to establish the null hypothesis to rule out a difference less than or equal to the MID. This is rarely done, and would have major implications for sample size."

Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.
December 2009

~~"If the MID is truly to be the smallest effect considered meaningful, however, it would be logical to establish the null hypothesis to rule out a difference less than or equal to the MID. This is rarely done, and would have major implications for sample size."~~

Planning for Clinical Trial Interpretation Using a Responder Definition

- pAVK Prüfrichtlinien (Fontaine II), 1995

"Für die Annahme einer relevanten therapeutischen Wirksamkeit von vasoaktiven Substanzen wird es für erforderlich gehalten, dass der beobachtete Unterschied für die schmerzfreie Gehstrecke bei Abschluss der Prüfung gegenüber Placebo mindestens 30% beträgt. Bei adäquater Planung der Fallzahl wird dies in der Regel dann der Fall sein, wenn die Nullhypothese, dass der irrelevante Unterschied zwischen Verum und Placebo höchstens 20% beträgt, auf dem einseitigen Signifikanzniveau alpha = 5% abgelehnt werden kann …"

Heidrich H, Cachovan M, Creutzig A, Rieger H, Trampisch HJ. Prüfrichtlinien für Therapiestudien im Fontaine Stadium II-IV bei peripherer arterieller Verschlusskrankheit. VASA 1995; 24: 107-113.

- pAVK Prüfrichtlinien (Fontaine II), 1995 - Kommentar

"This comment emphasizes again that testing should no longer be done against the null effect but against a 'non relevant' improvement of at least 5%."

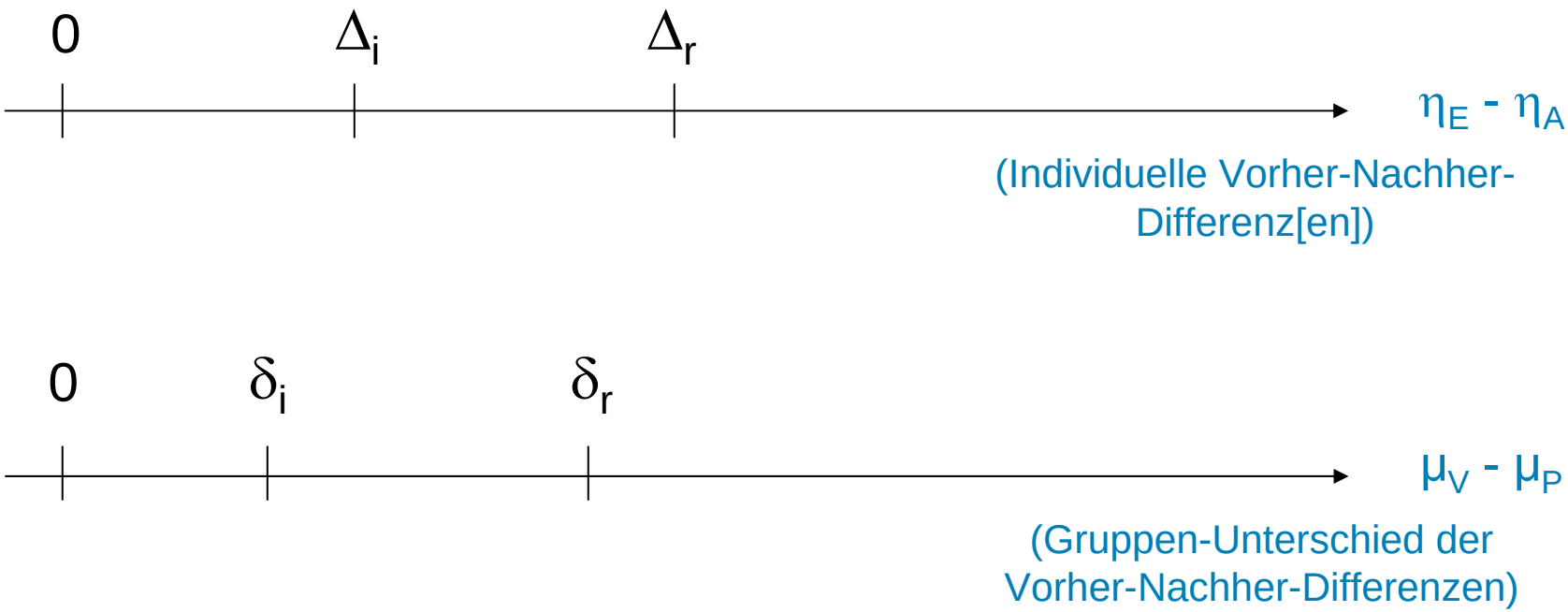
Heidrich H, Trampisch HJ, Röhmel J. Comments on Guidelines for therapeutic studies in Fontaine's stage II –IV peripheral arterial occlusive disease. VASA 1996; 25: 74-75.

"Deshalb wird vorgeschlagen, die klinische Relevanz und Bedeutung anhand von vier in Diskussionen mit dem Kliniker vor Studienbeginn festzulegenden Größen zu beurteilen und verschobene Nullhypothesen zu testen, in die die 'klinische relevante Differenz' als Verschiebungsparameter eingeht. Methodische Probleme, …, werden aufgezeigt und andere Möglichkeiten … (Konstruktion von Erfolgskriterien) diskutiert."

Victor N. On Clinically Relevant Differences and Shifted Nullhypotheses. Methods Inf Med 1987; 26: 109-16.

Schwartz et al. 1970: L'essai Thérapeutique chez l'Homme.

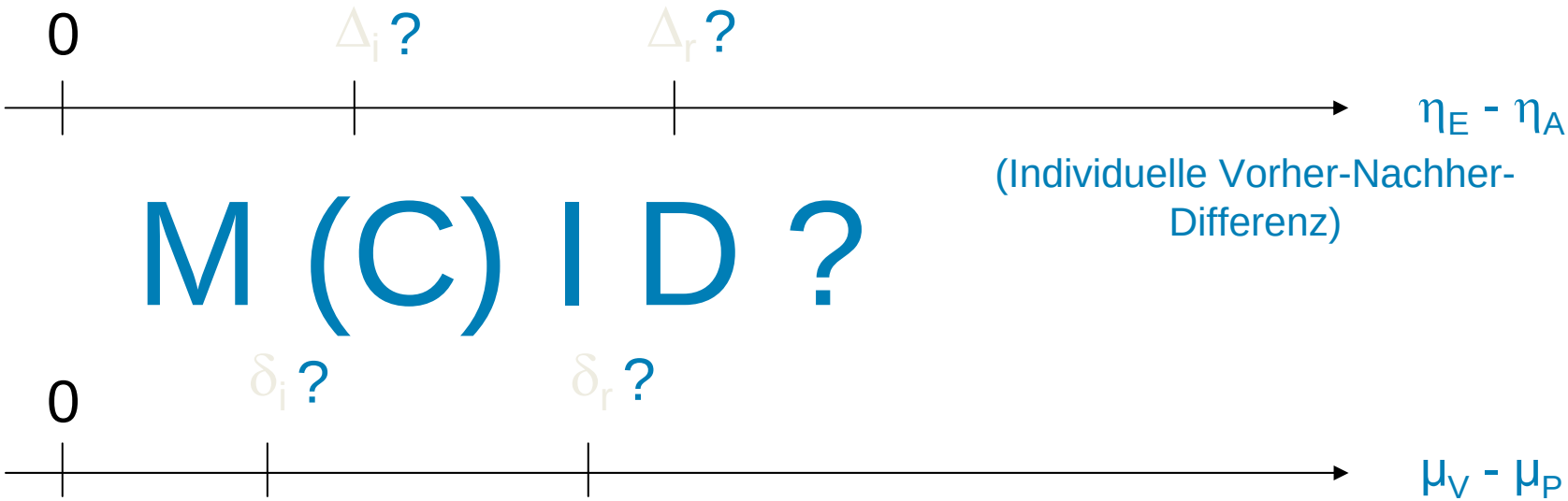
→ verschobene Nullhypothese → "Handicap"



Δ_i / δ_i = sicher irrelevant

Δ_r / δ_r = sicher relevant

Adaptiert nach Victor N. On Clinically Relevant Differences and Shifted Nullhypotheses. Methods Inf Med 1987; 26: 109-16.

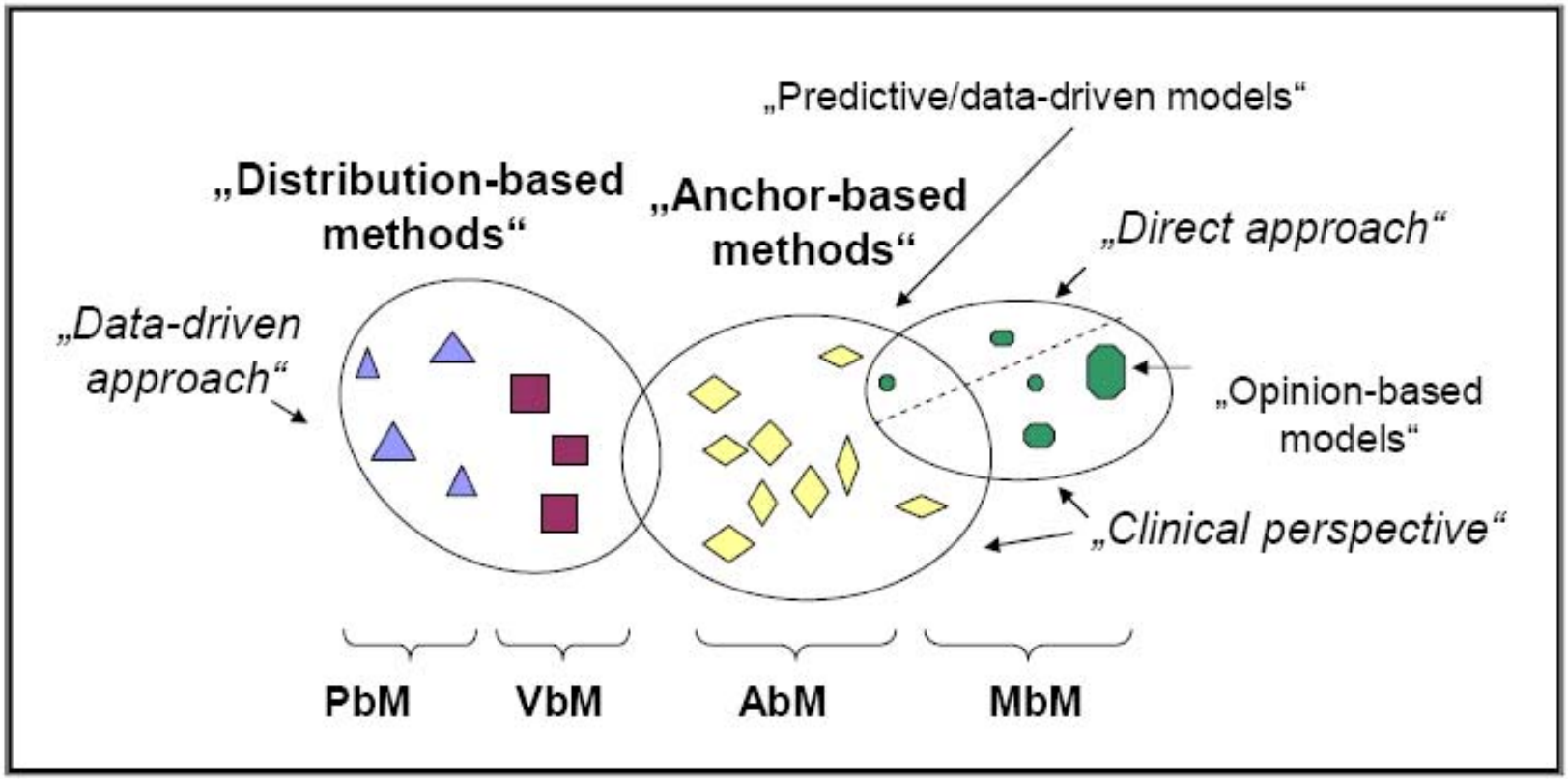


Δ_i / δ_i = sicher irrelevant

Δ_r / δ_r = sicher relevant

Adaptiert nach Victor N. On Clinically Relevant Differences and Shifted Nullhypotheses. Methods Inf Med 1987; 26: 109-16.

Vielfalt von Methodenvorschlägen für M(C)ID

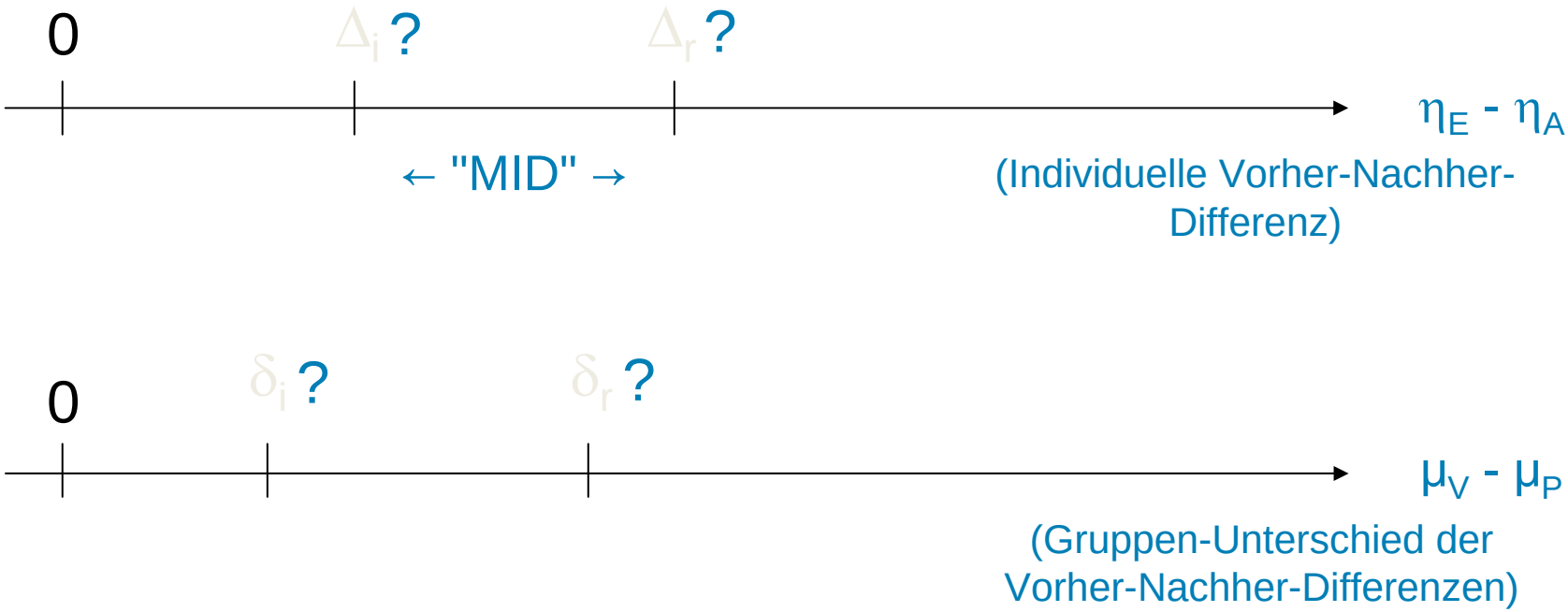


Thomas, S. Klinische Relevanz von Therapieeffekten - Systematische Sichtung, Systematisierung und Bewertung methodischer Konzepte. Dissertation, Universität Duisburg-Essen, 2009

Überblick: Intendierter Verwendungszweck pro Methoden- und Endpunkttyp

	Endpunkte	n	Häufigkeit angegebener Verwendungszwecke lt. Autoren*			Orientierend: Besonderheiten
PbM (n=30)	HRQoL	19	Re (10)	*unkl.* (5)	FZP (5)	**Reliabilitätskoeffizient** bei HRQoL-Skalen Cronbachs α /Test-Retest (n=10 bzw. n=9); für andere Skalen Test-Retest/Interrater (n=6 bzw. n=3), selten Cronbachs α: (n=1)‡
	Symptom/Funktion†	10	Re (10)	FZP (3)	*unkl.* (2)	
	Schmerz	1	FZP (1)			
VbM	HRQoL	12	*unkl.* (6)	FZP (4)	Re (3)	Nur multi-Item-Skalen;
AbM (n=120)	Ska-len HRQoL	82	*unkl.* (40)	FZP (25)	Re (20)	
	Symptom/Funktion	22	*unkl* (8)	FZP (7)	Re (7)	
	Schmerz¶	13	FZP (5)	*unkl.* (4)	Re (3)	
	sonstige Messgrößen für Symptome und Funktion	3	*unkl.* (3)			
	Symptom/Funktion	10	FZP (4)	Re (4)	*unkl.* (2)	**gungen nie** bei HRQoL, selten bei Symptom-/Funktionsskalen bzw. Schmerz (je n=1) und häufig bei binären Endpunkten (n=8)
	Schmerz	2	FZP (1)	Re (1)		
	re Endpunkte und wandte" intervallskalierte	11	FZP (10)	Re (1)		**Trade-off-Verfahren** nur bei binären Endpunkten (n=7)
MbM (n=30)						

Thomas, S. Klinische Relevanz von Therapieeffekten - Systematische Sichtung, Systematisierung und Bewertung methodischer Konzepte. Dissertation, Universität Duisburg-Essen, 2009



Δ_i / δ_i = sicher irrelevant

Δ_r / δ_r = sicher relevant

Adaptiert nach Victor N. On Clinically Relevant Differences and Shifted Nullhypotheses. Methods Inf Med 1987; 26: 109-16.

(Individuelle Vorher-Nachher-Differenz)

Annahme: δ_r sei bekannt

(Gruppen-Unterschied der Vorher-Nachher-Differenzen)

Δ_i / δ_i = sicher irrelevant

Δ_r / δ_r = sicher relevant

Verschobene Nullhypothese testen?

Punktschätzer bewerten?

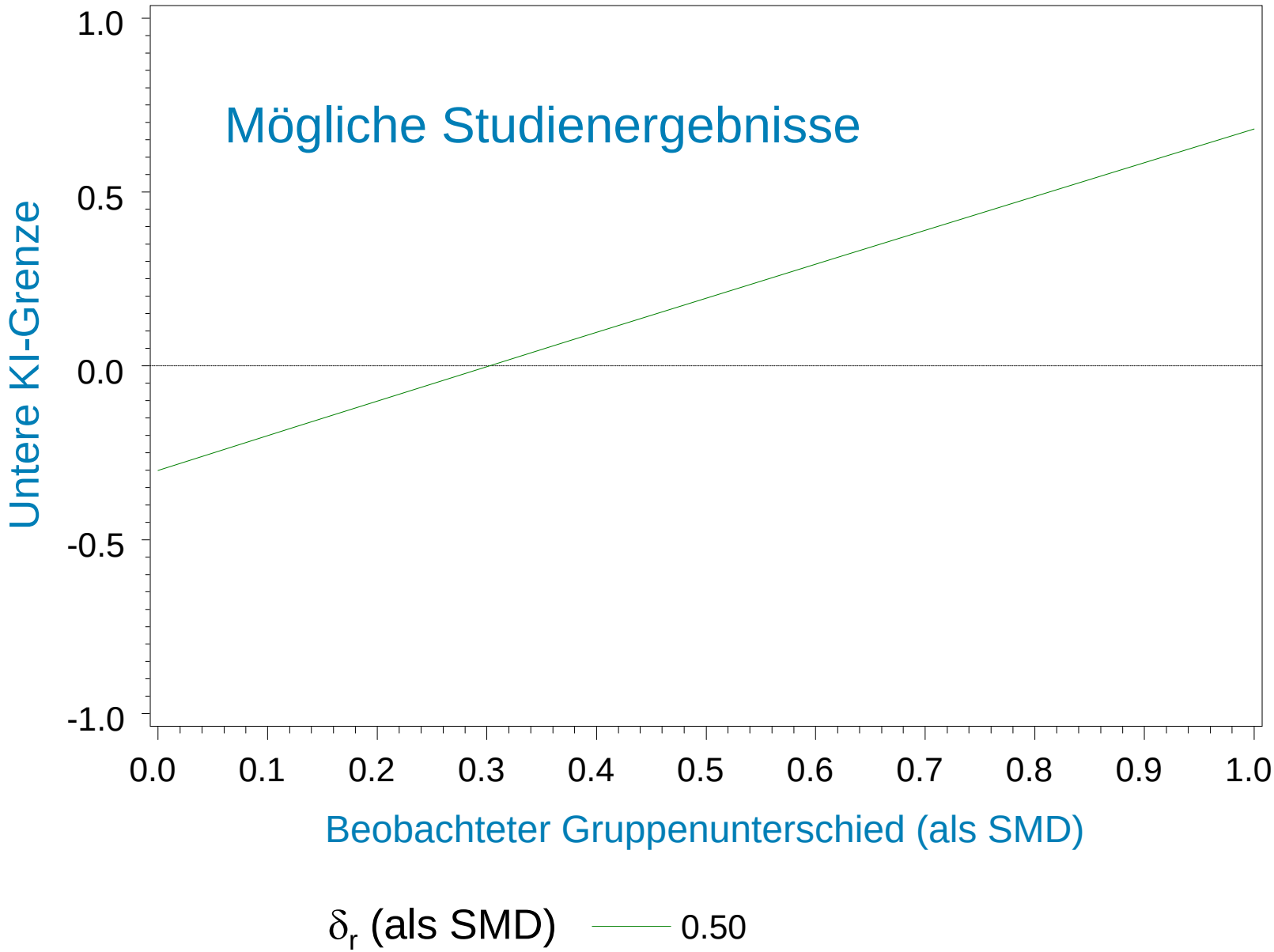
Adaptiert nach Victor N. On Clinically Relevant Differences and Shifted Nullhypotheses. Methods Inf Med 1987; 26: 109-16.

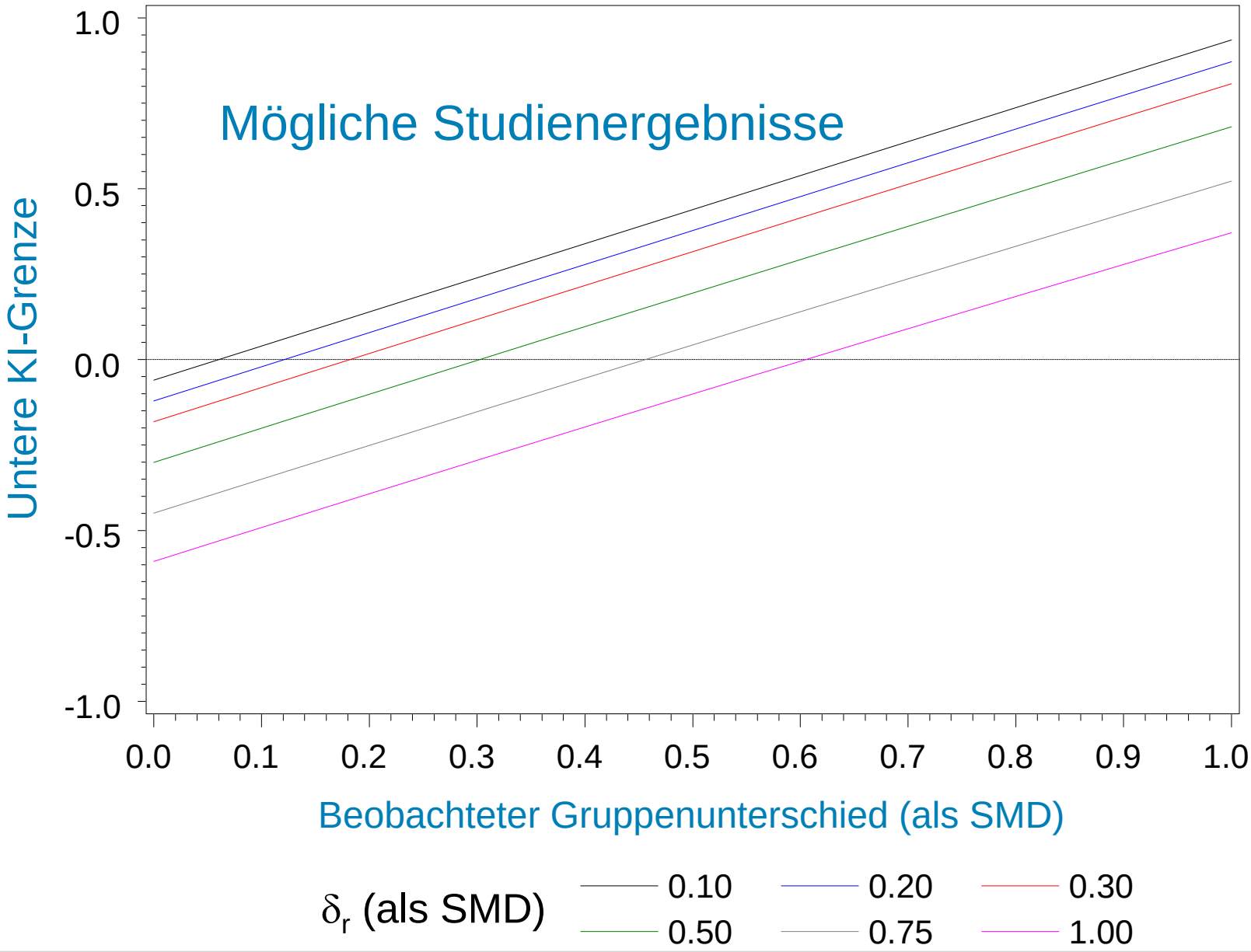
Annahme: δ_r sei bekannt

1. Durchführung einer Fallzahlplanung für eine hypothetische Studie mit folgenden Annahmen:

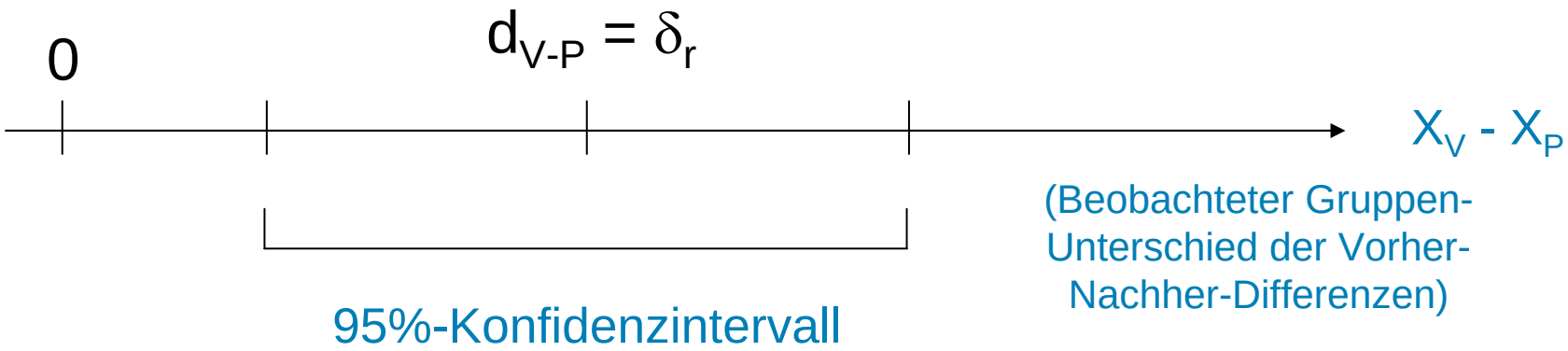
- normalverteilte Daten
- geplanter Gruppenunterschied (als SMD*), der δ_r entspricht
- $\alpha = 0{,}05$ (zweiseitig)
- $\beta = 0{,}1$ (d.h. Power = 90 %)
- gleichgroße Gruppen ($n_1 = n_2$)

*Standardisierte Mittelwertsdifferenz

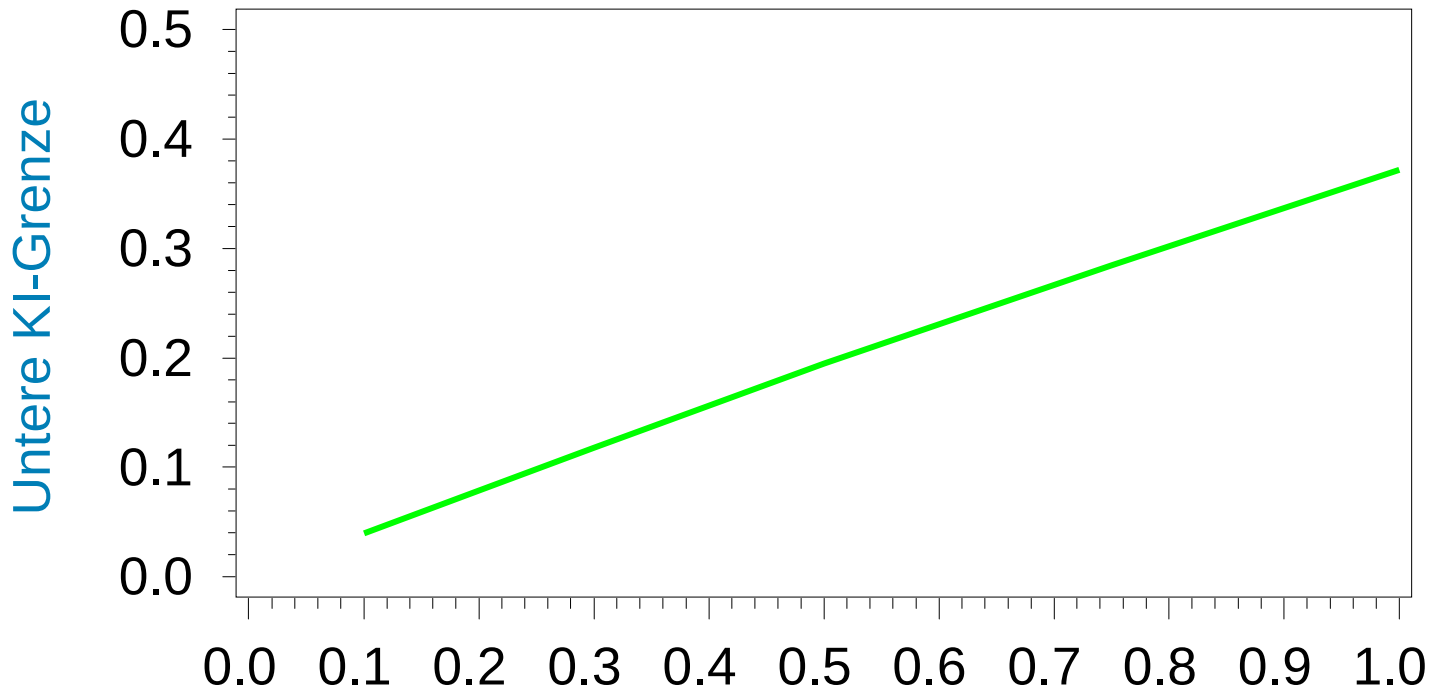

Mögliche Studienergebnisse
Untere KI-Grenze
1.0
0.5
0.0
-0.5
-1.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Beobachteter Gruppenunterschied (als SMD)
δr (als SMD)
0.50


Mögliche Studienergebnisse
Untere KI-Grenze
1.0
0.5
0.0
-0.5
-1.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Beobachteter Gruppenunterschied (als SMD)
δr (als SMD)
0.10
0.20
0.30
0.50
0.75
1.00

Beobachtetes Ergebnis = δ_r



Studienergebnis = Annahme für Fallzahlplanung

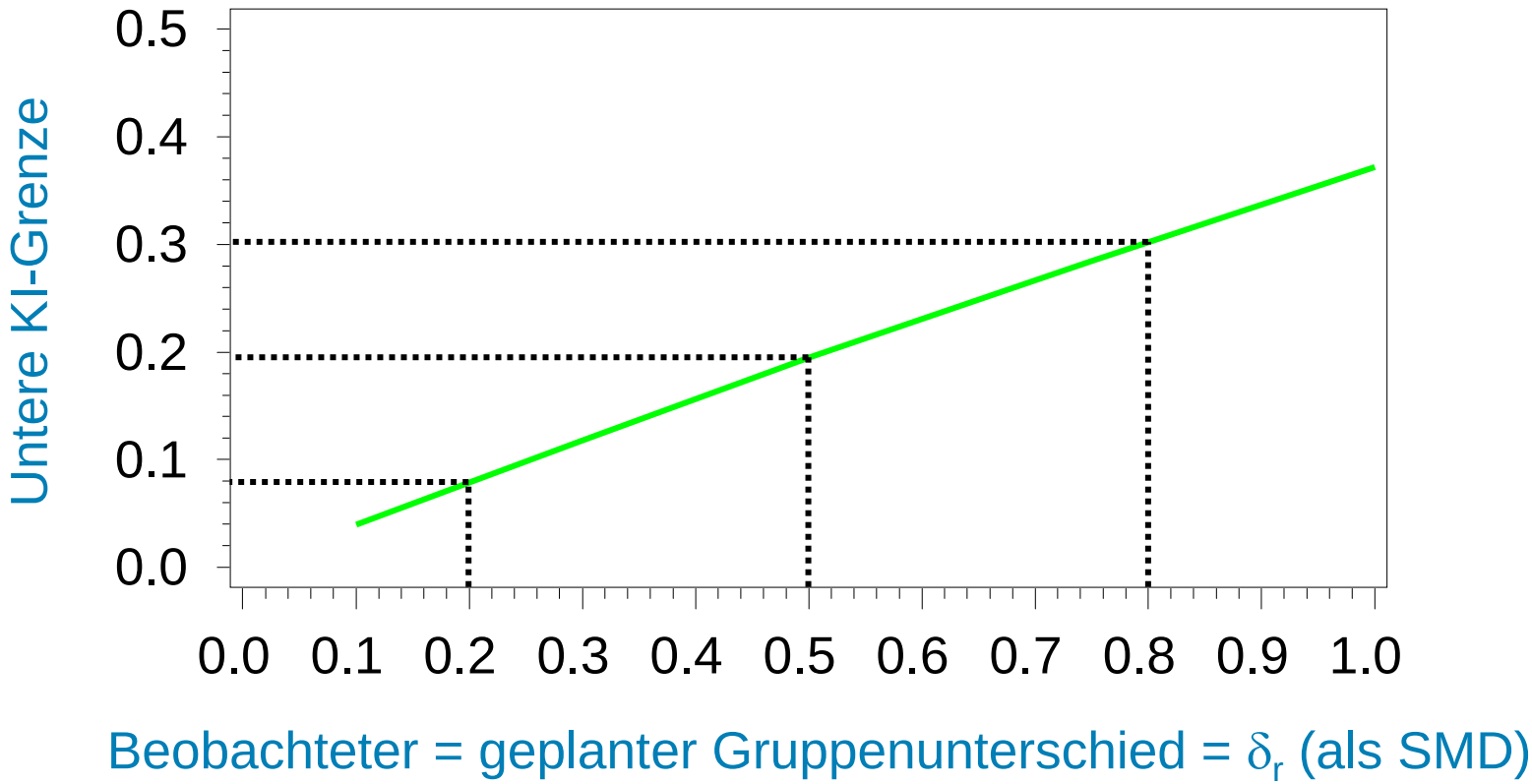


Beobachteter = geplanter Gruppenunterschied = δ_r (als SMD)

"Based upon his experience in the social sciences, he suggested that effect sizes of 0.2 to 0.5 have generally been regarded by investigators as being 'small', of 0.5 to 0.8 are 'moderate' and those of 0.8 or above are 'large'. These apparently arbitrary thresholds have stood the test of time very well. Perhaps surprisingly, the values 0.2, 0.5 and 0.8 have since been found to be broadly applicable in many fields of research as well as in social sciences from where Cohen had drawn his experience."

Fayers PM, Machin D. The assessment, analysis and interpretation of patient-reported outcomes. In: Quality of life (Fayers PM, Machin D, Hrsg.), 2. Ausgabe. Wiley 2007.

Studienergebnis = Annahme für Fallzahlplanung



Beobachteter = geplanter Gruppenunterschied = δ_r (als SMD)

Annahme: δ_r sei bekannt

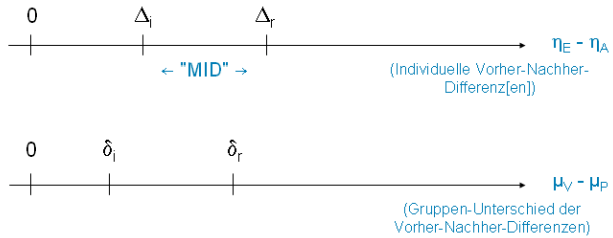
1. Durchführung einer Fallzahlplanung für eine hypothetische Studie mit folgenden Annahmen:

- normalverteilte Daten
- geplanter Gruppenunterschied (als SMD), der δ_r entspricht
- $\alpha = 0{,}05$ (zweiseitig)
- $\beta = 0{,}1$ (d.h. Power = 90 %)
- gleichgroße Gruppen ($n_1 = n_2$)

2. Berechnung eines (zweiseitigen) 95 %-KI für den Fall, dass der in dieser Studie beobachtete Gruppenunterschied δ_r entspricht.

3. Festlegung der unteren Grenze des resultierenden 95%-KI als δ_i

Hierarchie



1. Validierte δ_i vorhanden? —ja→ Teste verschobene Nullhypothese mit δ_i als Grenze

↓ nein

2. Validierte δ_r vorhanden? —ja→ Approximiere δ_i durch δ_r mit vorgeschlagenem Verfahren (hypothetische Fallzahlplanung)

↓ nein

3. Responder-Analyse mit validiertem Response-Kriterium vorhanden? —ja→ Verwende Responder-Analyse und teste mit üblicher Nullhypothese

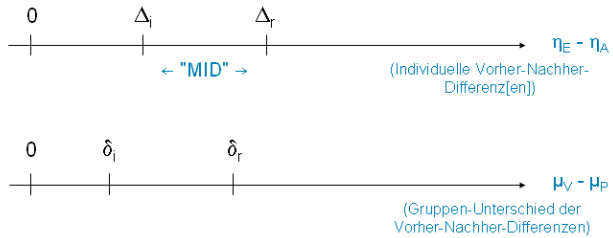
↓ nein

4. Validierte M(C)ID vorhanden? —ja→ Ersetze δ_r mit M(C)ID und approximiere δ_i mit vorgeschlagenem Verfahren

"In most circumstances, the threshold of discrimination for changes in health-related quality of life for chronic diseases appears to be approximately half a SD."

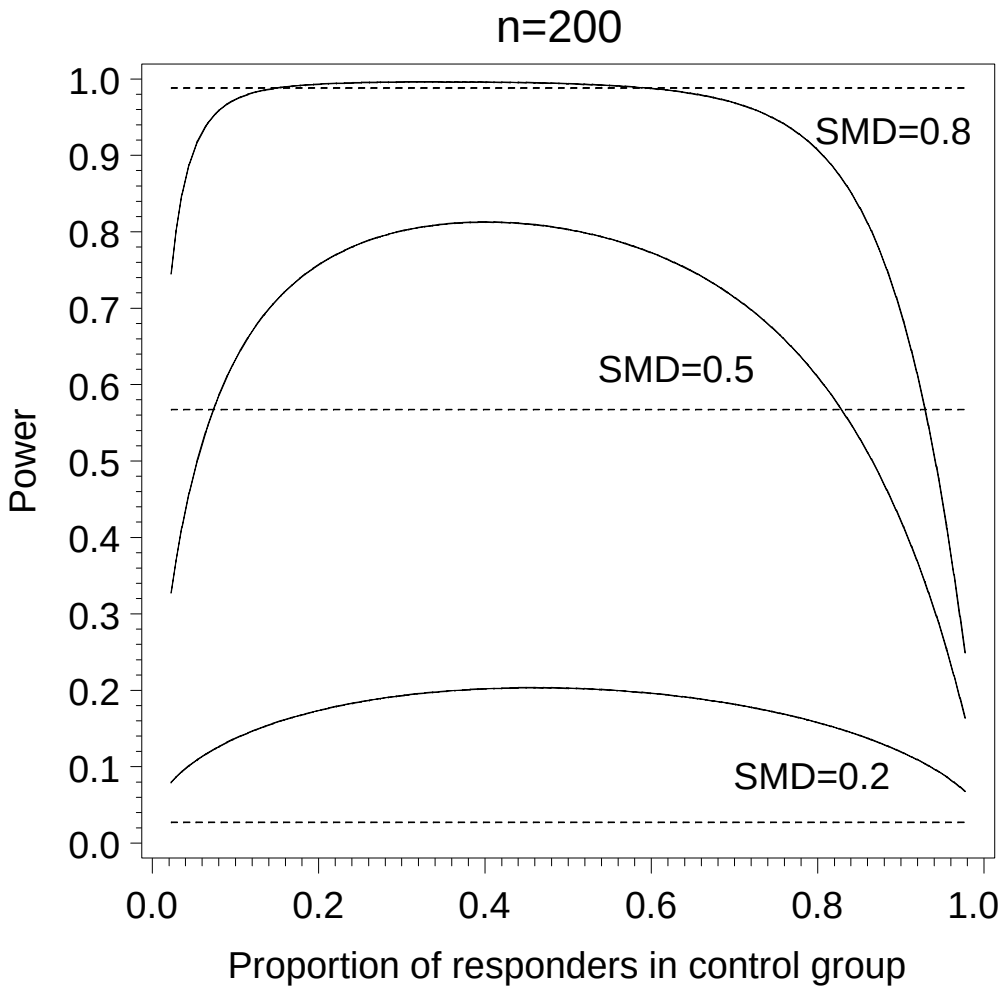
Norman GR, Sloan JF, Wyrwich KW. Interpretation of Changes in Health-related Quality of Life. The Remarkable Universality of Half a Standard Deviation. Med Care 2003; 41: 582-92.

Hierarchie

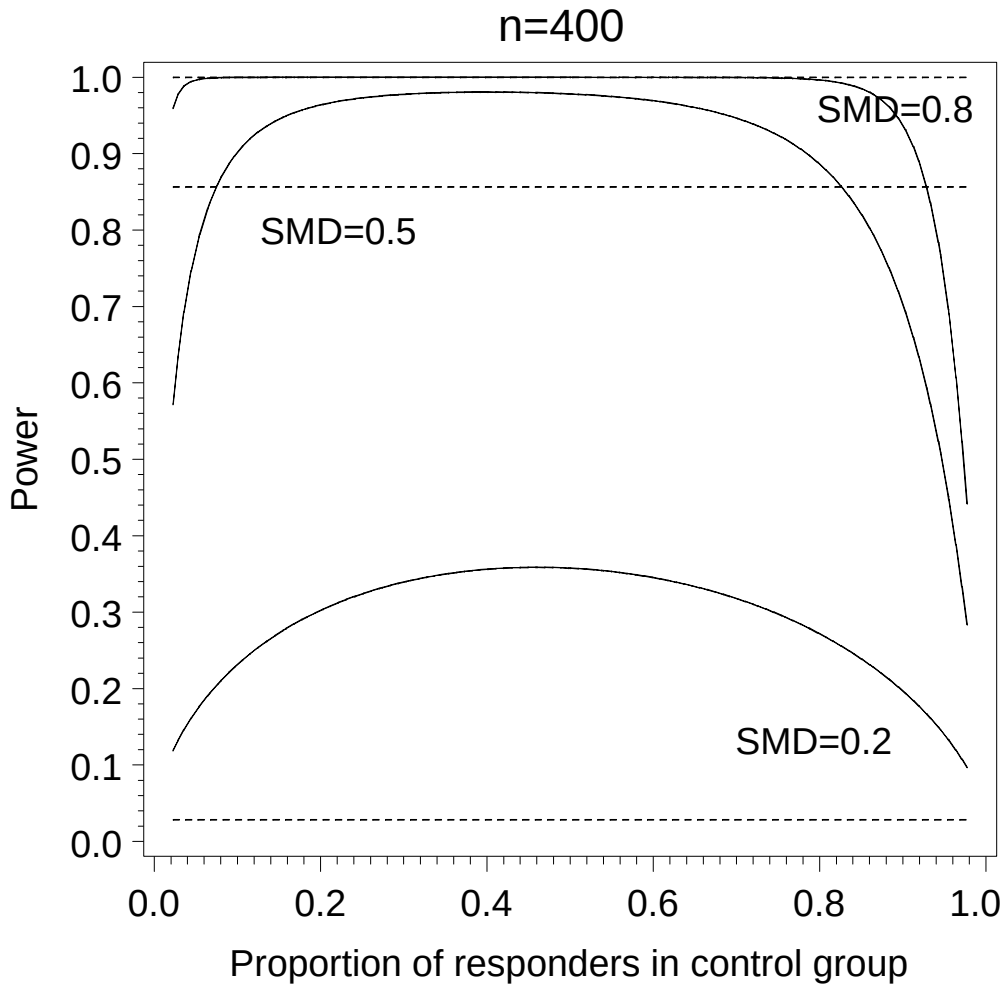


1. Validierte δ_i vorhanden? —ja→ Teste verschobene Nullhypothese mit δ_i als Grenze
   ↓ nein
2. Validierte δ_r vorhanden? —ja→ Approximiere δ_i durch δ_r mit vorgeschlagenem Verfahren (hypothetische Fallzahlplanung)
   ↓ nein
3. Responder-Analyse mit validiertem Response-Kriterium vorhanden? —ja→ Verwende Responder-Analyse und teste mit üblicher Nullhypothese
   ↓ nein
4. Validierte M(C)ID vorhanden? —ja→ Ersetze δ_r mit M(C)ID und approximiere δ_i mit vorgeschlagenem Verfahren
   ↓ nein
5. Keine validierte M(C)ID vorhanden? —ja→ Ersetze δ_r mit 0,5 und approximiere δ_i mit vorgeschlagenem Verfahren ($\rightarrow \delta_i \approx 0{,}2$)

$\delta_r = 0,5$


n=200
SMD=0.8
SMD=0.5
SMD=0.2
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of responders in control group

$\delta_r = 0{,}5$


n=400
SMD=0.8
SMD=0.5
SMD=0.2
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of responders in control group

Schlussfolgerungen

- Patientenrelevanz hat 2 Komponenten: eine qualitative und eine quantitative
- Wenn die qualitative Komponente unsicher ist, rückt die quantitative Komponente in den Vordergrund
- Es fehlt bislang an Standards, irrelevante und relevante Unterschiede zu etablieren
- Ersatzweise wird ein Verfahren vorgeschlagen, das empirische Relevanzbetrachtungen und statistische Entscheidungssicherheit verbindet

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

- Dillenburger Straße 27
- D-51105 Köln

- Telefon +49-221/3 56 85-0
- Telefax +49-221/3 56 85-1

- info@iqwig.de
- www.iqwig.de