

IQWiG im Dialog 2010

Ist der Effekt relevant?

Erfahrungen bei der Arzneimittelbewertung

Thomas Kaiser

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)
Dillenburger Str. 27 , 51105 Köln

Ausgangslage

Because statistical significance can sometimes be achieved for small changes in PRO measures that may not be clinically meaningful (i.e., do not indicate treatment benefit), we encourage sponsors to avoid proposing labeling claims based on statistical significance alone.

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

Agenda

- **Methoden – Best practice HTA / SR?**
- **Fallbeispiele**
 - **Diabetes mellitus**
 - **Depression**
 - **Alzheimer Demenz**

Methoden – Best practice HTA/SR?

- **NICE Methodendokumente**
- **Cochrane-Handbook**
- **IQWiG-Methoden**

Methoden - NICE

- **Guides to the Methods of Technology Appraisal**
 - **2004: nicht adressiert**
 - **2008: nicht adressiert**

- **The Guidelines Manual 2009: nicht adressiert**

Methoden - Cochrane

- **Cochrane Handbook**
 - **2002: nicht adressiert**
 - **2004: nicht adressiert**
 - **2008 (5.0.1): erstmals adressiert**

Methoden – Cochrane Handbook 5.0.1

- ... misinterpretation ... to assume that a result with a small P value for the summary effect estimate implies that an intervention has an important benefit.
- ...such [continuous] outcomes may be difficult to interpret, particularly when they relate to rating scales.
- Knowledge of the smallest change in instrument score that patients perceive is important – the minimal important difference – and can greatly facilitate the interpretation of results.

Methoden – Cochrane Handbook 5.0.1

- Rules of thumb exist for interpreting SMDs (or ‘effect sizes’) [...] 0.2 represents a small effect, 0.5 a moderate effect, and 0.8 a large effect (Cohen 1988). Variations exist (for example, <0.41 = small, 0.40 to 0.70 = moderate, >0.70 = large).
- Another, perhaps even less satisfactory, approach suggests that a standardized mean difference of 0.5 approximates, in many cases, to a minimal important difference
- However, some methodologists believe that such interpretations are problematic because patient importance of a finding is context-dependent and not amenable to generic statements.

Methoden – Cochrane Handbook 5.0.1

- For clinical interpretation, such an analysis [SMD] may be less helpful than dichotomizing responses and presenting proportions of patients benefiting.
- Methods are available for creating dichotomous data out of reported means and standard deviations, but require assumptions that may not be met.

Methoden – Cochrane Handbook 5.0.1

- **Ideally, review authors will present estimates of the MID in the abstract. While this is very helpful, it potentially tempts clinicians to make inappropriate inferences. If the MID is 0.5 and the mean difference between treatments is 0.4, clinicians may infer that nobody benefits from the intervention. If the mean difference is 0.6, they may conclude that everyone benefits.**

Methoden – Cochrane - Zusammenfassung

- **Bewertung der Relevanz ist wichtig, insbesondere für Skalen**
- **MID ist wichtig**
- **Die Bewertung des Effektschätzers allein ist unzureichend**
- **Selbst berechnete Responderraten sind eine Möglichkeit für die Relevanzbewertung**
- **Die Bewertung der SMD ist eine Möglichkeit für die Relevanzbewertung**
- **Alle Verfahren haben Nachteile**
- **Keine konkrete Handlungsanleitung**

Methoden - IQWiG

- **Allgemeine Methoden:**
 - **2005 (1.0): adressiert, aber ohne konkrete Handlungsanleitung**
 - **2006 (2.0): adressiert, aber ohne konkrete Handlungsanleitung**
 - **2008 (3.0): adressiert, aber ohne konkrete Handlungsanleitung**

Methoden – IQWiG

- Für die Beurteilung der klinischen Relevanz von Studienergebnissen gibt es noch kein breit akzeptiertes methodisches Vorgehen.
(2005: ...ist ein weitgehend ungelöstes Problem)

Methoden – IQWiG

- **Erster Ansatz: Bewertung des Effektschätzers und des dazugehörigen Konfidenzintervalls mit Hilfe medizinischer Sachkenntnis.**
- **Formales Relevanzkriterium: Beurteilung der unteren Konfidenzgrenze für den Effektschätzer beziehungsweise die Anwendung eines statistischen Tests mit Verschiebung der Nullhypothese.**
- **Weitere Möglichkeit: Responderdefinition.**
- **Darüber hinaus spielt die individuelle Einschätzung der Betroffenen eine wesentliche Rolle.**

Methoden – IQWiG

- **2008: Für die Frage, ob sich aus offenen Studien ein Hinweis für einen Nutzen einer Intervention bezüglich PRO ableiten lässt, ist die Größe des beobachteten Effekts ein wichtiges Entscheidungskriterium.**

Methoden – IQWiG - Zusammenfassung

- **Verschiedene Möglichkeiten der Relevanzbewertung (auch: verschobene Nullhypothese)**
- **Effektstärke bei PRO / offene Studien wichtiges Entscheidungskriterium**
- **Keine konkrete Handlungsanleitung**

Fallbeispiel – Diabetes mellitus

Fallbeispiel – Diabetes mellitus

■ 7 Abschlussberichte

Zielgröße	Krankheitsspezifisch	MID	Generisch	MID
Lebensqualität	6	0 (0%)	2	2 (100%)
Therapiezufriedenheit	3	0 (0%)	0	-

Fallbeispiel – Diabetes mellitus

- **Generische Instrumente (MID vorhanden):**
 - **Kein statistisch signifikanter Unterschied Mittelwert**
 - **Keine Responderanalysen**
- **Krankheitsspezifische Instrumente (keine MID):**
 - **Nur im Ausnahmefall statistisch signifikanter Unterschied**
 - **Dann geringe Effektstärke bei offenen Studien**



Fallbeispiel – Depression

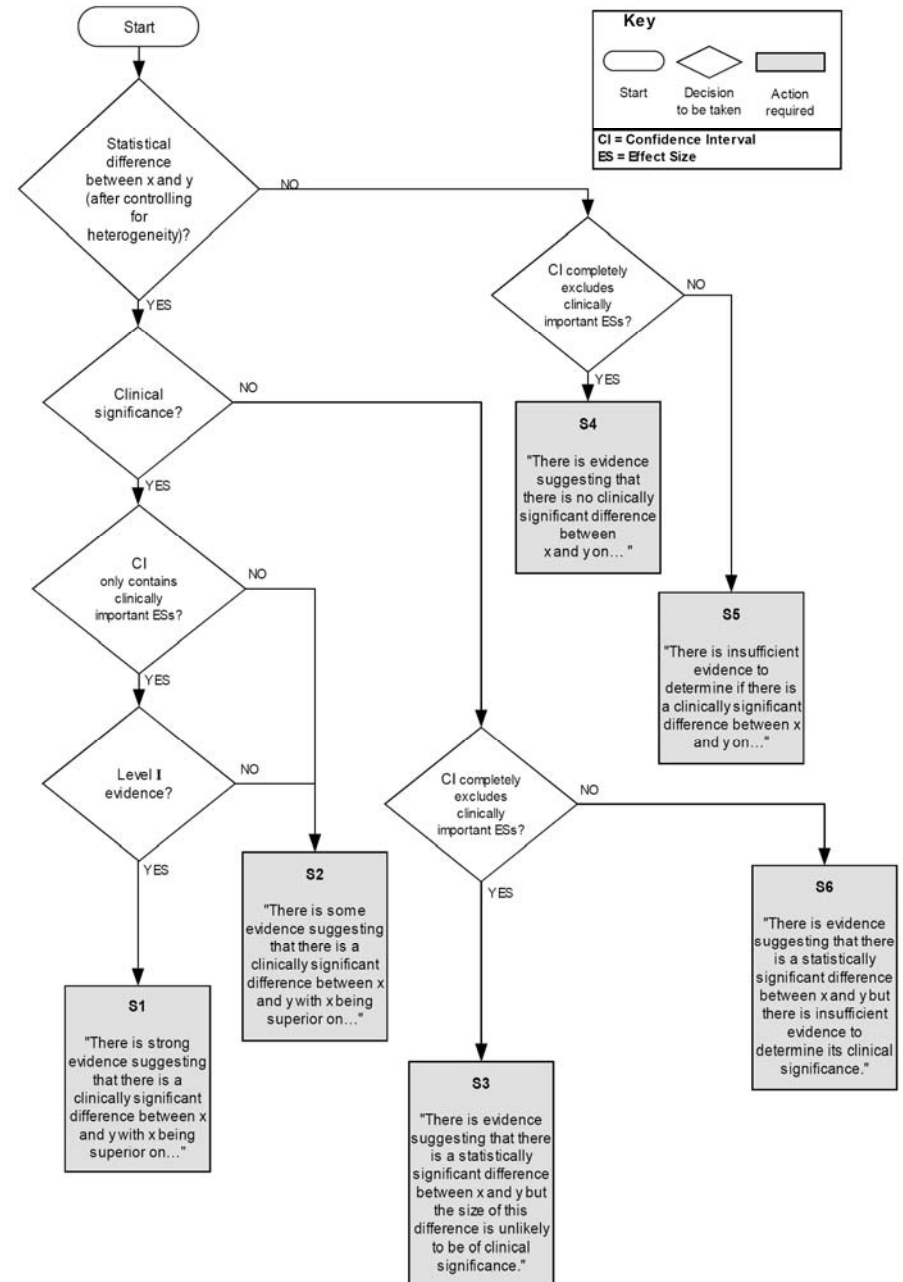
Fallbeispiel – Depression

- **2 Abschlussberichte**
 - **A05-20A: Duloxetin, Venlafaxin (2009)**
 - **A05-20C: Bupropion, Mirtazapin, Reboxetin (2009)**

NICE – CG23

- **95%CI > 0,5 SMD**
➔ **Beleg für Nutzen**
- **95%CI < 0,5 SMD**
➔ **Kein Nutzen**
- **0,5 SMD in 95%CI**
➔ **Hinweis auf Nutzen**

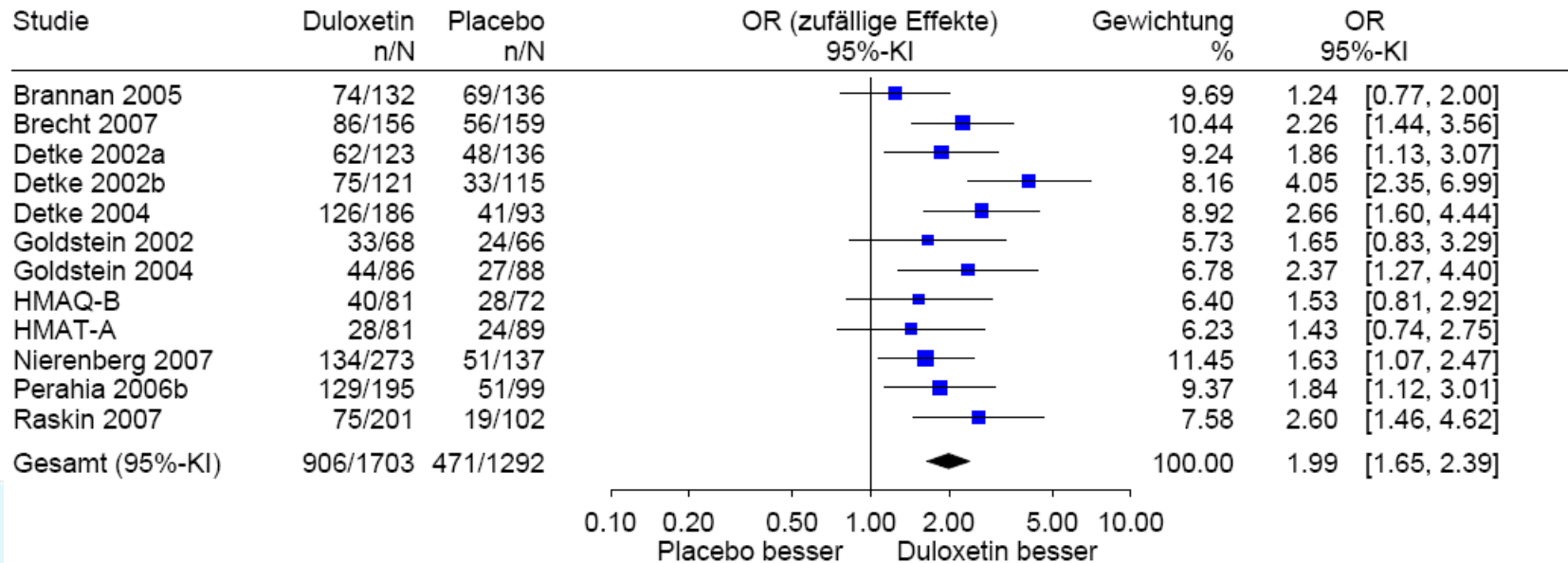
Flowchart 1: Guideline Statement Decision Tree



A05-20A – Duloxetine (antidepressive Symptomatik)

■ Response:

Meta-Analyse B1
Response
Duloxetine vs. Placebo

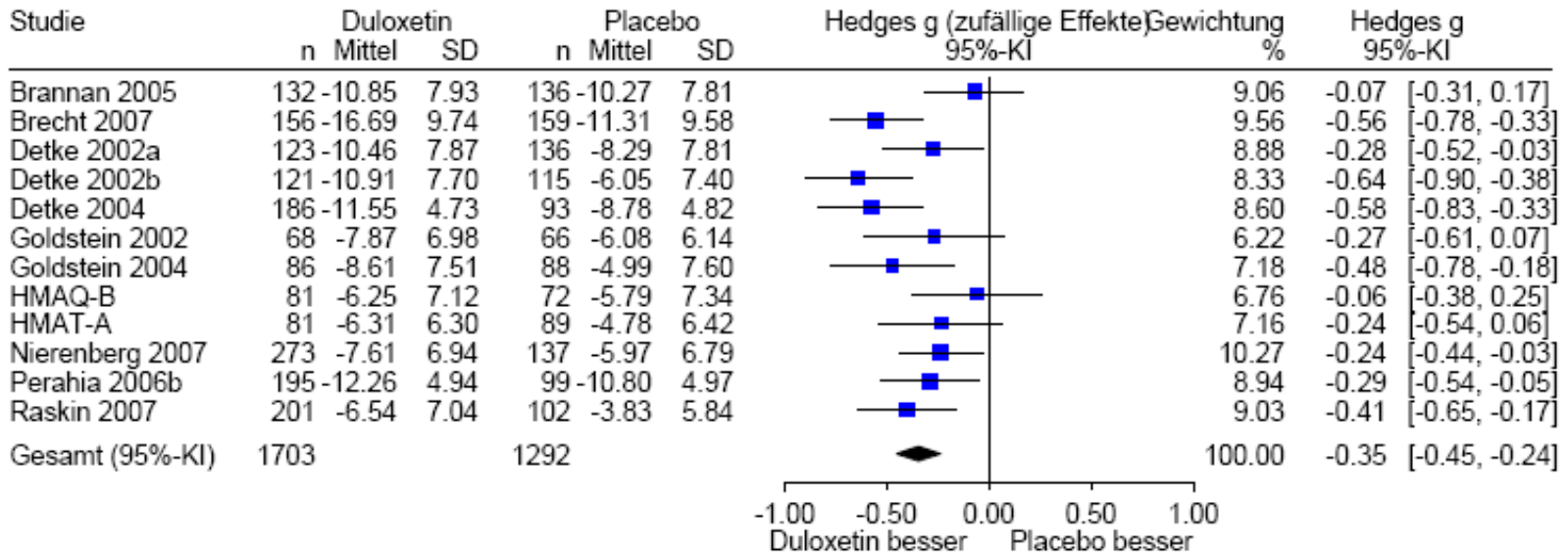


Heterogenität: $Q=15.87$, $df=11$ ($p=0.146$), $I^2=30.7\%$
Gesamteffekt: $Z\text{ Score}=7.24$ ($p=0.000$), $\tau^2=0.033$

A05-20A – Duloxetine (antidepressive Symptomatik)

■ SMD:

Gesamt-Score, Change from Baseline
Duloxetine vs. Placebo



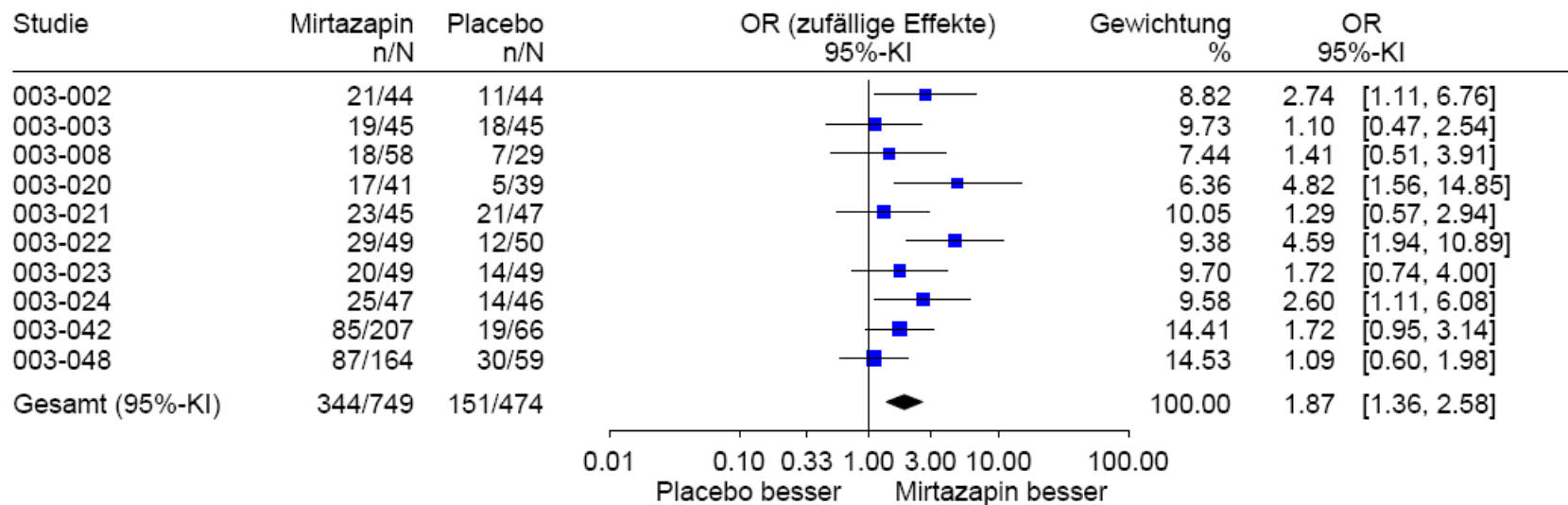
Heterogenität: $Q=22.67$, $df=11$ ($p=0.020$), $I^2=51.5\%$
Gesamteffekt: Z Score=-6.33 ($p=0.000$), $\tau^2=0.018$

■ Fazit: Beleg für einen Nutzen

A05-20C – Mirtazapin (antidepressive Symptomatik)

■ Response:

Response
Mirtazapin vs. Placebo

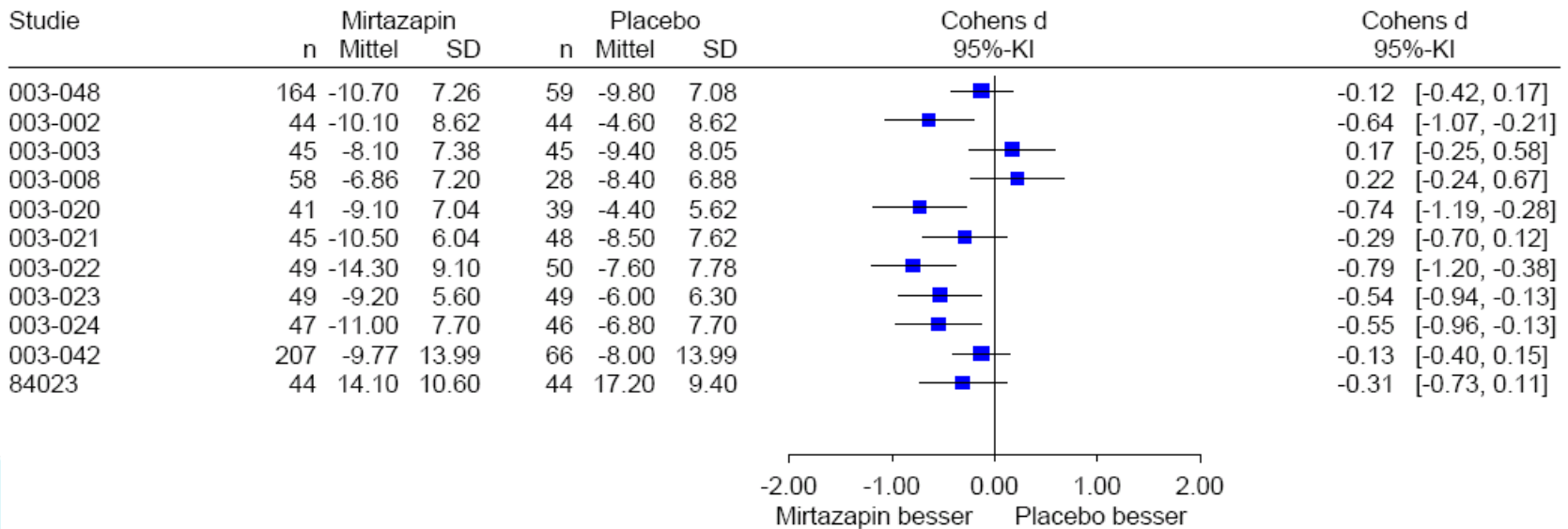


Heterogenität: $Q=13.91$, $df=9$ ($p=0.125$), $I^2=35.3\%$
Gesamteffekt: Z Score= 3.83 ($p=0.000$), $\tau^2=0.093$

A05-20A – Duloxetine (antidepressive Symptomatik)

■ SMD:

Gesamtscore
Mirtazapin vs. Placebo



Heterogenität: $Q=26.94$, $df=10$ ($p=0.003$), $I^2=62.9\%$

■ **Fazit: Beleg für einen Nutzen bzgl. Response**

Depression - Zusammenfassung

- Bei weitgehend homogenen Analysen Response / SMD kongruent
- Nutzenbeleg bei statistisch signifikantem Ergebnis bzgl. Response

Fallbeispiel – Alzheimer Demenz

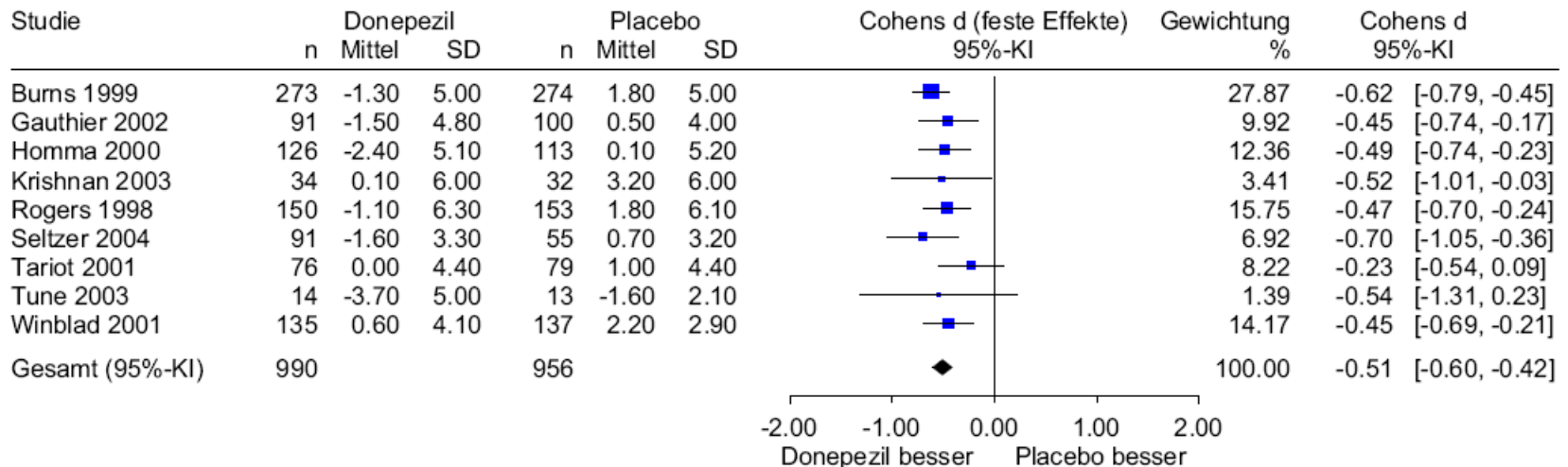
Fallbeispiel – Alzheimer Demenz

- **3 Abschlussberichte**
 - **A05-19A: Cholinesterasehemmer (2007)**
 - **A05-19B: Ginkgo (2008)**
 - **A05-19C: Memantin (2009)**

A05-19A - Cholinesterasehemmer

■ Beispiel Kognition (Donepezil)

Donepezil - kognitive Leistungsfähigkeit
Endpunkt: ADAS-cog, MMSE, SIB - Differenz zu Baseline
Distanzmaß: standardisierte Differenz der Mittelwerte



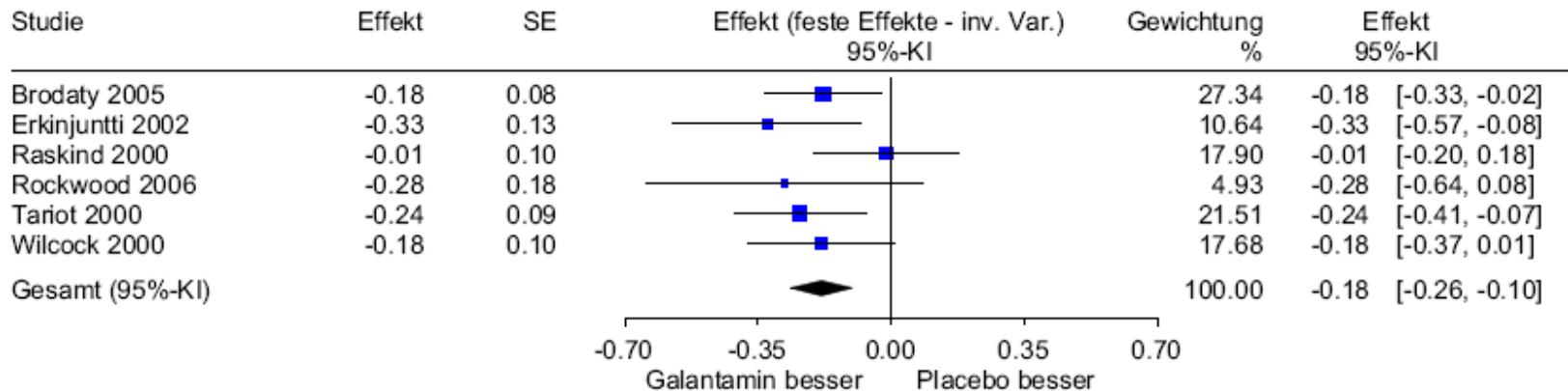
Heterogenität: $Q=6.43$, $df=8$ ($p=0.600$), $I^2=0\%$
Gesamteffekt: Z Score=-10.99 ($p=0.000$)

■ **Fazit: Beleg für Nutzen („...hat einen Nutzen...“)**

A05-19A - Cholinesterasehemmer

■ Beispiel Aktivitäten des tgl. Lebens (Galantamin)

Galantamin - Aktivitäten des täglichen Lebens
Endpunkt: ADCS-ADL, DAD - Differenz zu Baseline
Distanzmaß: standardisierte Differenz der Mittelwerte



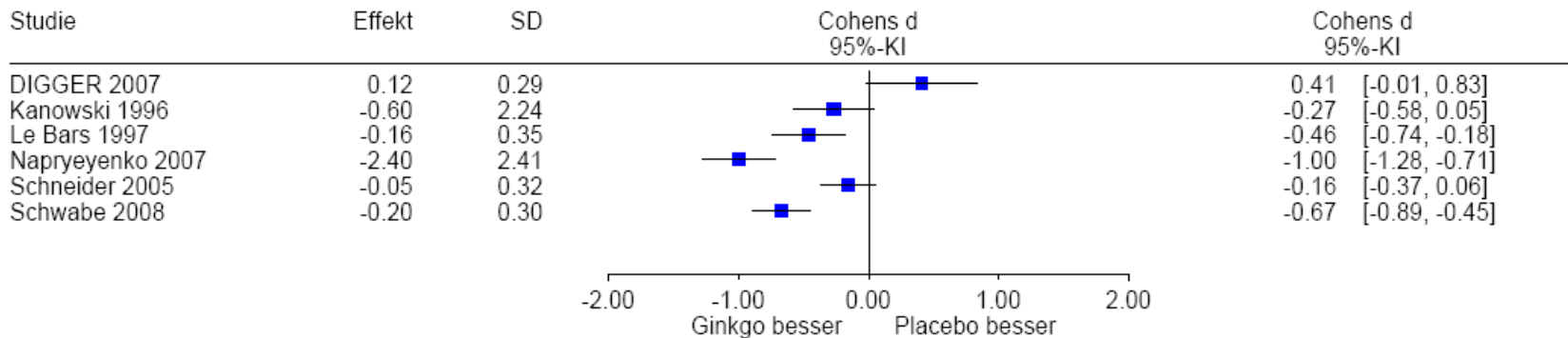
Heterogenität: $Q=5.03$, $df=5$ ($p=0.413$), $I^2=0.5\%$
Gesamteffekt: Z Score=-4.46 ($p=0.000$)

- **Fazit: Hinweis auf einen Nutzen
(→ kein Beleg für einen Nutzen)**

A05-19B - Ginkgo

■ Beispiel Aktivitäten des tgl. Lebens

Ginkgo, Aktivitäten des täglichen Lebens
Endpunkt: NAB, GERRI - Gruppenunterschied zu Placebo
Distanzmaß: standardisierte Mittelwertdifferenz

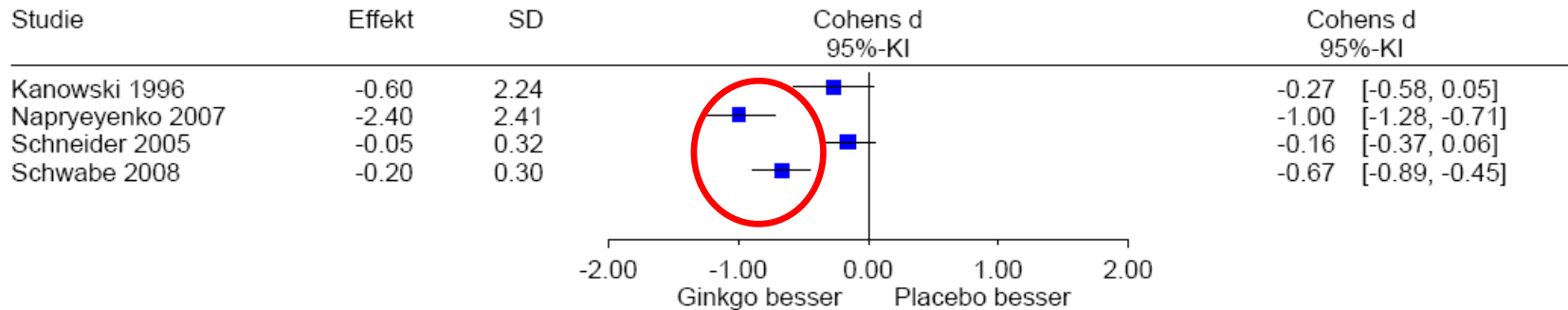


Heterogenität: $Q=42.37$, $df=5$ ($p=0.000$), $I^2=88.2\%$, $\tau^2=0.144$

A05-19B - Ginkgo

■ Beispiel Aktivitäten des tgl. Lebens (nur hohe Dosis)

Ginkgo, Aktivitäten des täglichen Lebens, Dosierung 240 mg
Endpunkt: NAB, GERRI - Gruppenunterschied zu Placebo
Distanzmaß: standardisierte Mittelwertdifferenz



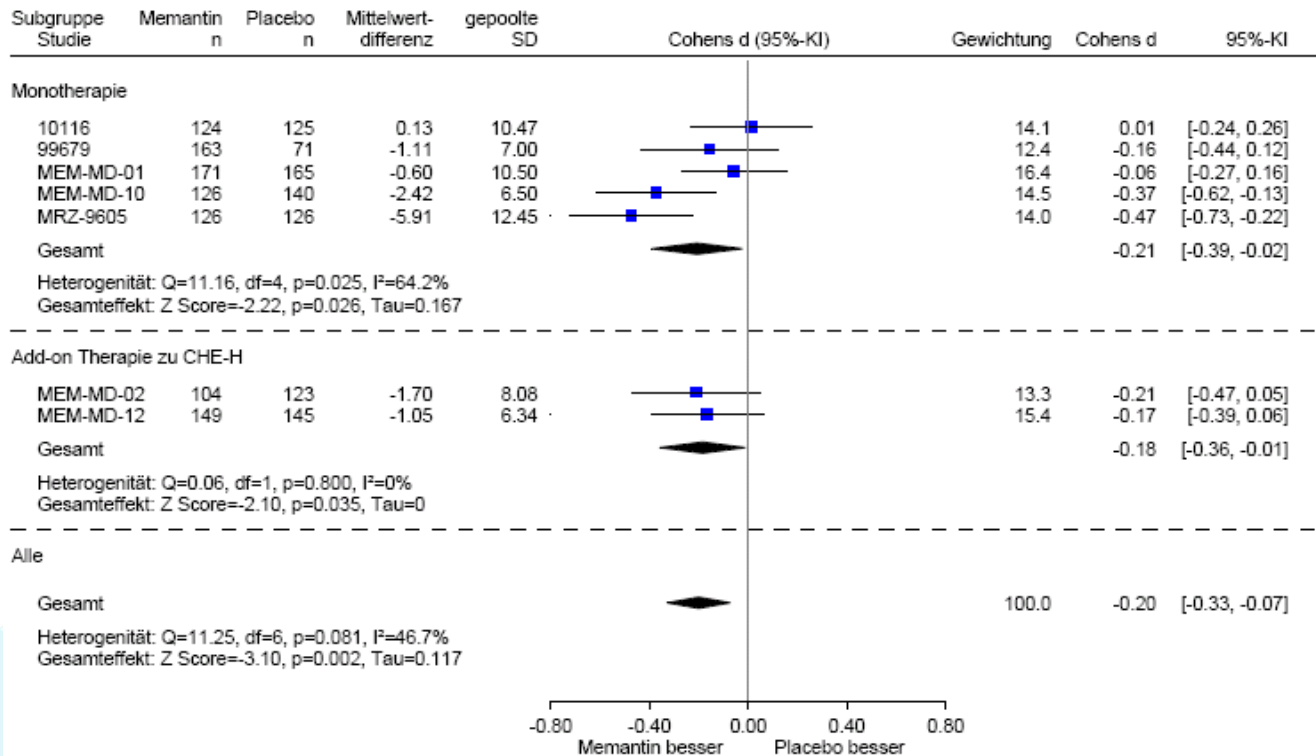
Heterogenität: $Q=26.11$, $df=3$ ($p=0.000$), $I^2=88.5\%$, $\tau^2=0.127$

- **Fazit: Beleg für einen Nutzen (hohe Dosis, Patienten mit begleitenden psychopathologischen Symptomen, Übertragbarkeit?)**

A05-19C – Memantin

■ Beispiel Kognition

Memantin vs. Placebo
Kognitive Leistung: ADAS-Cog, SIB,
Modell mit zufälligen Effekten - DerSimonian und Laird



■ **Fazit: Kein Beleg für einen Nutzen**

A05-19C – Memantin

EPAR 2005:

- The MAH provided additional data (a meta-analysis and a responders analysis)...
- The meta-analyses on the efficacy endpoints showed statistically significant and clinically relevant effects of memantine compared to placebo on the cognitive ($p < 0.001$) ... domains with effect sizes in the same range as AChEIs.

A05-19C – Memantin

- **Effekt Kognition Memantin / Donepezil (vs. Placebo):**
 - **Memantin: 0.20 [0.07; 0.33]**
 - **Donepezil: 0.51 [0.42; 0.60]**
- ➔ **„effect size in the same range“?**

A05-19C – Memantin

- **EPAR 2005:**
- **The triple responder analyses were performed based on definitions that identified patients whose condition concurrently worsened in all three domains during the 6-month treatment: Marked Clinical Worsening – a decline of ≥ 4 points on the ADAS-cog or ≥ 5 points on the SIB and a decline on the CIBIC-plus and a decline on the ADL.**
- **MID für ADL? Ist jegliche Änderung auf der ADL-Skala relevant? (Standardabweichung ADL₂₃ im Median ca. 10 Punkte)**

Systematic Review of Measures of Clinical Significance Employed in Randomized Controlled Trials of Drugs for Dementia

Frank J. Molnar, MSc, MD, MDCM,^{†‡} Malcolm Man-Son-Hing, MSc, MD,^{*†‡} and Dean Fergusson, PhD^{*§||}*

Despite being central to the interpretation of trial results ..., patient- and caregiver-centered measures of clinical significance have not been adequately studied and integrated into dementia drug RCTs.

The most commonly cited measures of clinical significance were a 4-point change in the Alzheimer's Disease Assessment Scale - Cognitive Subscale (ADAS-Cog)...

Systematic Review of Measures of Clinical Significance Employed in Randomized Controlled Trials of Drugs for Dementia

Frank J. Molnar, MSc, MDCM,^{*†‡} Malcolm Man-Son-Hing, MSc, MD,^{*†‡} and Dean Fergusson, PhD^{*§||}

LETTERS TO THE EDITOR

Despite being
and caregiv
been adequ

ults ..., patient-
nce have not
a drug RCTs.

IF SHRIMPS HAD LIPS, THEY COULD WHISTLE ...*

The most commo
4-point change in
Cognitive Subscale (ADAS-Cog)...

significance were a
ssment Scale -

*Julia Kreis, Dipl.-Psych
Thomas Kaiser, MD
Beate Wieseler, PhD
Institute for Quality and Efficiency in Health Care
Cologne, Germany*

Alzheimer Demenz - Zusammenfassung

- **Bezüglich Nutzenbeleg konsistente Schlussfolgerungen, obwohl die Effektstärke zunächst nur implizit herangezogen wurde**
- **Vorliegende Responderanalysen unzureichend**
- **Allgemein Forschungsbedarf hinsichtlich Relevanzkriterien**

Zusammenfassung

- **Methodik der Relevanzbewertung international unzureichend (Reviews / HTA / Studienebene)**
- **Validierte MIDs fehlen oft**
- **Responderanalysen häufig unzureichend**
- **Implizite / explizite Relevanzbewertung des IQWiG weitgehend kongruent**
- **(Responderanalysen / SMD in der Praxis kongruent)**