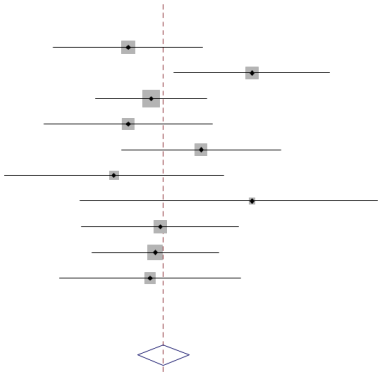
How to do a meta-analysis



Orestis Efthimiou
Dpt. Of Hygiene and Epidemiology, School of Medicine
University of Ioannina, Greece

Overview

- (*A brief reminder of…*) What is a Randomized Controlled Trial (RCT) and how to estimate treatment effects in an RCT?
- What is a meta-analysis? Why do a meta-analysis?
- What is heterogeneity, how to detect and quantify it?
- Fixed vs. Random effects meta-analysis
- When not to do a meta-analysis?
- Conclusions

Randomized Controlled Trials (RCTs)

Let’s assume we want to compare two treatment options A and B

Example: a (non-randomized) study to compare 2 interventions A and B on preventing infarction

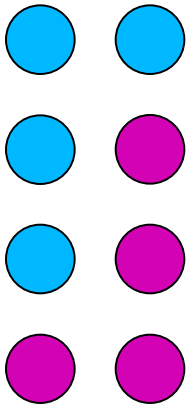
Group A

Group B

- **We give intervention A to the first group, intervention B to the second group.**
- **We compare the risk of infarction in the two groups after receiving the interventions.**

Randomization

Participants



Intervention A

Intervention B

Randomization

- By chance, all characteristics will be the same on average in the two treatment groups

- This means that the two groups we compare are similar to everything **except the treatment**

- Thus, all observed differences in the outcome will be due to treatment effects, and not due to **confounders** (such as age)

RCTs are generally considered to be the most reliable source of information regarding **relative treatment effects**

Estimating relative treatment effects from RCTs: Continuous vs. Binary outcomes

- The outcome can be continuous (e.g. change in symptoms using a scale, weight, etc.) or binary (e.g. response to treatment, remission, anything that can be measured with a Yes/No question) *

- Relative treatment effects for continuous outcomes can be measured using mean difference (and standardized mean difference)

- For binary outcomes we use risk ratio, odds ratio or risk difference

* There are also other types of outcomes (e.g. time-to-event and categorical outcomes)

Estimating relative treatment effects

A. Continuous outcomes

	Mean	Standard deviation	N
Intervention A	**4.7**	**2.1**	120
Intervention B	**2.5**	**2.7**	119

Mean difference (MD) = 2.2

Standardized Mean Difference (SMD): Is the MD divided by the standard deviation of the observations. Is useful in a meta-analysis because it can combine studies of same clinical outcome using different instruments (E.g. two different depression scales)

**Standard deviation measures the variability of individual outcomes of the included patients*

Estimating relative treatment effects

B. Binary outcomes

	response	non-response	total
Intervention A	**35**	**65**	100
Intervention B	**22**	**78**	100

Risk Ratio (RR): Probability of responding in treatment A **over** probability of responding in treatment B: (0.35/0.22=1.59)

Risk Difference (RD): Probability of responding in treatment A **minus** probability of responding in treatment B: (0.35-0.22=0.13=13%)

Odds Ratio (OR): Odds of responding in treatment A **over** odds of responding in treatment B: (35/65)/(22/78)=1.91

Estimating relative treatment effects

- ✓ The aim is to estimate the **true** relative treatment effects in the **general population of interest**
- ✓ But the RCT only includes a (small) **sample of patients**, not the general population
- ✓ Thus, we can never be sure that our estimates are correct
- ✓ This means that all estimates come with an **uncertainty**
- ✓ The **larger the sample** size of the RCT, the **smaller the uncertainty** of our estimates (usually…)

Standard error and 95% Confidence Interval

- ✓ Whenever we estimate the effect size, we must also estimate the corresponding **standard error** (SE)
- ✓ SE quantifies our uncertainty
- ✓ **Variance** is the square of the SE: $Variance=SE^2$
- ✓ Using the SE we can calculate the 95% Confidence Interval (95% CI)
- ✓ The CI gives a range of values within which we can be reasonably sure that the true effect actually lies.
- ✓ If the CI does not include the **null effect** (e.g. MD=0, OR=1, etc.) the finding is said to be "statistically significant".

Uncertainty vs. sample size

Odds Ratio

	response	non-response
A	9	18
B	4	15

1.88 (0.48, 7.32)

Statistically non-significant

	response	non-response
A	90	180
B	40	150

1.88 (1.22, 2.88)

Statistically significant

	response	non-response
A	900	1800
B	400	1500

1.88 (1.64, 2.15)

Statistically significant

Meta-analysis of RCTs

Question: is risperidone better than quetiapine for

Stefan Leucht et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis, The Lancet, Volume 382, Issue 9896

- Different RCTs may give different and often **conflicting** answers to the same question
    - Maybe due to **chance** (sampling error)?

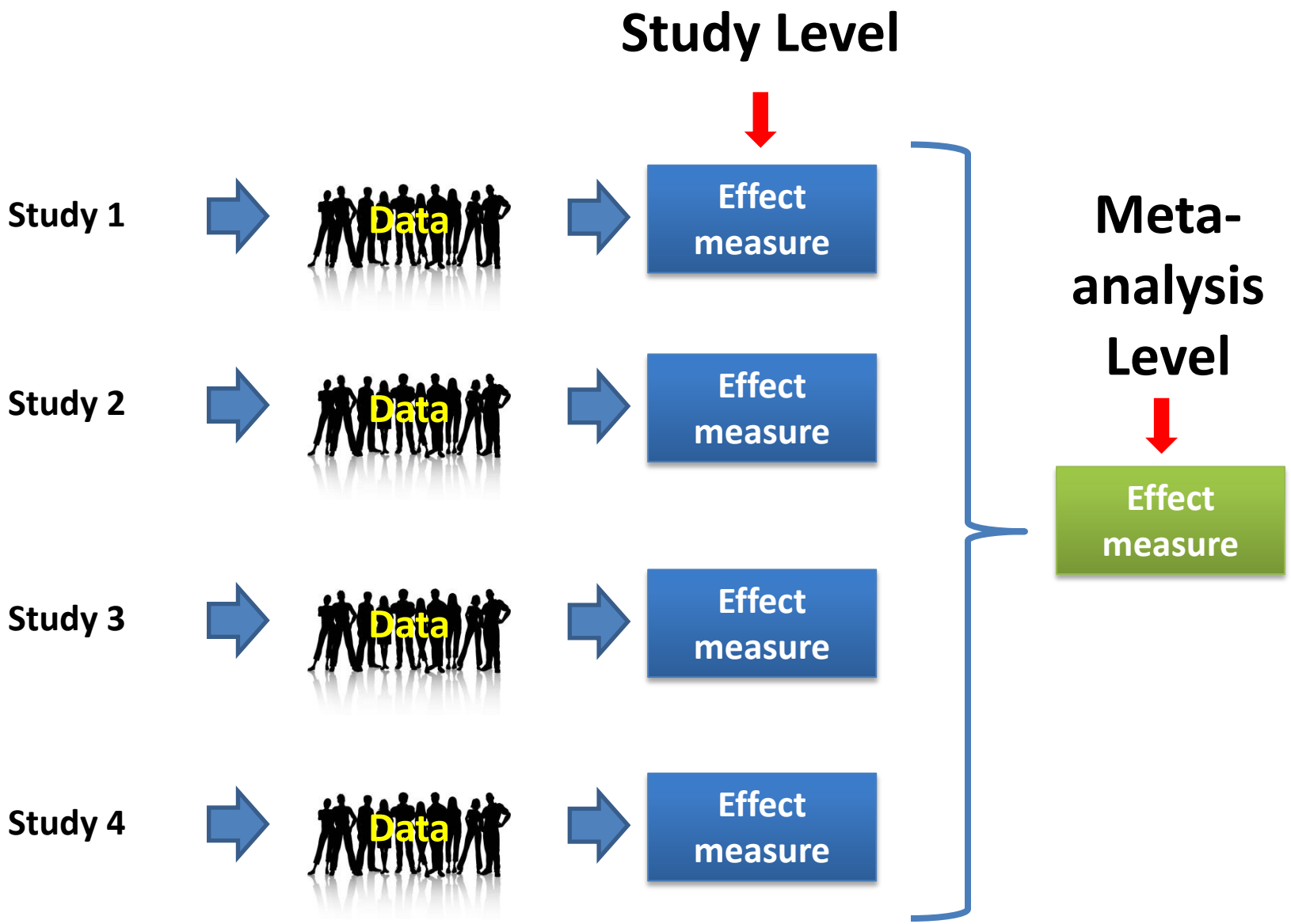
- But also maybe due to differences in populations?
    - …in interventions?
        - … in the way they measured the outcome?
            - …other reasons?

Q: How can find your way through this plethora of (conflicting) information?

Meta-analysis allows you to synthesize all this information into a meaningful answer

What is a meta-analysis?

- It is a statistical method for combining the results from two or more studies
- It allows the estimation of a ‘common’ effect size
- It is an *optional part* of a systematic review


Study Level
Study 1
Data
Effect measure
Study 2
Data
Effect measure
Study 3
Data
Effect measure
Study 4
Data
Effect measure
Meta-analysis Level
Effect measure

Why do a meta-analysis?

- To quantify treatment effects and their uncertainty
- To settle controversies between studies
- To increase power and precision
- To explore differences between studies

When can you do a meta-analysis?

- ❑ More than one study has measured an effect
- ❑ Studies are sufficiently similar
- ❑ The outcome has been measured in similar ways
- ❑ Data are available from each study

Steps in a meta-analysis

After you have identified all relevant studies:

- ✓ Identify the **outcome** you will use
- ✓ Collect the **data** from each study
- ✓ **Combine** the results to obtain a summary effect
- ✓ Explore the **differences** between the studies
- ✓ **Interpret** results

Q: is CBT effective for panic disorder in adults?

Study: Dow (2000)

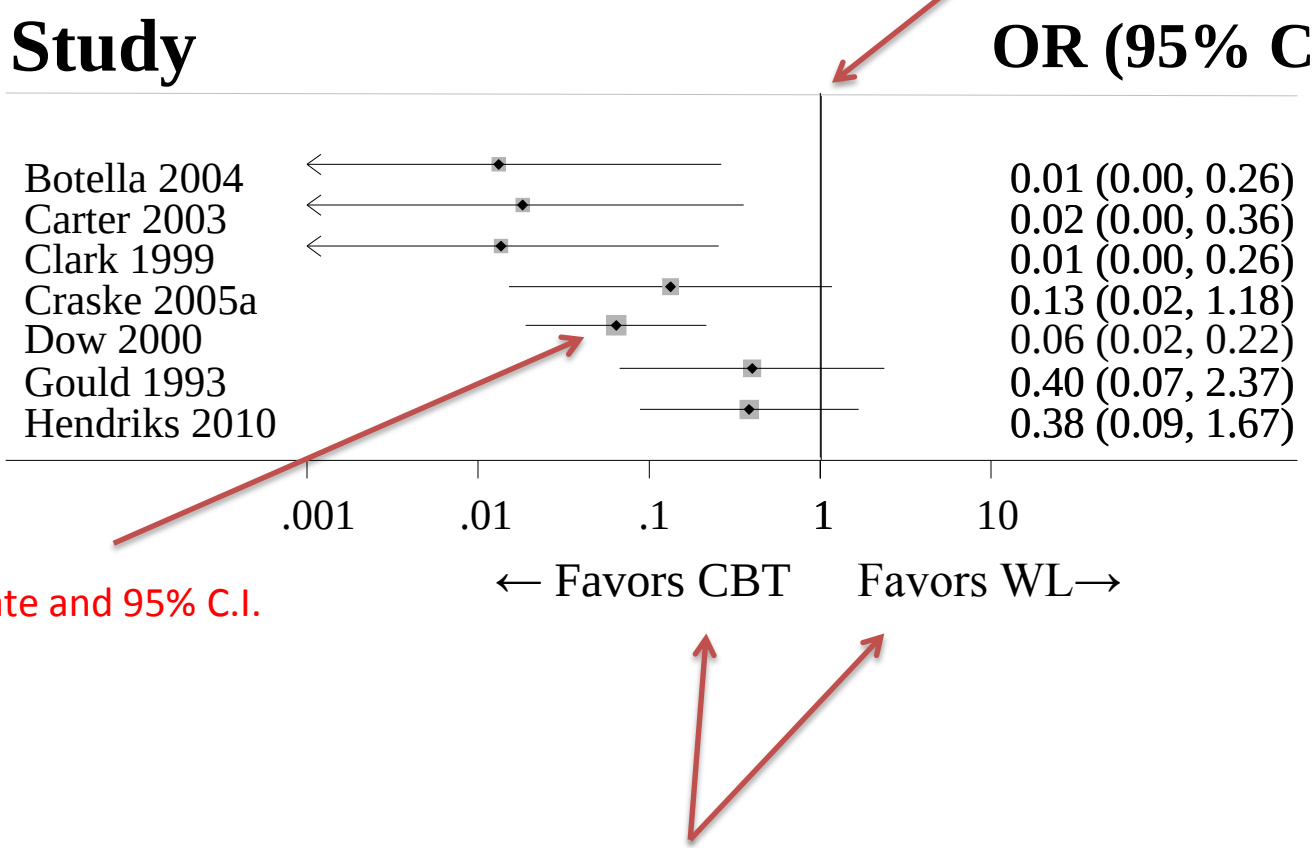
	Responders	Non-responders	Total
CBT	73	67	140
Waiting list	3	43	46

$$OR = 0.064\ (0.02, 0.22)$$

Pompoli et al. Psychological therapies for panic disorder with or without agoraphobia in adults, 2016

Q: is CBT effective for panic disorder in adults?

Line of no treatment effect (*OR = 1*)

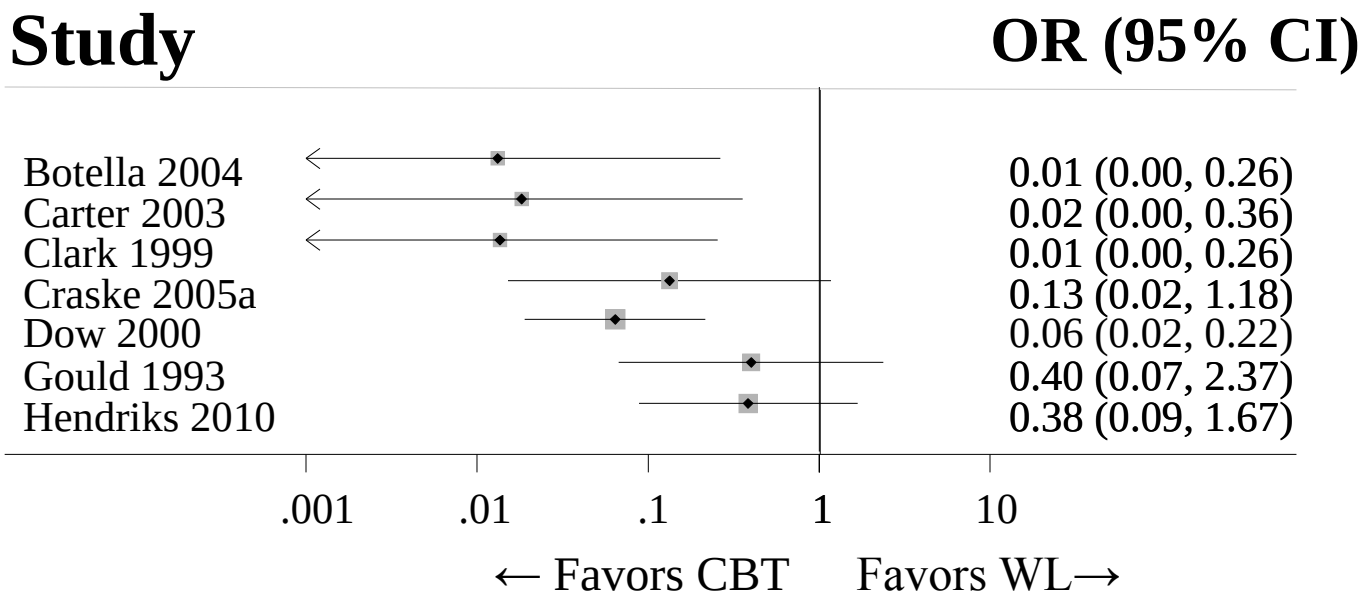


Estimate and 95% C.I.

Direction of effects

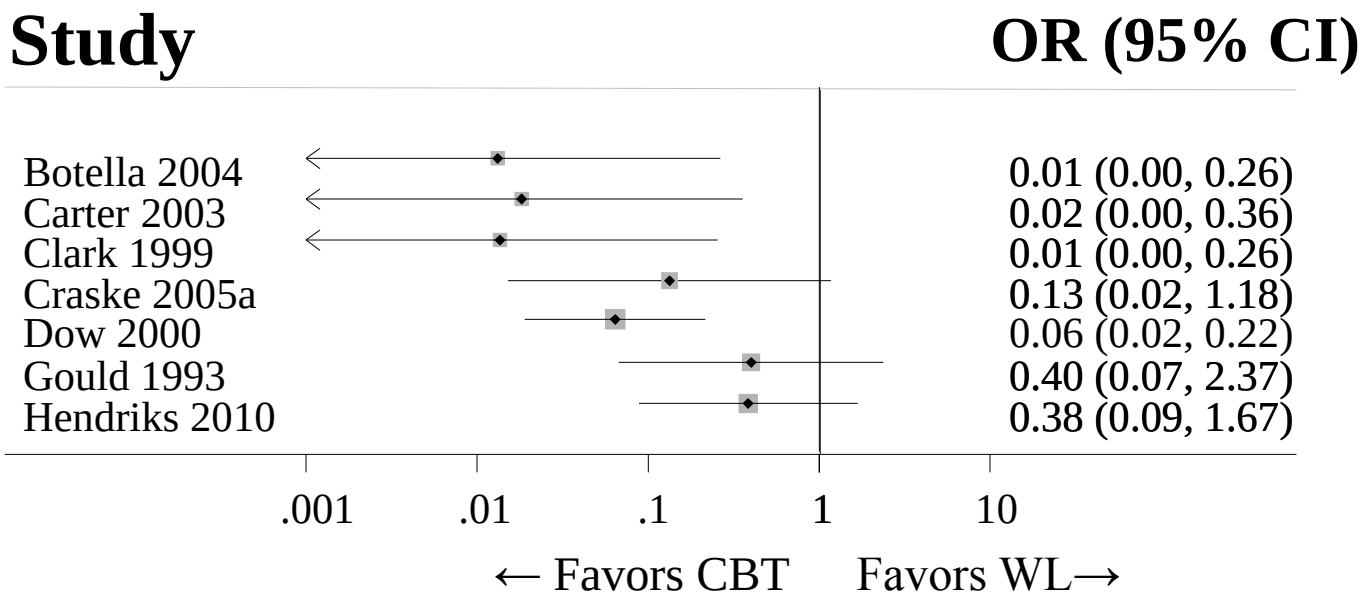
Q: is CBT effective for panic disorder in adults?

How can I synthesize this evidence?



Q: is CBT effective for panic disorder in adults?

How can I synthesize this evidence?



- **What if I just take the average of the effects across studies?**
- **This way all studies (big or small) will have the same influence on the result**

What if I pooled data in a single table and estimate the effect?

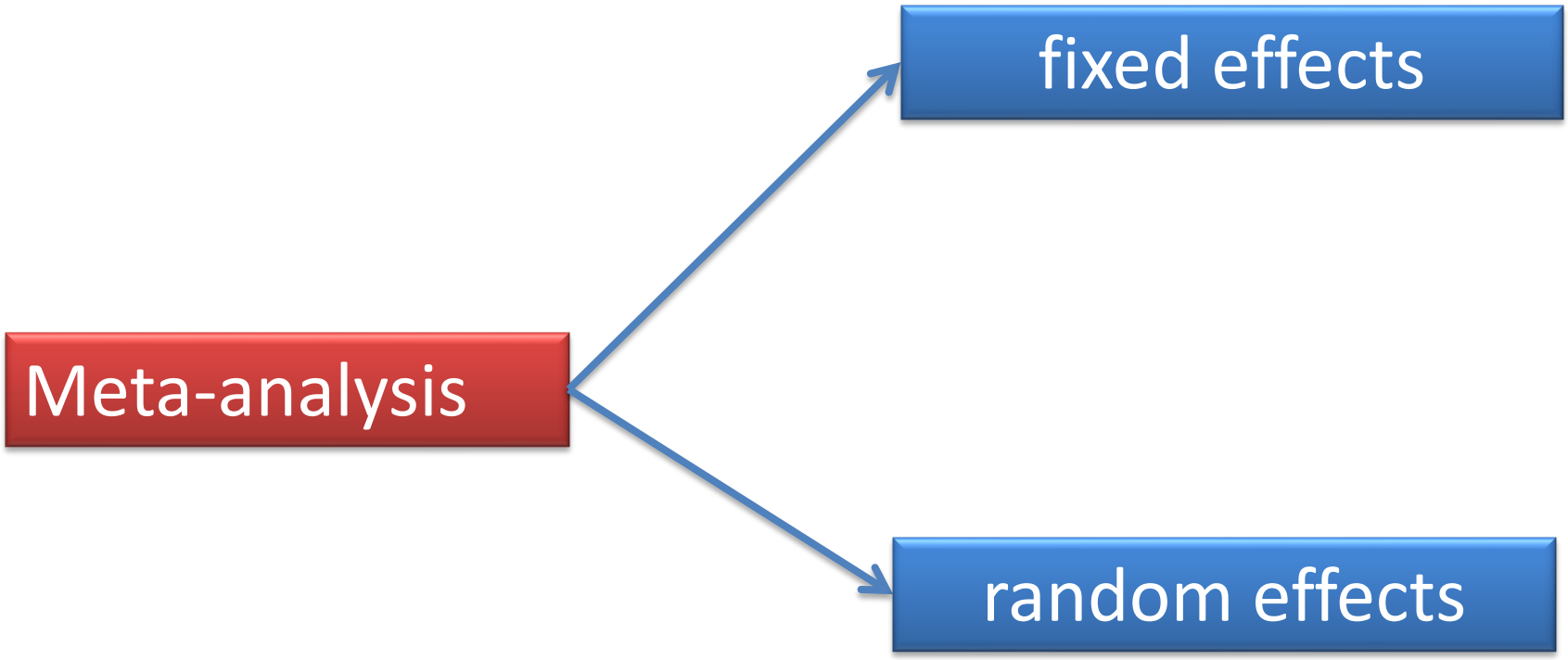
	Responders	Non-responders	Total
CBT	…	…	(All patients that received CBT in all studies)
Waiting list	…	…	(All patients that in WL in all studies)

$$OR = \cdots$$

- **This method ignores the fact that different patients come from different studies**
- **It can lead to paradoxical results and it should be avoided**

Meta-analysis principles

- We estimate the effect size in each study **separately**
- Patients from a study are **not directly compared** to patients from other studies
- We assign a **weight** to each study so that more precise studies (usually more precise=bigger) receive more weight
- We combine the estimators from the different studies in a **pooled result**


Meta-analysis
fixed effects
random effects

Fixed effects meta-analysis

The fixed effects assumption: the true treatment effect is **exactly the same** in all studies. All studies are trying to estimate this single effect.

"Under the fixed-effect model we assume that there is one true effect size […] and that all differences in observed effects are due to sampling error."

Introduction to Meta-Analysis, Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, Hannah R. Rothstein

Fixed effect meta-analysis: The inverse variance method

- In essence we calculate a **weighted average**
- From each study we have
  - The **effect size** (Mean difference, logRR, logOR etc.)
  - The **variance** of this estimate

The weight we assign to each study is inversely proportional to the variance. This way:

✓ more precise studies (smaller variance) receive larger weights

✓ Less precision→larger variance→smaller weight

CAUTION!

- For the case of binary outcomes meta-analysis using **Odds Ratio (OR)** or **Risk Ratio (RR)** we need to switch to the logarithmic scale
- We use the logOR or the logRR and the corresponding variances and not OR and RR directly!
- After the meta-analysis we can then go back to the natural scale

The inverse variance method (fixed effect)

Pooling the estimates from the different studies

$$\text{Meta-analysis estimate} = \frac{\sum(weight_i \times effect_i)}{\sum weight_i}$$

$$\text{Standard error} = \sqrt{\frac{1}{\sum weight_i}}$$

For each study i the weight is the inverse of the variance:

$$weight_i = \frac{1}{V_i}$$

The inverse variance method (fixed effect)

Pooling the estimates from the different studies

Meta-analysis estimate $= \frac{\sum(weight_i \times effect_i)}{\sum weight_i}$

e.g. Pooled logOR $= \frac{\frac{1}{V_1} * logOR_1 + \frac{1}{V_2} * logOR_2 + \cdots}{\frac{1}{V_1} + \frac{1}{V_2} + \ldots}$

How to do a meta-analysis?

There are many software options available:

- RevMan (by the Cochrane Collaboration, freely available at http://tech.cochrane.org/revman)
- R (packages epiR, meta, metafor, and rmeta).
- Stata (metan command)
- Other commercial programs
- …

example of fixed-effects meta-analysis in Stata (`metan` command)

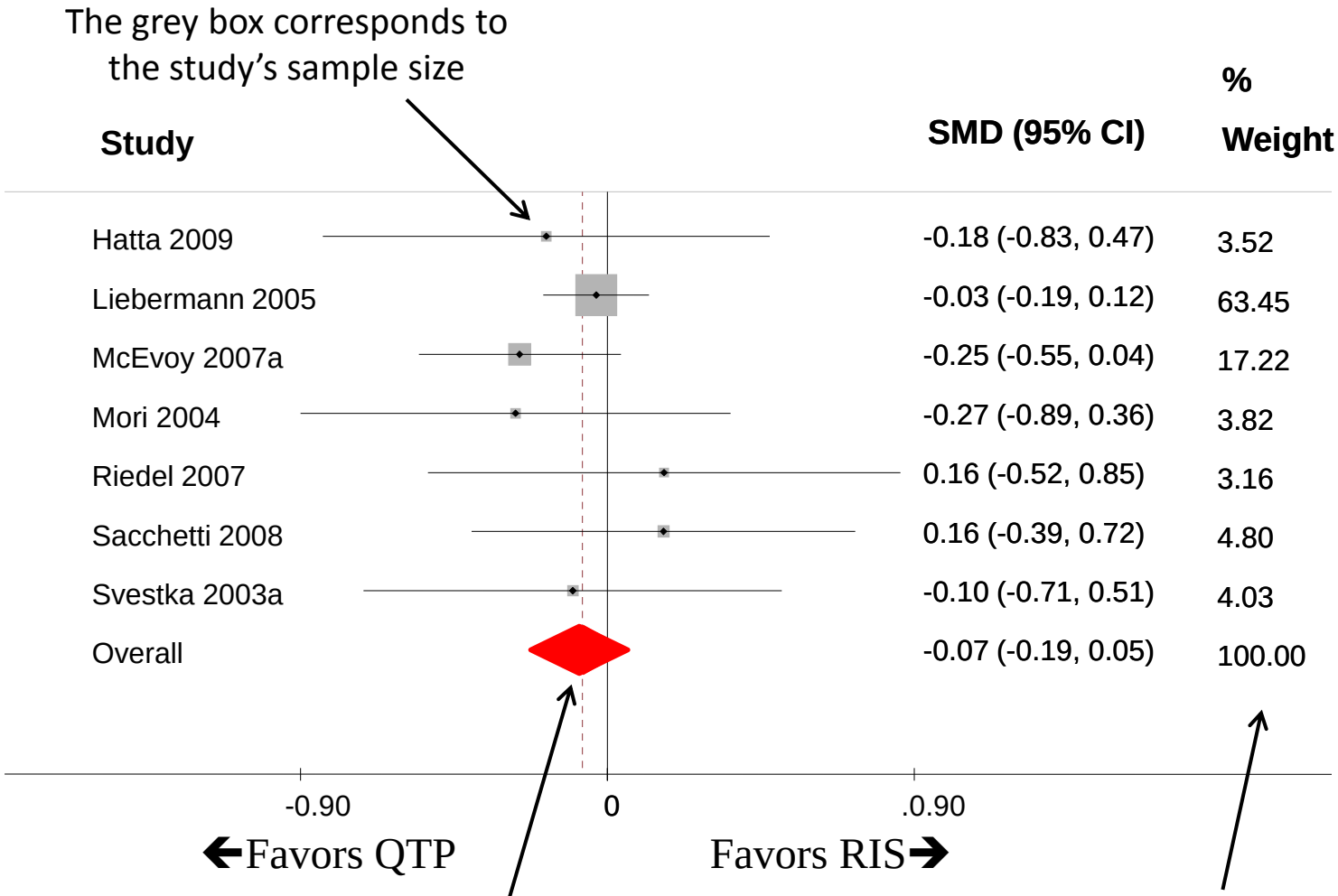
Olanzapine vs. Quetapine for schizophrenia

Study	m1	SD1	n1	m2	SD2	n2
Hatta 2009	-33.4	20.8	17	-28.9	28.6	20
Liebermann 2005	-4.8	21.45854	330	-4.1	21.45854	329
McEvoy 2007a	-14.3	10.33	85	-11.6	10.88	96
Mori 2004	69.4	10.8	20	72.9	15.1	20
Riedel 2007	-17.88	20.71	17	-21.5	23.39	16
Sacchetti 2008	-33.5	16	25	-36.4	19.6	25
Svestka 2003a	-45.65	11.96	20	-43.91	20.94	22

```
metan n1 m1 SD1 n2 m2 SD2, fixed lcols(Study)
```

Stefan Leucht et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis, The Lancet, Volume 382, Issue 9896

(Fixed effects) meta-analysis and **forest plot**



Random effects meta-analysis

The random effects assumption: the true treatment effect is **not the same** in all the studies.

"… under the random-effects model we allow that the true effect could vary from study to study. For example, the effect size might be higher (or lower) in studies where the participants are older, or more educated, or healthier than in others, or when a more intensive variant of an intervention is used…"

Introduction to Meta-Analysis, Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, Hannah R. Rothstein

Random effects meta-analysis

The variation in the **true effects** underlying the studies of a review is called **heterogeneity**

You might have heterogeneity due to:

- ***Differences in patients' characteristics across studies***
  - e.g. differences in mean age: studies performed in younger patients may show different results than studies in older patients; differences in the severity of illness etc.
- ***Interventions defined differently across studies***
  - e.g. intensity / dose / duration, sub-type of drug, mode of administration, experience of practitioners, nature of the control (placebo/none/standard care) etc.

Random effects meta-analysis

The variation in the **true effects** underlying the studies of a review is called **heterogeneity**

You might have heterogeneity due to:

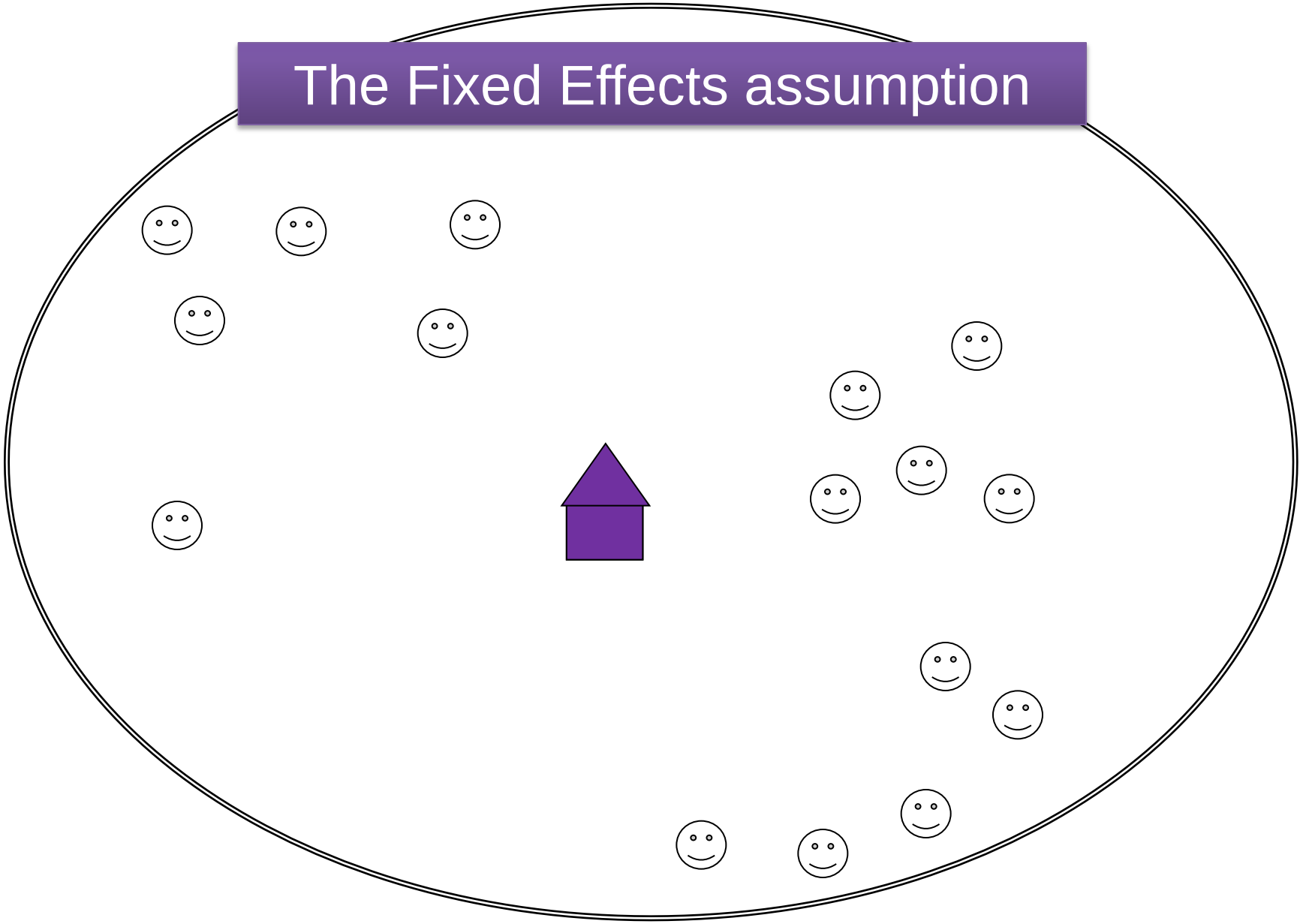
- ***Conduct of the studies***
  - e.g. allocation concealment, blinding etc., approach to analysis, imputation methods for missing data
- ***Definition of the outcome***
  - e.g. definition of an event, follow-up duration, ways of measuring outcomes, cut-off points on scales

Random effects meta-analysis

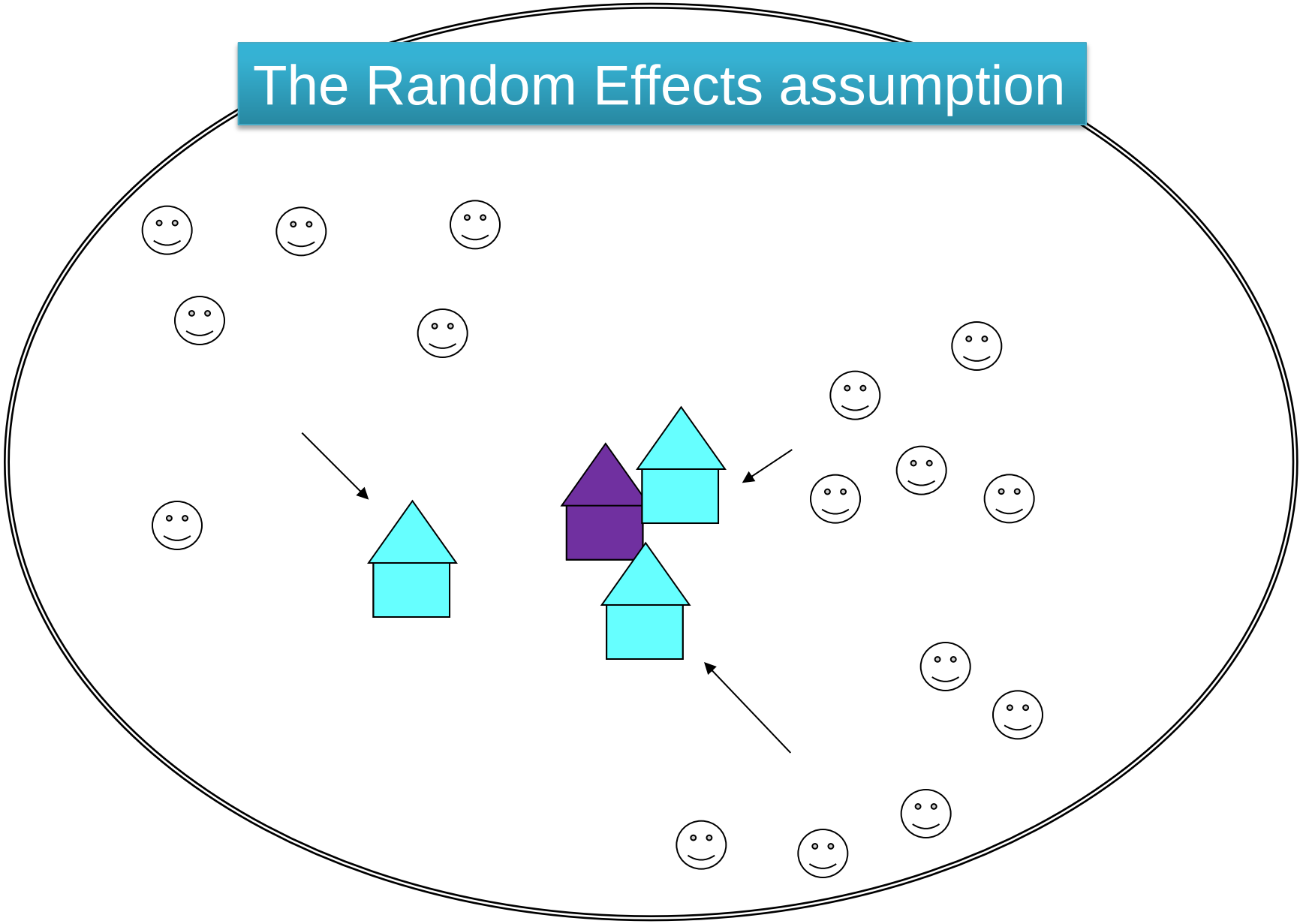
- Heterogeneity suggests that the studies have **important underlying differences**.
- We can allow the true effects underlying the studies to differ.
- We assume the true effects underlying the studies follow a distribution.
  - conventionally a normal distribution
- It turns out that we can use **a simple adaptation** of the inverse-variance weighted average.

DerSimonian and Laird (1986)

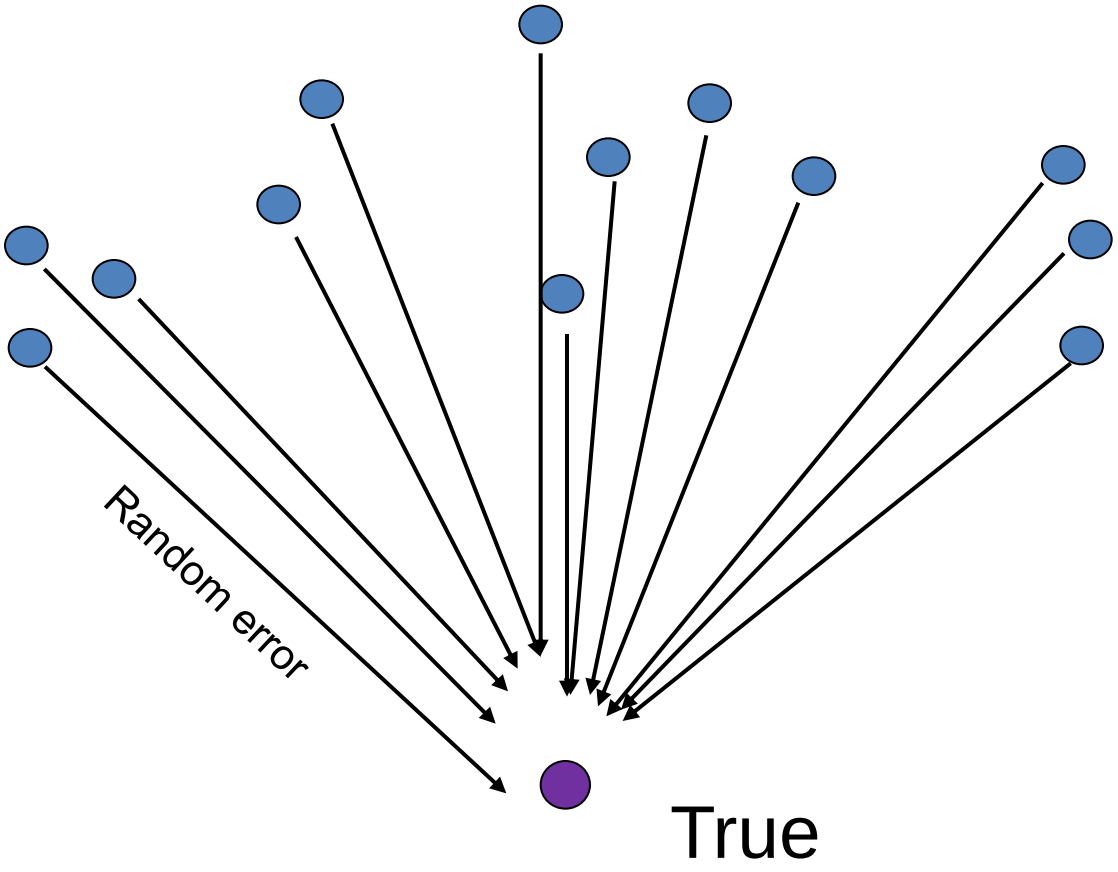
The Fixed Effects assumption



The Random Effects assumption

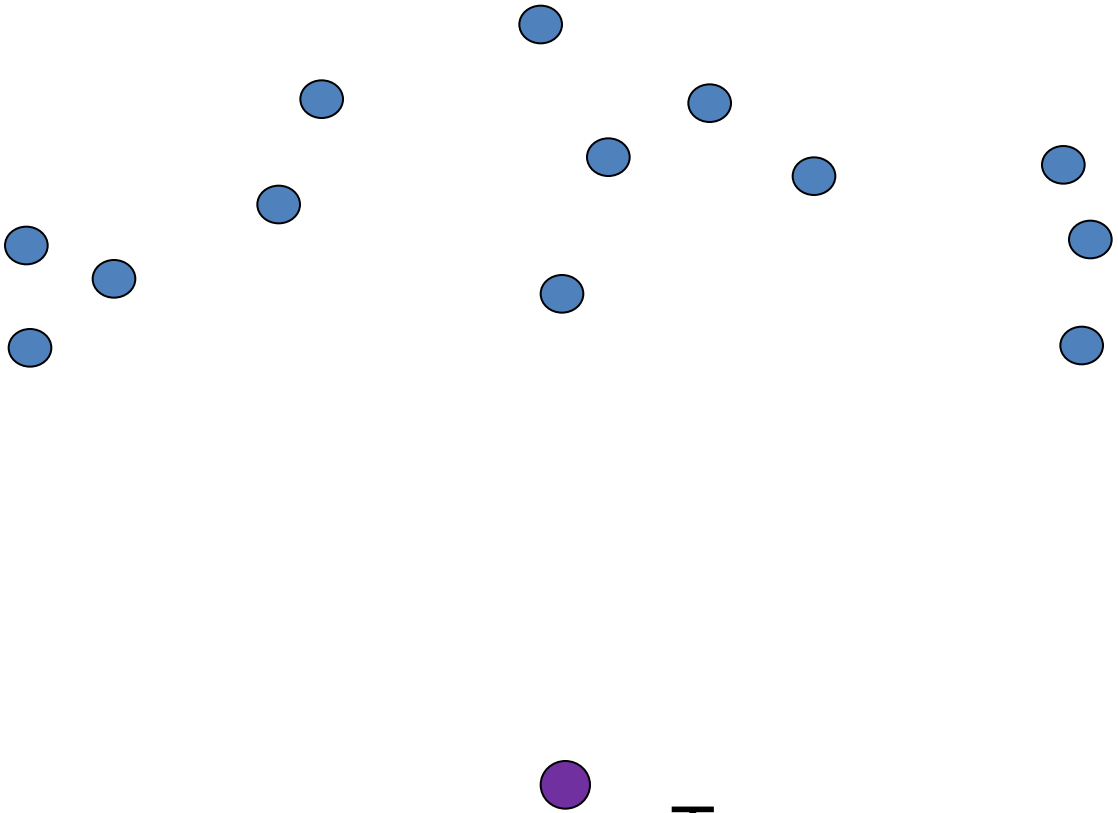


The Fixed Effects assumption

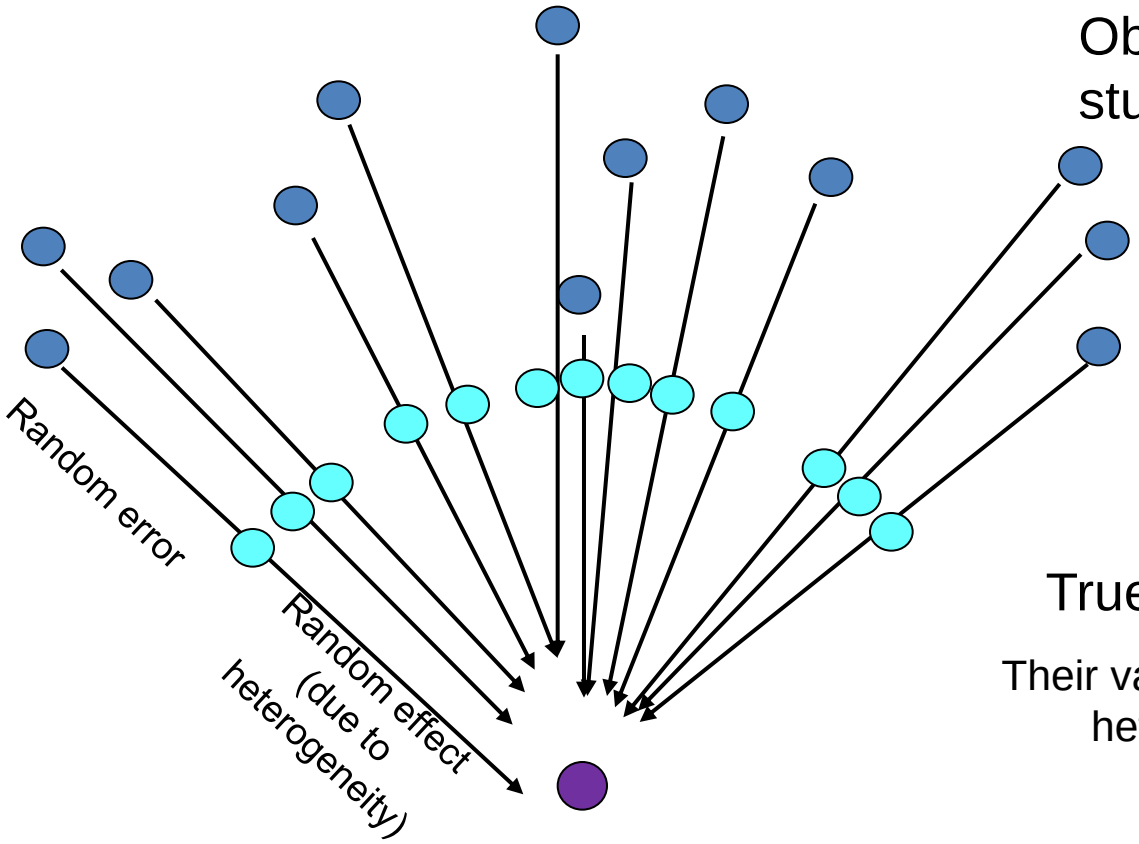


Observed in studies

If we could increase precision of all studies indefinitely (no random error)…

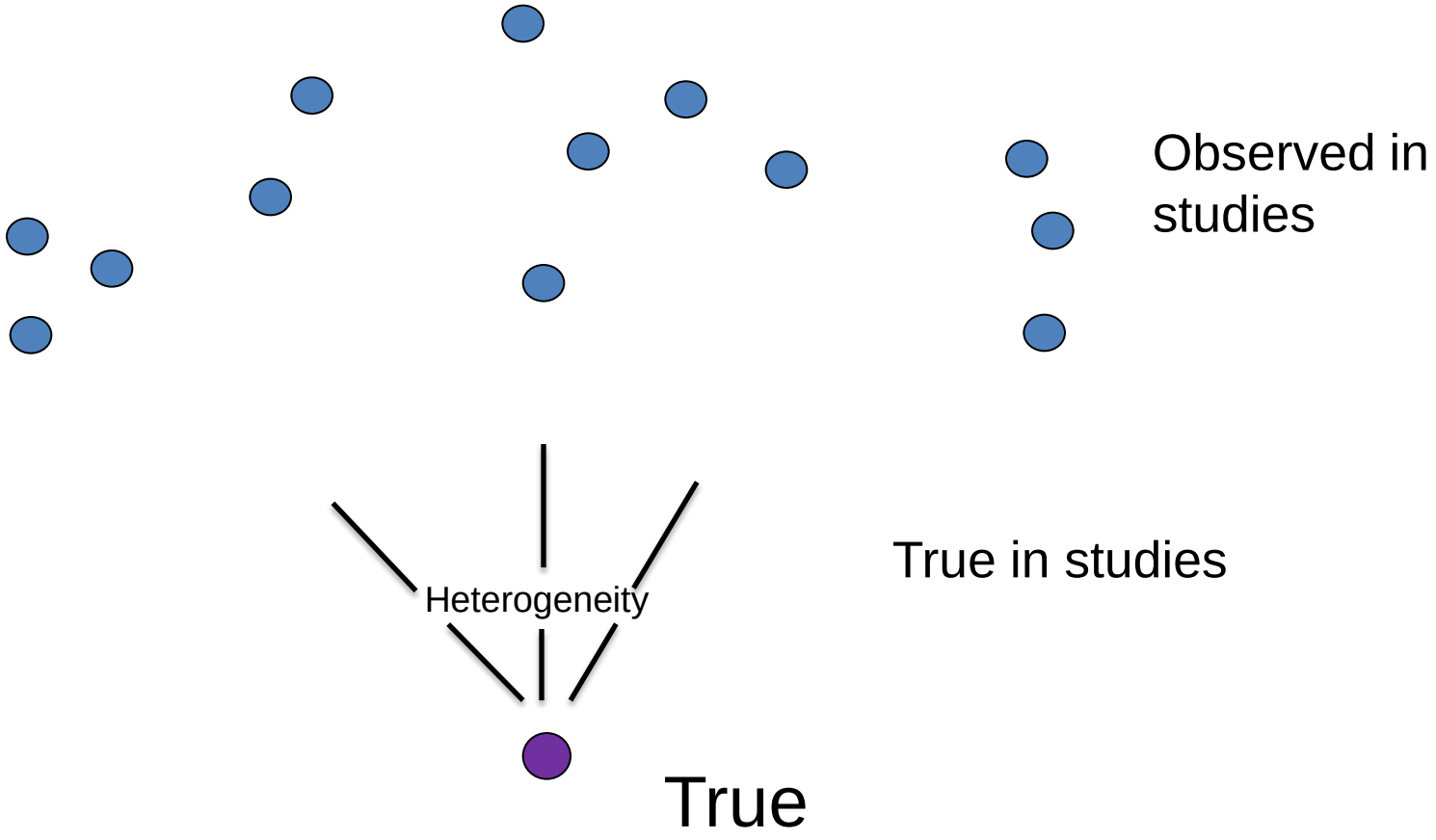


The Random Effects assumption

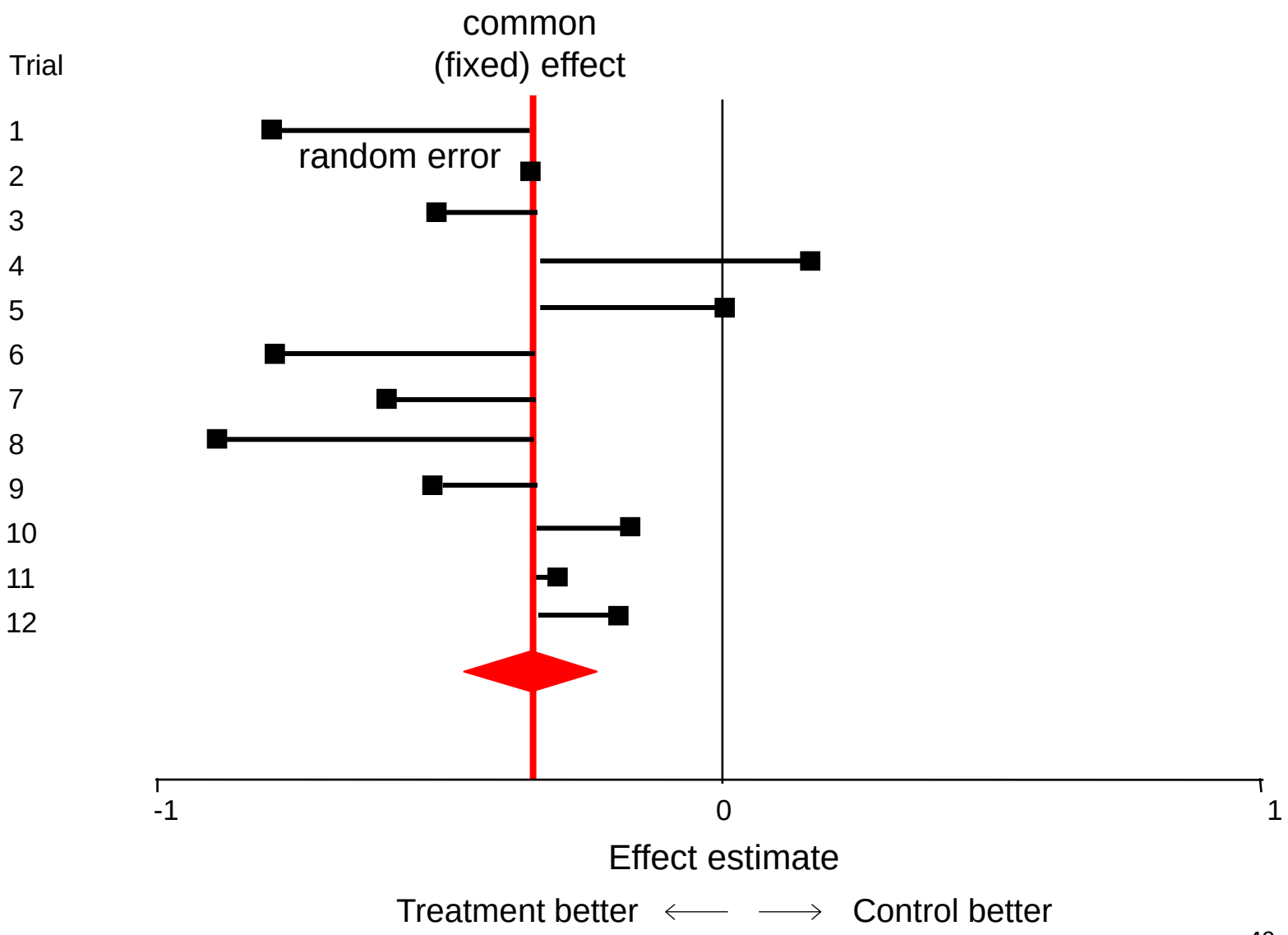


The Random Effects assumption

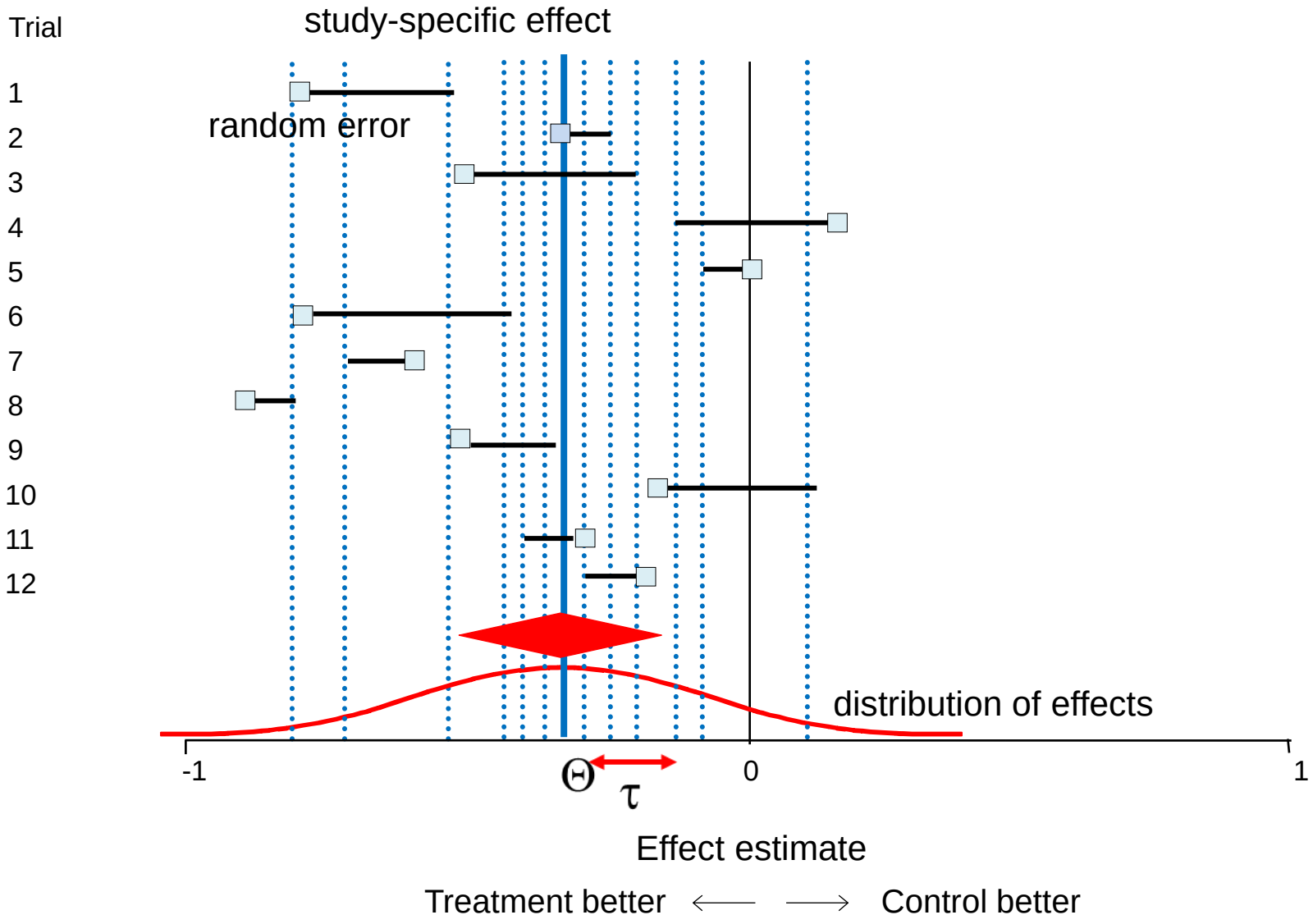
If we could increase precision of all studies indefinitely (no random error)…



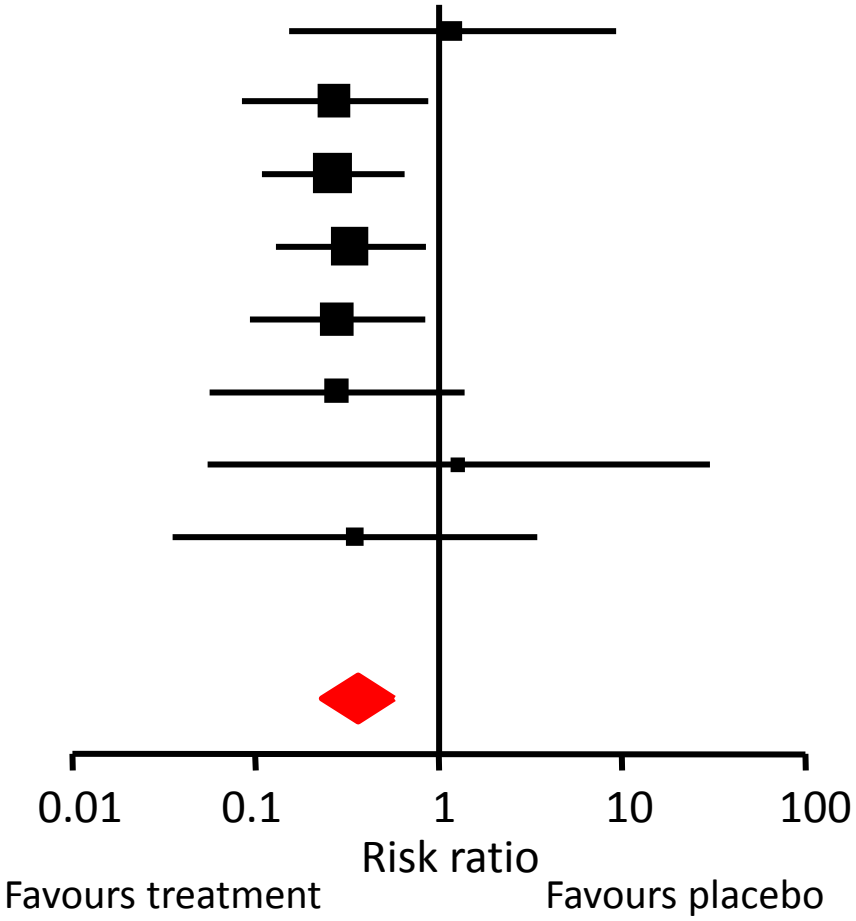
Fixed effect meta-analysis

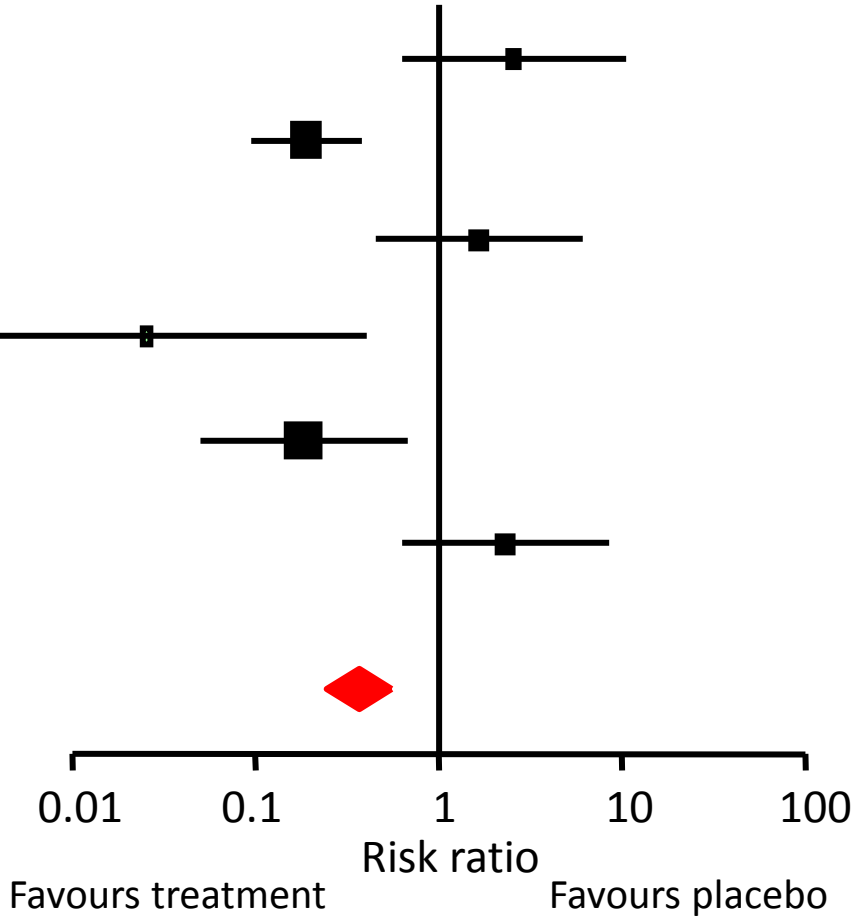


Random effects meta-analysis



Identifying heterogeneity: eyeballing





The lack of overlap in the CI's suggests the presence of heterogeneity

Identifying heterogeneity: the Q test

The Q test uses a χ^2 (chi-squared) distribution and can provide a yes-no answer to whether or not there is significant heterogeneity, but:

- Has low power since there are usually very few studies, i.e. test is not very good at detecting heterogeneity as statistically significant when it exists

- Has excessive power to detect clinically unimportant heterogeneity when there are many studies

Cochrane Handbook advises

*'… since clinical and methodological diversity **always occur** in a meta-analysis, statistical heterogeneity is inevitable (Higgins 2003). Thus the test for heterogeneity is irrelevant to the choice of analysis**; heterogeneity will always exist** whether or not we happen to be able to detect it using a statistical test.'*

Quantifying heterogeneity: the I^2 Statistic

- The Q-test is not asking a useful question if heterogeneity is inevitable

The I-square measure for heterogeneity

I^2 describes the proportion of variability that is due to heterogeneity rather than sampling error

Higgins and Thompson (2002)

Identifying heterogeneity
I^2 Statistic

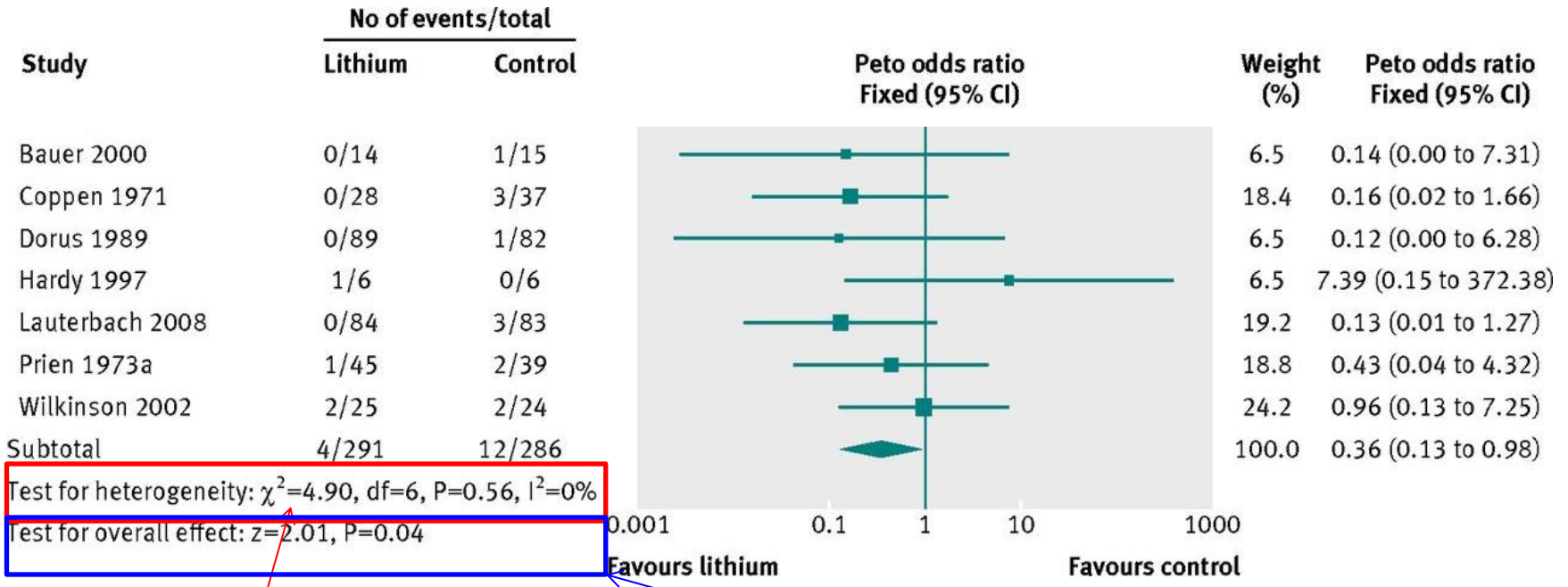
Interpreting I^2 (a rough guide*)

- 0% to 40% might not be important
- 30% to 60% may represent moderate heterogeneity
- 50% to 90% may represent substantial heterogeneity
- 75% to 100%: considerable heterogeneity

*depending on the magnitude and the direction of the effects and the strength of evidence.

Higgins and Thompson (2002)

Example: Lithium vs. placebo in the prevention of suicide mood disorders

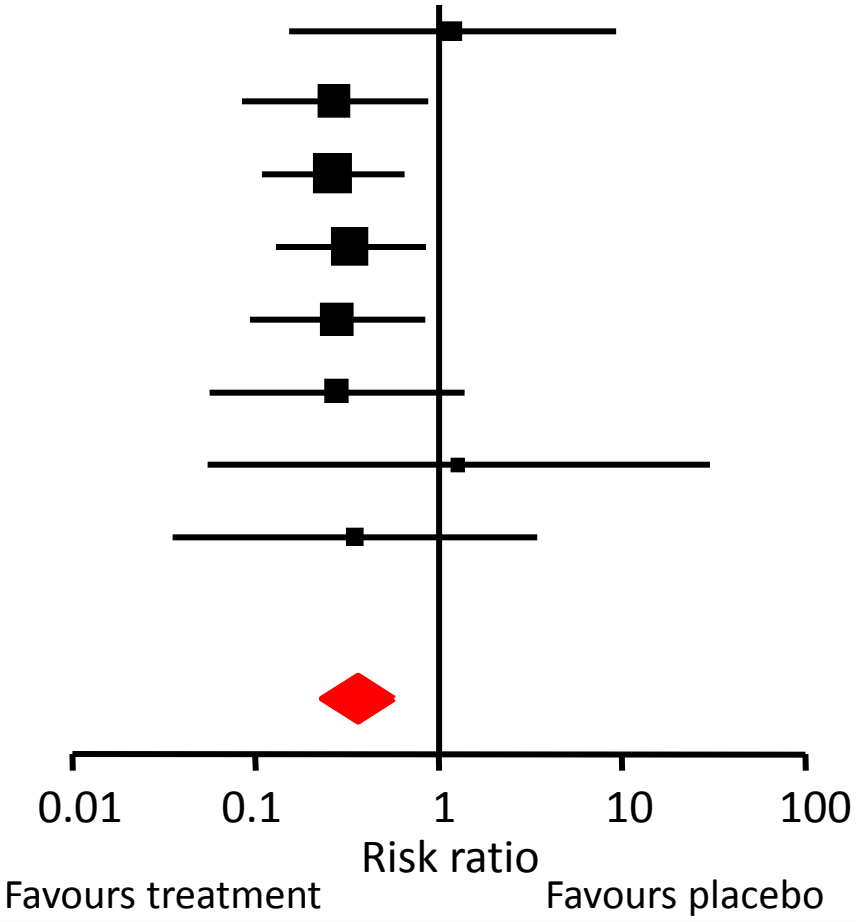


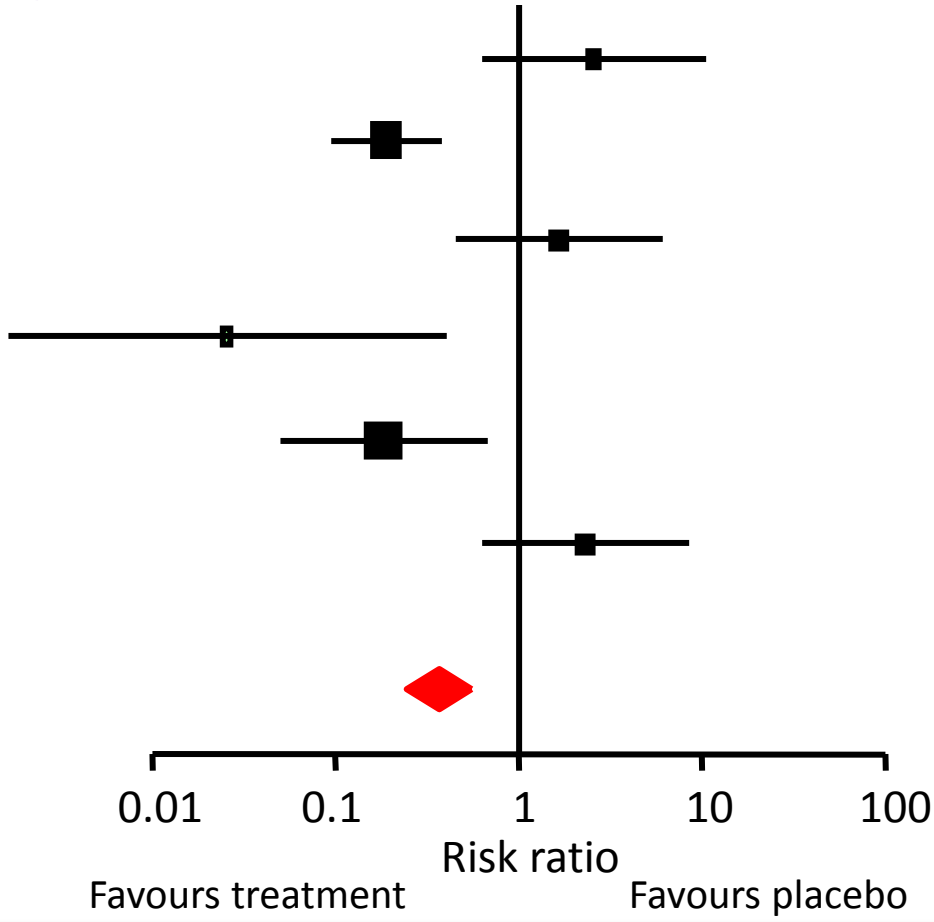
χ^2 and df correspond to the Q test. P is the p-value of the Q test.

This corresponds to the meta-analysis pooled effect

Lithium in the prevention of suicide in mood disorders: updated systematic review and meta-analysis, Cipriani et al. BMJ. 2013 Jun 27;346:f3646. doi: 10.1136/bmj.f3646.

Example: two fixed effects meta-analyses giving the same result





How to take account of heterogeneity into our pooled result?

Random effects meta-analysis model

- We use a **simple extension** of the inverse variance method, by taking into account the variance of the random effects τ^2.

Three steps:

1. Estimate τ^2 (also called the heterogeneity parameter)
2. Re-define the weights w_i^*
3. Estimate the pooled treatment effect and its variance using the weights new w_i^*

Random effects

- We incorporate the heterogeneity parameter in the study weights:

Fixed Effect Weights

$$w_i = \frac{1}{V_i}$$

Random Effects Weights

$$w_i^* = \frac{1}{V_i + \tau^2}$$

where V_i is the variance in study *i*

Random effects: estimation
Step 3: Calculate the pooled estimate

$$\Theta = \frac{\sum w_i^* y_i}{\sum w_i^*}$$

$$SE(\Theta) = \sqrt{\frac{1}{\sum w_i^*}}$$

where

$$w_i^* = \frac{1}{V_i + \tau^2}$$

Example: Five studies comparing Ziprasidone vs. Placebo for acute mania

StudyID	SMD	sd
10	-0.374	0.152
11	-0.501	0.153
12	-0.108	0.142
14	0.097	0.099
54	-0.403	0.131

Performing the meta-analysis in Stata:

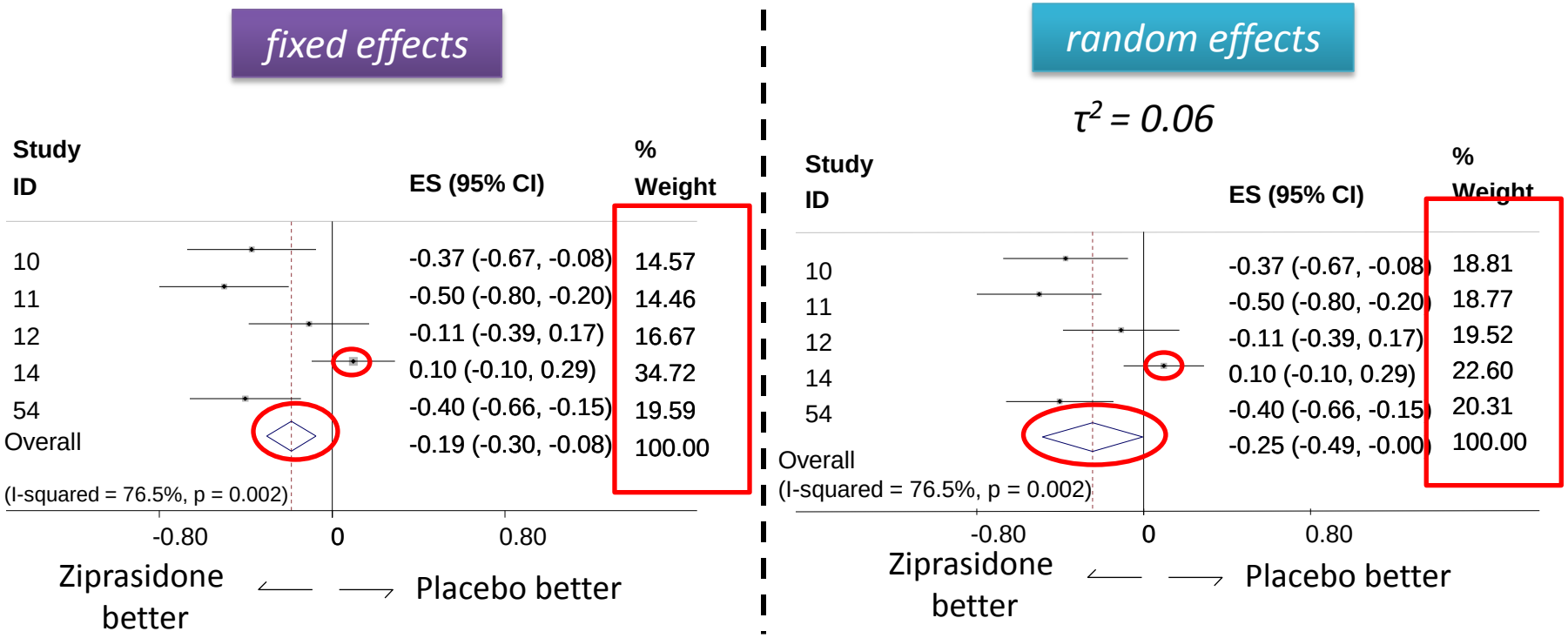
Fixed effects: `metan SMD sd`

Random effects: `metan SMD sd, randomi`

Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. Cipriani et al. Lancet. 2011 Oct 8;378(9799):1306-15.

Fixed vs. Random effects meta-analysis:
Find the differences!

Ziprasidone vs. Placebo for acute mania



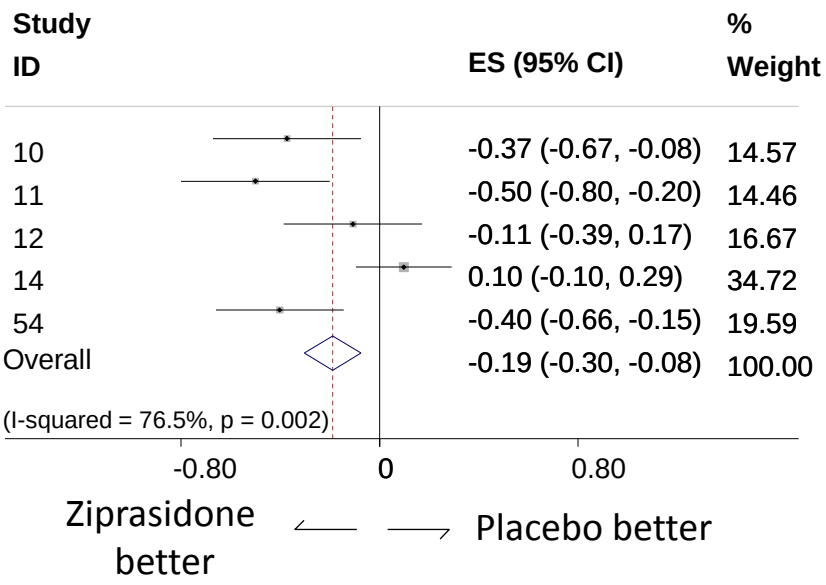
*analysis performed in Stata using the metan command

Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. Cipriani et al. Lancet. 2011 Oct 8;378(9799):1306-15.

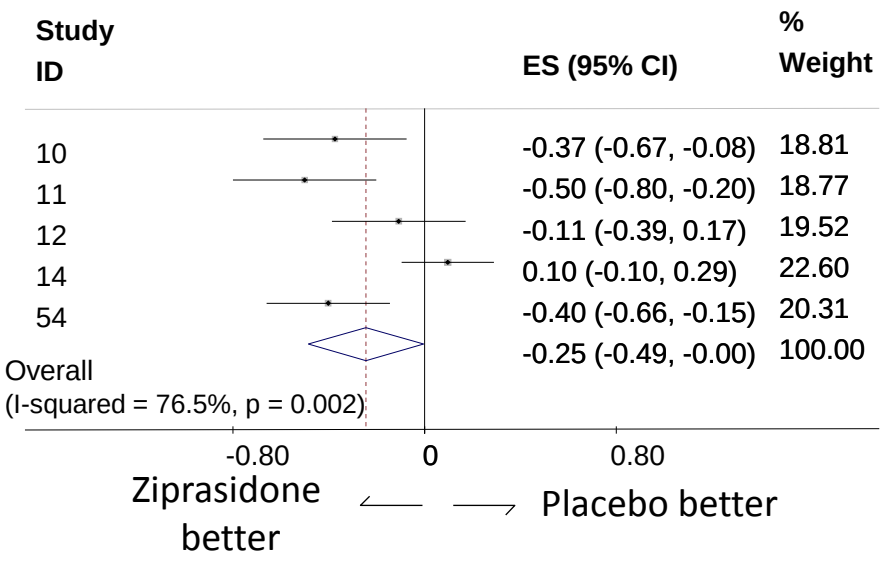
Fixed vs. Random effects meta-analysis: Find the differences!

Ziprasidone vs. Placebo for acute mania

fixed effects



random effects



- RE meta-analysis gives more **conservative results** compared to FE (wider CI)
- **Mean estimate may be** (slightly) different
- The weights are more evenly distributed in RE, **smaller studies get more weight compared to FE**

Fixed vs. random effects

- Fixed effect model is often unrealistic, random effects model might be easier to justify
- It is more sensible to extrapolate results from the random effects into general populations

But:

- If the number of studies is small it is impossible to estimate τ^2
- Random effects analysis may give spurious results when effect size depends on precision
  - (gives relatively more weight to smaller studies)
  - Important because
    - Smaller studies may be of lower quality (hence biased)
    - Publication bias may result in missing smaller studies

Fixed vs. random effects

Fixed or random effects meta-analysis should be specified *a priori*, based on the nature of studies and our goals and not on the basis of the Q test

What to do:

- Think about the question you asked, the available studies etc: do you expect them to be very diverse?

- You can always apply and present both fixed and random effects

Comparison of Fixed and Random Effects Meta-analyses

- Fixed and random effects inverse-variance meta-analyses may
  - be **identical** (when $\tau^2 = 0$)
  - give **similar** point estimate, **different** confidence intervals

What can we do with heterogeneity?

❑ Check the data	• Incorrect data extraction?
❑ Try to bypass it	• Change effect measure?
❑ Encompass it	• Random effects meta-analysis?
❑ Explore it	• Subgroup analysis? Meta-regression?
❑ Resign to it	• Do no meta-analysis?
❑ **Ignore it**	• **Don't do that!** 

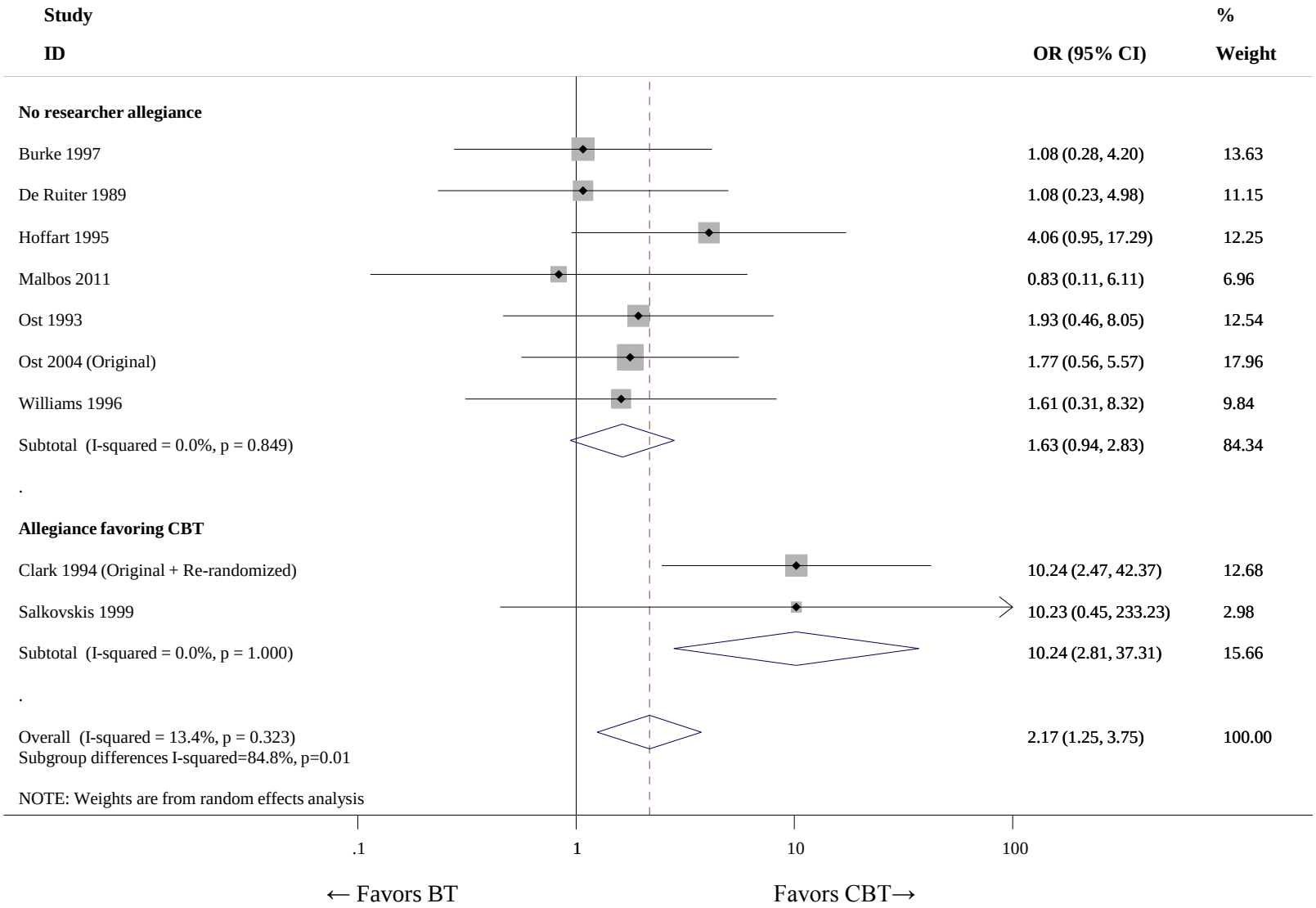
Subgroup analysis

- Using a subgroup analysis we split the studies in **two or more** groups in order to make comparisons between them
- This offers means for **investigating heterogeneity** in the results
- However, performing multiple subgroup analysis may give **misleading** results
- Thus, subgroup categories must be defined a priori (e.g. in the protocol), to avoid selective use of data.

Subgroup analysis

Example: CBT vs BT for panic disorder

Stata command: `metan SMD sd, by(allegiance)`



Meta-regression

- Meta-regression is an extension of subgroup analysis
- Using meta-regression an **outcome variable** is predicted according to the values of one or more **explanatory variables**.
- For example the outcome (e.g. logOR of treatment vs. placebo) may be influenced by a characteristic of the study (e.g. severity of illness of participants). Such characteristics are also called **effect modifiers**, i.e. they change the treatment effect
- Meta-regression can be performed using the `metareg` command in Stata

Alternative methods meta-analysis

- Apart from the inverse variance method, which is the most common, there are also 2 alternative methods for **dichotomous** outcomes meta-analysis:

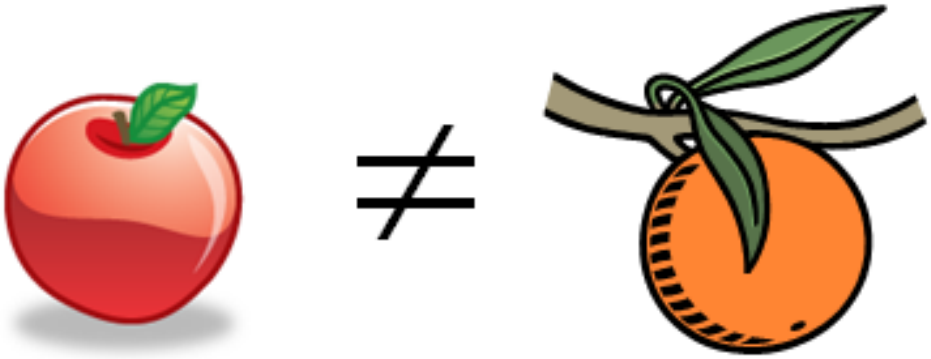
  - ✓ **Mantel-Haenszel** method (works well for small sample sizes and/or rare events), applicable to both FE and RE

  - ✓ **Peto method** (only for OR, works best for rare events, small treatment effects and balanced arms) applicable only to FE

When NOT to do a meta-analysis?

No point in '***mixing apples with oranges***'

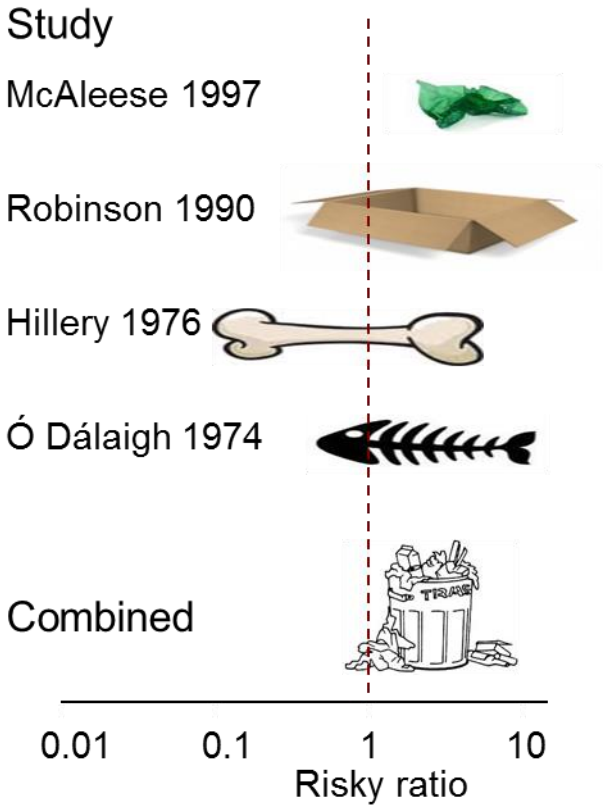
- ✓ Studies must address the same clinical question
- ✓ If you combine a mix of studies addressing a broad mix of different questions the answer you will get will be meaningless



When NOT to do a meta-analysis?

Beware of the
'*garbage in – garbage out*' rule

- ✓ A meta-analytical result is only as good as the included studies
- ✓ If included studies are biased results will be biased
- ✓ If studies are an unrepresentative set, results will be biased (eg. due to publication bias)

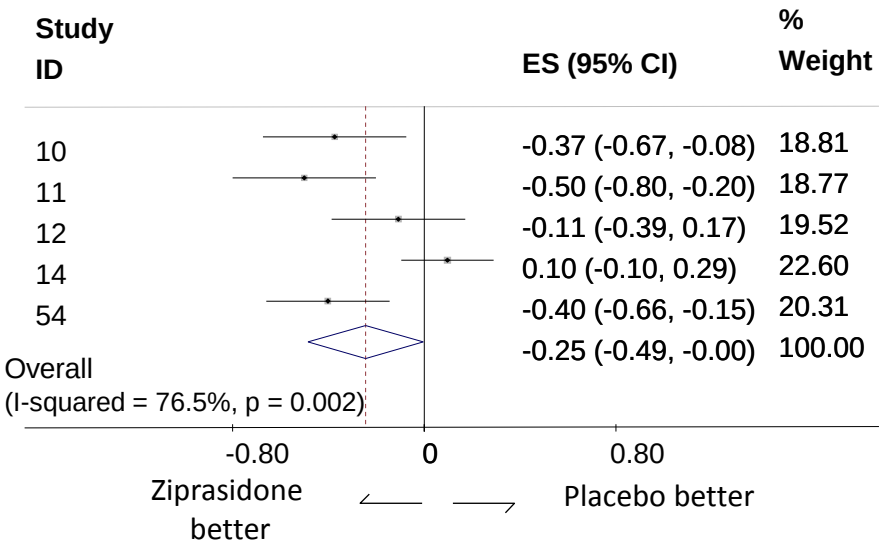


Summary

- There are many advantages in performing a meta-analysis (but it is not always possible or appropriate)
- 2 meta-analysis models: fixed and random effects. Usually an inverse variance approach is used for pooling results in both cases (but there are others)
- The choice between FE and RE should be guided by clinical considerations
- A forest plot is an essential part of any meta-analysis

Summary

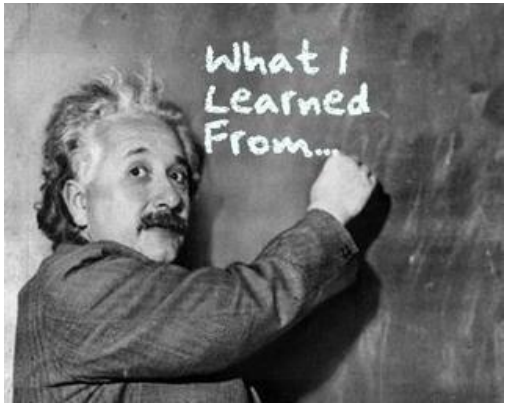
- A **forest plot** graphically displays:
  - ✓ The effect estimate from each individual study, along with the confidence intervals
  - ✓ The pooled, meta-analytical result (the "*diamond*")



  - ✓ The relative weight assigned to each study
  - ✓ An assessment of heterogeneity: a p-value for the Q-test, the value of I^2

Take home message

- Plan your analysis carefully, including comparisons, outcomes and meta-analysis methods
- Be clear about the statistical methods you use
- Present your results in a comprehensive manner
- Interpret your results with caution



References

- Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
- Cochrane online training material, available at http://training.cochrane.org/sites/training.cochrane.org/files/uploads/satms/public/english/10_Introduction_to_meta-analysis_1_1_Eng/story.html
- Introduction to Meta-Analysis, Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, Hannah R. Rothstein
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986; **7**: 177-188
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539-1558