Aussagekraft von Metaanalysen

Prof. Stefan Leucht
Klinik für Psychiatrie und Psychotherapie der TU-München

Disclosures

In the past 3 years:

- Consulting/advisory board honoraria from Alkermes, Bristol-Myers Squibb, Eli Lilly, Janssen, Johnson & Johnson, Lundbeck, MedAvante, Roche
- Lecture honoraria from AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Essex Pharma, Janssen, Johnson & Johnson, Lundbeck Institute, Pfizer, Sanofi-Aventis
- Eli Lilly has provided medication for a trial with Stefan Leucht as the primary investigator

Definitions

- **Systematic Review:** Means the systematic approach in terms of literature search, selection, presentation and analysis of the data

- **Meta-analysis:** Means the mathematical combination of the results of different studies on one question

Why do we need systematic reviews?

- In 10000 medical journals 2 million articles are published every year
- A general practicioner would have to read 19 articles everyday, 365 days per year to cover relevant reports
- More than 300 randomised controlled studies about the „atypical“ antipsychotics are available

The principal steps in the development of a systematic review

I. BEFORE: Writing a protocol

- Which patients
- Which interventions
- Which outcomes
- Literature search (databases, search strings)
- Statistical method

The protocol is reviewed by two editors of the Cochrane Schizophrenia Group and it is published in the Cochrane Library

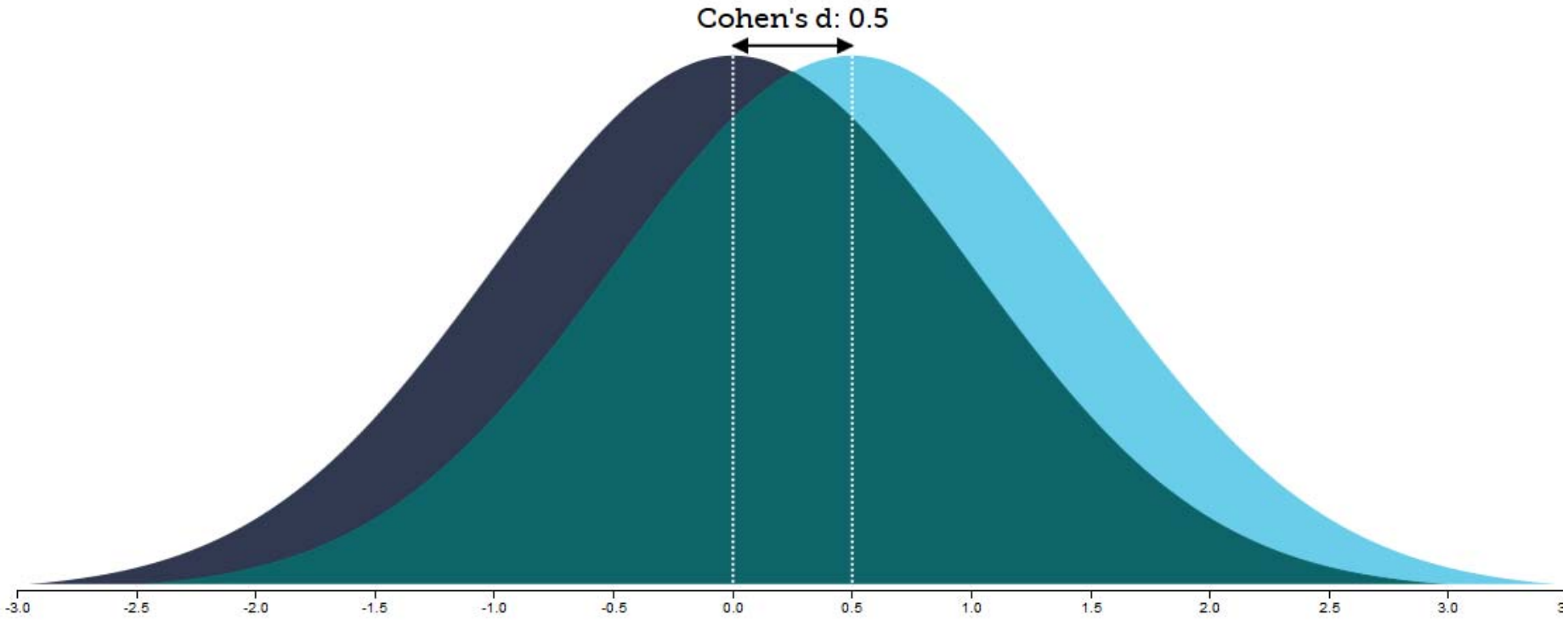
II. Literature search

- Not only MEDLINE
- Not only English/Dutch
- Electronic databases, conference abstract books, book chapters, contacting pharmaceutical companies, contacting study authors, FDA webpage

Calculation of Effect Sizes for Continuous Variables

- Effect size = (mean A – mean B)/pooled standard deviation
- Example: (90 – 80)/20 = 0.50

Ilustration of the meaning of effect size



source : http://rpsychologist.com/d3/cohend/

Cohen‘s rule

Standardised mean difference of

0.20 = small

0.50 = medium

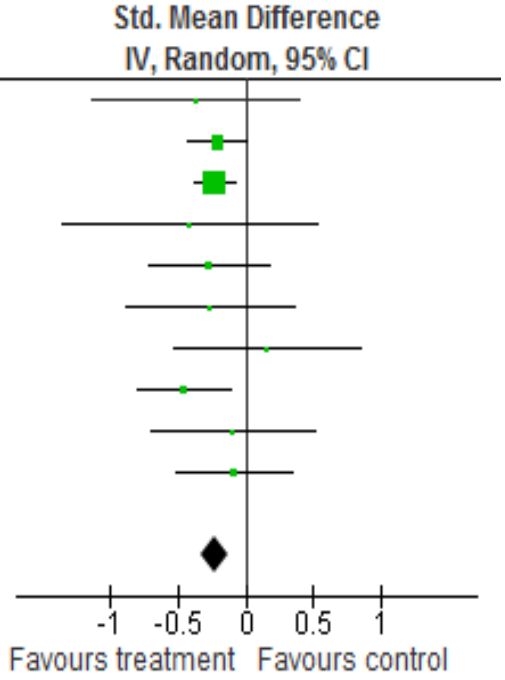
0.80 = large

Principle of meta-analysis, example: Olanzapine versus quetiapine for schizophrenia
(Komossa et al. Cochrane review 2009)

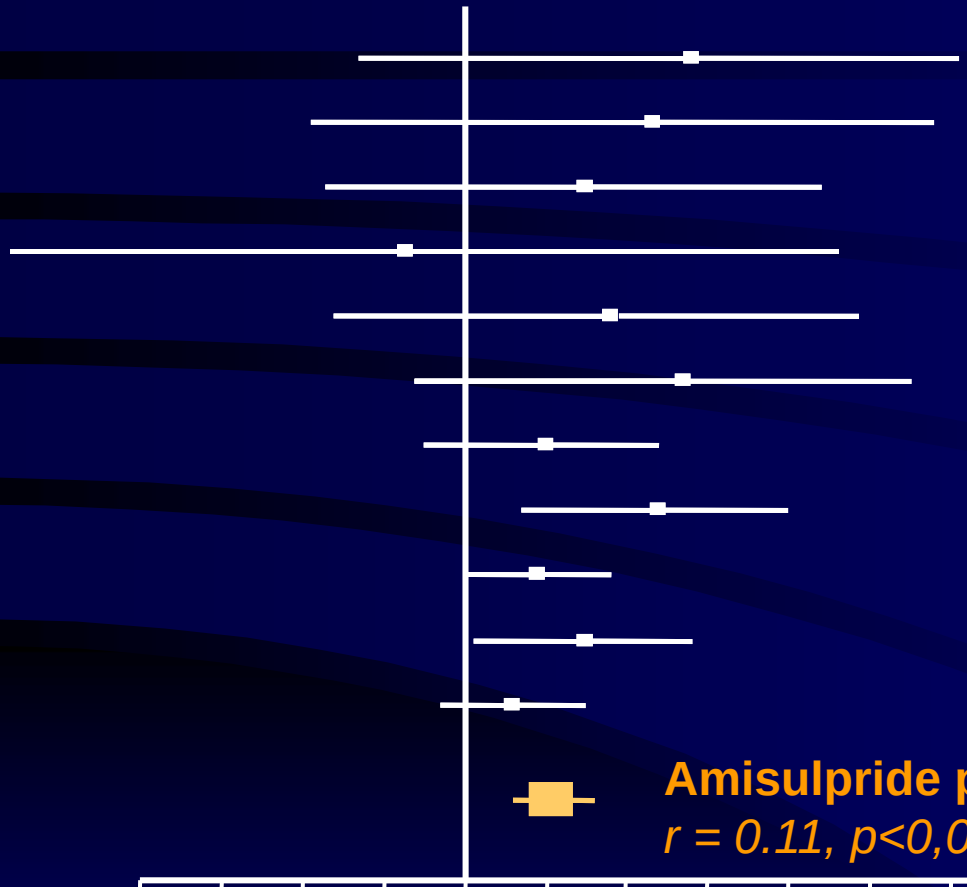
	Treatment			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Atmaca 2003	74.86	6.41	13	77.24	6.08	14	1.8%	-0.37 [-1.13, 0.39]
Kinon 2006b	-11.3	18.3	166	-7.2	21.2	169	23.2%	-0.21 [-0.42, 0.01]
Lieberman 2005	-11.27	22.31	330	-6.08	22.31	329	45.6%	-0.23 [-0.39, -0.08]
McEvoy 2006	-7.7	9.8	10	-1.3	19.23	8	1.2%	-0.41 [-1.36, 0.53]
McEvoy 2007	-18.4	9.73	37	-15.6	10.68	44	5.5%	-0.27 [-0.71, 0.17]
Mori 2004	69.4	10.8	20	72.9	15.1	20	2.8%	-0.26 [-0.88, 0.36]
Riedel 2007	-17.88	20.71	17	-21.5	23.39	16	2.3%	0.16 [-0.52, 0.84]
Stroup 2006	-8.2	22.31	66	2	22.31	63	8.7%	-0.45 [-0.80, -0.10]
Svestka 2003b	-45.65	11.96	20	-43.91	20.94	22	2.9%	-0.10 [-0.70, 0.51]
Voruganti 2007	48.5	9.9	42	49.4	12	43	5.9%	-0.08 [-0.51, 0.34]
Total (95% CI)			**721**			**728**	**100.0%**	**-0.23 [-0.34, -0.13]**

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 3.85$, df = 9 (P = 0.92); $I^2 = 0\%$

Test for overall effect: Z = 4.39 (P < 0.0001)



BPRS: Amisulpride vs. typical antipsychotics



Rüther (1989) **vs. perazine**[1]
Pichot (1988)
Ziegler (1989)
Klein (1985)
Costa e Silva (1989)
Delcker (1990)
Möller (1997)
Wetzel (1999) **vs. flupentixol**
Puech (1998)
Carrière (2000)
Colonna (2000)

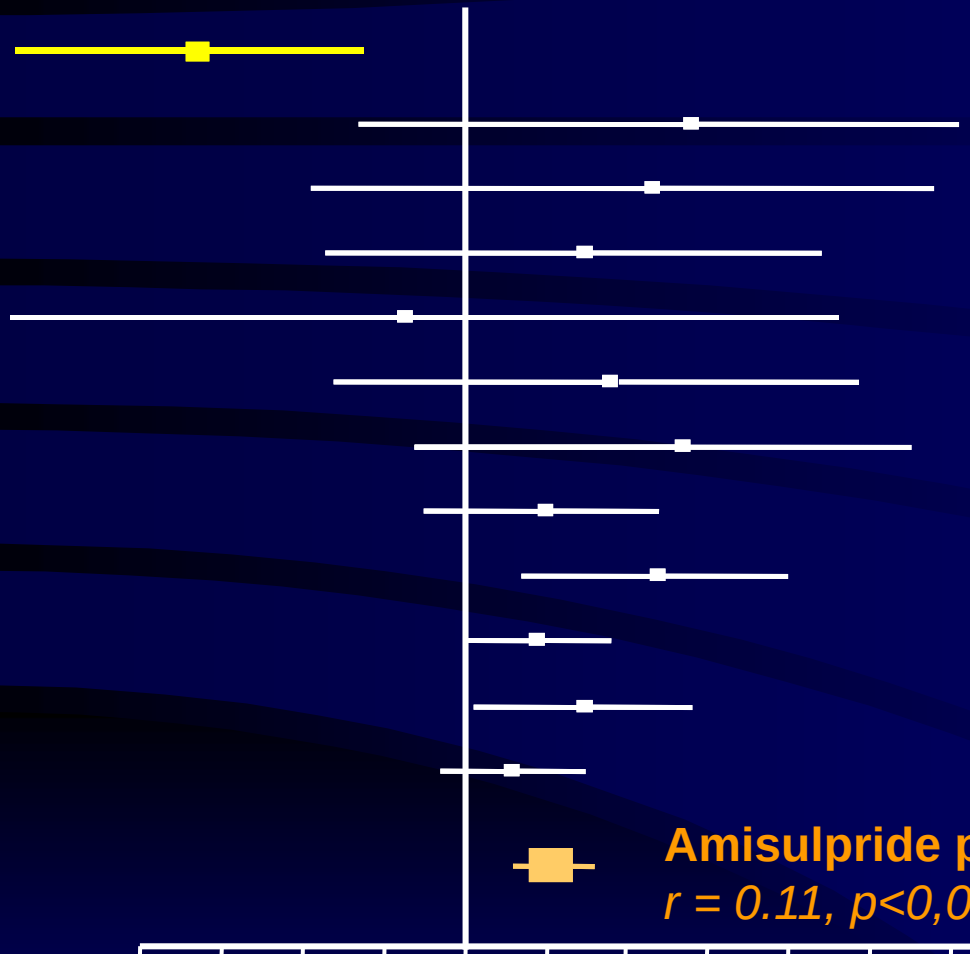
Amisulpride pooled:
r = 0.11, p<0,0001, 10 studies, n = 1654

-0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 ***r** (95% CI)*

[1] **endpoint analysis, not used for mean effect size**

Leucht et al. Am J Psychiatry 2002

BPRS: Amisulpride vs. typical antipsychotics



Leucht 2008

Rüther (1989) **vs. perazine**[1]
Pichot (1988)
Ziegler (1989)
Klein (1985)
Costa e Silva (1989)
Delcker (1990)
Möller (1997)
Wetzel (1999) **vs. flupentixol**
Puech (1998)
Carrière (2000)
Colonna (2000)

Amisulpride pooled:
r = 0.11, p<0,0001, 10 studies, n = 1654

-0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 ***r** (95% CI)*

[1] **endpoint analysis, not used for mean effect size**

Leucht et al. Am J Psychiatry 2002

III. Data extraction

2. Dichotomous Variables

(yes – no Outcomes, e.g. stroke, relapse)

For both groups:

Number of events (e.g. relapse), total N

Effect size measures:

Absolute risk difference, relative risk difference, Odds Ratio

Calculation of Effect Sizes for Dichotomous Variables

- Risk: 1 out of 10 patients relapsed, i.e. 1/10 = 0.1 (or 10%)
- Absolute risk (response) difference
  - Risk of the intervention group – Risk of the control group, e.g. 2%-4% = (-) 2%
- Relative risk reduction (or response ratio)
  - Risk of the intervention group divided by the risk of the control group, e.g. 2% / 4% = 0.5 = 50%
- Number needed to treat
  - How many patients must be treated to have one relapse less?
  - Inverse of the absolute risk difference here 1/2% = 1/0.02 = 50

Reduction in the Risk of Dying From Breast Cancer

	100,000 Women without mammography	100,000 Women with mammography
Breast cancer mortality in 10 years	**0.36% (360/100,000)**	**0.29% (290/100,000)**

Relative mortality reduction = 20% (1 – [0.29%/0.36%])

Absolute mortality reduction = 0.07% (0.36%-0.29%)

Data from Kürzl Deutsches Ärzteblatt 9/2004

Meta-analyses are often the only way to objectively summarise the evidence if there are many studies

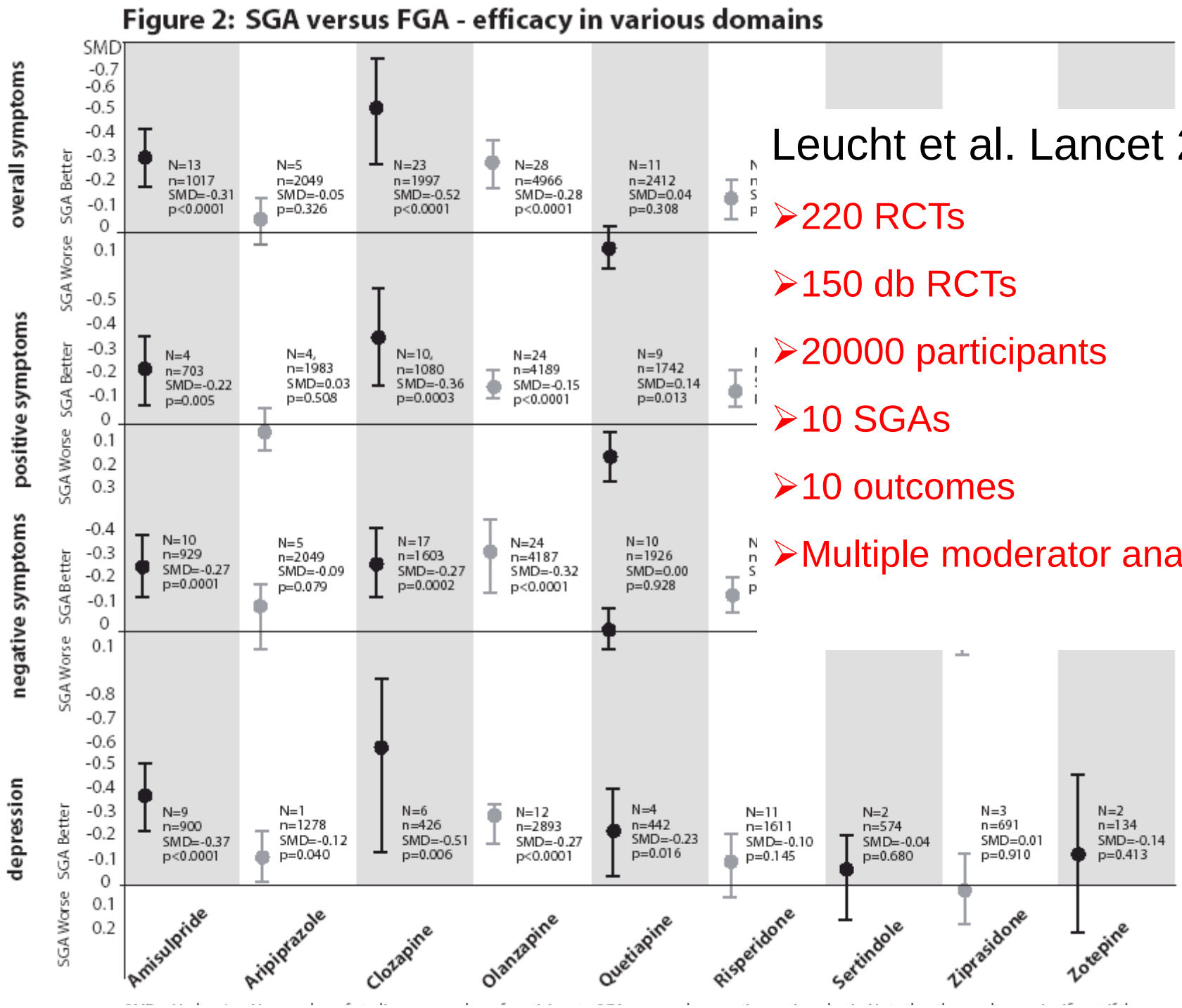
Reviews and Overviews

Why Olanzapine Beats Risperidone, Risperidone Beats Quetiapine, and Quetiapine Beats Olanzapine: An Exploratory Analysis of Head-to-Head Comparison Studies of Second-Generation Antipsychotics

The overall outcome reported in the abstract of head to head comparisons of atypical antipsychotics strongly depends on the sponsor

In a blinded analysis of the abstracts of 33 head to head comparisons of atypical antipsychotics in about 90% the overall outcome was in favour of the sponsor

Heres et al. Am J Psych 2006



Leucht et al. Lancet 2009

- 220 RCTs
- 150 db RCTs
- 20000 participants
- 10 SGAs
- 10 outcomes
- Multiple moderator analyses

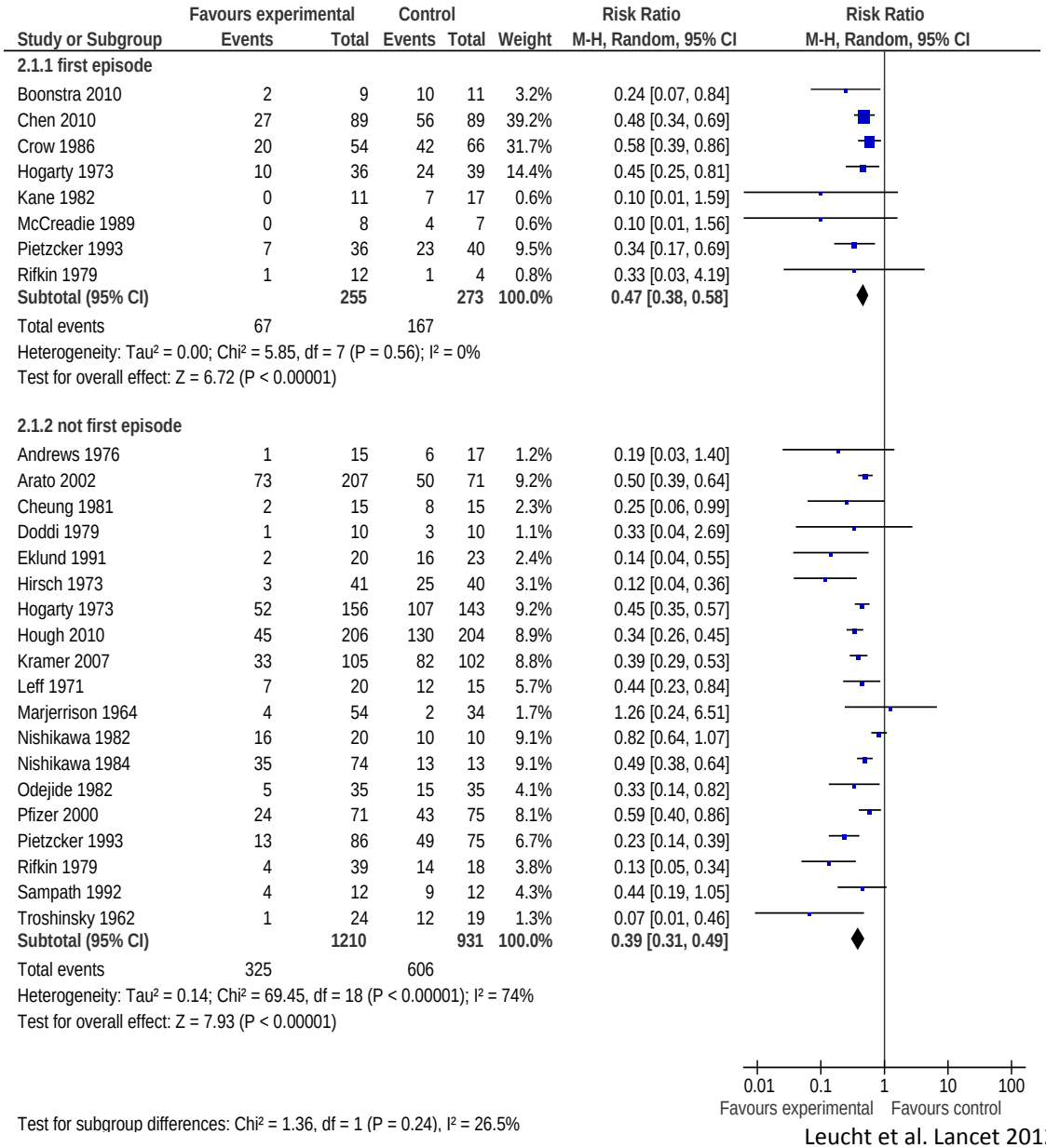
First-episode versus multiple episode patients (relapse 7-12 months)

Rationale:
- 20% of first-episode patients will not have a 2nd episode within 5 years (Robinson et al. Arch Gen Psych 1999, Shepherd et al. BrJPsych Suppl. 1994)

- They are thought to have a better prognosis

- Do they need maintenance treatment?

- Problem: 20% can not be identified in advance



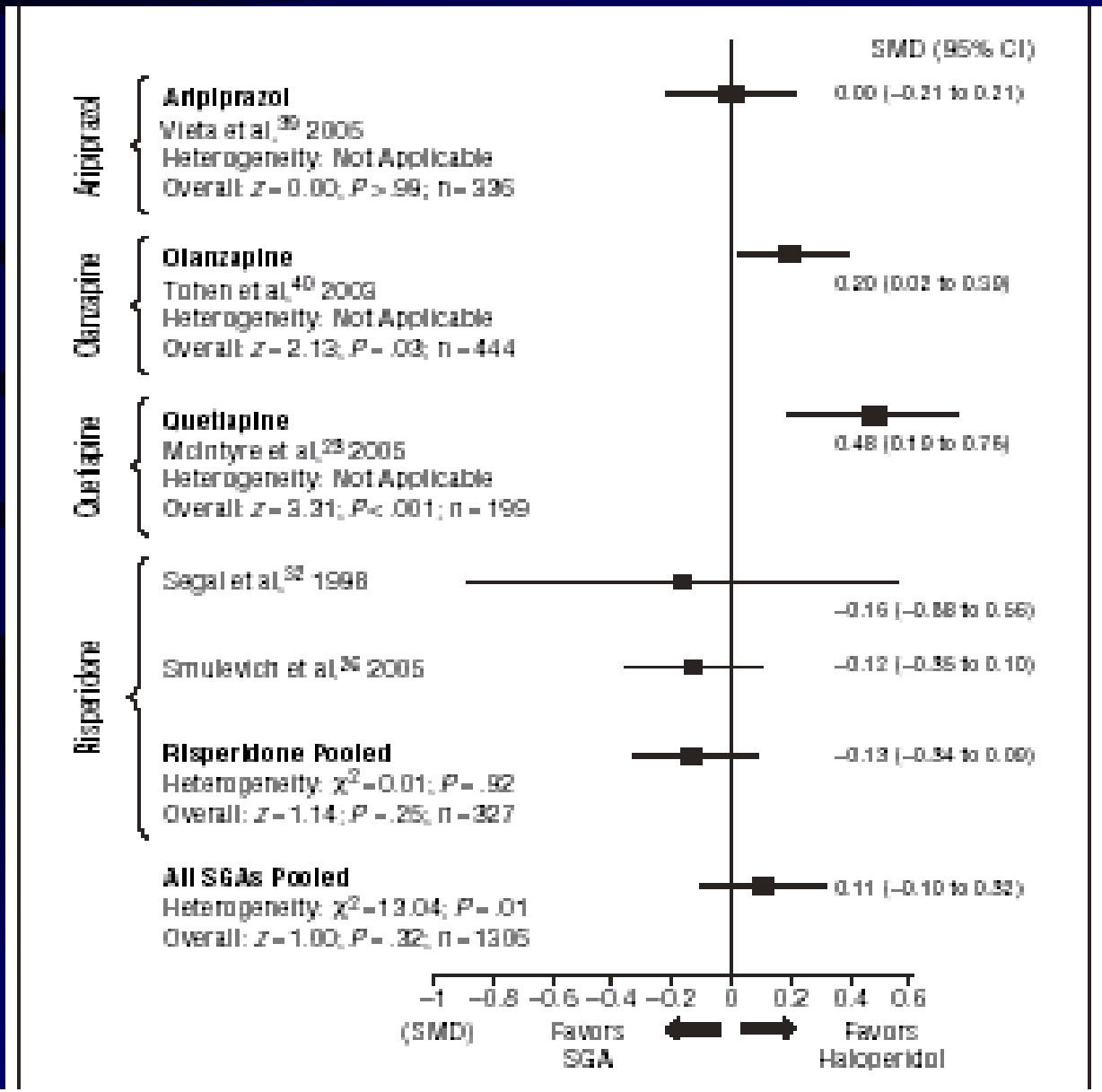
Study or Subgroup	Favours experimental Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
2.1.1 first episode						
Boonstra 2010	2	9	10	11	3.2%	0.24 [0.07, 0.84]
Chen 2010	27	89	56	89	39.2%	0.48 [0.34, 0.69]
Crow 1986	20	54	42	66	31.7%	0.58 [0.39, 0.86]
Hogarty 1973	10	36	24	39	14.4%	0.45 [0.25, 0.81]
Kane 1982	0	11	7	17	0.6%	0.10 [0.01, 1.59]
McCreadie 1989	0	8	4	7	0.6%	0.10 [0.01, 1.56]
Pietzcker 1993	7	36	23	40	9.5%	0.34 [0.17, 0.69]
Rifkin 1979	1	12	1	4	0.8%	0.33 [0.03, 4.19]
Subtotal (95% CI)		**255**		**273**	**100.0%**	**0.47 [0.38, 0.58]**
Total events	67		167			
Heterogeneity: Tau² = 0.00; Chi² = 5.85, df = 7 (P = 0.56); I² = 0%						
Test for overall effect: Z = 6.72 (P < 0.00001)						
2.1.2 not first episode						
Andrews 1976	1	15	6	17	1.2%	0.19 [0.03, 1.40]
Arato 2002	73	207	50	71	9.2%	0.50 [0.39, 0.64]
Cheung 1981	2	15	8	15	2.3%	0.25 [0.06, 0.99]
Doddi 1979	1	10	3	10	1.1%	0.33 [0.04, 2.69]
Eklund 1991	2	20	16	23	2.4%	0.14 [0.04, 0.55]
Hirsch 1973	3	41	25	40	3.1%	0.12 [0.04, 0.36]
Hogarty 1973	52	156	107	143	9.2%	0.45 [0.35, 0.57]
Hough 2010	45	206	130	204	8.9%	0.34 [0.26, 0.45]
Kramer 2007	33	105	82	102	8.8%	0.39 [0.29, 0.53]
Leff 1971	7	20	12	15	5.7%	0.44 [0.23, 0.84]
Marjerrison 1964	4	54	2	34	1.7%	1.26 [0.24, 6.51]
Nishikawa 1982	16	20	10	10	9.1%	0.82 [0.64, 1.07]
Nishikawa 1984	35	74	13	13	9.1%	0.49 [0.38, 0.64]
Odejide 1982	5	35	15	35	4.1%	0.33 [0.14, 0.82]
Pfizer 2000	24	71	43	75	8.1%	0.59 [0.40, 0.86]
Pietzcker 1993	13	86	49	75	6.7%	0.23 [0.14, 0.39]
Rifkin 1979	4	39	14	18	3.8%	0.13 [0.05, 0.34]
Sampath 1992	4	12	9	12	4.3%	0.44 [0.19, 1.05]
Troshinsky 1962	1	24	12	19	1.3%	0.07 [0.01, 0.46]
Subtotal (95% CI)		**1210**		**931**	**100.0%**	**0.39 [0.31, 0.49]**
Total events	325		606			
Heterogeneity: Tau² = 0.14; Chi² = 69.45, df = 18 (P < 0.00001); I² = 74%						
Test for overall effect: Z = 7.93 (P < 0.00001)						
Test for subgroup differences: Chi² = 1.36, df = 1 (P = 0.24), I² = 26.5%						

Leucht et al. Lancet 2012
Cochrane Database Syst Rev 2012

Meta-analyses are often the only way to objectively summarise the evidence if there are many studies

They help to clarify „hidden“ issues

Haloperidol reduces mania more than some second generation antipsychotics



Scherk et al.

Arch Gen Psych 2007

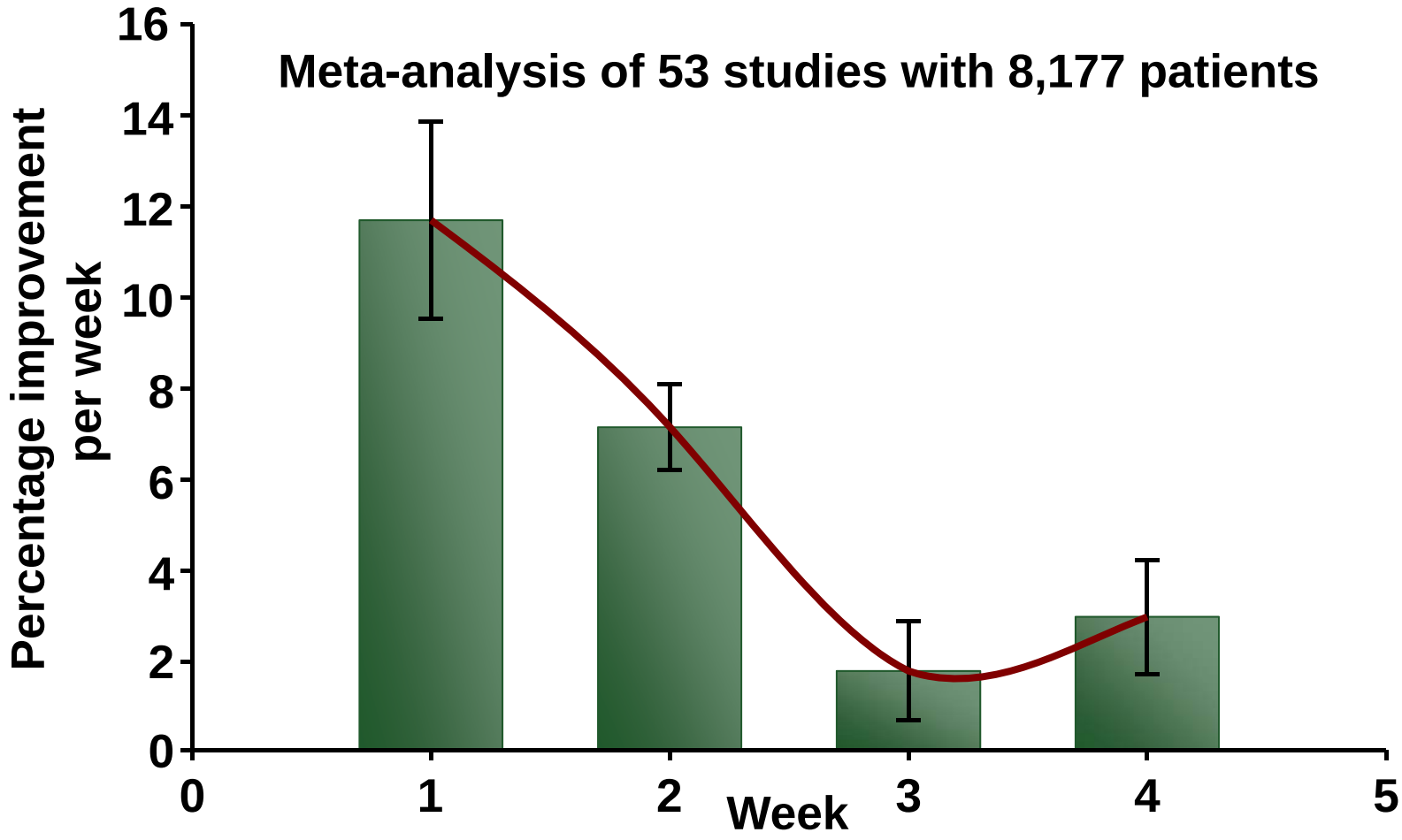
Meta-analyses are often the only way to objectively summarise the evidence if there are many studies

They help to clarify „hidden“ issues

They sometimes reject dogmas

Time course of antipsychotic effect

Psychotic symptoms after subtraction of placebo effect



Slide obtained with generous permission from Ofer Agid

Agid et al. Arch Gen Psychiatry 2003; 60: 1228–1235

Meta-analysis of the cognitive effects of conventional antipsychotics (Mishara and Goldberg Biol Psych 2004)

	Citation	CogFunction	N1	N2	Effect	Lower	Upper	PValue
	Allen et al, 1997	Multiple Cog Domains	25	22	-.27	-.86	.33	.36
	Allen et al, 2000	Executive/ Motor	19	19	-.21	-.87	.45	.52
	Braff & Saccuzzo, 1982	Speed of Proc/Verbal Ability	20	16	.28	-.41	.96	.41
	Chapman & Knowles, 1964	Conceptual Breadth	24	24	-.32	-.90	.27	.27
	Cleghorn et al, 90	Multiple Cog Domains	37	27	.04	-.47	.55	.87
	David, A. 1995	Semantic Priming	6	40	.17	-.72	1.05	.70
	Earle-boyer et al, 1991	Attention (Sustained)	17	17	.29	-.42	.99	.40
	Epstein et al, 1996	Attention (Sustained)	23	23	.41	-.19	1.01	.16
	Finkelstein et al, 1997	Attention (Sustained)/WM	44	44	-.11	-.54	.31	.59
	Gilbertson&Kemp, 1997	Attention/ Memory	12	9	.14	-.79	1.06	.75
	Gold and Hurt, 1990	General intelligence	19	19	.49	-.18	1.16	.13
	Harvey & Pedley, 1989	Attention (Selective)	18	18	.43	-.26	1.12	.20
	Heilizer, 1959	RT/Perceptual/Practice	22	22	.21	-.40	.82	.48
	Himmelhoch et al., 1996	Multiple Cog Domains	10	16	-.01	-.85	.82	.97
	Hommer et al., 1991	Oculomotor	8	7	.15	-.97	1.27	.76
	Killian, Holzman et al, 1984	Multiple Cog Domains	13	34	.37	-.29	1.04	.25
	Litman et al, 1989	Oculomotor	11	10	.95	-.03	1.92	.04
	Marder et al, 1984	Attention (Visual)	13	13	.56	-.27	1.39	.15
	Mass, 1983	Multiple Cog Domains	40	43	-.27	-.71	.17	.22
	Nestor et al, 1991	Attention (Sustained)	12	13	.58	-.27	1.43	.15
	Oltmanns et al, 1978	Attention (Selective)	15	16	.88	.10	1.65	.02
	Pigache, 1993	Attention (auditory)	20	20	.37	-.28	1.01	.24
	Purdon et al, 2001	Multiple Cog Domains	27	27	.48	-.08	1.03	.08
	Serper et al, 1990	Learning/Auto Proc/Atten	4	4	1.43	-.69	3.55	.06
	Serper et al, 1994	Learning/Auto Proc/Atten	12	13	.89	.01	1.76	.03
	Spohn et al, 1977	Attention (Selective)	20	20	.61	-.05	1.27	.06
	Spohn et al., 1989	Oculomotor	36	64	.10	-.31	.52	.62
	Sweeney et al., 1994	Oculomotor	24	24	-.05	-.63	.53	.86
	Sweeney et al., 1997	Oculomotor	10	10	-.09	-1.03	.85	.83
	Sumiyoshi et al., 1999	Language	11	11	.31	-.59	1.21	.46
	Tomer & Flor-Henry, 1989	Hemispheric Proc	15	15	.27	-.48	1.03	.45
	Tomer, 1990	Verbal Ability/Hemispher Proc	12	12	1.06	.14	1.97	.01
	Ulferts et al, 1999	Time Perception	12	12	-.12	-.97	.73	.76
	Zahn et al, 1994	Atten/Sensory Dominance	14	14	.60	-.19	1.40	.11
Random	Combined (34)		625	698	.22	.10	.34	.00

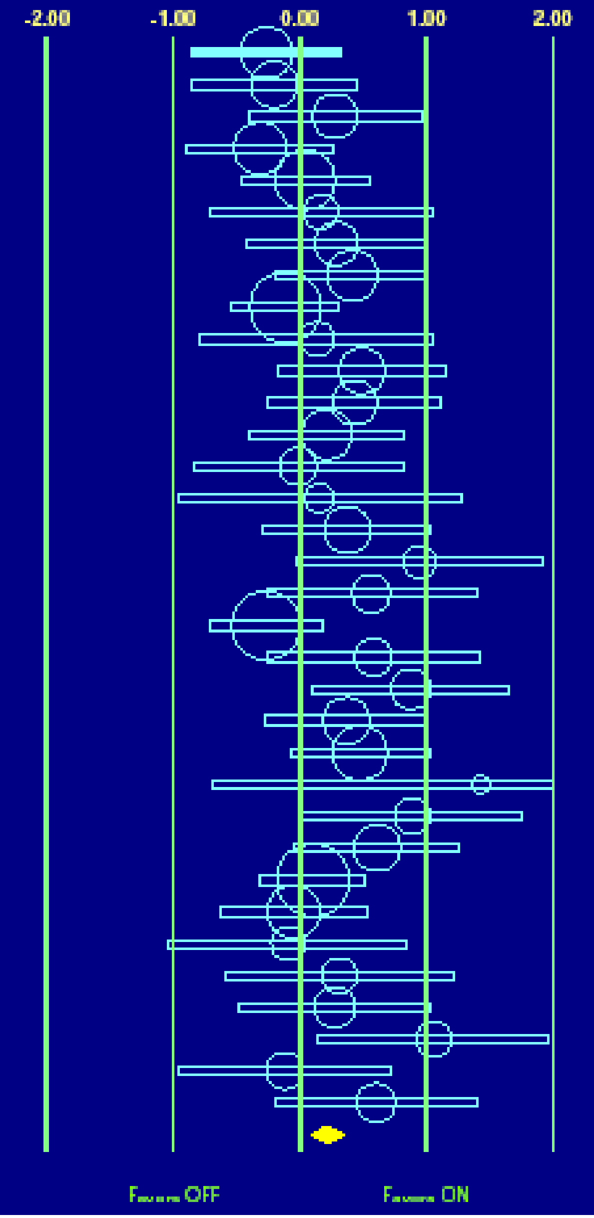
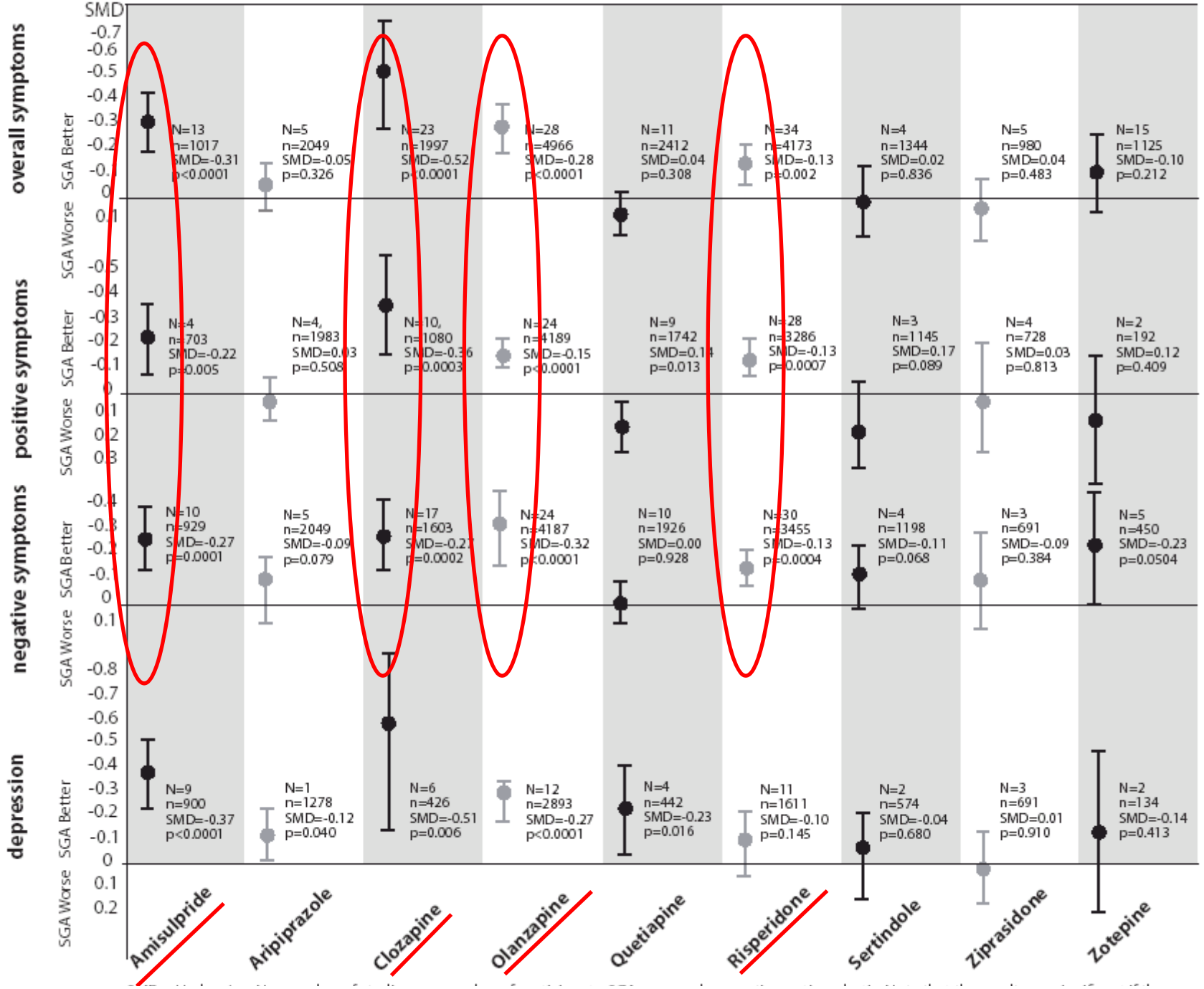
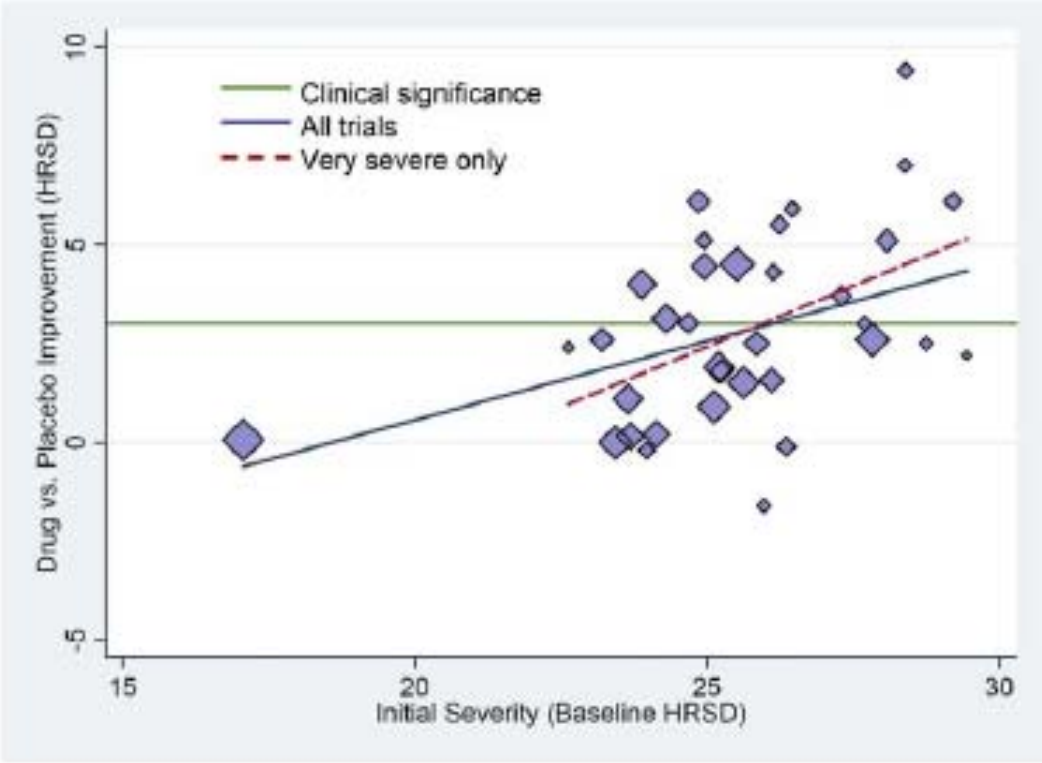


Figure 2: SGA versus FGA - efficacy in various domains



THE NEW YORKER

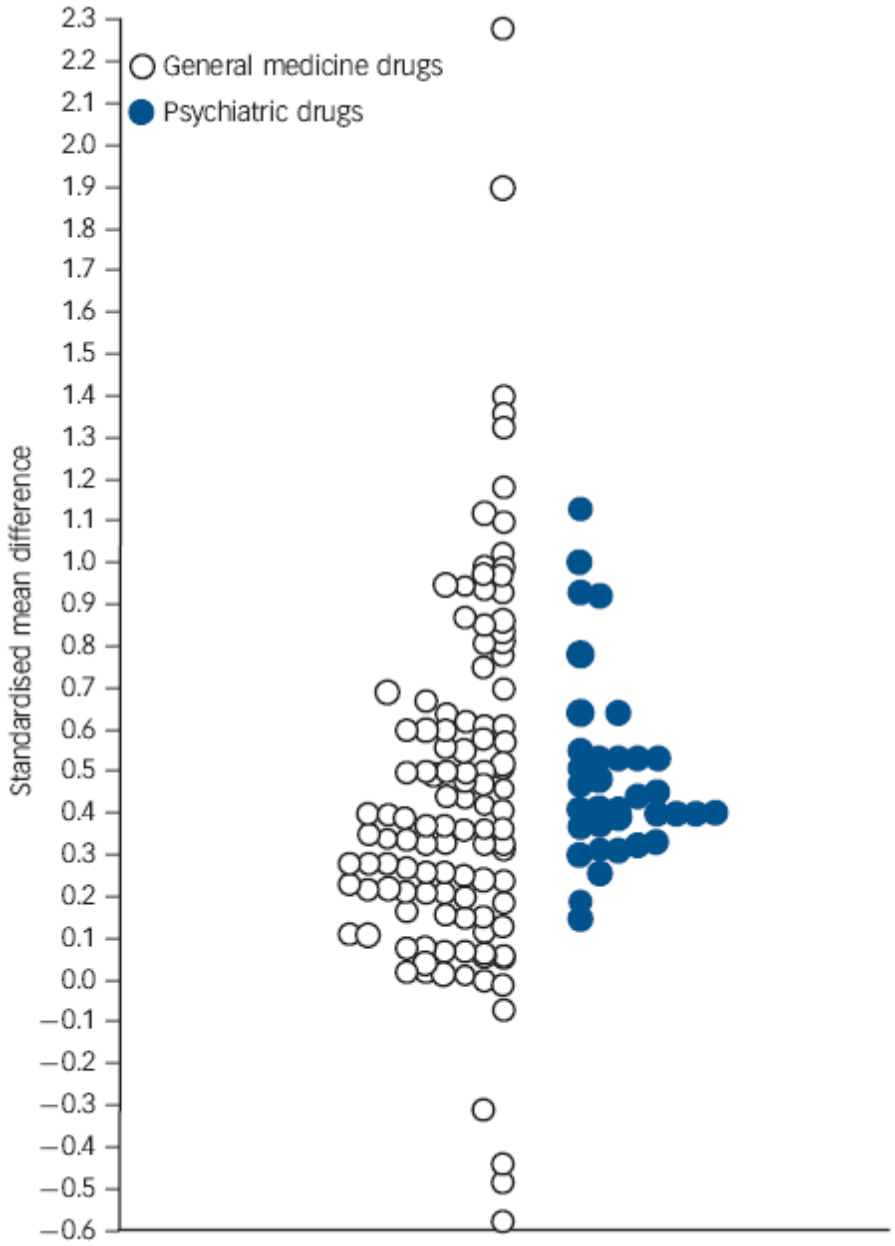


Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 2008; 5: e45.

The psychiatric literature is so confusing that even the dissidents disagree.

Menand L. Head case: can psychiatry be a science? The New Yorker 2010; 1 March: 68–74.

Vergleich von 94 Metaanalysen somatischer Krankheitsbilder mit 33 Metaanalysen psychiatrischer Krankheitsbilder



	Nicht-Psycho-pharmaka	Psycho-pharmaka
SMD (median)	0.37	0.41
SMD (mean)	0.45	0.49
95% Konfidenz-intervall	0.37-0.53	0.41-0.57

Leucht S, Hierl S, Kissling W, Dold M, Davis JM. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. Br J Psychiatry 2012, 200: 97-106.

Meta-analyses are often the only way to objectively summarise the evidence if there are many studies

They help to clarify „hidden“ issues

They sometimes reject dogmas

They sometimes show that there is no evidence

Meta-analyses of old antipsychotic drugs with regionally restricted use

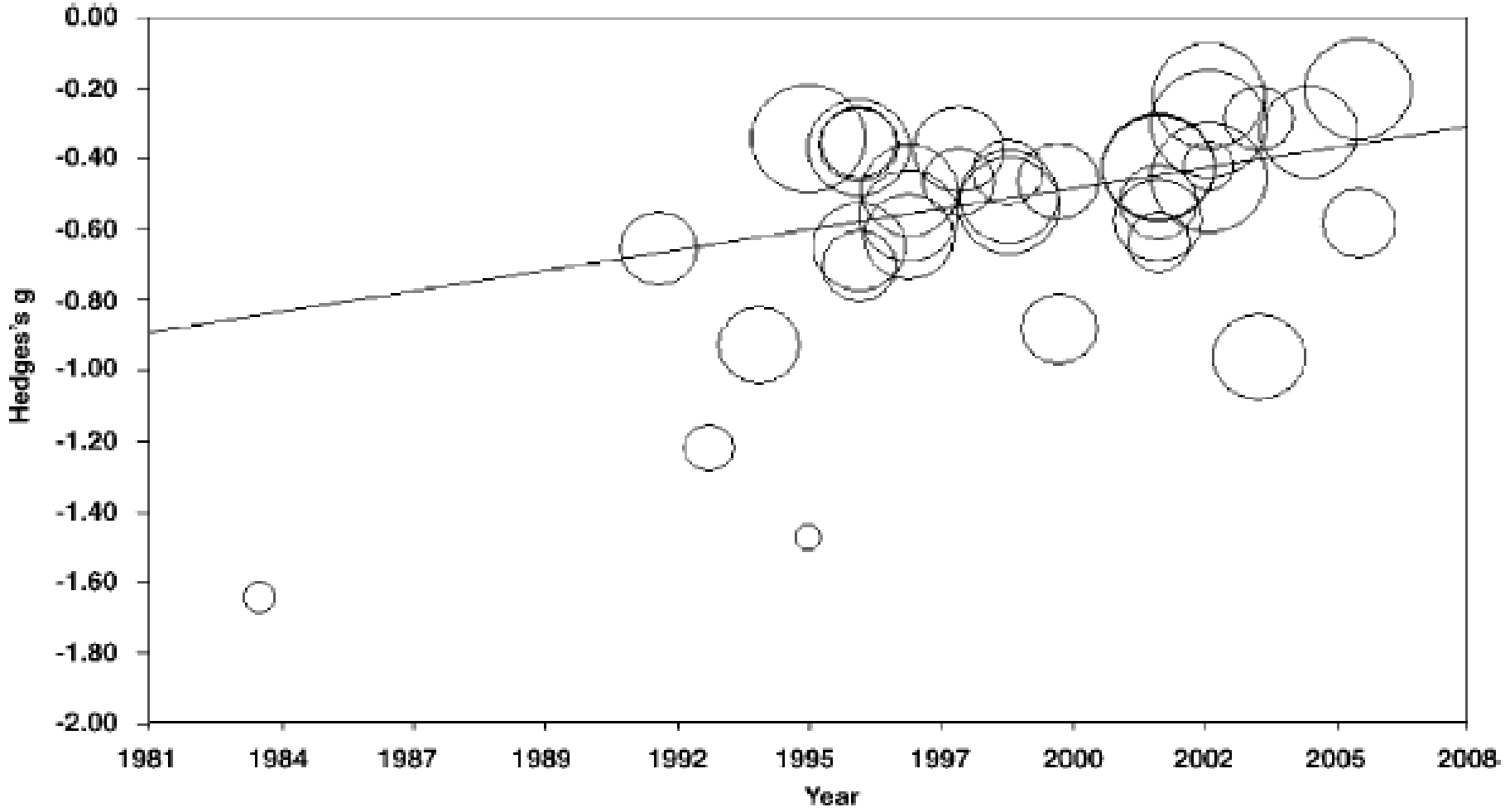
- ➔ Perazine only 5 small randomised studies,n = 286!
- ➔ Benperidol only one small RCT (unpublished manuscript)

Leucht and Hartung Cochrane Database of systematic reviews 2006

- Meta-analyses are often the only way to objectively summarise the evidence if there are many studies
- They help to clarify „hidden“ issues
- They sometimes reject dogmas
- They sometimes show that there is no evidence
- They can resolve heterogeneity by metaregression

Example for metaregression: decreasing antipsychotic drug versus placebo difference over time

(Leucht et al. Molecular Psychiatry 2009)



Meta-analyses are often the only way to objectively summarise the evidence if there are many studies

They help to clarify „hidden“ issues

They sometimes reject dogmas

They sometimes show that there is no evidence

They can resolve heterogeneity by metaregression

Latest methodological development: Multiple Treatments Meta-analysis ("network meta-analysis")

Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis

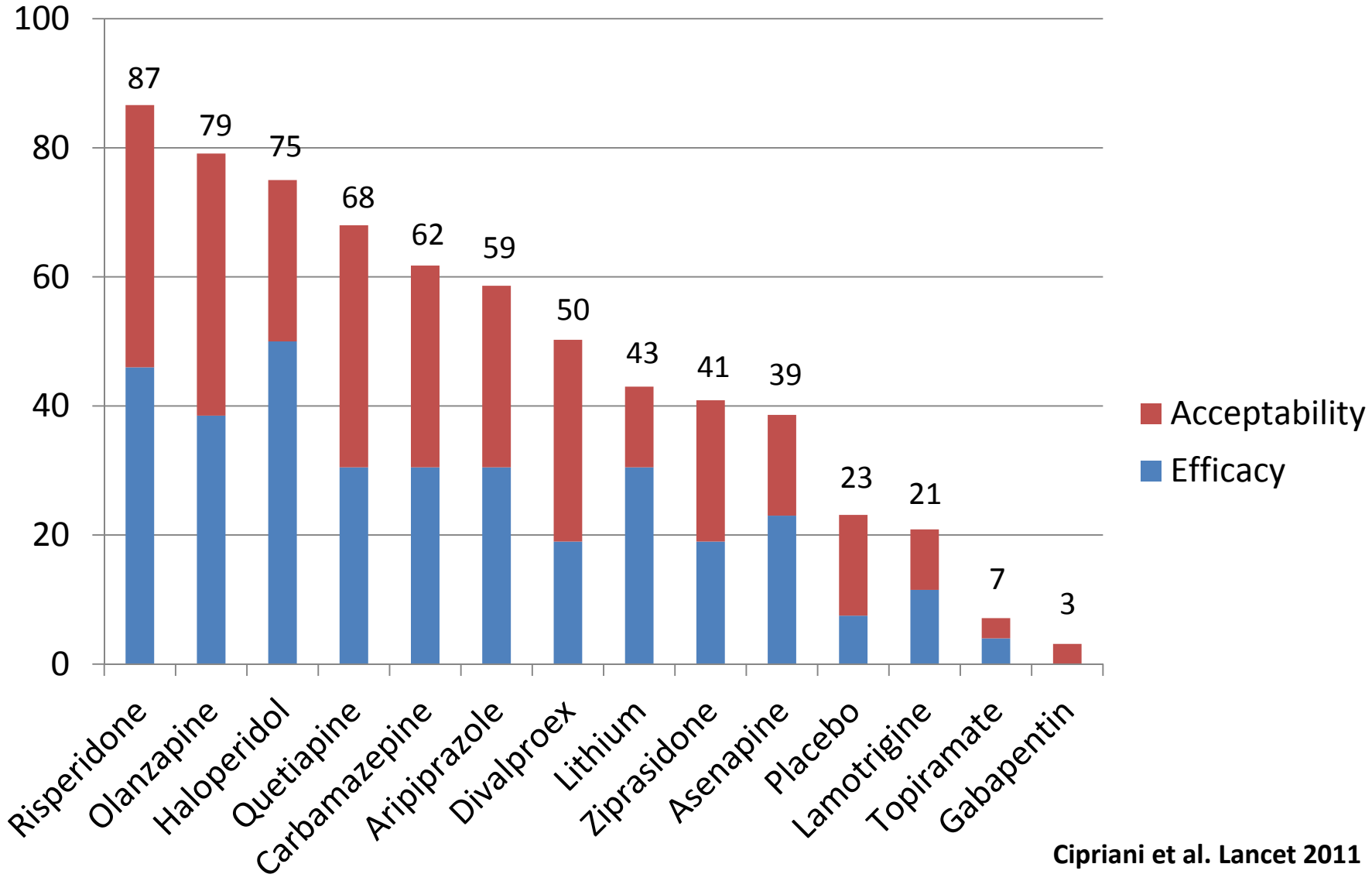
Andrea Cipriani, Toshiaki A Furukawa, Georgia Salanti, John R Geddes, Julian P T Higgins, Rachel Churchill, Norio Watanabe, Atsuo Nakagawa, Ichiro M Omori, Hugh McGuire, Michele Tansella, Corrado Barbui

Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis

Andrea Cipriani, Corrado Barbui, Georgia Salanti, Jennifer Rendell, Rachel Brown, Sarah Stockton, Marianna Purgato, Loukia M Spineli, Guy M Goodwin, John R Geddes

1. Cipriani et al. Lancet 2009;373:746–758;
2. Cipriani et al. Lancet 2011;378:1306–1315

Multiple treatments meta-analysis. Overall probability of antimanic treatments to be the best terms of both efficacy and acceptability, showing the separate contributions to the overall scores of efficacy (blue) and acceptability (red).



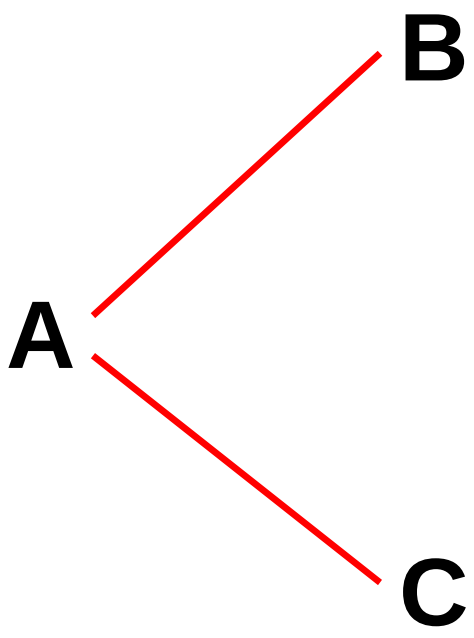
Cipriani et al. Lancet 2011

Advantages of Multiple-Treatments Meta-analysis

1. Uses all the data (direct and indirect)

2. Allows to calculate a hierarchy of drugs for an outcome: this is what guidelines need!

Principle of Multiple Treatments Meta-analysis („network meta-analysis“)

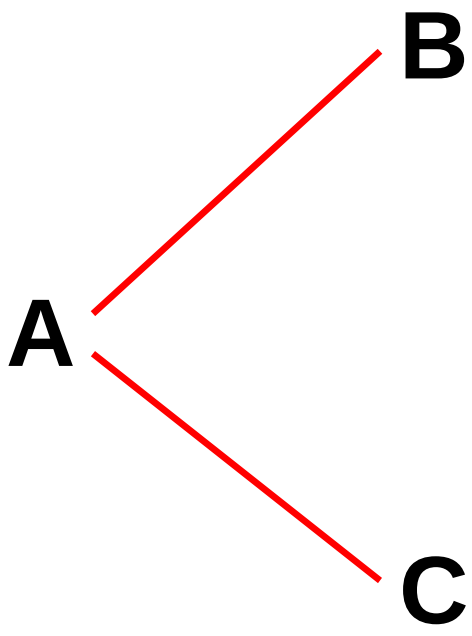


Examples:
1. Cipriani et al. Lancet 2009;373:746–758;
2. Cipriani et al. Lancet 2011;378:1306–1315

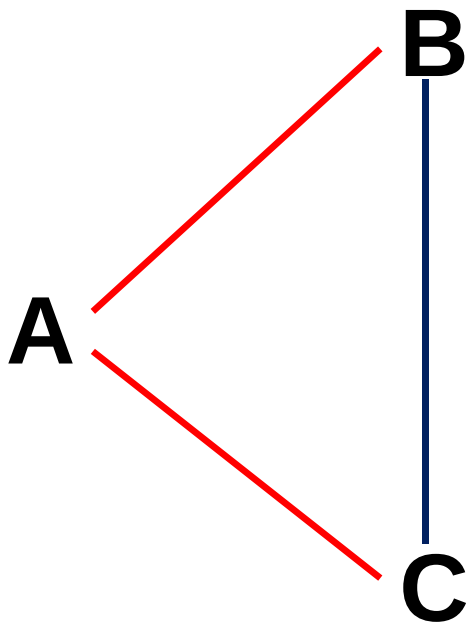
There are trials for A vs B and A vs C but none for B vs C

MTM, multiple treatments meta-analysis

Principle of MTM



There are trials to compare A vs B and A vs C, but none to compare B vs C



Trial results to compare B vs C can be estimated from those of A vs B and A vs C

MTM, multiple treatments meta-analysis

S Leucht, personal communication

Main problem of Multiple-Treatments Meta-analysis

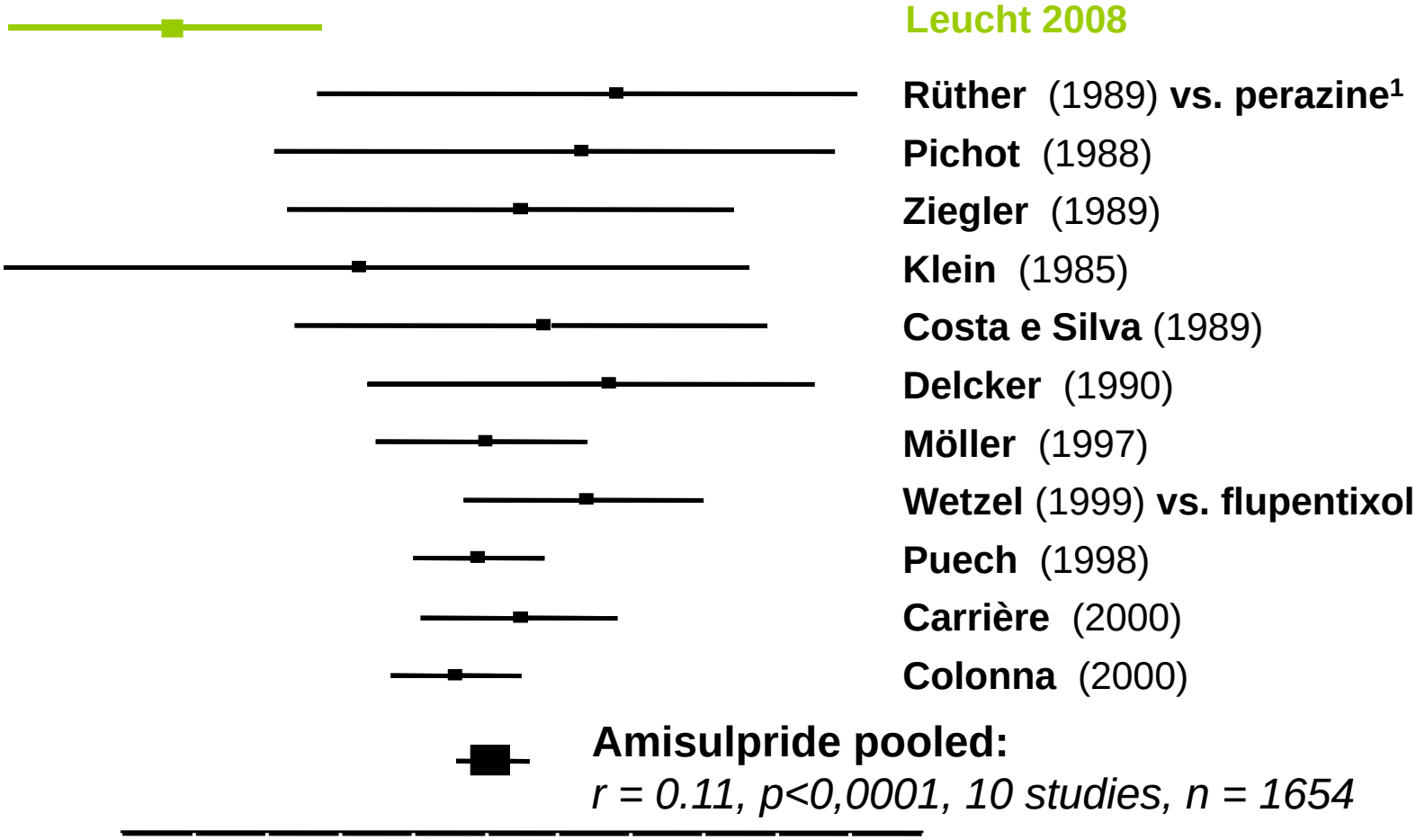
Can direct and indirect evidence be combined?

The coherence of direct and indirect evidence is examined by statistical tests

Limitations of Meta-analyses

- The „apples and oranges problem“ - heterogeneity, different study quality etc.
- In meta-analysis there are many judgement calls
- The original studies are frequently so poorly reported that meta-analytic procedures are not possible
- Publication bias

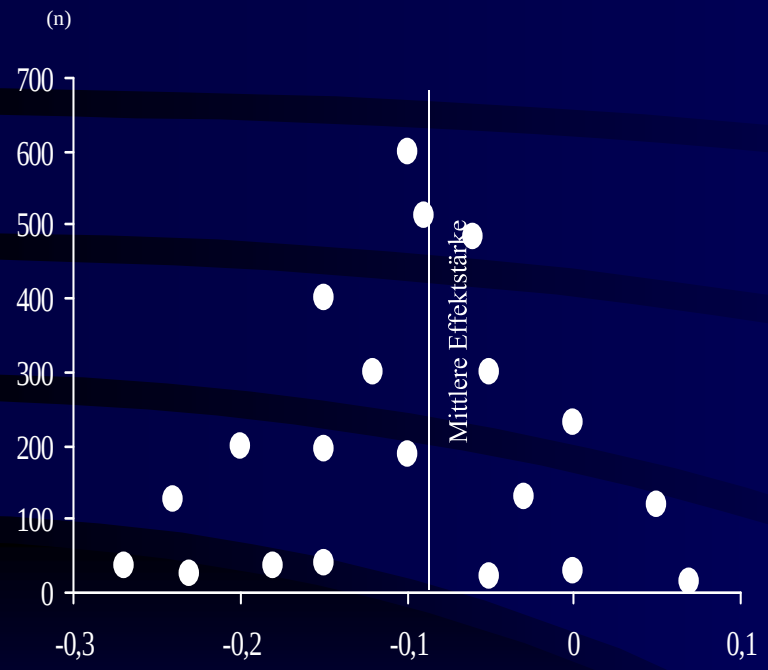
BPRS: Amisulpride vs. typical antipsychotics



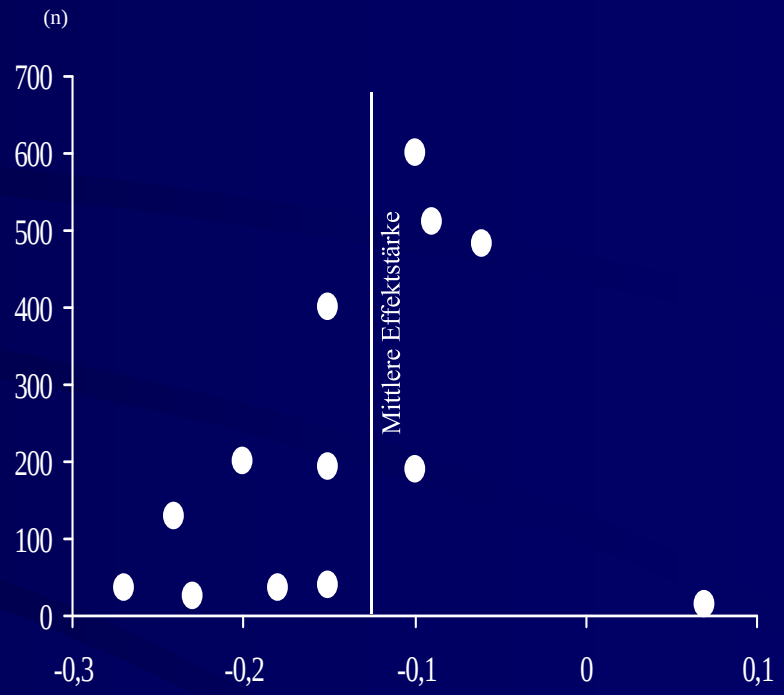
[1] endpoint analysis, not used for mean effect size

Leucht et al. Am J Psychiatry 2002

„Funnel-plot“ without publication bias



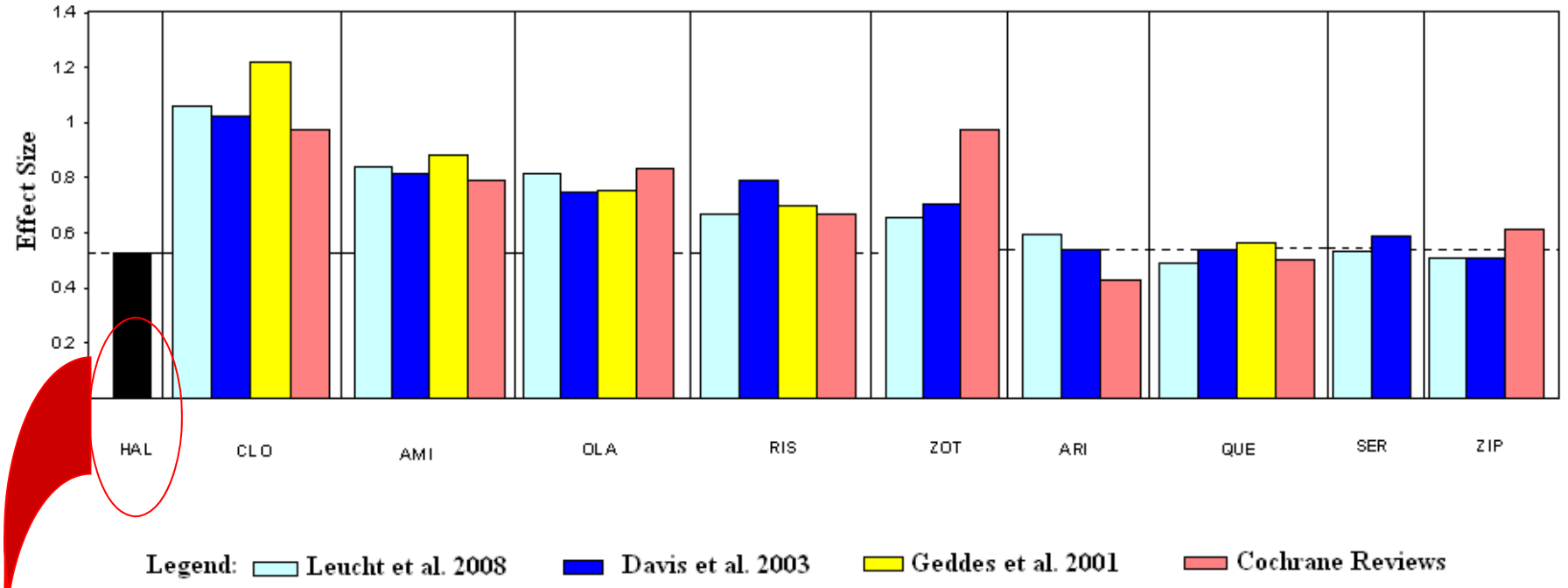
„Funnel-plot“ showing possible publication bias



Limitations of Meta-analyses

- The „apples and oranges problem“ - heterogeneity, different study quality etc.
- In meta-analysis there are many judgement calls
- The original studies are frequently so poorly reported that meta-analytic procedures are not possible
- Publication bias
- Although meta-analyses are methodologically objective, the results can be interpreted differently

The results of meta-analyses are consistent



The effect size of haloperidol versus placebo derived from 11 double-blind trials with 1540 participants, which is used as a benchmark. The effect sizes of the SGAs compared to FGAs have been added to haloperidol versus placebo for illustration.

HAL, haloperidol; AMI, amisulpride; ARI, aripiprazole; CLO, clozapine; OLA, olanzapine; QUE, quetiapine; RIS, risperidone; SER, sertindole; ZIP, ziprasidone; ZOT, zotepine; SGA, second-generation antipsychotics; FGA, first-generation antipsychotics

Leucht et al. Psychological Med 2009;39:1591–1602

Thank you for your attention