How Much Can We Trust Meta-Analyses: The Good, the Bad, and the Ugly (with case examples)

Tullio Palmerini

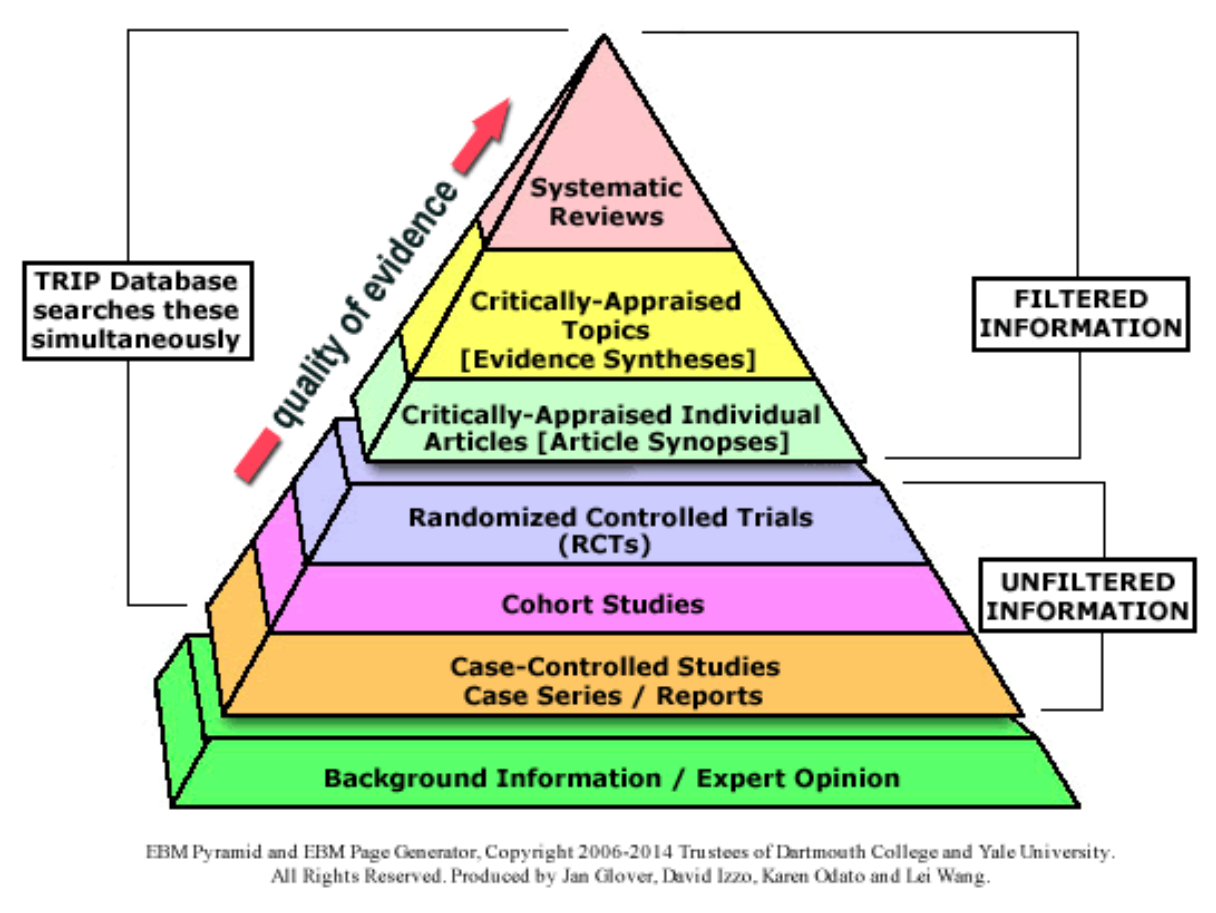
University of Bologna

Italy

Disclosure

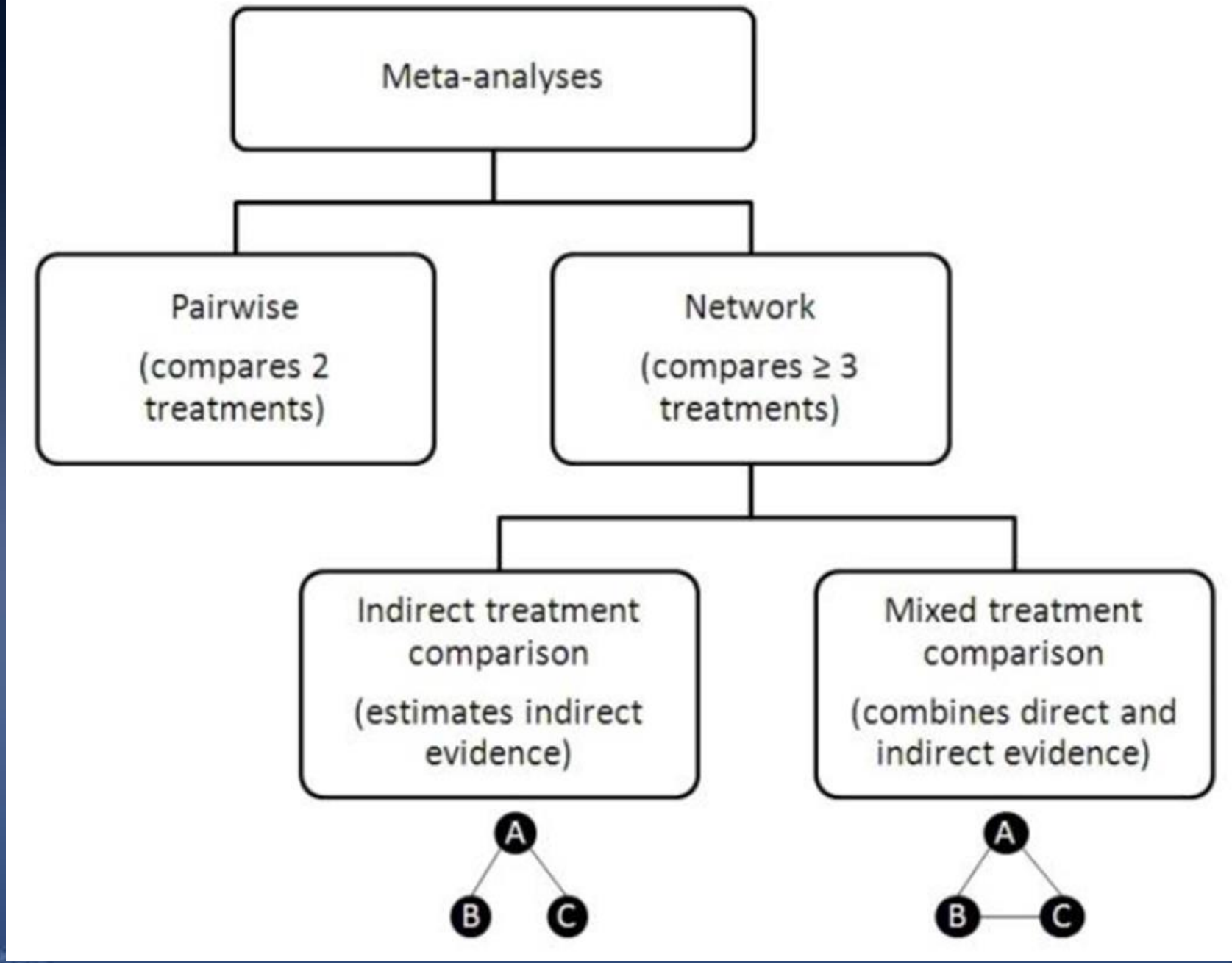
- None

The pyramid of evidence based medicine

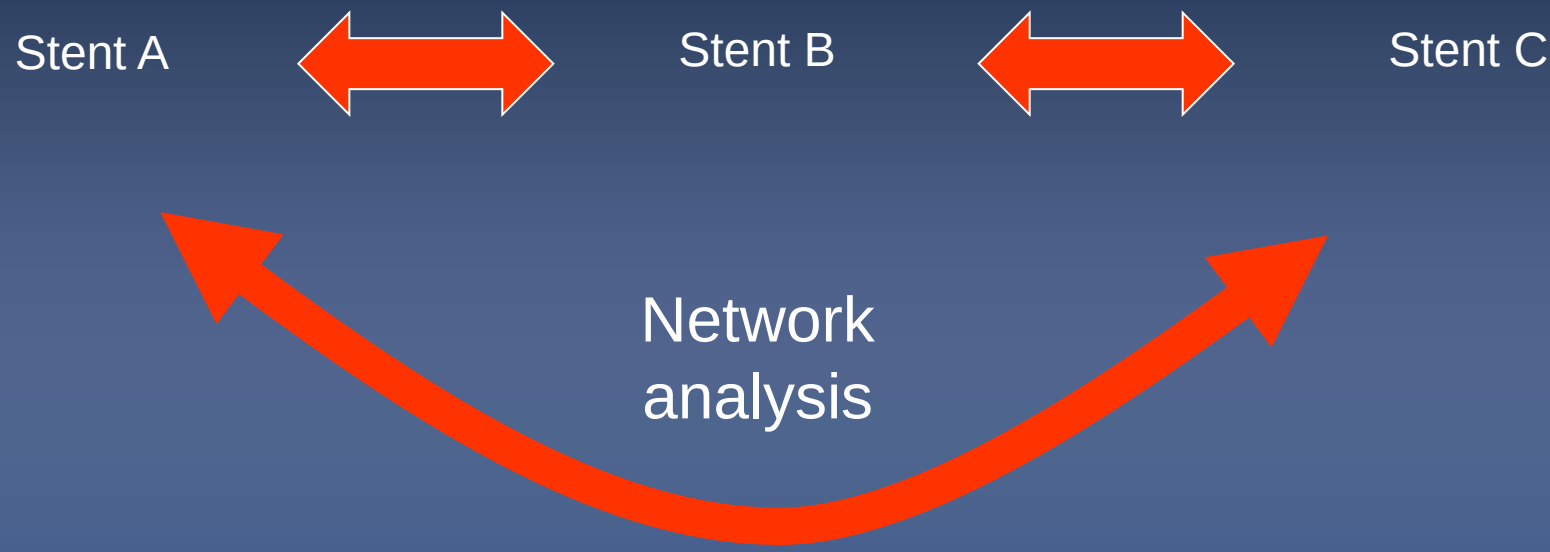


Sackett DL, Straus SE, Richardson WS, et al. *Evidence-based medicine: how to practice and teach EBM*. 2nd ed. Edinburgh: Churchill Livingstone, 2000.

ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT	SIZE OF TREATMENT EFFECT: CLASS I *Benefit >>> Risk* Procedure/Treatment **SHOULD** be performed/ administered	CLASS IIa *Benefit >> Risk* *Additional studies with focused objectives needed* **IT IS REASONABLE** to perform procedure/administer treatment	CLASS IIb *Benefit ≥ Risk* *Additional studies with broad objectives needed; additional registry data would be helpful* Procedure/Treatment **MAY BE CONSIDERED**	CLASS III *No Benefit* or CLASS III *Harm* Procedure/ Test / Treatment COR III: No benefit — Not Helpful / No Proven Benefit COR III: Harm — Excess Cost w/o Benefit or Harmful / Harmful to Patients
LEVEL A Multiple populations evaluated* Data derived from multiple randomized clinical trials or meta-analyses	■ Recommendation that procedure or treatment is useful/effective ■ Sufficient evidence from multiple randomized trials or meta-analyses	■ Recommendation in favor of treatment or procedure being useful/effective ■ Some conflicting evidence from multiple randomized trials or meta-analyses	■ Recommendation's usefulness/efficacy less well established ■ Greater conflicting evidence from multiple randomized trials or meta-analyses	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Sufficient evidence from multiple randomized trials or meta-analyses
LEVEL B Limited populations evaluated* Data derived from a single randomized trial or nonrandomized studies	■ Recommendation that procedure or treatment is useful/effective ■ Evidence from single randomized trial or nonrandomized studies	■ Recommendation in favor of treatment or procedure being useful/effective ■ Some conflicting evidence from single randomized trial or nonrandomized studies	■ Recommendation's usefulness/efficacy less well established ■ Greater conflicting evidence from single randomized trial or nonrandomized studies	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Evidence from single randomized trial or nonrandomized studies
LEVEL C Very limited populations evaluated* Only consensus opinion of experts, case studies, or standard of care	■ Recommendation that procedure or treatment is useful/effective ■ Only expert opinion, case studies, or standard of care	■ Recommendation in favor of treatment or procedure being useful/effective ■ Only diverging expert opinion, case studies, or standard of care	■ Recommendation's usefulness/efficacy less well established ■ Only diverging expert opinion, case studies, or standard of care	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Only expert opinion, case studies, or standard of care


Meta-analyses
Pairwise
(compares 2 treatments)
Network
(compares ≥ 3 treatments)
Indirect treatment comparison
(estimates indirect evidence)
Mixed treatment comparison
(combines direct and indirect evidence)
A
B
C
A
B
C

What is an indirect comparison ?



Problem formulation according to the PICO approach

- **Population of interest** - eg elderly male >2 weeks after myocardial infarction)
- **Intervention (or exposure)** – eg intracoronary infusion of progenitor blood cells
- **Comparison** – eg patients treated with progenitor cells vs standard therapy
- **Outcome(s)** – eg change in echocardiographic left ventricular ejection fraction from discharge to 6-month control

Study selection: Systematic review

- ***1st - screening of titles and abstracts (*BioMedCentral, clinicaltrials.gov, EMBASE, LILACS, and PubMed, Conference proceedings, Website)**

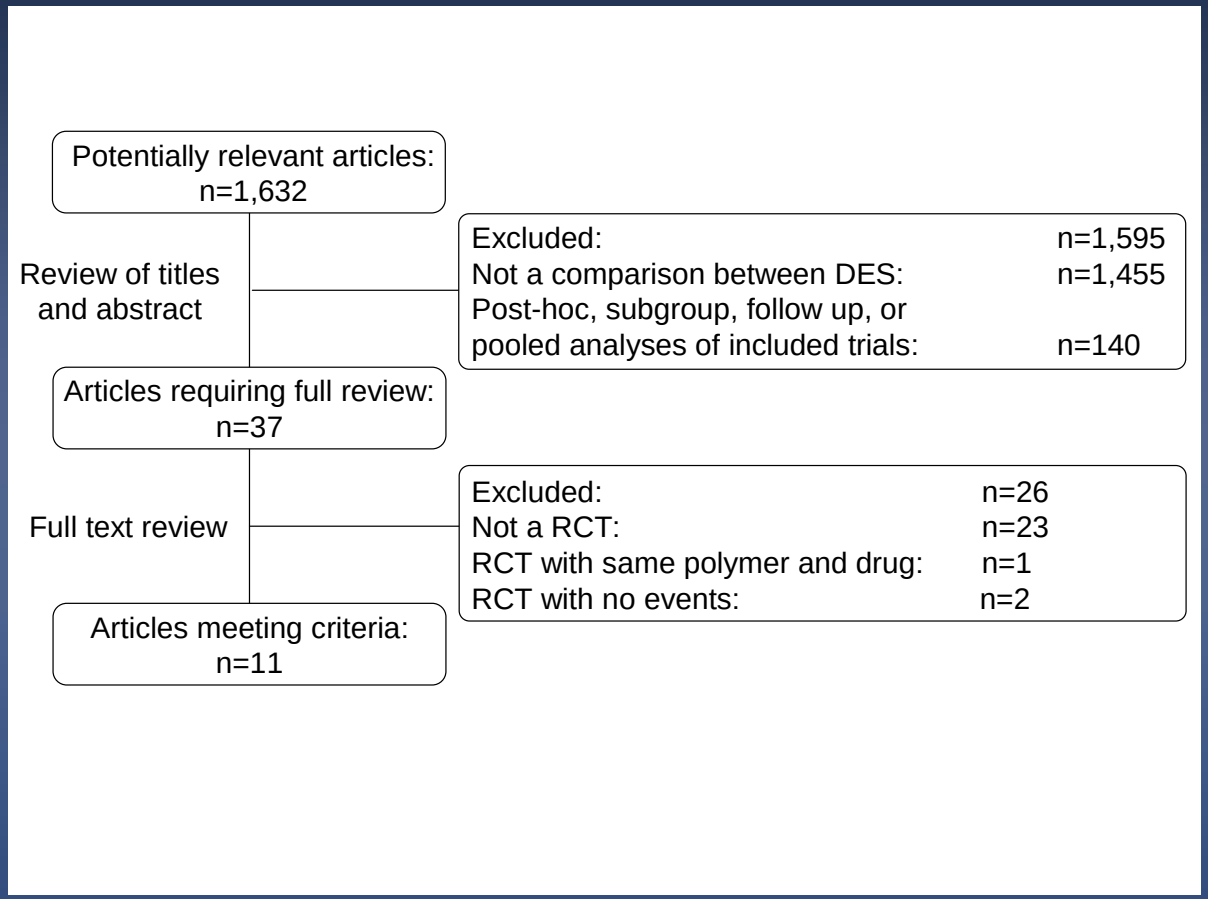
- ***2nd* – potentially pertinent citations are then retrieved as full reports and appraised according to prespecified and explicit inclusion/exclusion criteria**

- ***3rd* – studies fullfilling both inclusion and exclusion criteria, are then included in the systematic review**

Stent Thrombosis With Everolimus-Eluting Stents

Meta-Analysis of Comparative Randomized Controlled Trials

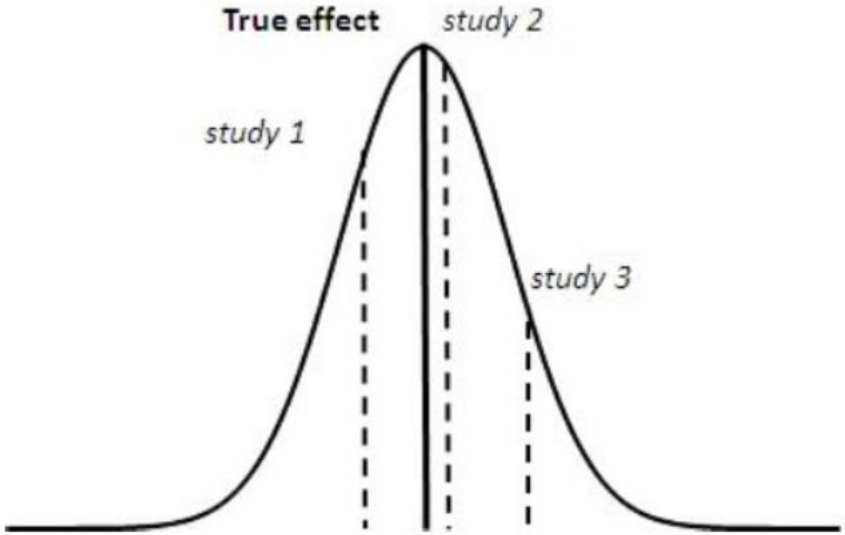
Tullio Palmerini, MD; Ajay J. Kirtane, MD, SM; Patrick W. Serruys, MD; Pieter C. Smits, MD; Elvin Kedhi, MD; Dean Kereiakes, MD; Diego Sangiorgi, MStat; Letizia Bacchi Reggiani, MSc; Christoph Kaiser, MD; Hyo-Soo Kim, MD; Antoinette De Waha, MD; Flavio Ribichini, MD; Gregg W. Stone, MD



Palmerini et al. Circ Cv Int 2012

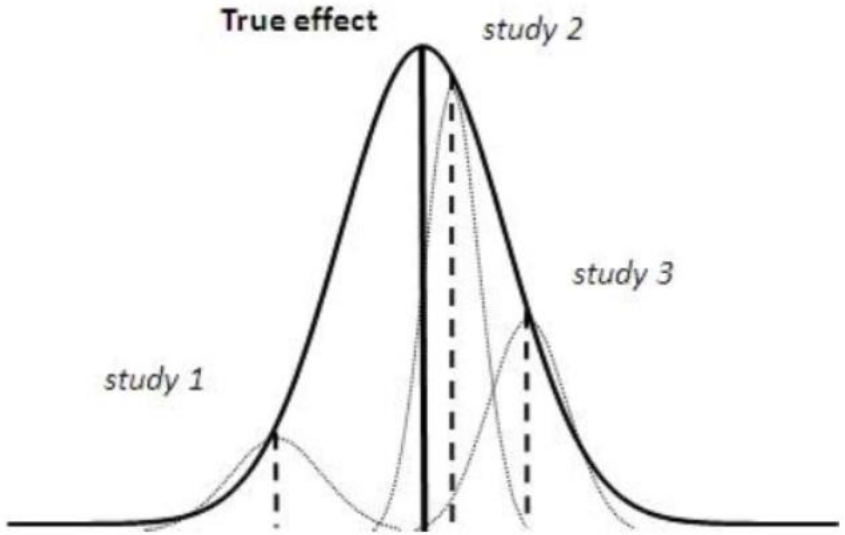
Statistical pooling: fixed versus random effect method

Fixed effect method



One distribution of true effect
Each trial is measuring the true effect
Natural random sampling variation

Random effect method



Acknokledges the occurrence of variation of true effects among studies

Mean effect of studies with their own distribution

Biondi-Zoccai, Network Meta-Analysis: Evidence Synthesis with Mixed Treatment Comparison 2014

Data synthesis: effect size for dichotomus events

- **Relative risks (RR)** are defined as the ratio of incidence rates, and are thus used for dichotomic variables).
- **Odds ratios (OR)** are defined as the ratio of the odds (P/[1-P]) and also used for dichotomic variables.
- **Hazard ratio (HR)** adjust for different duration of follow up and allow for data censoring.
- **The absolute risk difference (RD)** is the difference between the incidence of events in the experimental vs control groups. Depends on the prevalence of disease and allows calculation of NNT (1/RD).


THE GOOD, THE BAD & THE UGLY
Complete
Original Score
Extended
Version
musiche di
ENNIO
MORRICONE
IL BUONO, IL BRUTTO, IL CATTIVO

Meta-analysis

- …..the good

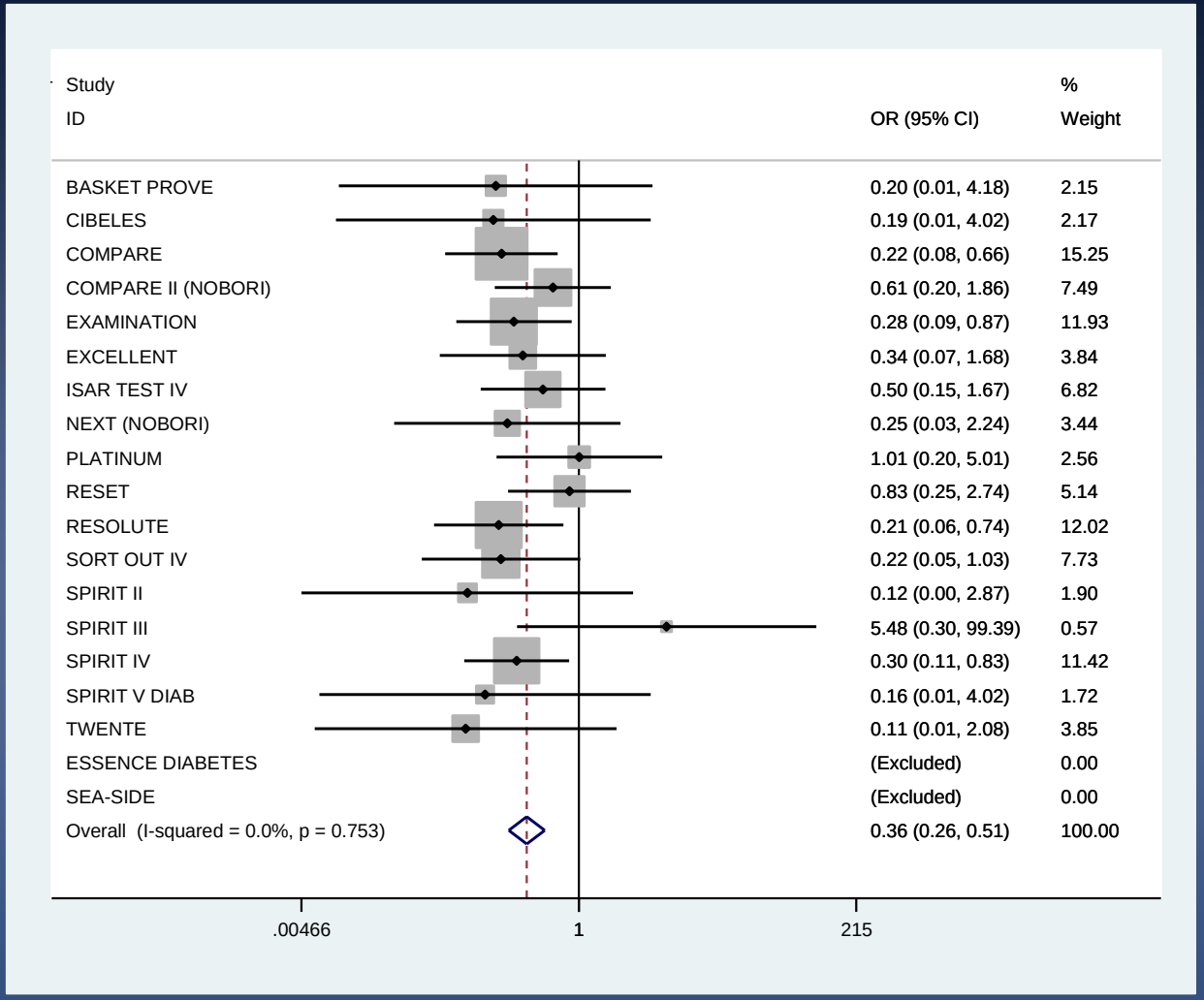
Advantages of meta-analyses

- **Quantitative synthesis with increased statistical power of the available evidence**

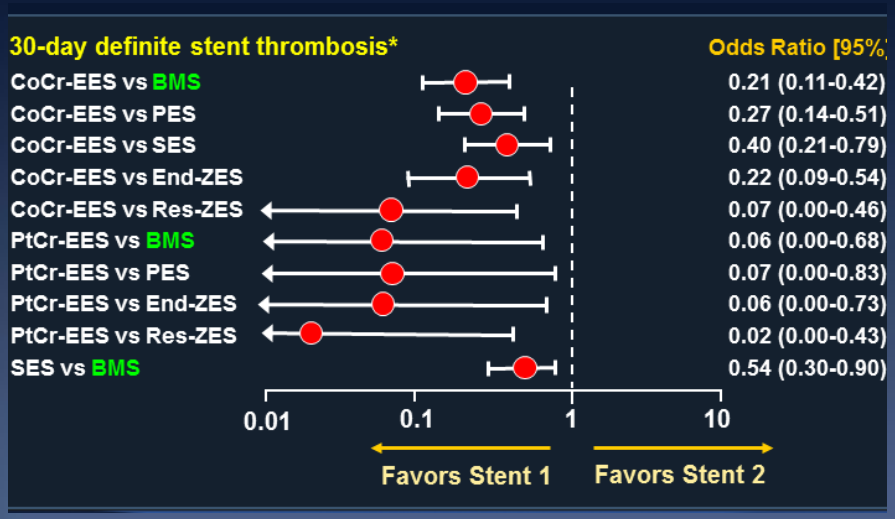
Meta-analysis in DES trials

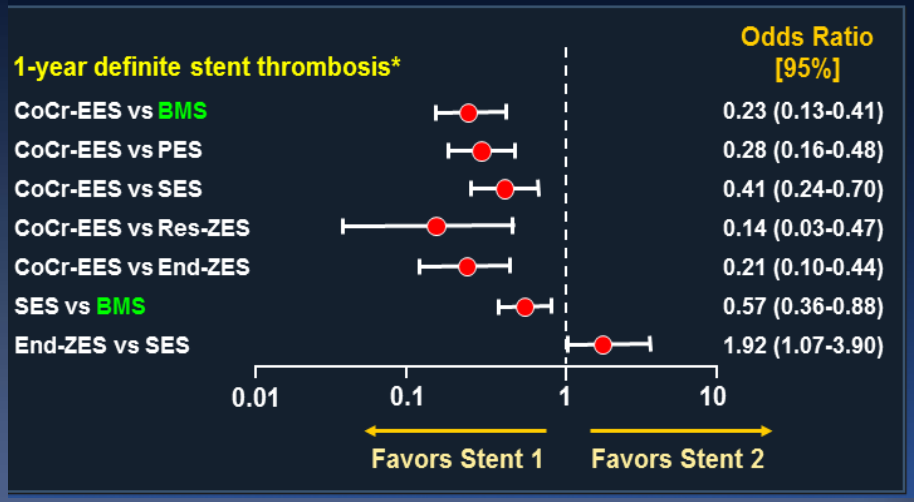
- At least 70 RCTs with almost 90.000 randomized patients
- Most studies have an nonferiority design
- Many of them are underpowerd because of an imbalance between observed and expected events
- Most of them have a composite endpoint
- None was powered for stent thrombosis

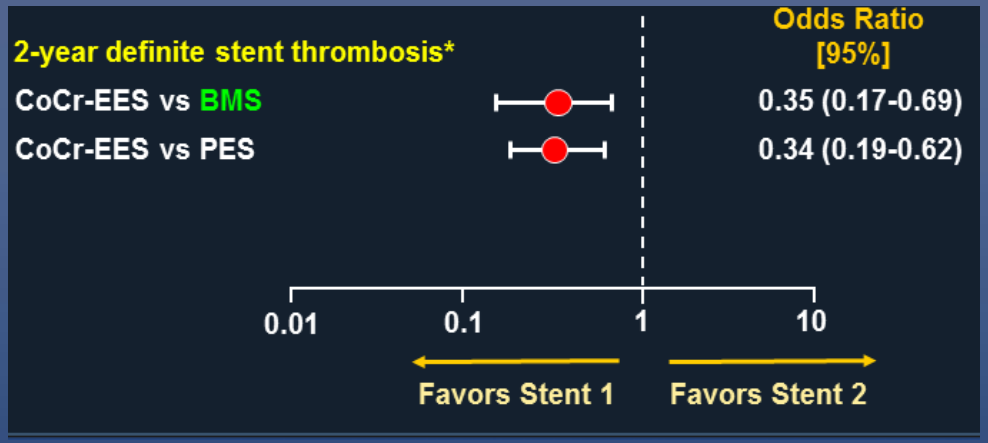
1-year definite ST



Network meta-analysis: 49 RCTs and 50,844 pts

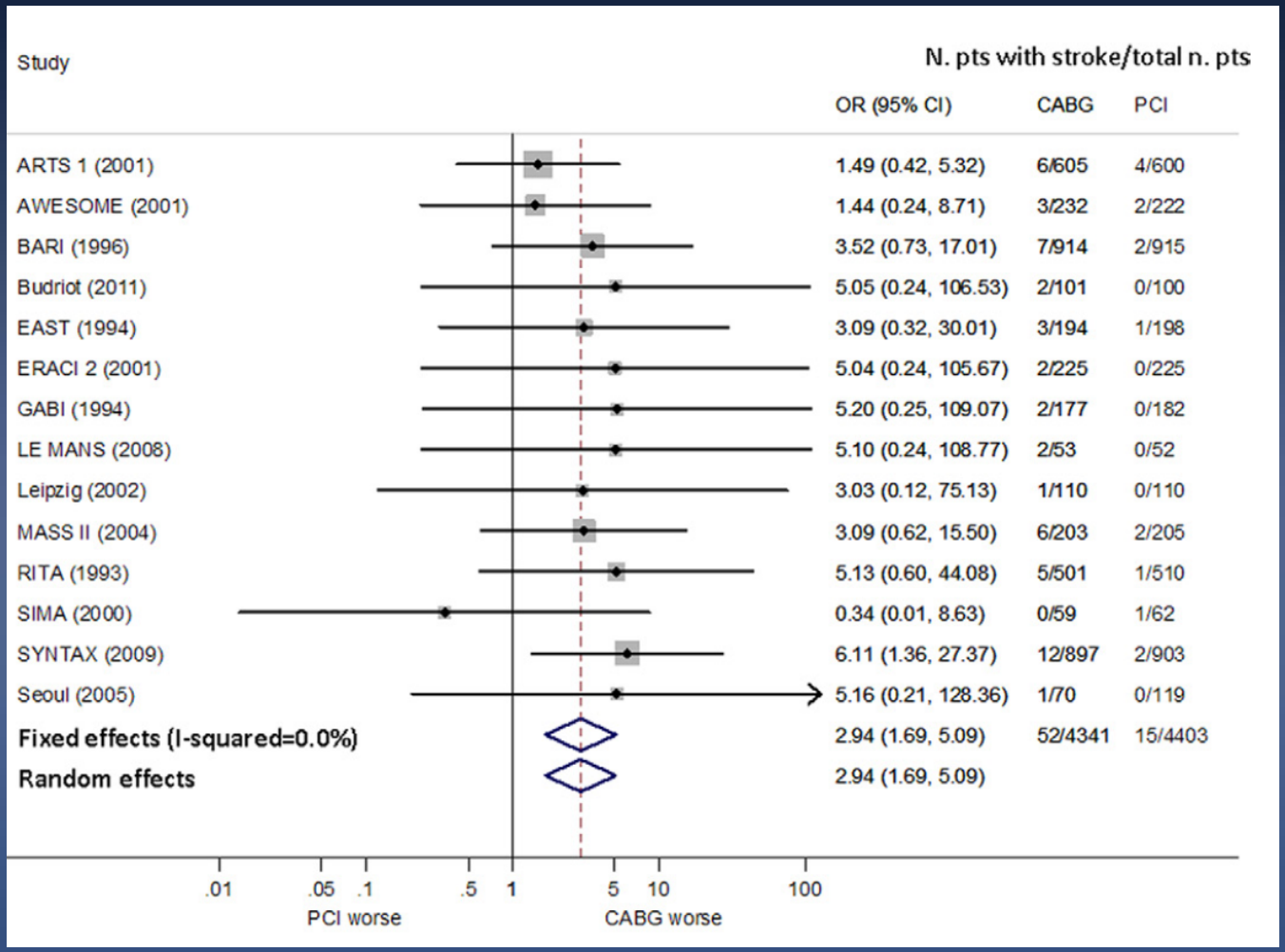






Palmerini et al. Lancet 2012;379:1393-402

Stroke with PCI versus CABG: 14 RCTs



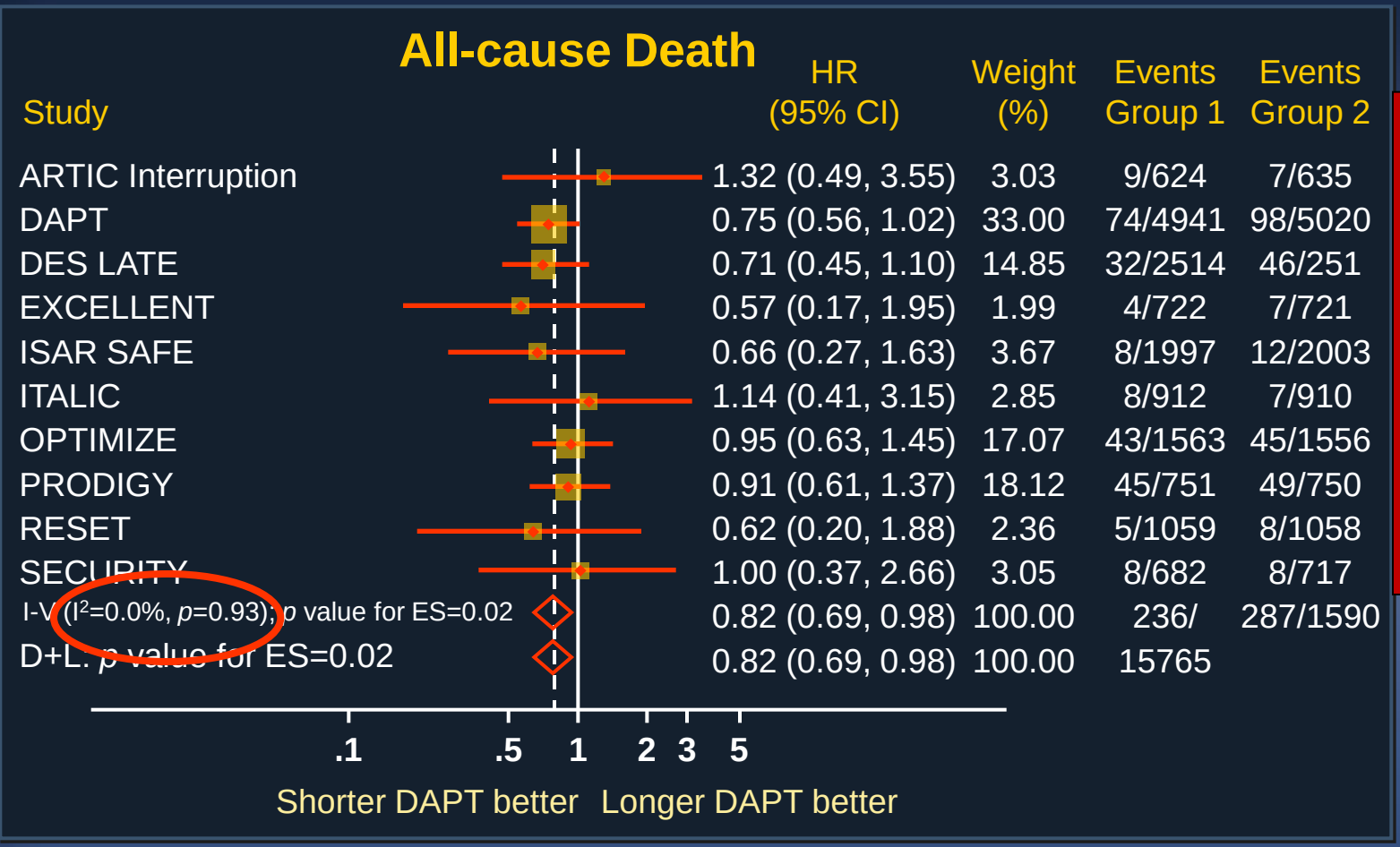
Palmerini et al; JACC 2012

DAPT trial

Outcome	Continued Thienopyridine (N=5020)	Placebo (N=4941)	Hazard Ratio, Thienopyridine vs. Placebo (95% CI)†	P Value†
	no. of patients (%)			
Stent thrombosis‡	19 (0.4)	65 (1.4)	0.29 (0.17–0.48)	<0.001
Definite	15 (0.3)	58 (1.2)	0.26 (0.14–0.45)	<0.001
Probable	5 (0.1)	7 (0.1)	0.71 (0.22–2.23)	0.55
Major adverse cardiovascular and cerebrovascular events§	211 (4.3)	285 (5.9)	0.71 (0.59–0.85)	<0.001
Death	98 (2.0)	74 (1.5)	1.36 (1.00–1.85)	0.05
Cardiac	45 (0.9)	47 (1.0)	1.00 (0.66–1.52)	0.98
Vascular	5 (0.1)	5 (0.1)	0.98 (0.28–3.39)	0.98
Noncardiovascular	48 (1.0)	22 (0.5)	2.23 (1.32–3.78)	0.002
Myocardial infarction	99 (2.1)	198 (4.1)	0.47 (0.37–0.61)	<0.001
Stroke	37 (0.8)	43 (0.9)	0.80 (0.51–1.25)	0.32
Ischemic	24 (0.5)	34 (0.7)	0.68 (0.40–1.17)	0.16
Hemorrhagic	13 (0.3)	9 (0.2)	1.20 (0.50–2.91)	0.68
Type uncertain	0	1 (<0.1)	—	0.32

Mauri et al, NEJM 2014

Mortality with Extended Duration DAPT After DES: A Pairwise and Bayesian Network Meta-Analysis of 10 RCTs and 31,666 Pts



22% ↑ mortality with prolonged DAPT (p=0.02)

ES=effect size

Palmerini T, Stone GW. Lancet 2015:on-line

Advantages of meta-analyses

- Application to any clinical research question
- Thorough appraisal of the internal validity of primary studies
- Explore clinical and statistical heterogeneity

Meta-analysis

- ……..the bad

Interpreting a meta-analysis

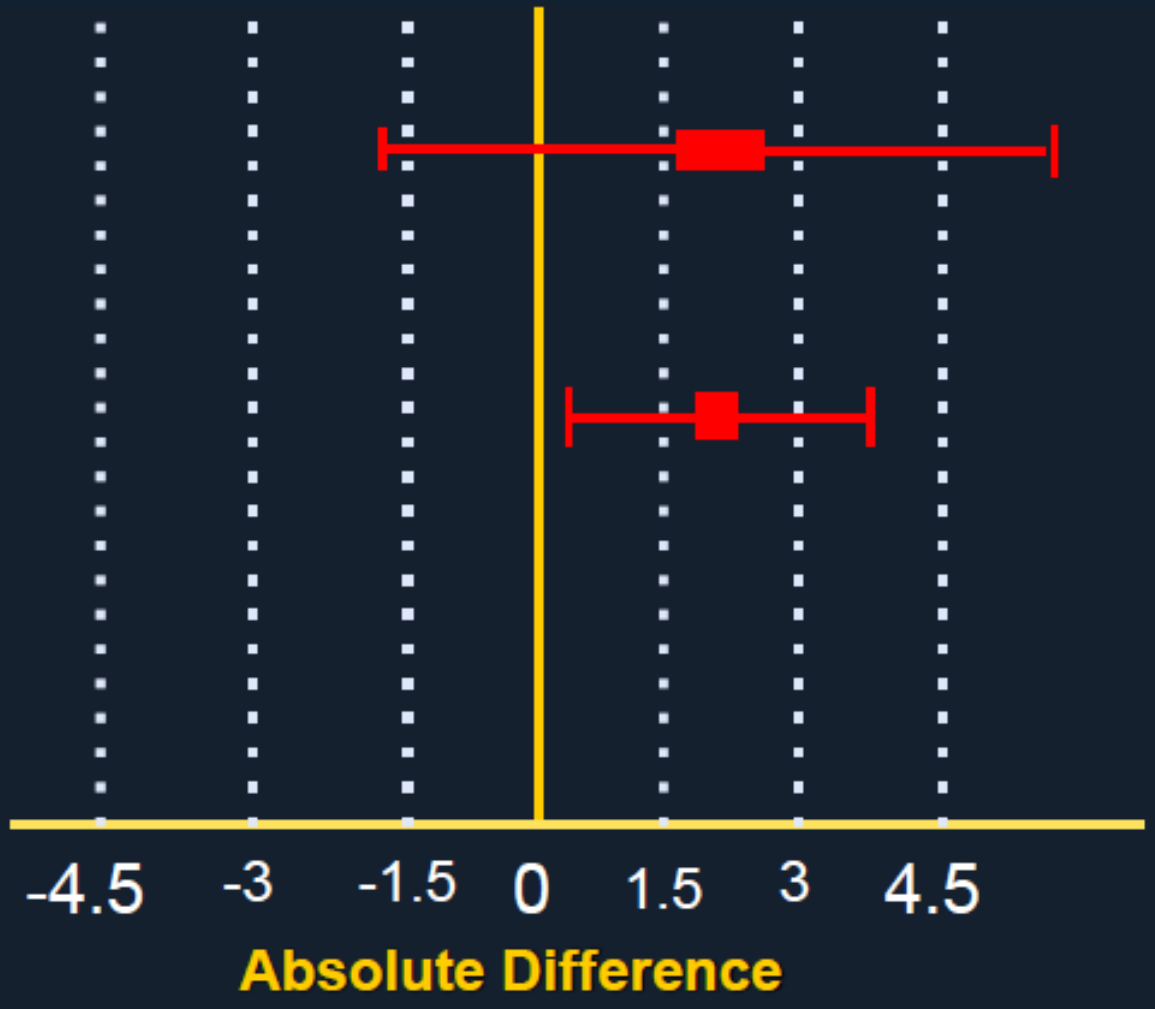
- **The strength of a meta-analysis is the strength of individual trials and of methods used**

Potential problem

Individual Small Trial

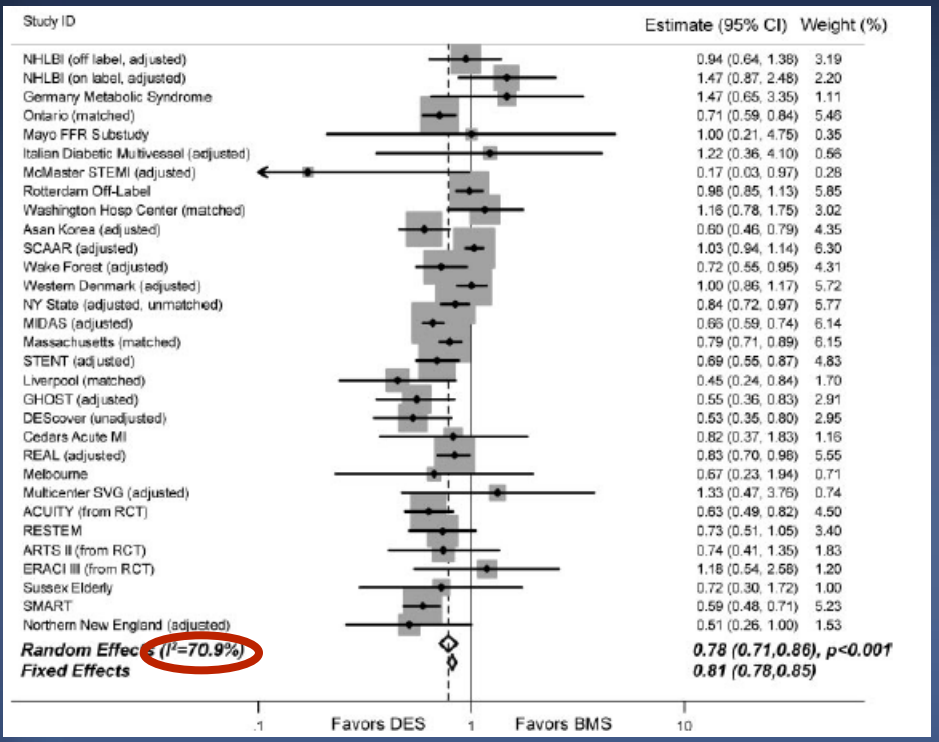
Meta-analysis of several such trials

If all the trials have the same systemic bias, all we have done is to tighten the confidence intervals!

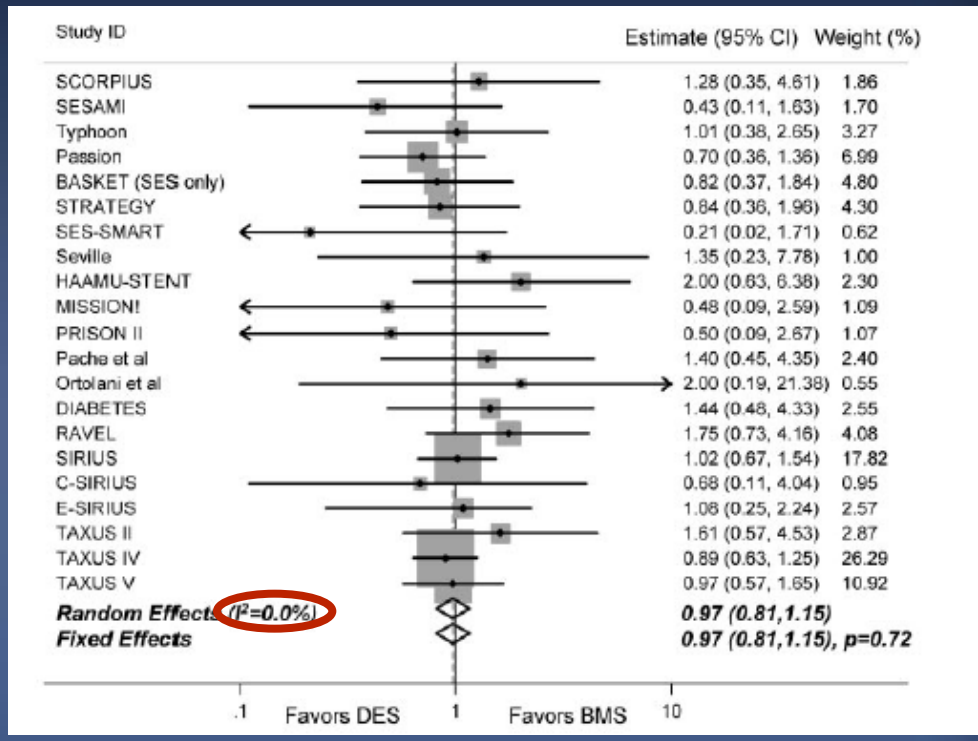


Mortality with DES vs BMS

Observational studies



RCT



Kirtane et al Circulation 2009

Type of bias	Description	Relevant domains in the Collaboration's 'Risk of bias' tool
Selection bias.	Systematic differences between baseline characteristics of the groups that are compared.	• Sequence generation; • Allocation concealment.
Performance bias.	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	• Blinding of participants, personnel and outcome assessors; • Other potential threats to validity.
Attrition bias.	Systematic differences between groups in withdrawals from a study.	• Incomplete outcome data; • Blinding of participants, personnel and outcome assessors.
Detection bias.	Systematic differences between groups in how outcomes are determined.	• Blinding of participants, personnel and outcome assessors; • Other potential threats to validity.
Reporting bias.	Systematic differences between reported and unreported findings.	• Selective outcome reporting; • (see also Chapter 10).

Meta-analysis: misleading results

- …..and the ugly

- Mixing apples with pears!

  Dfferent type of study (obs. vs RCT)

  Different clinical setting

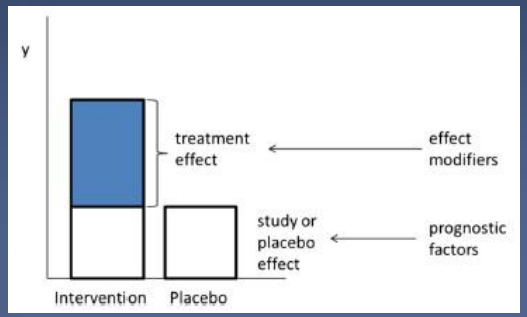
  Difference in endpoint definition

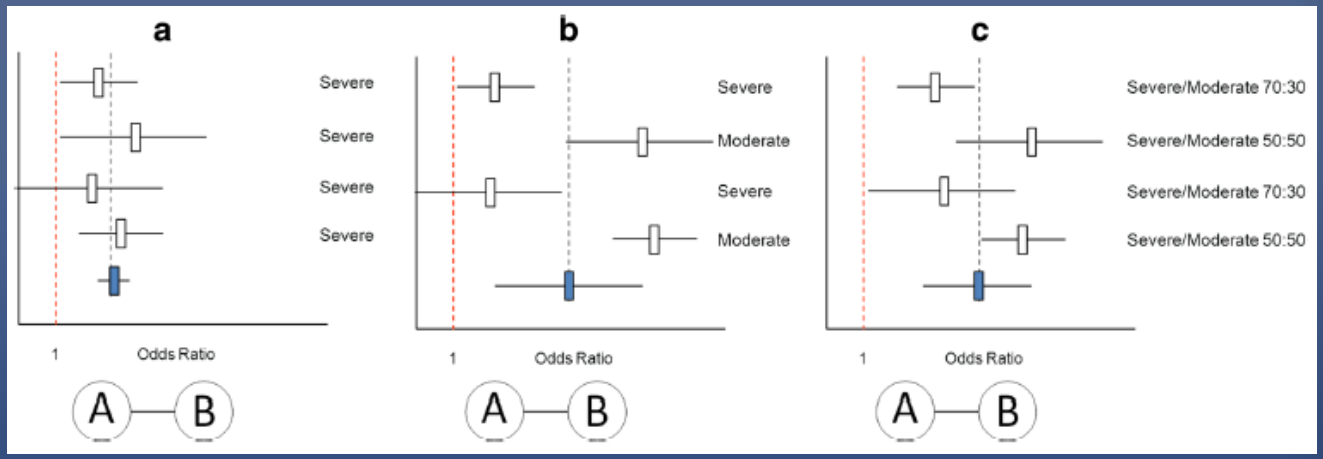
  Different follow up duration

  Different drug formulation

Exploring statistical heterogeneity

- **Statistical heterogeneity is the variation in true treatment effect among patients within a single study or between different studies**





Statistical inconsistency

- Statistical inconsistency (I^2) has been recently introduced to overcome the risk of alpha and beta error of standard tests for statistical heterogeneity
- It is computed as [(Q – df)/Q] x 100%, where Q is the chi-squared statistic and df is its degrees of freedom

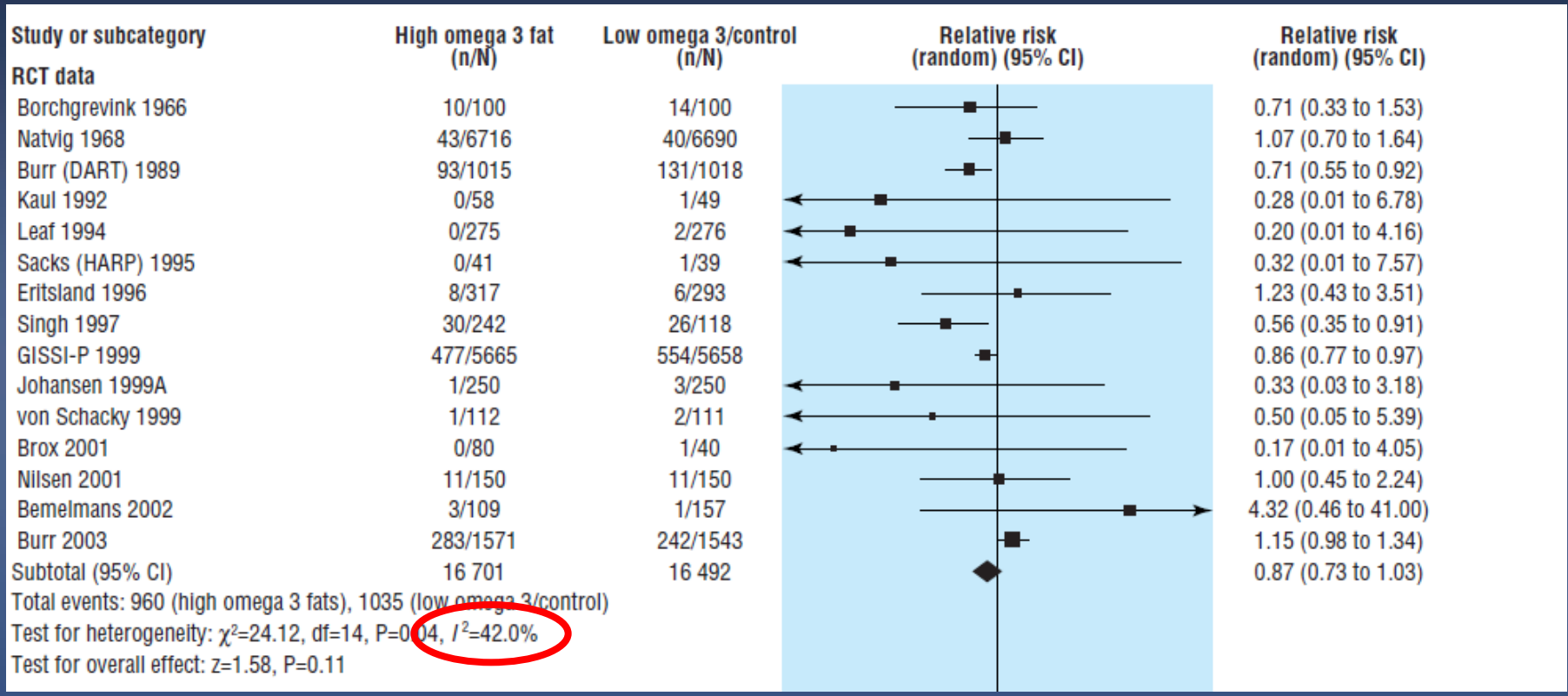
  I^2< 25%: low inconsistency

  I^2 between 25% and 50%: moderate inconsistency

  I^2 > 50%: severe inconsistency

Mortality with Omega 3 fats

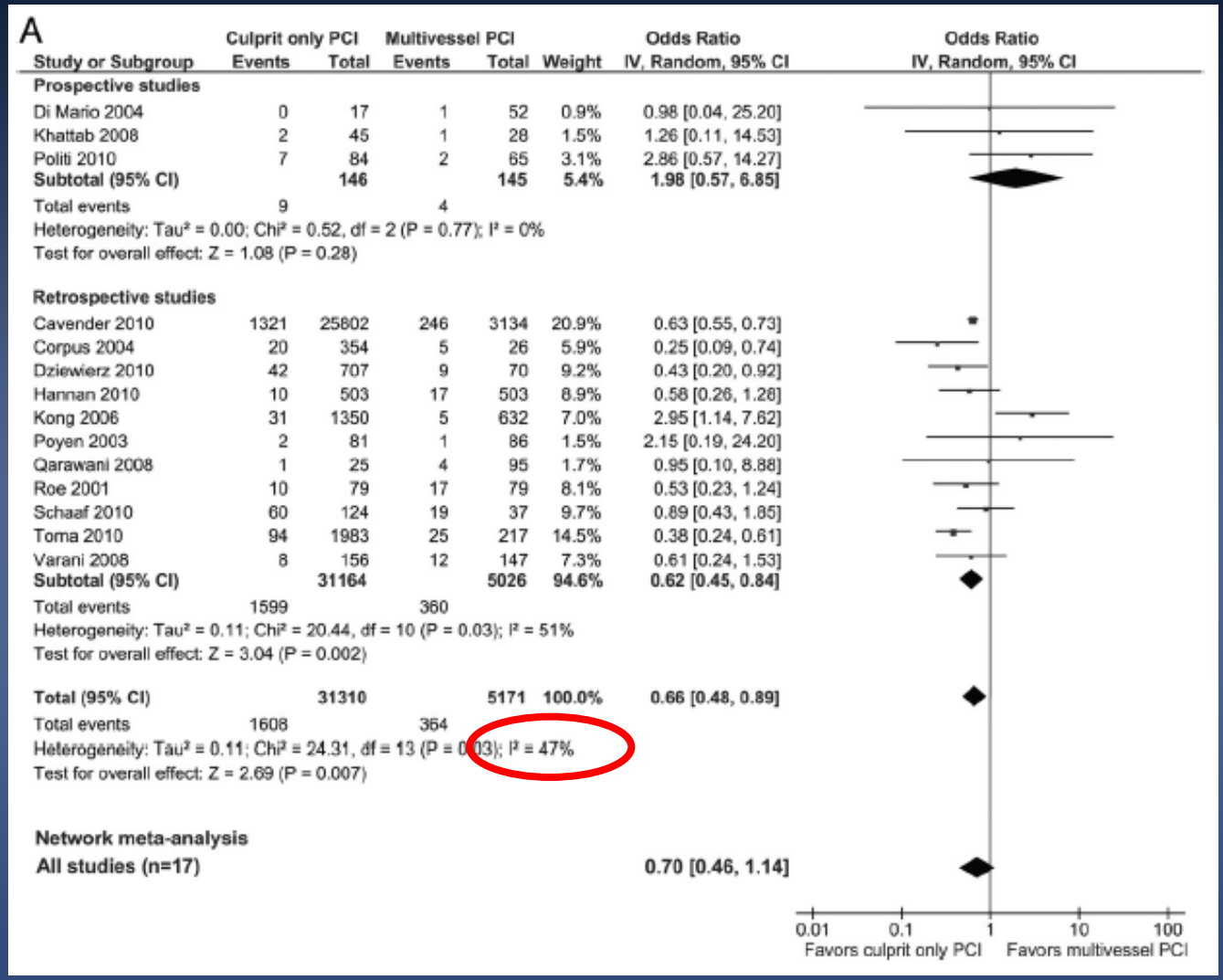
Pts with or without CAD
Purifed supplement/fish oil capsules



Study or subcategory	High omega 3 fat (n/N)	Low omega 3/control (n/N)	Relative risk (random) (95% CI)
RCT data			
Borchgrevink 1966	10/100	14/100	0.71 (0.33 to 1.53)
Natvig 1968	43/6716	40/6690	1.07 (0.70 to 1.64)
Burr (DART) 1989	93/1015	131/1018	0.71 (0.55 to 0.92)
Kaul 1992	0/58	1/49	0.28 (0.01 to 6.78)
Leaf 1994	0/275	2/276	0.20 (0.01 to 4.16)
Sacks (HARP) 1995	0/41	1/39	0.32 (0.01 to 7.57)
Eritsland 1996	8/317	6/293	1.23 (0.43 to 3.51)
Singh 1997	30/242	26/118	0.56 (0.35 to 0.91)
GISSI-P 1999	477/5665	554/5658	0.86 (0.77 to 0.97)
Johansen 1999A	1/250	3/250	0.33 (0.03 to 3.18)
von Schacky 1999	1/112	2/111	0.50 (0.05 to 5.39)
Brox 2001	0/80	1/40	0.17 (0.01 to 4.05)
Nilsen 2001	11/150	11/150	1.00 (0.45 to 2.24)
Bemelmans 2002	3/109	1/157	4.32 (0.46 to 41.00)
Burr 2003	283/1571	242/1543	1.15 (0.98 to 1.34)
Subtotal (95% CI)	16 701	16 492	0.87 (0.73 to 1.03)
Total events: 960 (high omega 3 fats), 1035 (low omega 3/control)			
Test for heterogeneity: χ^2=24.12, df=14, P=0.04, I^2=42.0%			
Test for overall effect: z=1.58, P=0.11			

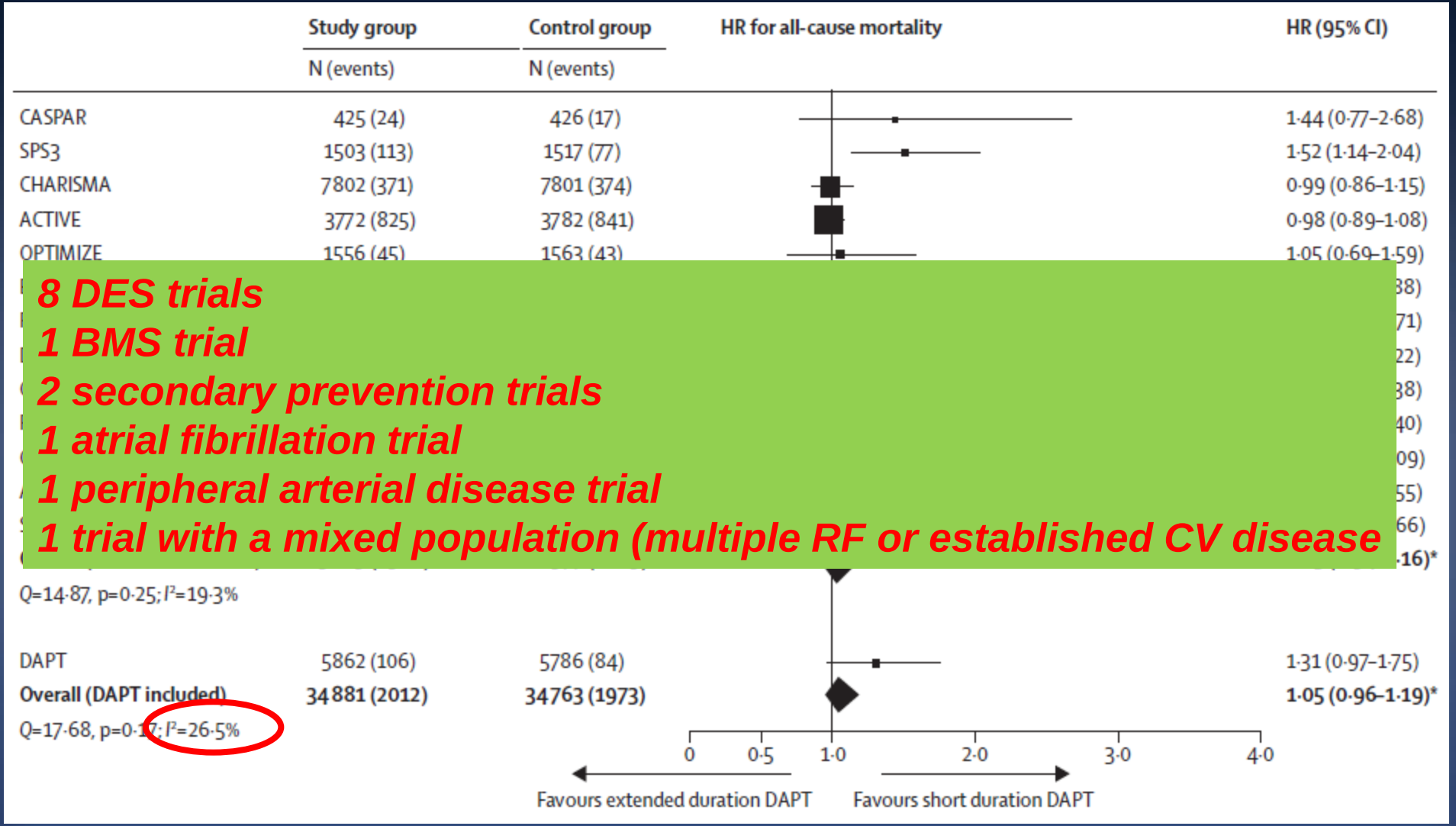
Hooper et al; BMJ 2006

Mortality in pts with STEMI: Different type of studies pooled



Vlaar et al; JACC 2011

Different population and study design



Elmariah et al; Lancet 2014

Potential problems

- **Meta-analysis is as strong as the studies that are included**

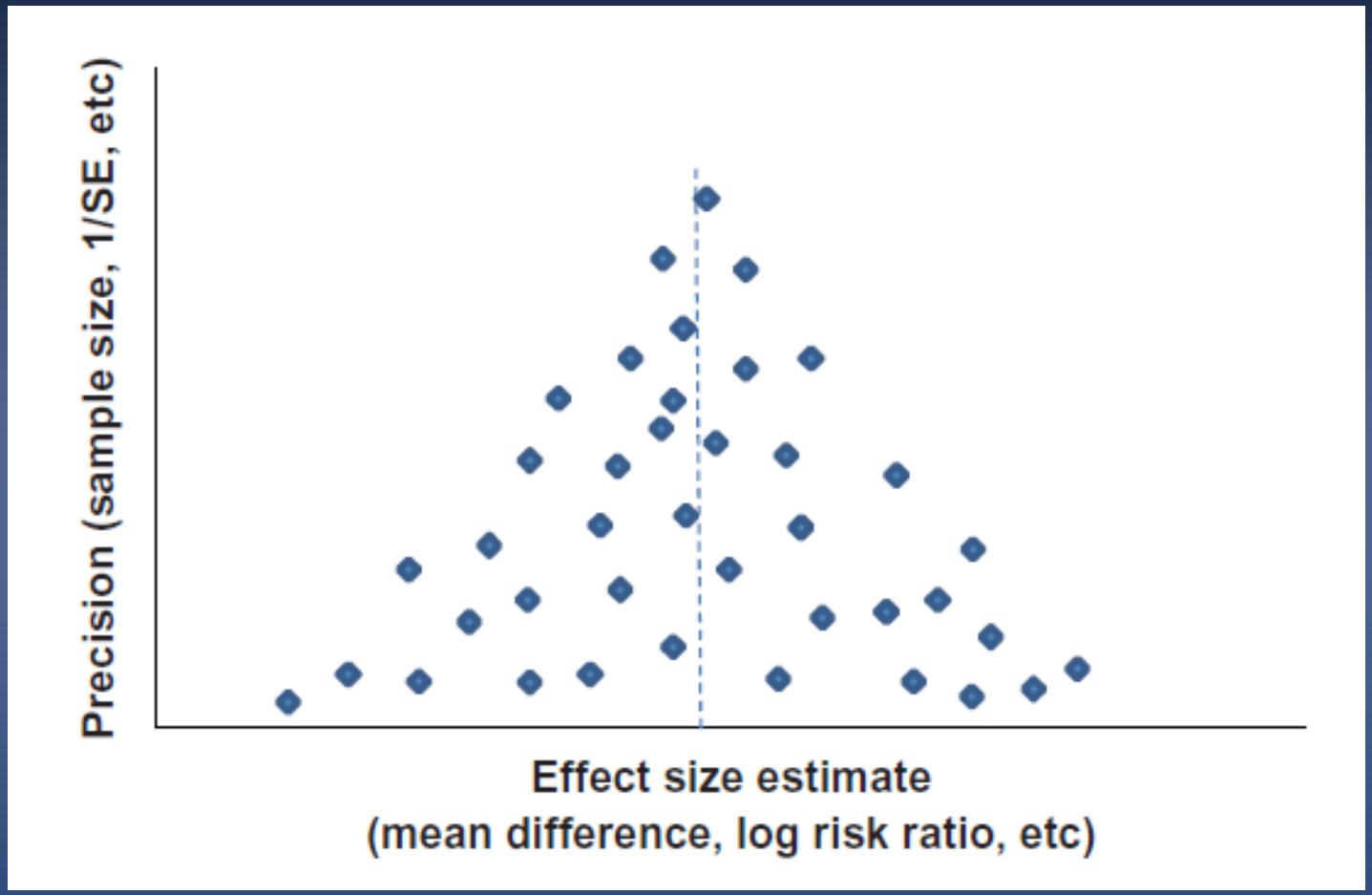
A good meta-analysis of badly designed studies will still result in bad statistics

- **The sources of bias are not controlled by the method**

What publication bias is: the file drawer problem

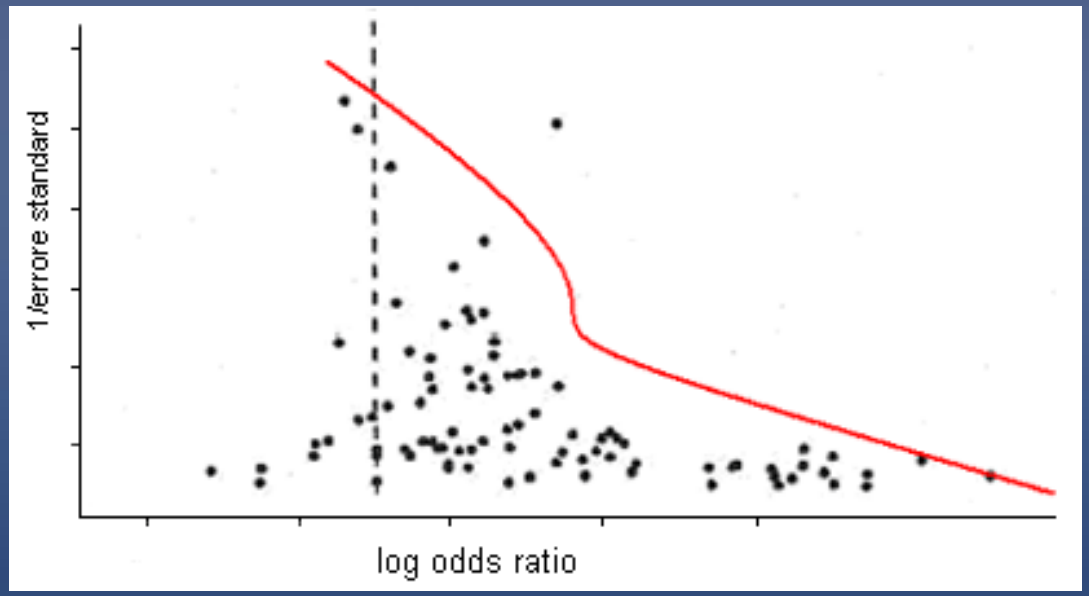
- The publication of studies depends on the nature and direction of results
- Positive studies are more likely to be published than negative studies
- The results is that published studies are systematically different from unpublished studies

Detection of publication bias: smaller studies have larger random error



Publication bias

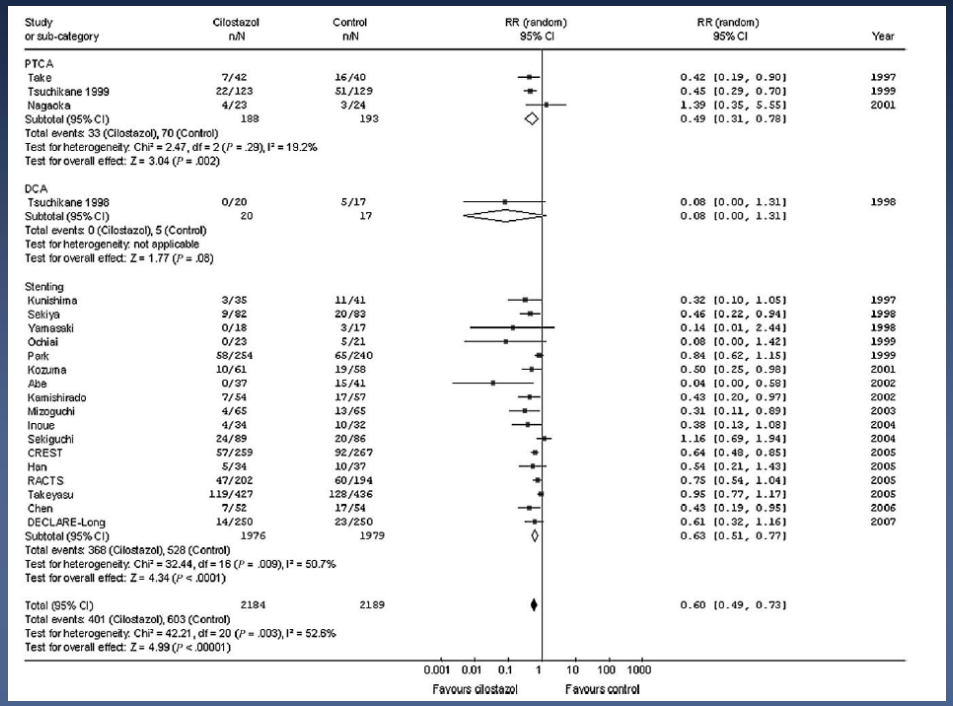
- **However, when small studies are predominately in one direction (usually the direction of larger effect sizes), asymmetry will ensue and this may be indicative of publication bias**

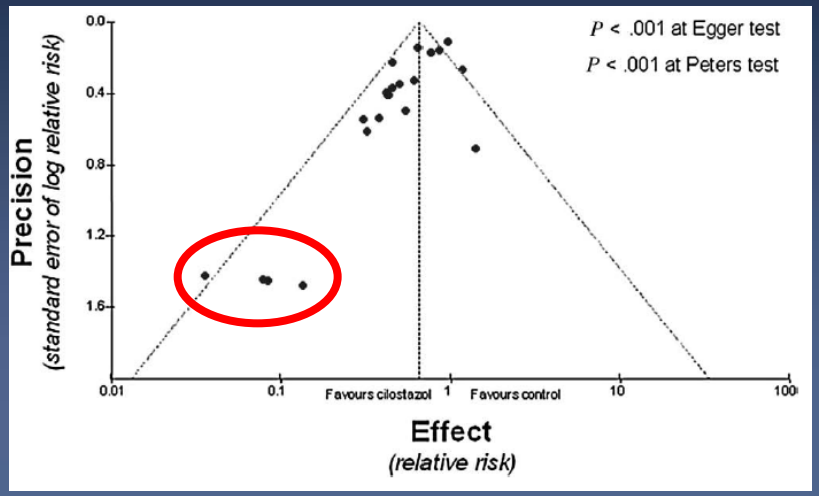


Exploring publication bias

- Begg’s test
- Egger’s test
- Meta-regression on study size

Effect of cilastazol on restenosis





Biondi Zoccai et al, AHJ 2008

Conclusions

- **The validity of a meta-analysis refers to the soundness of the original studies and the procedures used to combine them.**
- **Although several potential limitation have been identified in these procedures, systematic reviews and meta-analyses succeed when researchers implement the correct methodology and enforce sound validity checklists.**