Introduction to meta-analysis 2: Meta-analysis of binary and continuous outcomes

Dr Joanne McKenzie,
Monash University, Australia

Dr Areti Angeliki Veroniki,
University of Ioannina, Greece

September 17, 2018

We have no actual or potential conflicts of interest in relation to this presentation

Acknowledgements

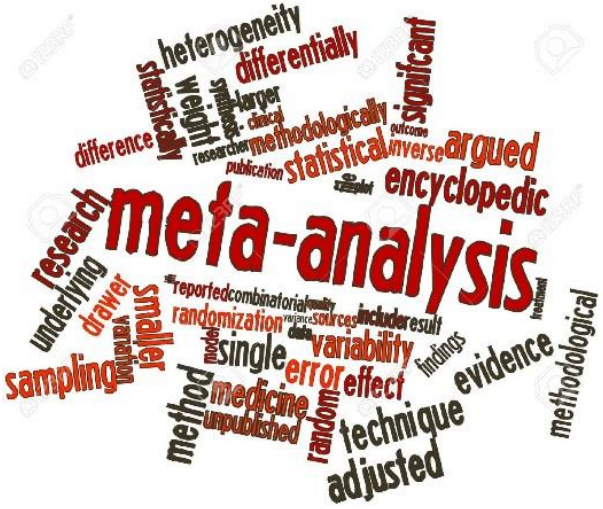
- Georgia Salanti
- Julian Higgins
- Dimitris Mavridis
- Orestis Efthimiou

Learning objectives

To provide an introduction to:

- Effect measures for dichotomous and continuous outcomes
- Meta-analysis of dichotomous and continuous outcomes
- Considerations for choosing an effect measure
- Interpretation of the effect measures and potential problems
- Data extraction and identifying errors
- Other issues that arise in practice

The systematic review process



1. Formulation of a clear question and inclusion criteria
2. Search for relevant studies
3. Data extraction and assessment of included studies
4. Synthesis of findings
5. Interpretation

Meta-analysis

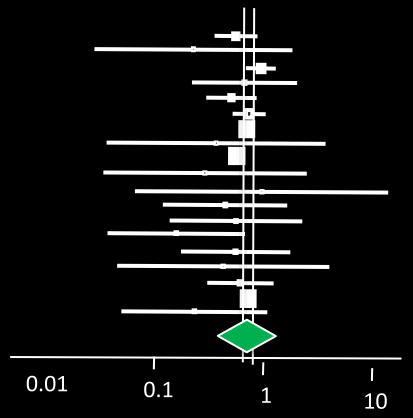
http://handbook.cochrane.org/

Synthesis of findings – meta-analysis

- Statistically synthesize the study results in a meta-analysis

Meta-analysis can be thought of as "conducting research about previous research"

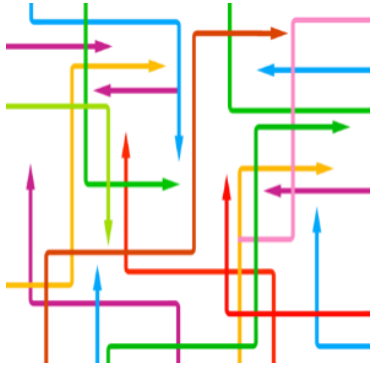
Meta-analysis is a statistical technique for combining the findings from independent studies



It is most often used to assess the clinical effectiveness or safety of healthcare interventions

It combines data from 2 or more randomized controlled trials

Synthesis of findings - Why perform a meta-analysis?

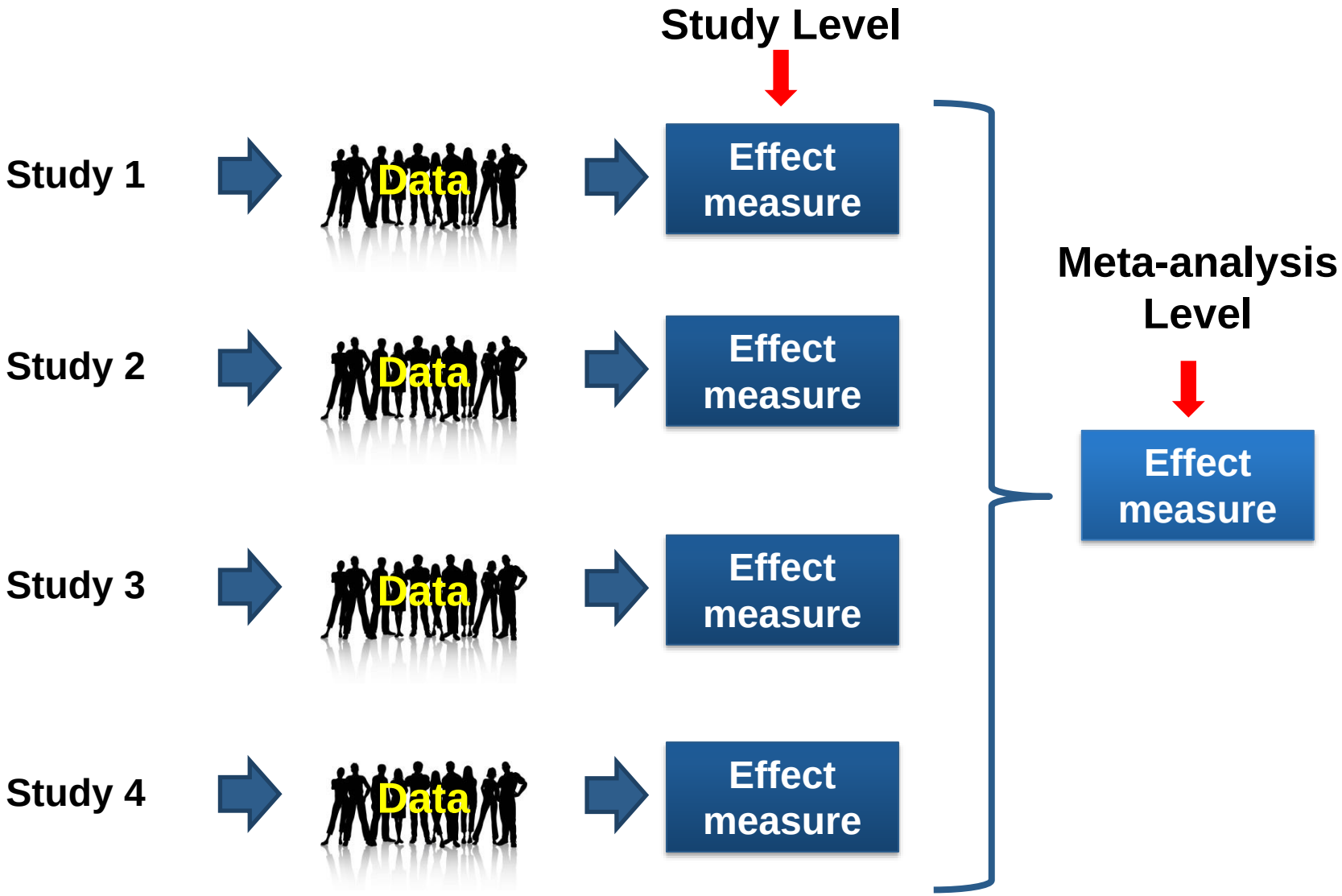


- To increase power/precision
- To reduce problems of interpretation due to sampling variation
- To answer questions not posed by the individual studies
- To settle controversies arising from conflicting studies

Synthesis of findings - Basic principles of meta-analysis

- Participants in one study are *not* directly compared with those in another
- Each study is analysed separately
- Summary statistics are combined to give the meta-analysis estimate
- Each study is weighted according to the information it provides (usually the inverse of its variance)
- Larger studies are given greater weight, and hence their influence on the meta-analysis effect estimate is greater

To apply a meta-analysis



To apply a meta-analysis

1. Require from each study
   - estimate of treatment effect
   - variance of estimate

$$\textbf{weight of study} = \frac{1}{\text{variance}}$$

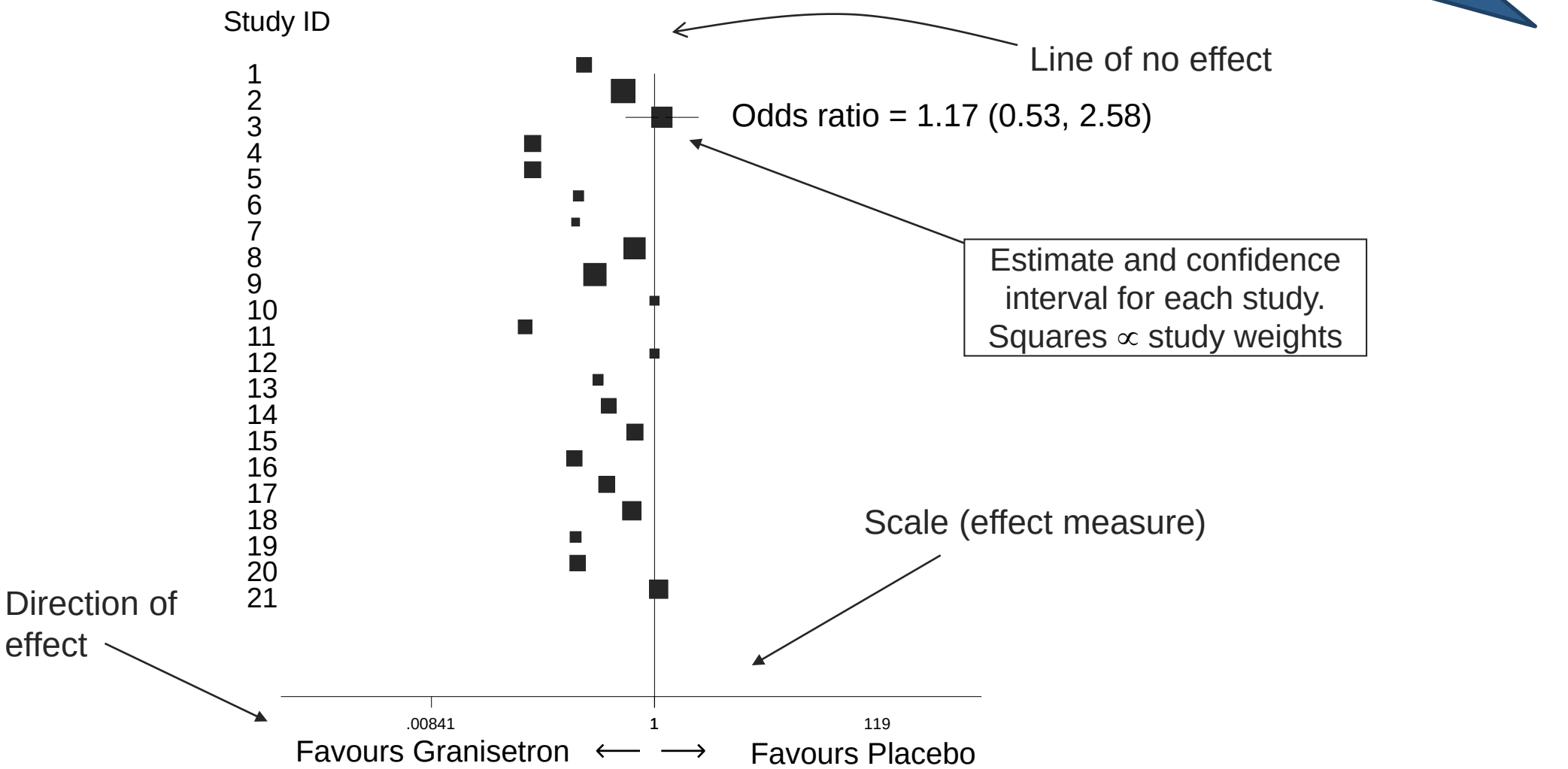
2. Combine these using a weighted average:

$$\textbf{pooled estimate} = \frac{\text{sum of (estimate} \times \text{weight)}}{\text{sum of weights}}$$

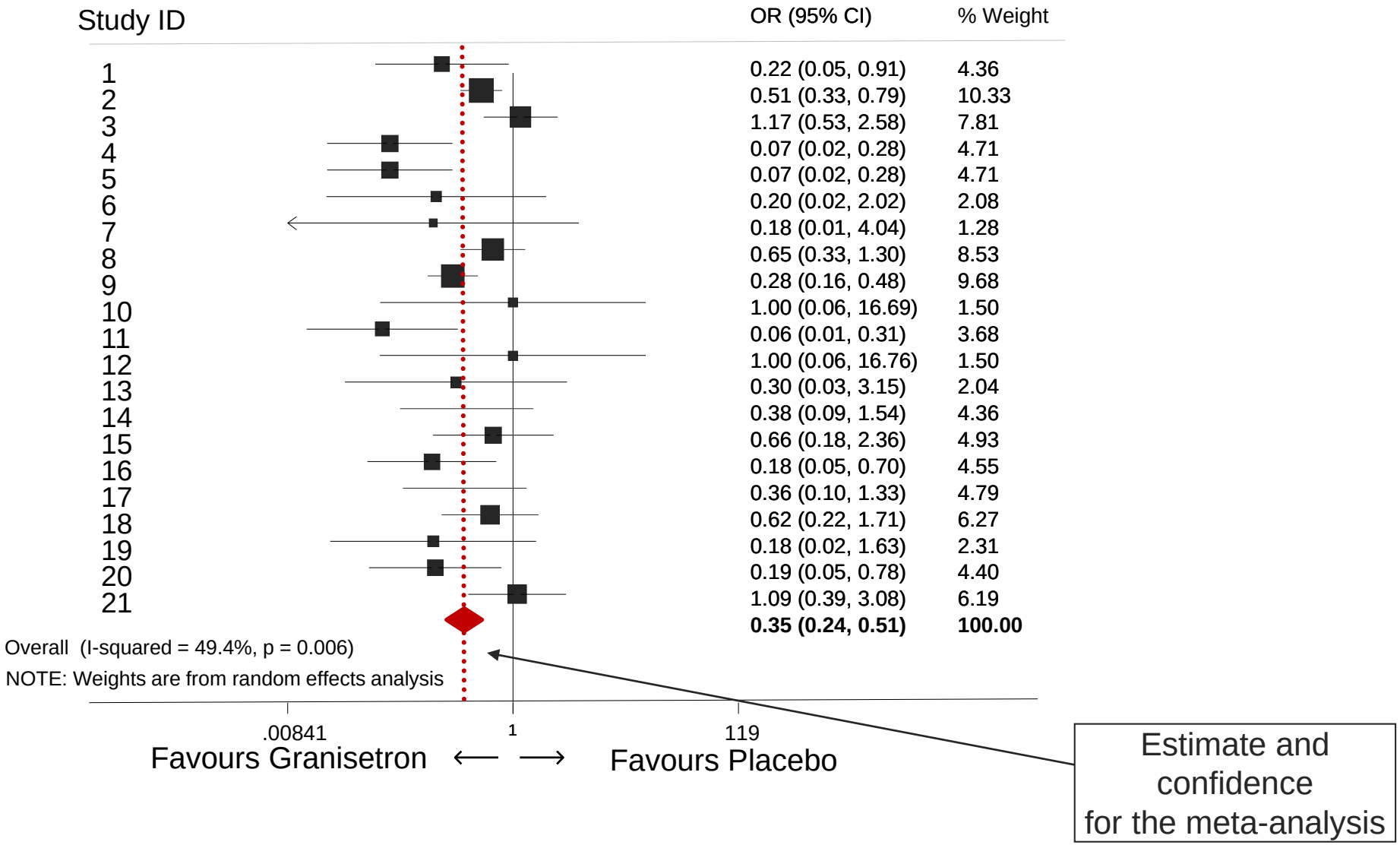
$$\textbf{with variance} = \frac{1}{\text{sum of weights}}$$

Synthesis of findings - Forest plot

Does granisetron prevent nausea?

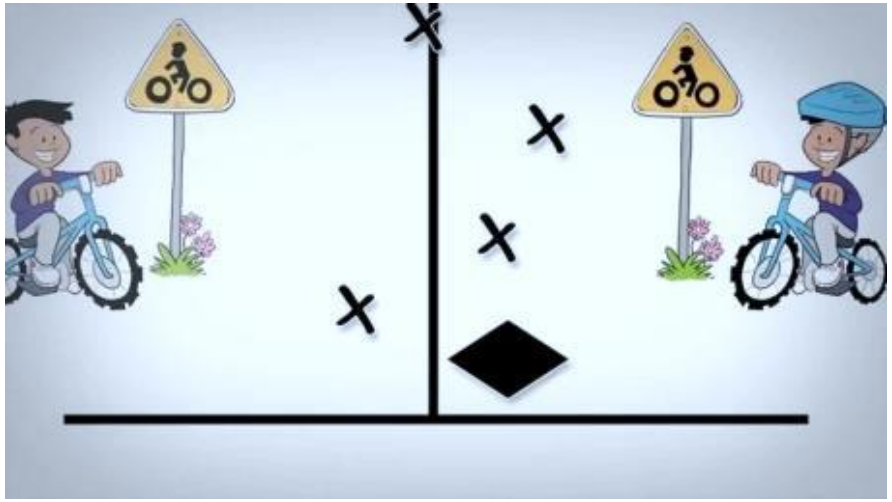


Synthesis of findings - Forest plot



How to start a meta-analysis

1. Identify the data type for the outcome measurements
2. Use an effect size to compare the outcomes between the interventions



Results of experiments or observations

- Studies usually compare outcomes between intervention groups
  - The risk of nausea with and without granisetron

	Nausea	**Non-nausea**	Total
Granisetron	**3**	**27**	30
Placebo	**10**	**20**	30
Total	13	47	200

Question: How can we **compare the outcomes** between the interventions?

→ Using **Effect Sizes**

Results of experiments or observations

- Effect size: a value reflecting the magnitude of the treatment effect

	Nausea	Non-nausea	Total
Granisetron	**3**	**27**	30
Placebo	**10**	**20**	30
Total	13	47	200

Dichotomous data

What are dichotomous (or binary) outcomes?

- When the outcome for every participant is one of two possibilities or events
  - alive or dead
  - healed or not healed
  - pregnant or not pregnant
  - …

Dichotomous data

Consider a single study:

	Event	*Non-Event*	*Total*
Treatment	**a**	**b**	m_1
Control	**c**	**d**	m_2
Total	N_1	N_2	N

Dichotomous data

- Two components
  - Number of events per group
  - Sample size per group

	Dead	*Alive*	*Total*
Treatment	**10**	**90**	100
Control	**14**	**86**	100
Total	24	176	200

Effect measures for dichotomous data

- We can compare the two groups in several ways:
  - Odds ratio (OR)
  - Risk ratio (RR) = Relative Risk
  - Risk difference (RD) = Absolute Risk Reduction (ARR)

- All estimates are uncertain and should be presented with a confidence interval, variance or standard error

Risk vs Odds

Risk and odds are just different ways of expressing how likely an event is

Risks and odds

- Risk is defined as the probability of having an event

$$\text{risk} = \frac{\text{number of events of interest}}{\text{total number of observations}}$$

- Example: What is the probability that today is Monday?
  - 1 day of the week is Monday / 7 days of the week = 1/7

- Odds is defined as the ratio of two probabilities: the probability of having an event over the probability of not having an event

$$\text{odds} = \frac{\text{number of events}}{\text{number of non events}}$$

- Example: What are the odds that today is Monday?
  - 1/6

Risk

- 20 people drank vodka during the Cochrane Colloquium Gala Dinner, 5 developed a headache

- **Risk** of a headache:

  = 5 headaches / 20 who drank vodka

  = 5/20 = 0.25 = 25%

$$Risk = \frac{number\ of\ events\ of\ interest}{total\ number\ of\ observations}$$

- The risk of having a headache is 25% if you drink vodka

Odds

- 20 people drank vodka, 5 developed a headache

- **Odds** of a headache

  = 5 headaches / 15 without headaches

  = 5/15 = 1/3

$$Odds = \frac{number\ of\ events}{number\ of\ non\ events}$$

- The chances of a headache are one third the chances of no headache when drinking vodka

- One person will have a headache for every three that will not

Odds

- I throw a dice, what is the risk and what are the odds of rolling a 4?
  - ✓ Risk = 1 / 6; Odds = 1 / 5

- I throw a dice, what is the risk and what are odds of rolling an even number (2, 4 or 6)?
  - ✓ Risk = 3 / 6; Odds = 3 / 3 = 1

- In an RCT, out of 100 people treated, 90 were cured. What is the “risk” and what are the odds of being cured?
  - ✓ Risk = 90 / 100 = 0.9; Odds = 90 / 10 = 9

Comparison between Risk and Odds

The difference between risk and odds is small when the event is rare but can be large for common events

$$Risk = \frac{Odds}{1+Odds}$$

$$Odds = \frac{Risk}{1-Risk}$$

Event	Total	Risk	Odds
5	100	0.05	0.0526
50	100	0.5	1
95	100	0.95	19

Q: Does vodka lead to an **increased** chance of having a headache?

A: We need to **compare** people drinking vodka with a control group, e.g. people drinking water

Relative effect measures: FROM risk and odds TO risk ratio and odds ratio

- Risk and odds measure the likelihood of an event (e.g. of having a headache when drinking too much vodka)

- In order to compare between groups (e.g. people drinking vodka vs. people drinking water, or in an RCT patients receiving treatment vs. placebo) we need to use relative effect measures

Relative effect measures: FROM risk and odds TO risk ratio and odds ratio

Group A (treatment)

Group B (control)

Risk (treatment group)

Risk (control group)

$$Risk\ Ratio = \frac{Risk\ in\ treatment\ group}{Risk\ in\ control\ group}$$

Risk Ratio allows us to compare between two groups

Relative effect measures: FROM risk and odds TO risk ratio and odds ratio

Group A (treatment)

Group B (control)

Odds (treatment group)

Odds (control group)

$$Odds\ Ratio = \frac{Odds\ in\ treatment\ group}{Odds\ in\ control\ group}$$

Odds Ratio allows us to compare between two groups

Risk ratio and odds ratio

$$\text{risk ratio} = \frac{\text{risk in treatment group}}{\text{risk in control group}}$$

$$\text{odds ratio} = \frac{\text{odds in treatment group}}{\text{odds in control group}}$$

	Event	*Non-Event*	*Total*
Treatment	**a**	**b**	a+b
Control	**c**	**d**	c+d
Total	a+c	b+d	N

$$Risk\ Ratio = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

$$Odds\ Ratio = \frac{\frac{a}{b}}{\frac{c}{d}}$$

Risk ratio

	Dead	*Alive*	*Total*
Treatment	**10**	**90**	100
Control	**14**	**86**	100
Total	24	176	200

Risk of event in treatment
= 10/100

Risk of event in control
= 14/100

$$\text{Risk Ratio} = \frac{10/100}{14/100} = \frac{0.10}{0.14} = 0.71$$

Odds ratio

	Dead	*Alive*	*Total*
Treatment	**10**	**90**	100
Control	**14**	**86**	100
Total	24	176	200

Odds of event in treatment
= 10/90

Odds of event in control
= 14/86

$$\text{Odds Ratio} = \frac{10/90}{14/86} = \frac{0.11}{0.16} = 0.69$$

Risk ratio

	Event	*Non-Event*	*Total*
Treatment	**90**	**10**	100
Control	**30**	**70**	100
Total	120	80	200

A risk ratio of 3 ($RR = 3$) implies:

- Risk of the event in the treatment group is 3 times that of the control
- The treatment increased the risk of the event by:

$$100 \times (RR - 1)\% = 200\%$$

$$RR = \frac{90/(90+10)}{30/(30+70)}$$

A risk ratio of 0.25 ($RR = 0.25$) implies:

- The risk of an event in the treatment group is 1/4 of the risk in the control group
- The treatment reduces the risk of events by

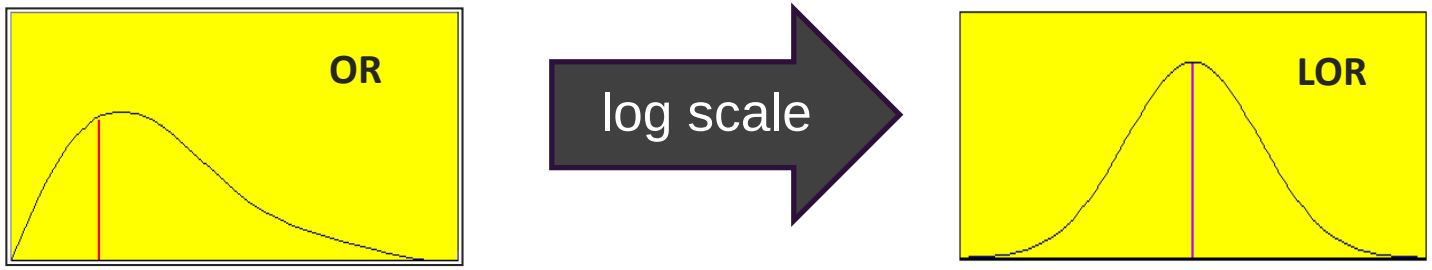
$$100 \times (1 - RR)\% = 75\%$$

$$RR = \frac{10/(10+90)}{40/(40+60)}$$

	Event	*Non-Event*	*Total*
Treatment	**10**	**90**	100
Control	**40**	**60**	100
Total	50	150	200

Treatment effects on the log-scale: Why?

- RR, OR are *not* symmetric (take values from 0 to +∞, with the value of no-effect being at 1)



- RR, OR are *not* additive

Ex: halving the odds (OR=0.5) and doubling the odds (OR=2) do not cancel out by summing (average OR=1.25)

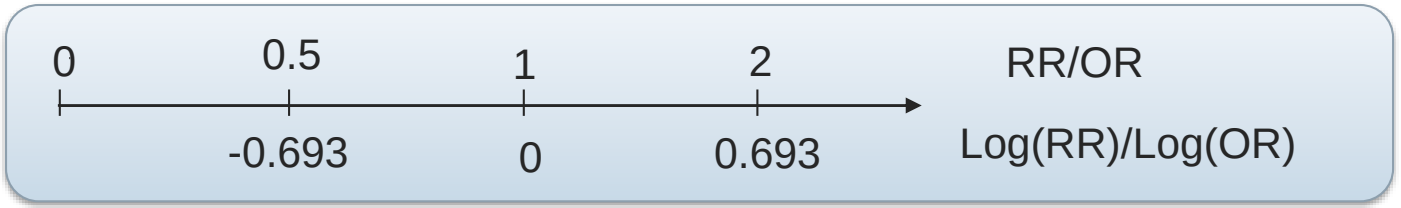
- …so we need a transformation in order to make them symmetric

- The most commonly used transformation is the natural logarithm, denoted by ln or log (but other transformations could be used)

Treatment effects on the log-scale: Why?

- log(OR) and log(RR)
  - are symmetric
    - log(OR) follows the normal distribution
    - log(RR) has a better approximation with the normal distribution than RR
  - no effect at zero (neutral value)
  - easier to compare positive with negative values
    - Log(OR) takes values in (-∞, ∞)
    - Log(RR) takes values in (-∞, $\log(1/CGR)$)

✓ **Typically the natural log transformation (log base e, written 'ln') is used**



Log-risk ratio (LogRR)

	Event	*Non-Event*	*Total*
Treatment	**a**	**b**	a+b
Control	**c**	**d**	c+d
Total	a+c	b+d	N

$$logRR = log\frac{\frac{a}{a+b}}{\frac{c}{c+d}} = log\left(\frac{a(c+d)}{c(a+b)}\right)$$

$$var(logRR) = \frac{1}{a} + \frac{1}{a+b} + \frac{1}{c} + \frac{1}{c+d}$$

- **When $logRR = 0$, there is no difference between the groups**

Log-Risk Ratio (LogRR)

	Dead	Alive	Total
Treatment	10	90	100
Control	14	86	100
	24	176	200

Calculate Risk Ratio

$$\boldsymbol{RR} = \frac{\frac{10}{100}}{\frac{14}{100}} = \frac{10}{14} = 0.71$$

- Where risk ratio = 1, this implies no difference in effect

Include in meta-analysis

$\boldsymbol{log(RR)} = log(0.71) = -0.34$ and $\boldsymbol{var(logRR)} = \frac{1}{10} + \frac{1}{100} + \frac{1}{14} + \frac{1}{100} = 0.15$

or $\boldsymbol{SE(logRR)} = \sqrt{var(logRR)} = \sqrt{0.15} = 0.39$

Calculate a 95% C.I. for logRR

$\boldsymbol{95\%\ CI\ for\ logRR}$: $logRR \pm 1.96 \times SE(logRR) = (-1.10, 0.42)$

Back-calculate to the original scale

$\boldsymbol{95\%\ CI\ for\ RR}$: $(e^{-1.10}, e^{0.42}) = (0.33, 1.53)$

Log-odds ratio (LogOR)

	Event	*Non-Event*	*Total*
Treatment	**a**	**b**	a+b
Control	**c**	**d**	c+d
Total	a+c	b+d	N

$$logOR = log\frac{\frac{a}{b}}{\frac{c}{d}} = log\left(\frac{a}{b}\right) - log\left(\frac{c}{d}\right) = log\left(\frac{ad}{bc}\right)$$

$$var(logOR) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

- **When $logOR = 0$, there is no difference between the groups**

Risk difference (RD)

- The difference in the risk between the treated and control groups

$$RD = \textbf{Treatment Group Risk} - \textbf{Control Group Risk} = \frac{a}{a+b} - \frac{c}{c+d}$$

- A measure easy to understand but clinical interpretation depends on context
  - A treatment reduces the risk of an adverse event by RD= 2%:
    - From 70% risk to 68% or from 3% to 1%?
- May give impossible values if applied in different populations
  - RD of -10% applied to a population with 7% Control Group Risk gives –3% Treatment Group Risk

Risk difference (RD)

	Event	*Non-Event*	*Total*
Treatment	**a**	**b**	a+b
Control	**c**	**d**	c+d
Total	a+c	b+d	N

$$RD = \frac{a}{a+b} - \frac{c}{c+d}$$

$$var(RD) = \frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}$$

- **When $RD = 0$, there is no difference between the groups**

Beware of the direction of effects!

When reading a result it might be unclear to the reader how the effect measures were defined

- Is a RR (or an OR) defined as treated over control **or** control over treated?
- Is RD defined as treated minus control **or** control minus treated?

- Make sure you are absolutely clear about how you define your effect measure
- Note which direction favors which treatment
  - ✓ *E.g. write "an OR>1 favours control"*

Example on effect measures

Dichotomous outcome data:

	Dead	Alive	Total
Treatment	10	90	100
Control	20	80	100
	30	170	200

$$\mathrm{RR} = \frac{\frac{10}{100}}{\frac{20}{100}} = 0.50$$

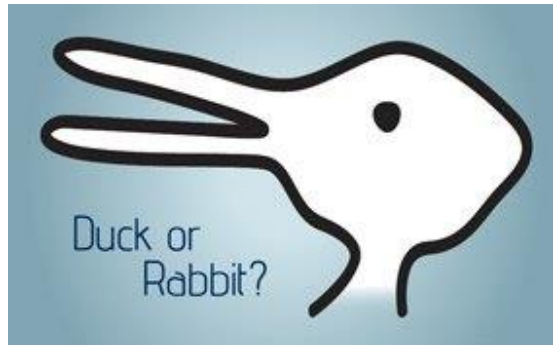
$$\mathrm{OR} = \frac{\frac{10}{90}}{\frac{20}{80}} = 0.44$$

$$\mathrm{RD} = \frac{10}{100} - \frac{20}{100} = -10\%$$

Choosing an effect measure

- Communication of effect
  - ✓users must be able to understand and apply the result
- Consistency of effect
  - ✓applicable to all populations and contexts
- Mathematical properties

Communication



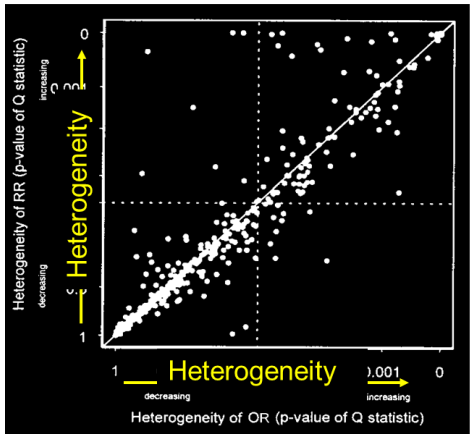
- OR is hard to understand, often misinterpreted
- RR is easier, but
  - ✓It can mean a very small or very big change, depending on the underlying risk
  - ✓It can be very different if we switch the outcome
- RD is easiest
  - ✓absolute measure of actual change in risk

Consistency

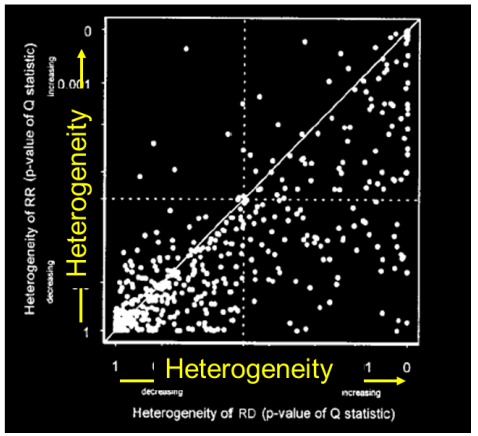
- Event rates usually vary from study to study within a review
- Study of meta-analyses in *The Cochrane Library*:
  - ✓RR and OR are less variable across different populations
  - ✓RD is more variable, dependent on baseline risk

Readers will apply results to their own population, which may be different

(ORvsRR)



(RDvsRR)



Source: Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2002; 21:1575-1600

Mathematical Properties

- Defining the event

  ✓Good or bad, presence or absence?

  ✓Think carefully and choose in advance

  ✓OR and RD are stable either way, RR varies

Summary

	OR	RR	RD
Communication	✗	✓	✓✓
Consistency	✓	✓	✗
Mathematics	✓✓	✗	✗

- No measure is uniformly best
- Consider meta-analysing using one measure, and interpreting using another

Methods for meta-analysis of dichotomous outcomes

- Inverse Variance Method (IV)
- Mantel-Haenszel (MH)
- Peto Odds Ratio

Inverse Variance Method

To perform an inverse variance meta-analysis:

- Use the natural log scale (log) for OR and RR, but not for RD
- Let's call y_i the effect size in study *i*
- What we have from each study is y_i and SE_i of the effect size
- Meta-analysis will summarize all y_i to estimate the pooled Θ_{IV} and $SE(\Theta_{\mathrm{IV}})$

Inverse Variance Method

$$\Theta_{IV} = \frac{\sum w_i y_i}{\sum w_i}$$

$w_1 y_1 + w_2 y_2 + \cdots$

$w_1 + w_2 + \cdots$

Where w_i is the inverse of the variance*: $\boldsymbol{w_i = \frac{1}{SE_i^2}}$

$$\boldsymbol{SE(\Theta_{IV}) = \frac{1}{\sqrt{\sum w_i}}}$$

- $\boldsymbol{y_i}$ can be $\boldsymbol{logOR}$, $\boldsymbol{logRR}$ or $\boldsymbol{RD}$ for each study (be sure to use the same effect measure for each study!) and can be pooled if we know the SE_i.

*for a fixed-effect meta-analysis. For a random-effects meta-analysis see the workshop *Introduction to meta-analysis 3: Dealing with heterogeneity*

Group Exercise: Inverse Variance Method Example

Let's say we have the logOR and the variance for each of the 3 studies we identified in a systematic review

	y_i (logOR)	variance
study 1	2	1
study 2	2	2
study 3	4	2

Group Exercise: Inverse Variance Method Example

Try to calculate by hand the IV meta-analysis pooled result

	y_i (logOR)	variance	w_i	$w_i y_i$
study 1	2	1	?	
study 2	2	2		
study 3	4	2		

$\Sigma w_i =$? $\quad \Sigma(w_i y_i) =$?

$$\Theta_{\mathrm{IV}} = \frac{\sum w_i y_i}{\sum w_i}$$

w_i is the inverse of the variance

$\Sigma(w_i y_i) / \Sigma w_i =$?

Group Exercise: Inverse Variance Method Example

Try to calculate by hand the IV meta-analysis pooled result

	y_i (logOR)	variance	w_i	$w_i y_i$
study 1	2	1	1	2
study 2	2	2	0.5	1
study 3	4	2	0.5	2
			Σw_i= 2	$\Sigma(w_i y_i)$ = 5

$$SE(\Theta_{IV}) = \frac{1}{\sqrt{\sum w_i}}$$

$\Sigma(w_i y_i)$ /Σw_i = 5/2 = 2.5

1/Sqrt(Σw_i)= 0.7

Group Exercise: Inverse Variance Method Example

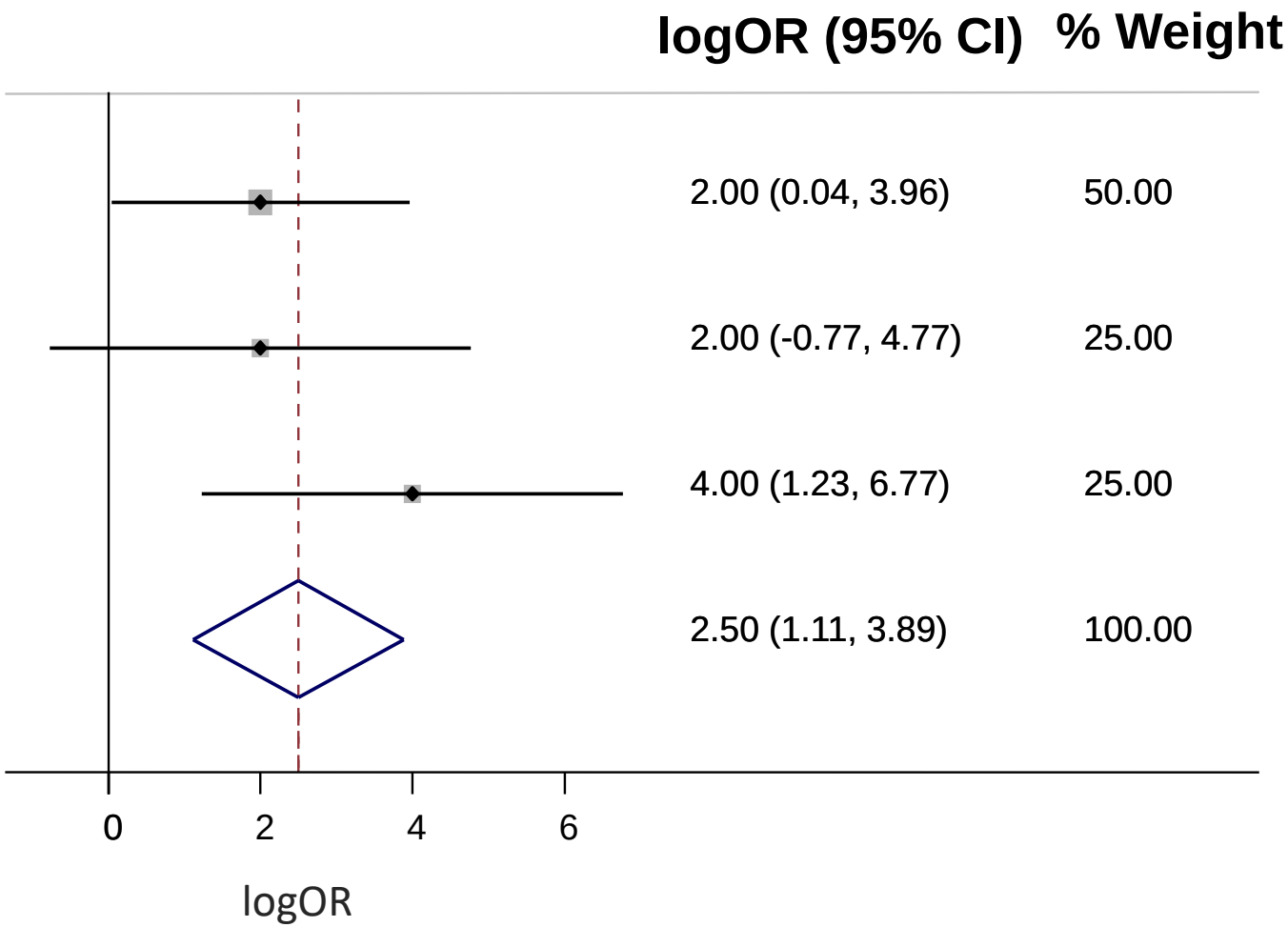
Try to calculate by hand the IV meta-analysis pooled result

	y_i (logOR)	variance	w_i	$w_i y_i$	% weight
study 1	2	1	1	2	1/2=50%
study 2	2	2	0.5	1	0.5/2=25%
study 3	4	2	0.5	2	0.5/2=25%
			Σw_i= 2	$\Sigma(w_i y_i)$ = 5	**100%**

$\Sigma(w_i y_i)$ /Σw_i = 5/2 = 2.5

1/Sqrt(Σw_i)= 0.7

Group Exercise: Inverse Variance Method Example



Mantel-Haenszel (MH) method

- Is more robust for
  - ✓ few events
  - ✓ sparse data (event rates being low or study size being small)
- In practice, meta-analysis follows the same logic as the inverse variance method, but now it uses the weighted mean of the effect (not the logarithm) and the weights are different
- The default method in RevMan

Peto method

- Developed for large trials with small treatment effects
- Biased when odds ratios are far from 1 or sizes of groups being compared are very different
- May be best method for rare events (≤1%) and when group sizes are balanced
- No correction needed for zero cells

Summary: Meta-analysis for Dichotomous Data

- Inverse-variance weighted averages are fine for studies with large sample sizes
  - For both fixed-effect and random-effects

- Mantel-Haenszel offers an improvement for sparse data
  - For both fixed-effect and random-effects

- Peto method may be best for rare events
  - ✗ only applicable for OR
  - ✗ only for fixed-effect model
  - ✓ if numbers of subjects in each arm are reasonably balanced and the effect size is not too large

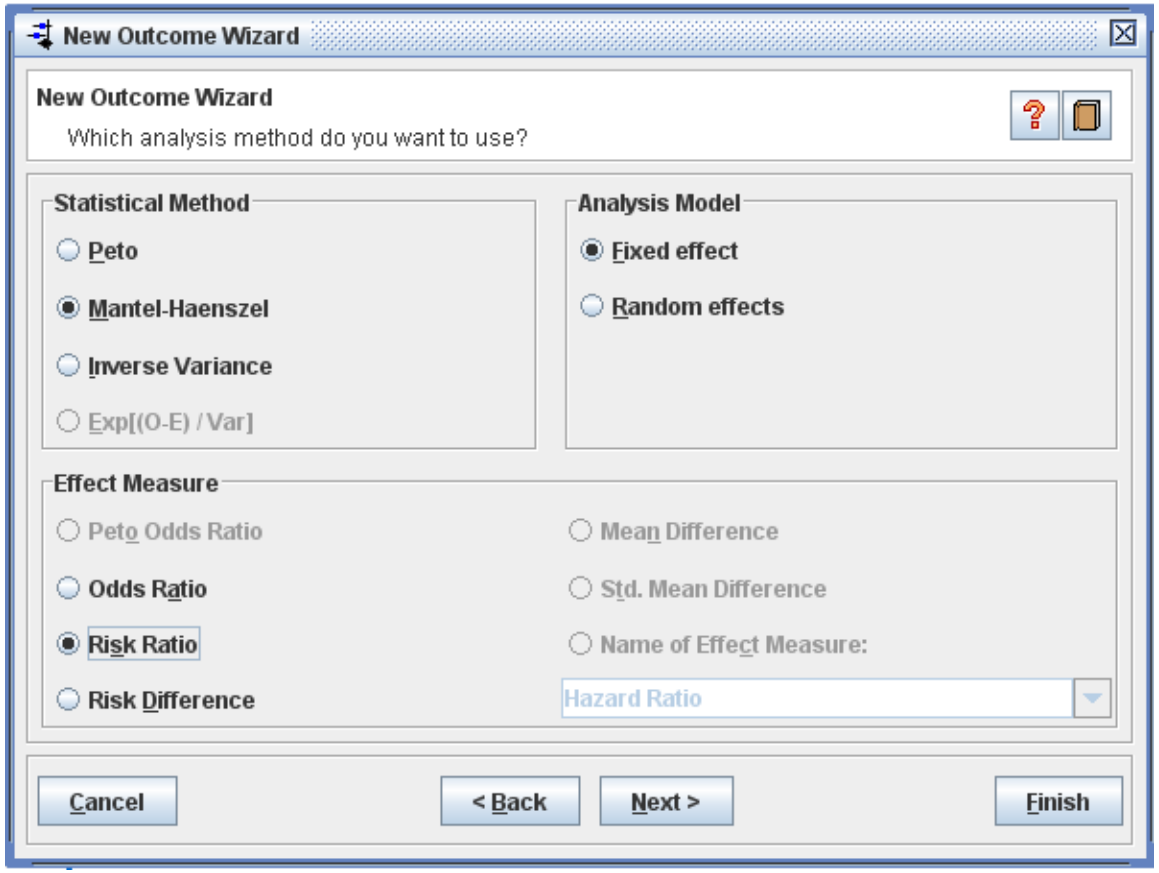
RevMan can:

Perform a meta-analysis using

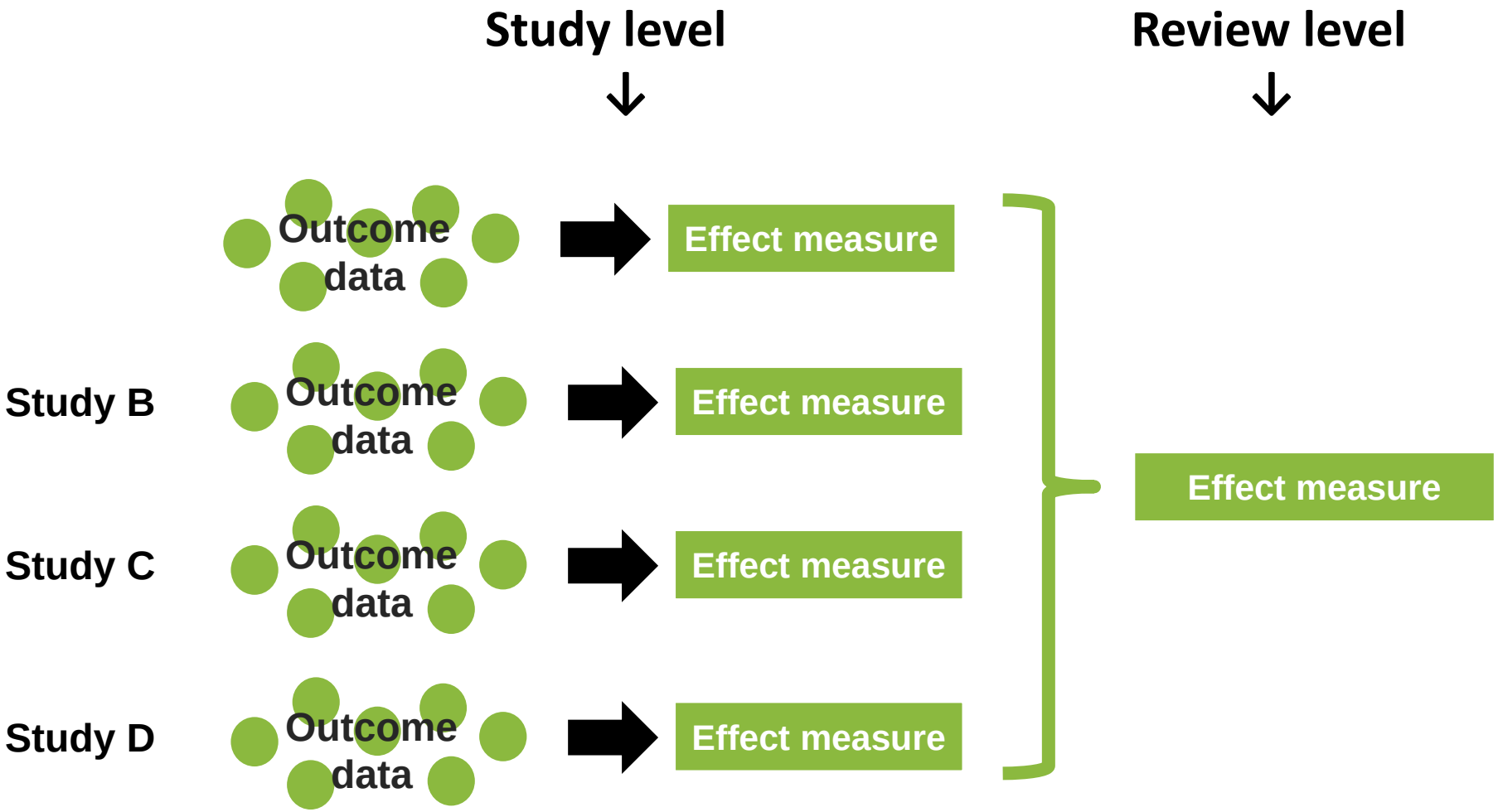
- Inverse Variance
- Mantel- Haenszel methods
- Peto

<u>It needs:</u>

1. The 2×2 table (a,b,c,d) [data entry as dichotomous outcome]
2. A treatment effect (e.g. $logOR$) and its standard error (SE) [data entry as generic inverse variance outcome]



Continuous data



Source: Jo McKenzie & Miranda Cumpston

Types of continuous data

- Continuous data
  - Outcome is a measurement of a numerical quantity
  - E.g. systolic and diastolic blood pressure, weight, height

- Ordinal data
  - Outcome is one of several ordered categories
  - E.g. pain scale (none/mild/moderate/severe), depression scales (long ordinal scales)

- Counts
  - Counting the number of events that each individual experiences
  - E.g. number of episodes of school absence, number of asthma attacks

- Time-to-event
  - Time until an event occurs
  - E.g. time to relapse of cancer, time to walk 100m

How common are continuous outcomes?

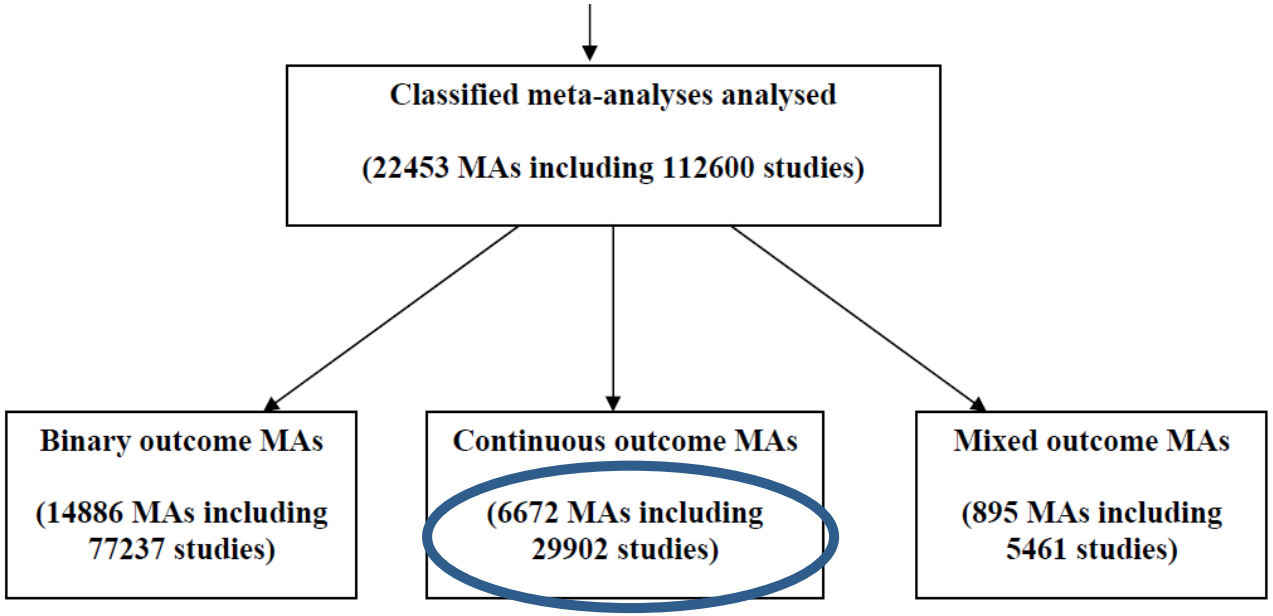


Figure 2 Flow diagram showing eligibility and exclusions of meta-analyses and studies.

Davey *et al*: Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. BMC Med Res Methodol 2011, 11:160.

Summary statistics of continuous data

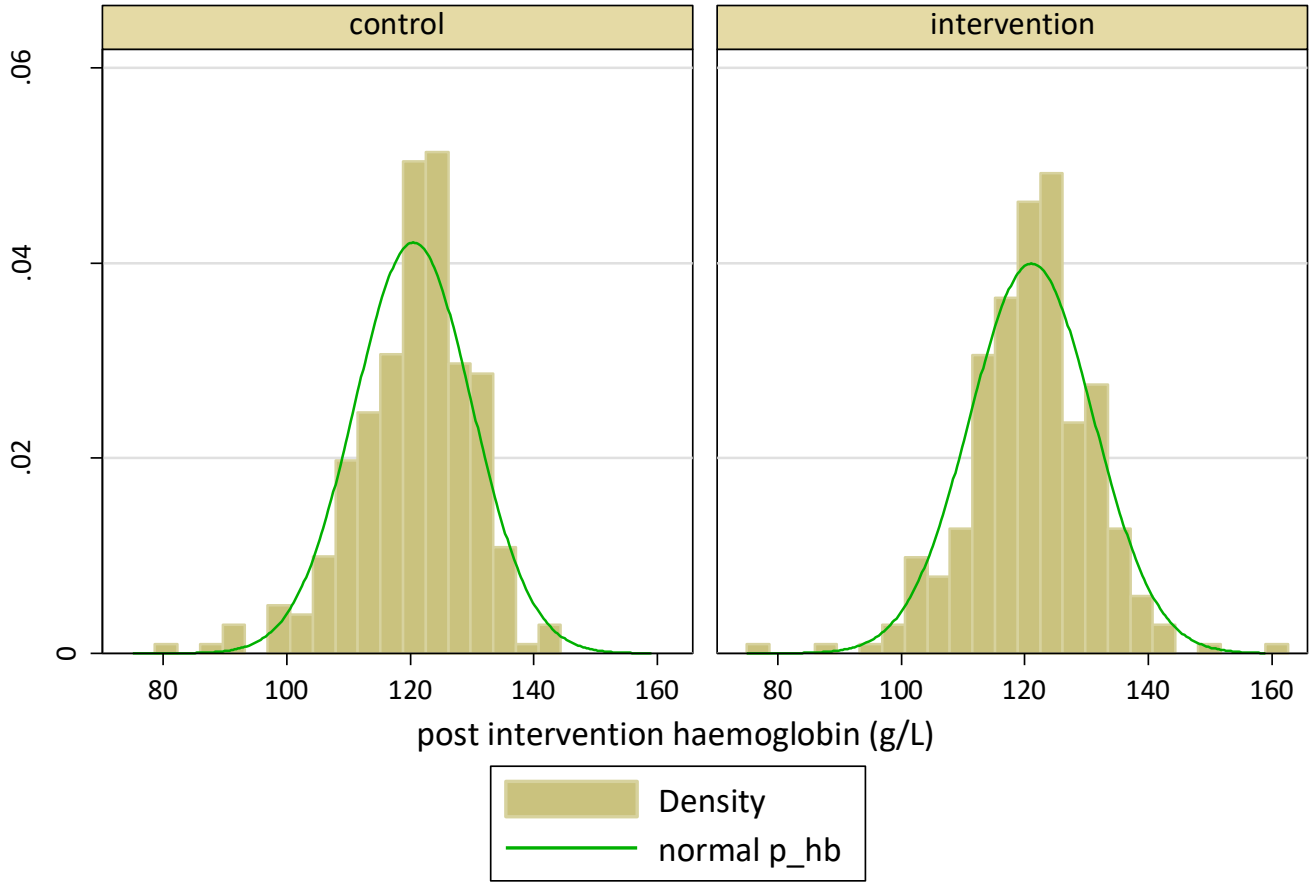
RCT carried out to test the effectiveness of a seasoning powder fortified with micronutrients on biochemical indices, anthropometry, functional health, and cognitive outcomes.

	Fortified group	Unfortified group
n	271	269
Mean Hb (g/L)	121.0	120.5
Standard deviation	10.1	9.5

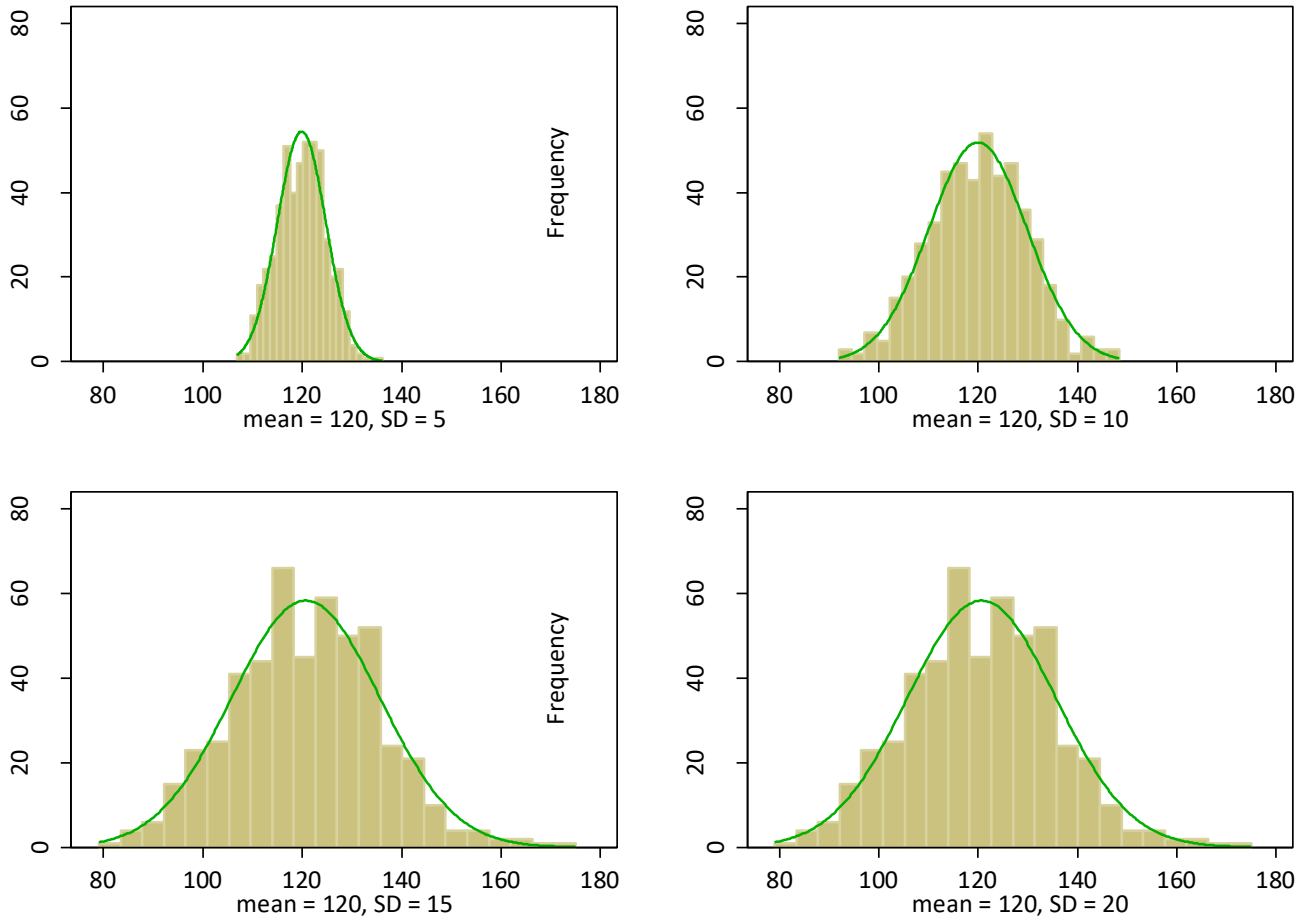
What is a standard deviation (SD)?

- A SD describes the variability in the data.
- For a particular outcome, a larger SD implies more variability than a smaller SD.
- A measure of how far, on average, an individual's value is from the mean.

Fortification randomised trial



Comparing different SDs



Effect measures: Mean difference (MD)

- The MD is calculated as:

  MD = mean outcome in intervention group – mean outcome in control group

	Fortified group	Unfortified group
n	271	269
Mean Hb	121.0	120.5
SD	10.1	9.5
Mean Difference (95% CI)		0.5 (-1.1, 2.2)

- Haemoglobin levels are on average 0.5 g/L larger in the fortified group compared with the unfortified group.

Effect measures: Standardised Mean difference (SMD)

- The SMD is calculated as:

$$\text{SMD} = \frac{\text{mean outcome in intervention group} - \text{mean outcome in control group}}{\text{standard deviation of outcome among participants}}$$

- SMD expresses the size of the intervention effect relative to the variability observed

Effect measures: Standardised Mean difference (SMD)

	Fortified group	Unfortified group
n	271	269
Mean Hb	121.0	120.5
SD	10.1	9.5
SMD (95% CI)		0.05 (-0.12, 0.22)

Haemoglobin levels are on average 0.05 of a standard deviation larger in the fortified group compared with the unfortified group.

Effect measures: Ratio of means (RoM)

- The RoM is calculated as:

$$\text{RoM} = \frac{\text{mean outcome in intervention group}}{\text{mean outcome in control group}}$$

	Fortified group	Unfortified group
n	271	269
Mean Hb	121.0	120.5
SD	10.1	9.5
Ratio of means (95% CI)		1.005 (0.991, 1.018)

- Mean haemoglobin levels are 0.5% larger in the fortified group compared with the unfortified group.
- Not currently implemented in RevMan (see Friedrich 2008, 2011).

Practical exercise

Table 1: Extracted data from reports of trials examining the effect of domperidone versus placebo on milk production in mothers of newborns. Da Silva also reports individual patient data (see Table 2)

Trial	Campbell-Yeo 2010				Da Silva 2001				Rai 2016				Blank 2003			
Intervention group	Domperidone		Placebo		Domperidone		Placebo		Domperidone		Placebo		Domperidone		Placebo	
	n	Stats	n	Stats	n	Stats	n	Stats	n	Stats	n	Stats	n	Stats	n	Stats
Baseline daily milk volume (mls)																
Mean (SD or SE, as indicated)	21	184.4 (SD 167.0)	24	217.7 (SD 154.5)	7	112.8 (SD 128.7)	9	48.2 (SD 63.3)					9	120 (SE 27)	9	143 (SE 19)
Median (lower quartile, upper quartile)									16	18.0 (12.0, 32.0)	14	21.0 (14.0, 41.0)				
Follow-up daily milk volume (mls)																
Mean (SD or SE, as indicated)	21	380.2 (SD 201.6)	24	250.8 (SD 171.6)	6	183.5 (NR)	8	66.1 (NR)					9	239 (SE 35)	9	172 (SE 39)
Median (lower quartile, upper quartile)									16	219 (116, 287)	14	102 (96, 113)				
Difference in mean change between groups (mls)																
Difference (95%CI)									88.6 (95%CI 32.5 to 144.8)							

SD = standard deviation; SE = standard error; CI = confidence interval

Practical exercise

Table 2: Individual patient data from Da Silva 2001

Table 2: Daily volumes of breast milk recorded by subjects in the domperidone and placebo groups

	Study day; milk volume, mL						
Patient	Baseline*	2	3	4	5	6	7
Domperidone							
1	13†	3	49	66	51	68	45
2	262	286	291	342	323	337	346
3	6	21	19	35	39	45	45
4	17	49	81	121	146	170	190
5	120	135	167	132	–	–	–
6	51	69	95	96	125	126	125
7	320	350	356	385	378	323	350
Mean volume	112.8	130.4	151.1	168.1	177.0	178.2	183.5
Placebo							
1	201†	152	180	102	133	80	110
2	12†	14	15	16	12	14	14
3	32†	36	36	32	36	40	40
4	14	37	62	55	42	57	58
5	11	28	18	8	4	0	0
6	50	92	82	120	121	150	180
7	89	102	94	111	95	101	107
8	23	21	22	22	–	–	–
9	< 1	10	20	25	27	15	19
Mean volume	48.2	54.7	58.8	54.6	58.8	57.1	66.1

*Defined as the volume of milk produced during the 24 hours before the start of the study medication.
†Baseline volume was unavailable; instead, volume produced within the 24 hours following enrolment was considered as the baseline.

Practical exercise: Questions to consider

- What are some of the features of the data reported?
- Is there enough information provided to calculate an estimate of intervention effect for each trial?
- What effect measure would you use?
- How would you calculate the effect estimate?
- What information might you need to calculated the standard error (or confidence interval) of the effect estimate?

Meta-analysis (1)

- Meta-analysis is a statistical analysis of the intervention effects from several studies leading to a quantitative summary.

- Two stages of meta-analysis:
    1. An observed intervention effect is calculated for each study (e.g. MD, SMD).
    2. A pooled intervention effect estimate is calculated as a weighted average of the intervention effects estimated in the individual studies.

Meta-analysis (2)

- A weighted average is calculated as:

$$\text{Weighted average} = \frac{\text{sum of (estimate x weight)}}{\text{sum of weights}}$$

- The weights reflect the amount of information that each study contributes
- The calculation of the weights differs depending on the effect measure (e.g. MD, SMD)
- Different types of meta-analysis models can be fitted (fixed effect versus random effects), and the weights can differ depending on the model

Meta-analysis: MD

- Use the mean difference when studies all report outcomes using the same scale.
- The weighting a study receives is based on the **SDs and sample size**.
  - For studies of the same size, those studies with smaller SDs will be given relatively more weight compared with studies with larger SDs.
  - For studies with the similar SDs, those studies with larger sample sizes will be given relatively more weight compared with studies with smaller sample sizes.

Review: Systemic corticosteroids for acute exacerbations of chronic obstructive pulmonary disease

Comparison: 01 Corticosteroid vs Placebo

Outcome: 05 Early FEV1 (litres)

Study	Corticosteroid N	Corticosteroid Mean(SD)	Placebo N	Placebo Mean(SD)	Weighted Mean Difference (Fixed) 95% CI	Weight (%)	Weighted Mean Difference (Fixed) 95% CI
Albert 1980	22	0.96 (0.24)	22	0.86 (0.31)		15.1	0.10 [-0.06, 0.26]
Bullard 1996	60	0.66 (0.59)	53	0.57 (0.74)		6.6	0.09 [-0.16, 0.34]
Davies 1999	29	0.96 (0.48)	25	0.77 (0.41)		7.2	0.19 [-0.05, 0.43]
Maltais 2002	57	1.15 (0.49)	58	0.90 (0.41)		14.9	0.25 [0.08, 0.42]
Niewoehner 1999	160	0.97 (0.39)	111	0.87 (0.39)		45.5	0.10 [0.01, 0.19]
Thompson 1996	13	1.27 (0.52)	14	0.91 (0.31)		3.8	0.36 [0.03, 0.69]
Wood-Baker 1997	18	0.82 (0.38)	10	0.71 (0.27)		6.9	0.11 [-0.13, 0.35]
Total	359		293			100.0	0.14 [0.08, 0.20]

Test for heterogeneity chi-square=4.75 df=6 p=0.58 I^2 =0.0%

Test for overall effect z=4.27 p=0.00002

Analysis 2.1. Comparison 2 Single dose: end value, Outcome 1 Weight (kg).

Review: Deworming drugs for treating soil-transmitted intestinal worms in children: effects on growth and school performance

Comparison: 2 Single dose: end value

Outcome: 1 Weight (kg)

Study or subgroup	Deworming N	Mean(SD)	Control N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
1 Albendazole							
Adams 1994	28	21.1 (2.43)	27	21.6 (3.9)		1.3 %	-0.50 [-2.22, 1.22]
Awasthi 2000	592	10.7 (2.1)	395	10.8 (1.9)		62.2 %	-0.10 [-0.35, 0.15]
Stephenson 1989	78	23.4 (3.8)	72	22.5 (4.41)		2.3 %	0.90 [-0.42, 2.22]
Stephenson 1993	96	31.3 (7.45)	93	31.3 (7.71)		0.8 %	0.0 [-2.16, 2.16]
Sur 2005	342	12.6 (5.73)	340	12.1 (4.61)		6.5 %	0.50 [-0.28, 1.28]
Watkins 1996	116	24.36 (3.99)	110	24.93 (4.2)		3.5 %	-0.57 [-1.64, 0.50]
Subtotal (95% CI)	**1252**		**1037**			**76.6 %**	**-0.05 [-0.27, 0.18]**

Heterogeneity: $Chi^2 = 5.21$, df = 5 (P = 0.39); $I^2 = 4\%$

Test for overall effect: Z = 0.40 (P = 0.69)

Problems with the MD

- Requires all the randomised trials to use the same scale for measuring the outcome.
- Can be unclear what constitutes a clinically important difference.
- The SDs are used to compute the weights
  - *this can be problematic if differences in SDs across trials is due to variability among study populations.*
- Examples include:
  - Trials with restricted eligibility criteria may be given more weight compared with trials with loose criteria (e.g. pragmatic vs explanatory trials).
  - Trials with longer term follow-up may be given less weight compared with those with short term follow-up.
- Measurements on the same scale may not always be comparable (e.g. health care costs between countries).

Meta-analysis: SMD

- Use the standardised mean difference when studies measure the same outcome using different scales.
- Note: sometimes “scale factors” are known and transformations can be directly made e.g. energy measured in kilojoules and calories.
- The SMD will be the same between trials if the difference in means is the same proportion of the SD.
  - Trial 1: MD = 2, SD = 10, SMD = 2/10 = 0.2
  - Trial 2: MD = 10, SD = 50, SMD = 10/50 = 0.2

Measurements on different scales

Trials comparing mental state at 12 months between ACT and standard care

Trial	ACT		Standard Care	
	N	mean (SD)	N	mean (SD)
Audini	30	41.4 (14.0)	28	42.3 (12.4)
Morse	37	0.95 (0.76)	35	0.89 (0.65)
Lehman	67	4.10 (0.83)	58	3.80 (0.87)

Measurements on different scales

Trials comparing mental state at 12 months between ACT and standard care

Trial	ACT		Standard Care		
	N	mean (SD)	N	mean (SD)	
Audini	30	41.4 (14.0)	28	42.3 (12.4)	Brief psychiatric rating scale
Morse	37	0.95 (0.76)	35	0.89 (0.65)	Brief symptom inventory
Lehman	67	4.10 (0.83)	58	3.80 (0.87)	Colorado symptom index

Measurements on different scales

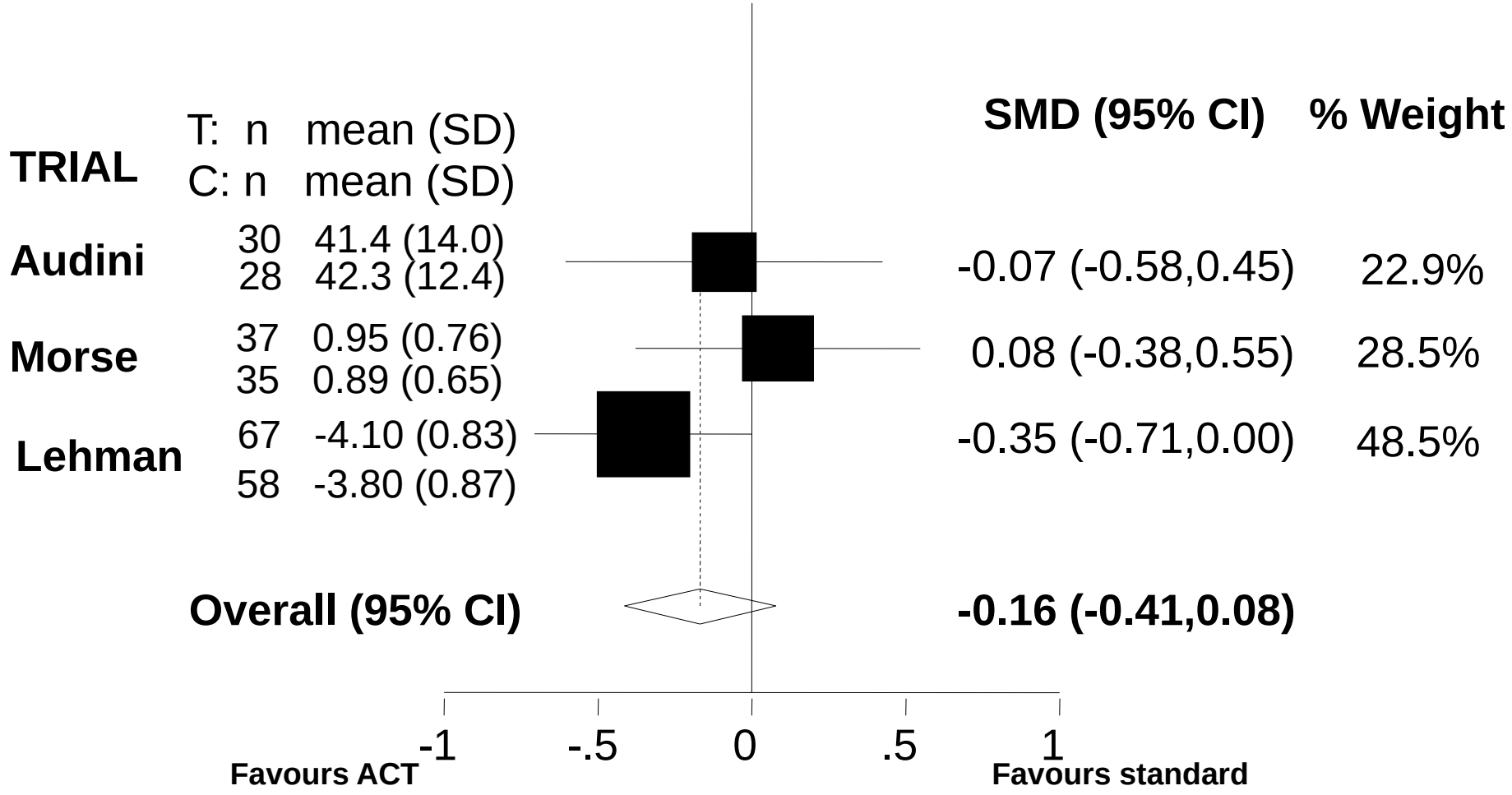
Trials comparing mental state at 12 months between ACT and standard care

Trial	ACT		Standard Care		
	N	mean (SD)	N	mean (SD)	
Audini	30	**41.4 (14.0)**	28	42.3 (12.4)	Brief psychiatric rating scale ↓
Morse	37	0.95 (0.76)	35	**0.89 (0.65)**	Brief symptom inventory ↓
Lehman	67	**4.10 (0.83)**	58	3.80 (0.87)	Colorado symptom index ↑

- High scores on the Colorado symptom index indicate good outcomes
- Low scores on the other two scales are good outcomes

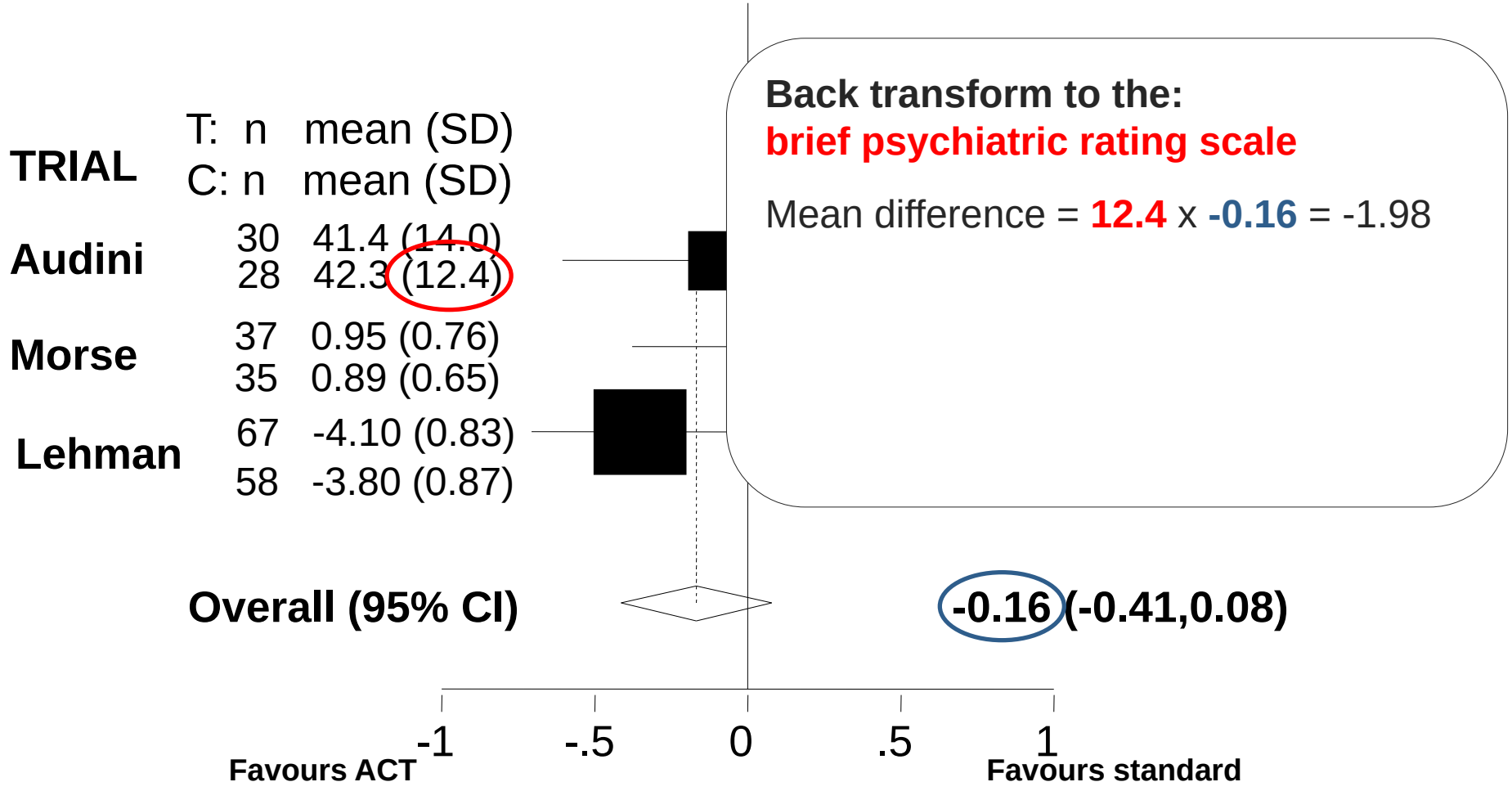
Comparison: ACT care versus Standard Care

Outcome: Measures of Mental State at 12 months



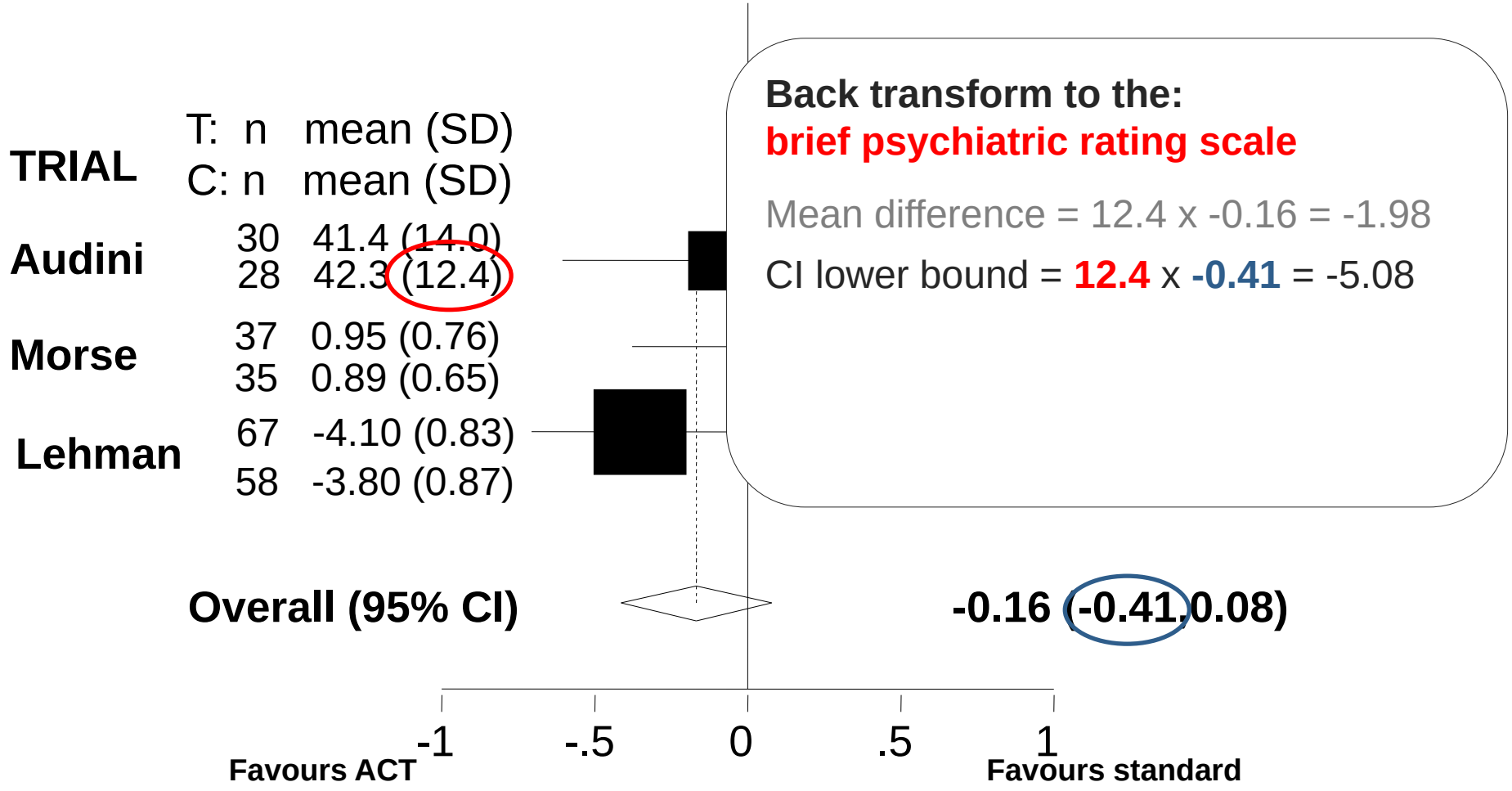
Comparison: ACT care versus Standard Care

Outcome: Measures of Mental State at 12 months



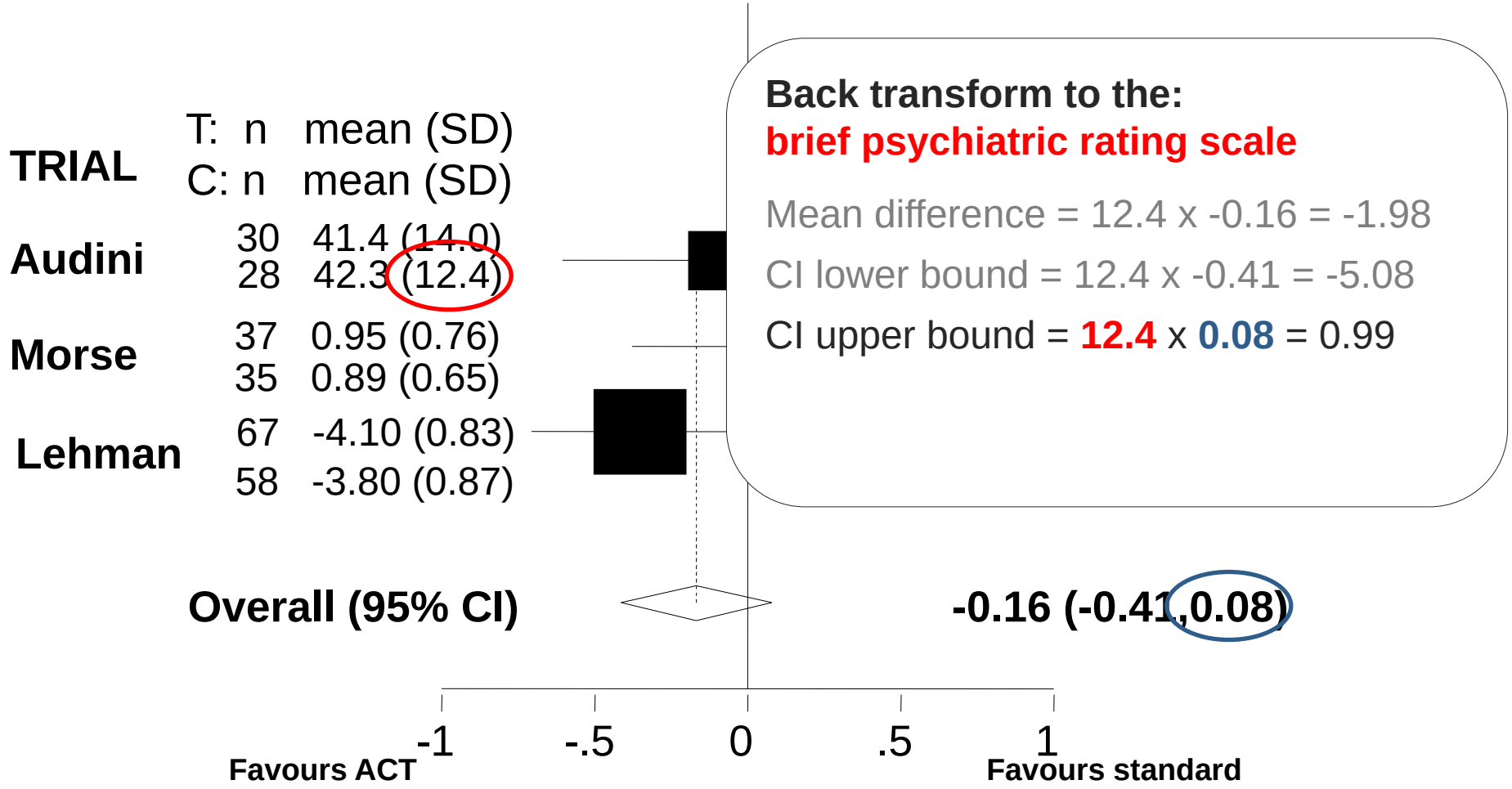
Comparison: ACT care versus Standard Care

Outcome: Measures of Mental State at 12 months



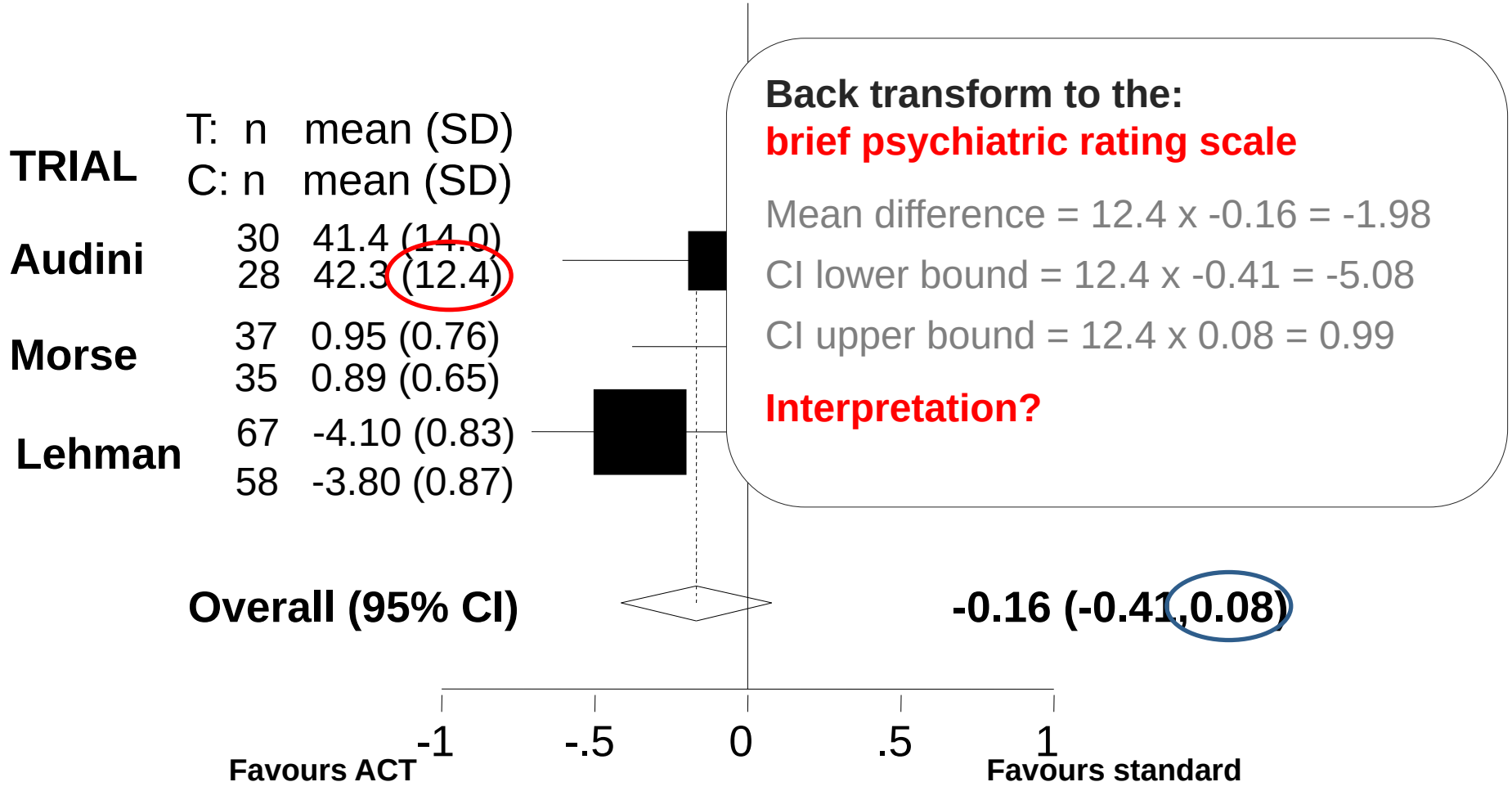
Comparison: ACT care versus Standard Care

Outcome: Measures of Mental State at 12 months



Comparison: ACT care versus Standard Care

Outcome: Measures of Mental State at 12 months



Back transform to the:
brief psychiatric rating scale

Mean difference = 12.4 x -0.16 = -1.98

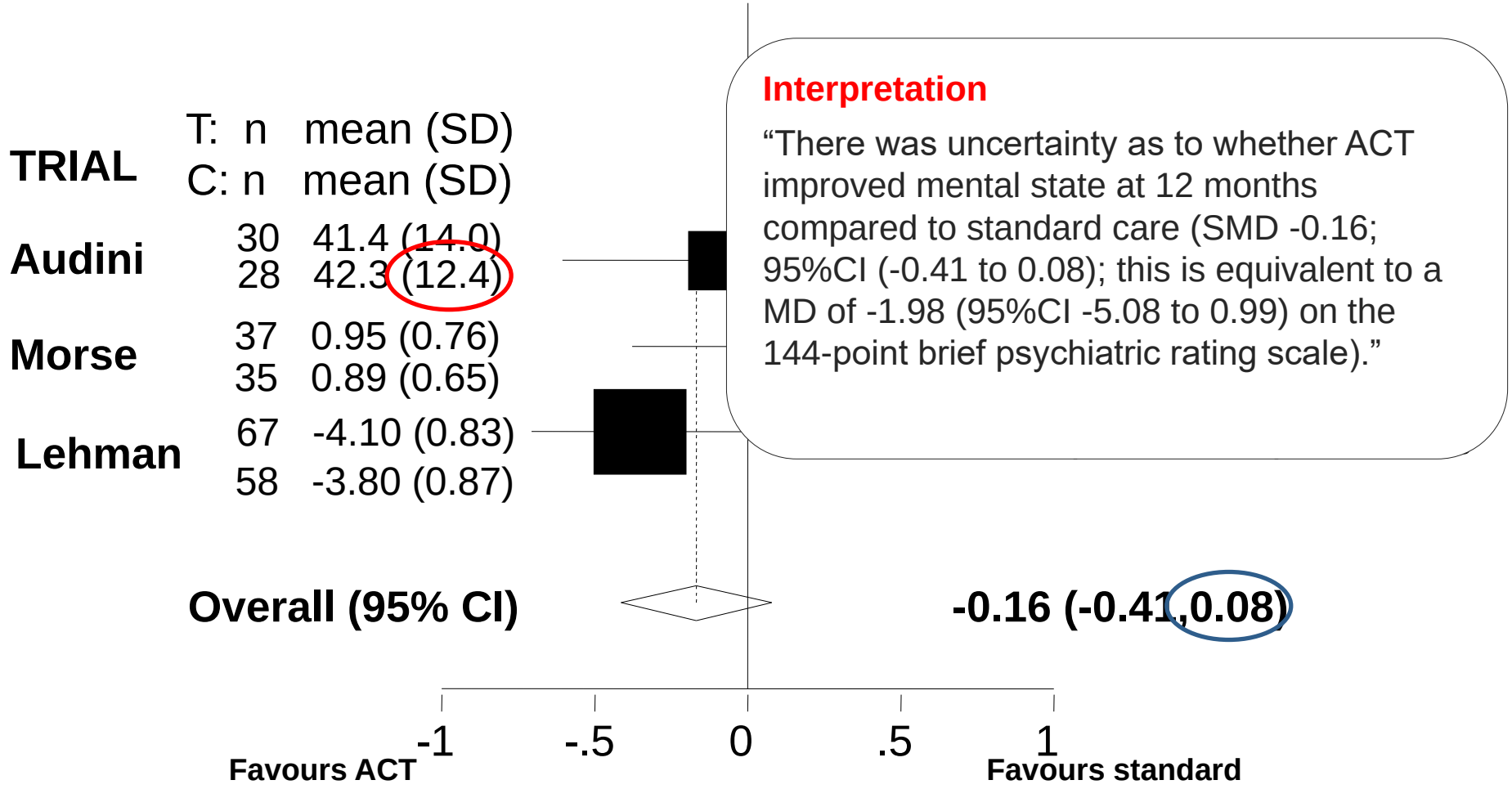
CI lower bound = 12.4 x -0.41 = -5.08

CI upper bound = 12.4 x 0.08 = 0.99

Interpretation?

Comparison: ACT care versus Standard Care

Outcome: Measures of Mental State at 12 months



Interpretation

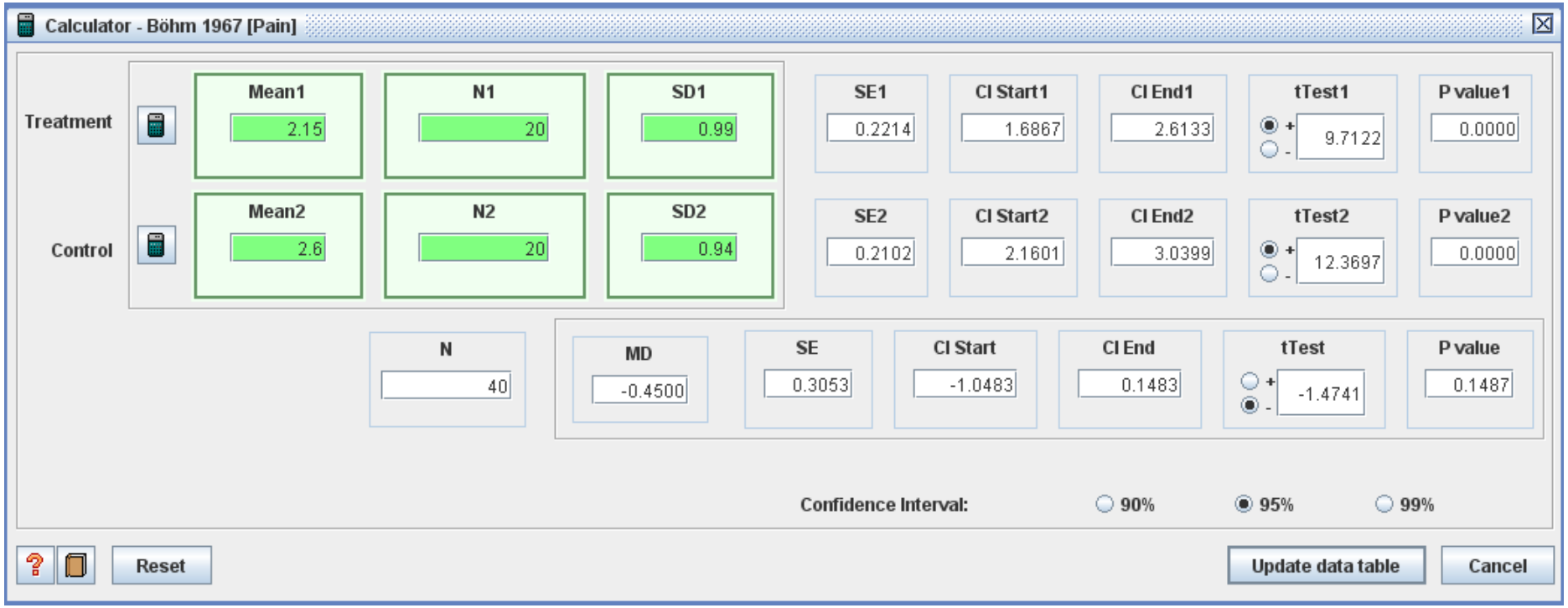
"There was uncertainty as to whether ACT improved mental state at 12 months compared to standard care (SMD -0.16; 95%CI (-0.41 to 0.08); this is equivalent to a MD of -1.98 (95%CI -5.08 to 0.99) on the 144-point brief psychiatric rating scale)."

Problems with the SMD

- Can be difficult to interpret outcomes in units of SD
  - *but can transform back to different scales (Section 12.6 of the CHSRI).*
- The SD may not always be a good scaling factor.
  - The method assumes that variation in SDs reflects only differences in measurement scales and *not* differences in the i) reliability of outcome measures or ii) variability among study populations.
  - Trials with restrictive eligibility criteria may have increased effect sizes as an artefact resulting from less variability between participants (i.e. smaller SDs).
- Need to remember to correct for differences in the direction of the scale (either through multiplication by -1, or subtraction of the mean from the max possible value (Section 9.2.3.2 CHSRI).

Data extraction

- Chapter 7 (Section 7.7.3) CHSRI.
- Two ways of entering continuous data in RevMan:
  - Entering means, SDs, and number of participants for the intervention and control groups.
  - Entering the intervention effect and its standard error.
- These methods cannot be used in combination. But …
  - RevMan 5.1 has a calculator which facilitates transformation between various statistics.


Calculator - Böhm 1967 [Pain]
Treatment
Mean1
2.15
N1
20
SD1
0.99
SE1
0.2214
CI Start1
1.6867
CI End1
2.6133
tTest1
9.7122
P value1
0.0000
Control
Mean2
2.6
N2
20
SD2
0.94
SE2
0.2102
CI Start2
2.1601
CI End2
3.0399
tTest2
12.3697
P value2
0.0000
N
40
MD
-0.4500
SE
0.3053
CI Start
-1.0483
CI End
0.1483
tTest
-1.4741
P value
0.1487
Confidence Interval:
90%
95%
99%
Reset
Update data table
Cancel

Data extraction

- Be careful to extract all reported statistics.
- While SDs may not be directly reported, they can be computed from:
  - standard errors
  - confidence intervals
  - t-tests
  - p-values from t and z tests.
- May need to search tables, text, and graphs for SDs.
- It may be the case that the SD needs to be imputed for some of the trials.
  - Contact the publication authors.
  - Use information on SDs from other trials.
  - Carry out sensitivity analyses investigating the effect of imputation.
  - Be careful to note in the review which SDs are imputed.
  - See Wiebe 2006 for options.

Detecting errors with SDs

- Confusion between the SD and the standard error is common (SE).
- The standard error is a measure of accuracy of an estimate of the mean.
- It is dependent on the sample size and variability of the data.

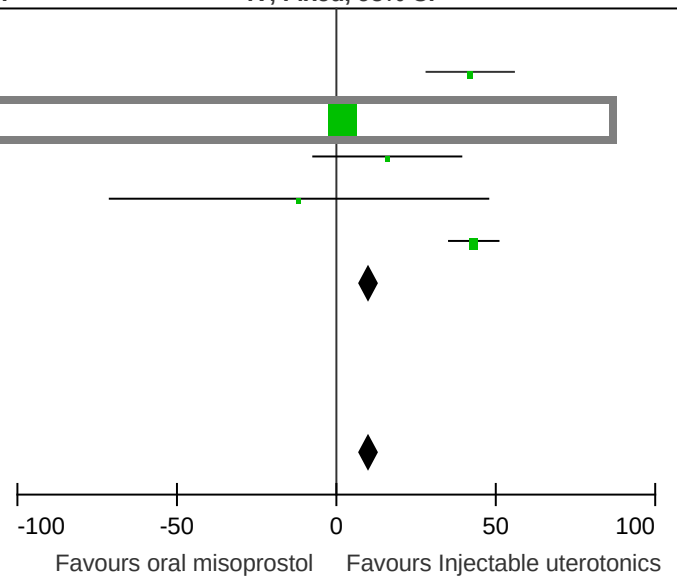
$$SE = \frac{SD}{\sqrt{n}} \qquad SD = \sqrt{n} \times SE$$

- Published reports may not specify whether a statistic is the SD or the SE, or use inappropriate terminology.

Prostaglandins for prevention of postpartum haemorrhage

Outcome = blood loss (ml)

Study or Subgroup	Oral misoprostol Mean	SD	Total	Injectable uterotonics Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.1.1 600 mcg								
Hong Kong 2001	296	160	1026	254	157	1032	4.7%	42.00 [28.30, 55.70]
Nigeria 2003	341	19.3	247	339	18.9	249	78.6%	2.00 [-1.36, 5.36]
Turkey 2003	328	152	388	312	176	384	1.6%	16.00 [-7.21, 39.21]
WHO 1999	340.9	295.08	199	352.6	309.59	200	0.3%	-11.70 [-71.04, 47.64]
WHO 2001	332.8	274.6	9213	289.7	262.1	9227	14.8%	43.10 [35.35, 50.85]
Subtotal (95% CI)			**11073**			**11092**	**100.0%**	**10.17 [7.19, 13.15]**
Heterogeneity: Chi² = 113.57, df = 4 (P < 0.00001); I² = 96%								
Test for overall effect: Z = 6.69 (P < 0.00001)								
Total (95% CI)			**11073**			**11092**	**100.0%**	**10.17 [7.19, 13.15]**
Heterogeneity: Chi² = 113.57, df = 4 (P < 0.00001); I² = 96%								
Test for overall effect: Z = 6.69 (P < 0.00001)								
Test for subgroup differences: Not applicable								



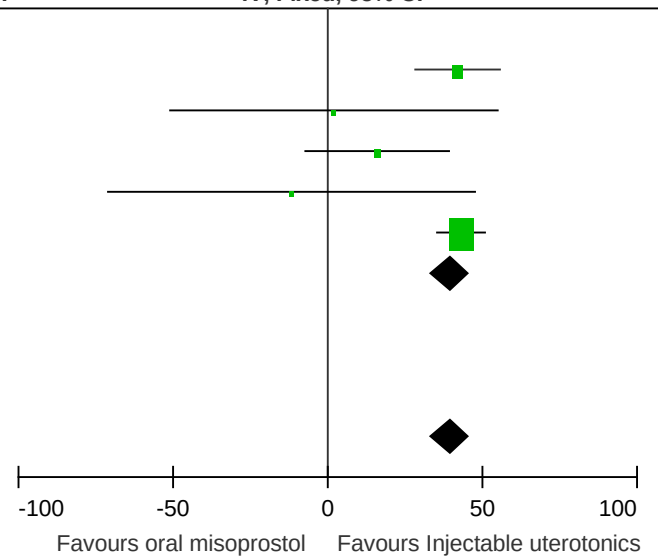
Oral misoprostol group: 19.3 x √247 = 303

Injectable uterotonics: 18.9 x √249 = 298

Prostaglandins for prevention of postpartum haemorrhage

Outcome = blood loss (ml)

Study or Subgroup	Oral misoprostol Mean	SD	Total	Injectable uterotonics Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.2.1 600 mcg								
Hong Kong 2001	296	160	1026	254	157	1032	21.8%	42.00 [28.30, 55.70]
Nigeria 2003	341	303.3233	247	339	298.237	249	1.5%	2.00 [-50.94, 54.94]
Turkey 2003	328	152	388	312	176	384	7.6%	16.00 [-7.21, 39.21]
WHO 1999	340.9	295.08	199	352.6	309.59	200	1.2%	-11.70 [-71.04, 47.64]
WHO 2001	332.8	274.6	9213	289.7	262.1	9227	68.0%	43.10 [35.35, 50.85]
Subtotal (95% CI)			**11073**			**11092**	**100.0%**	**39.57 [33.18, 45.96]**
Heterogeneity: Chi² = 9.68, df = 4 (P = 0.05); I² = 59%								
Test for overall effect: Z = 12.14 (P < 0.00001)								
Total (95% CI)			**11073**			**11092**	**100.0%**	**39.57 [33.18, 45.96]**
Heterogeneity: Chi² = 9.68, df = 4 (P = 0.05); I² = 59%								
Test for overall effect: Z = 12.14 (P < 0.00001)								
Test for subgroup differences: Not applicable								



MA1: 10ml (95%CI: 7, 13) greater blood loss when receiving OM

MA2: 40ml (95%CI: 33, 46) greater blood loss when receiving OM

MA1 I^2 = 96%; MA2 I^2 = 59%

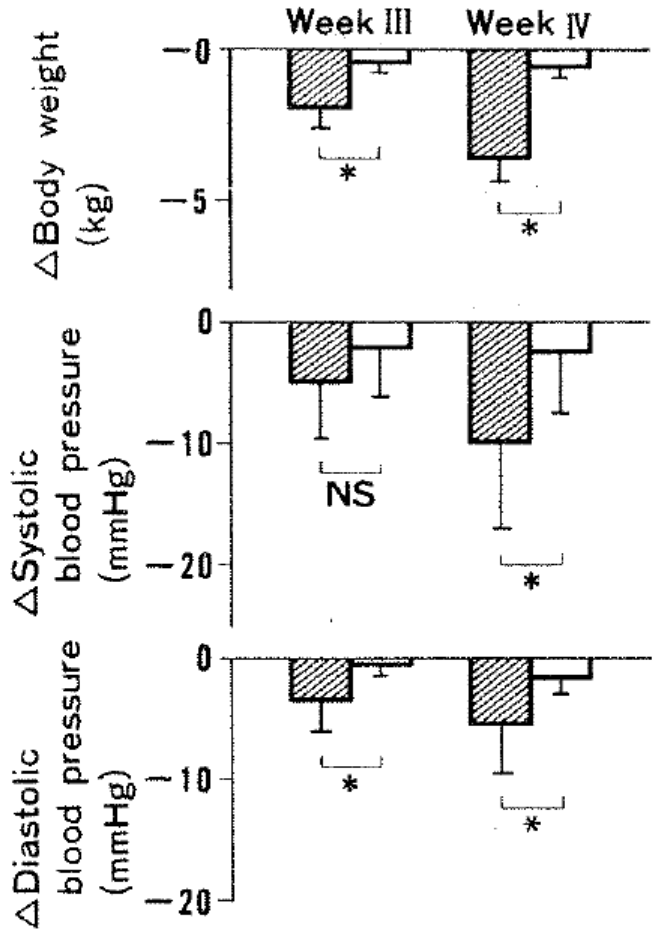


Fig. 1. Changes in body weight and blood pressure from baseline (week II) to week IV. ▨, Restricted-calorie group; □, unrestricted-calorie group. *$P<0.05$, restricted-calorie versus unrestricted-calorie group.

Detecting errors with SDs

- Clinical knowledge and common sense suggests SD is wrong.
- Size of SDs varies enormously across RCTs.
- Multiplying entered “SD” by √n results in more sensible (& consistent) SD (i.e. entered data likely to be SE).
- Examination of the statistics reported in the publication (e.g. p-values) may suggest “SDs” are really SEs.
- Weighting of one trial appears strangely large.
- The meta-analysis exhibits severe heterogeneity.

Other issues: FV and CS

- In some randomised trials the outcome will be collected at both baseline and follow-up.
- Comparisons between groups can then be made in several ways.
  - Calculating a mean difference of final values.
  - Calculating a mean difference of change from baseline.
  - Calculating an adjusted mean difference (using analysis of covariance).
- All methods estimate the same underlying intervention effect.

Analysing continuous outcomes in a trial

Three common approaches:

$$\hat{\theta}_{FV} = \bar{y}_{\text{int}} - \bar{y}_{ctrl}$$

$$\hat{\theta}_{CS} = (\bar{y}_{\text{int}} - \bar{y}_{ctrl}) - (\bar{x}_{\text{int}} - \bar{x}_{ctrl})$$

$$\hat{\theta}_{ANCOVA} = (\bar{y}_{\text{int}} - \bar{y}_{ctrl}) - \beta(\bar{x}_{\text{int}} - \bar{x}_{ctrl})$$

where $\beta = \rho \frac{\sigma_y}{\sigma_x}$

Other issues

Fortification randomised trial

	Hb (mean and SD)			
	Intervention (n = 271)	Control (n = 269)	Diff in means (95% CI)	P-value
Baseline	118.2 (10.5)	120.0 (9.0)	-1.8	
Analysis				
Follow-up	121.0 (10.1)	120.5 (9.5)	0.5 (-1.1, 2.2)	0.540
Change score	2.8 (8.3)	0.5 (8.4)	2.3 (0.9, 3.7)	0.002
ANCOVA			1.6 (0.4, 2.9)	0.012

Other issues: FV and CS

- In some randomised trials the outcome will be collected at both baseline and follow-up.
- Comparisons between groups can then be made in several ways.
  - Calculating a mean difference of final values.
  - Calculating a mean difference of change from baseline.
  - Calculating an adjusted mean difference (using analysis of covariance).
- All methods estimate the same underlying intervention effect.
- Therefore, we can combine the results from the different methods in one meta-analysis.
- The precision of the estimates will differ, depending on the correlation between the baseline measure of the outcome and the outcome.
- Can not use a mixture of the methods when using the SMD.

Other issues: non-normality

- Standard meta-analytic methods assume normality in the distribution of the means.
- Many outcomes are not normally distributed.

Other issues

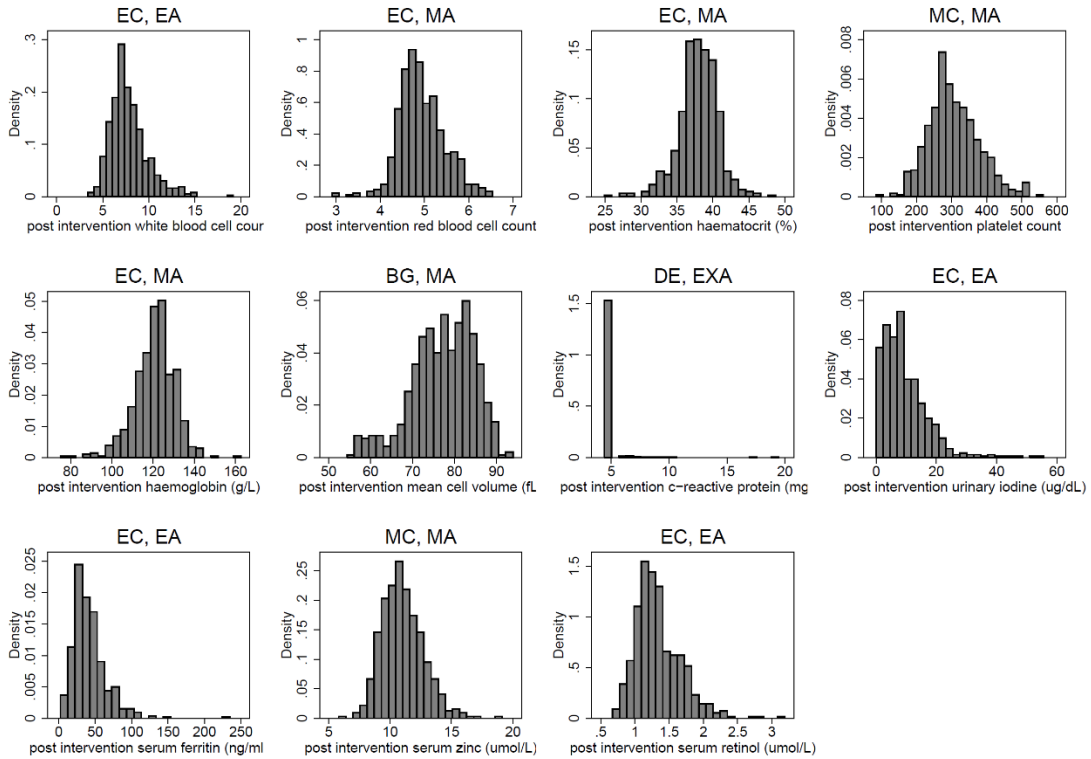


Figure 3: Histograms of post biochemical variables from Thailand RCT

First abbreviation = classification based on tail weight (U = Uniform, BG = Below Gaussian, G = Gaussian, MC = Moderate Contamination, EC = Extreme Contamination, DE = Double Exponential), second abbreviation = classification based on asymmetry (S = Near Symmetry, MA = Moderate Asymmetry, EA = Extreme Asymmetry, EXA = Exponential Asymmetry).

Other issues: non-normality

- Standard meta-analytic methods assume normality in the distribution of the means.
- Many outcomes are not normally distributed.
- Indications of skew include:
  - Geometric means, medians, interquartile ranges reported.
  - Large SD compared with the mean.
    - (mean – lowest possible score)/SD < 2 indicates skew
    - (highest possible score – mean)/SD < 2 indicates skew
- Methods are available to estimate parametric statistics (mean, SD) from non-parametric statistics (median, inter-quartile, range) (e.g. Hozo 2005; Wan 2014, Luo 2016), Weir 2018)

Other issues: non-normality

- In large trials, skewed distributions are not likely to be problematic.
- In small trials, may conduct the meta-analysis on the log-transformed scale (if this scale is believed to be less skewed) (Higgins 2008)
- Seek statistical support.

The Cochrane Collaboration

- CENTRAL provides the most comprehensive database of trials

www.cochrane.org

- Provides a free software for systematic reviews and meta-analyses (Review Manager; RevMan)

  – For a practical on RevMan see: https://www.youtube.com/watch?v=I6gqY5GkwMs

- See also the **Cochrane Handbook** (http://community.cochrane.org/handbook) that describes in detail the process of preparing and maintaining Cochrane systematic reviews on the effects of healthcare interventions.
  - For video about systematic reviews, also visit: http://www.cochrane.org/what-is-cochrane-evidence

Resources

- *Cochrane Handbook for Systematic Reviews of Interventions*
  -Higgins and Green (eds); Wiley 2008, updated online
- *RevMan Tutorial* and *User Guide*
  -http://tech.cochrane.org/revman/documentation
- *Introduction to Meta-analysis*
  -Borenstein, Hedge, Higgins and Rothwell; Wiley 2009
- *Meta-Analysis of Controlled Clinical Trials*
  -Whitehead; Wiley 2002
- 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
- *Handbook of Research Synthesis and Meta-analysis*
  -Cooper, Hedges and Valentine; Sage 2009

Resources

- Friedrich JO, Adhikari NK, Beyene J: The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. BMC Med Res Methodol 2008, 8:32.
- Friedrich JO, Adhikari NKJ, Beyene J: Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. Journal of Clinical Epidemiology 2011, 64:556-564.
- Higgins JP, White IR, Anzures-Cabrera J: Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales. Stat Med 2008.
- Hozo SP, Djulbegovic B, Hozo I: Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005, 5(1):13.
- Luo D, Wan X, Liu J, Tong T: Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res 2016.
- Wan X, Wang W, Liu J, Tong T: Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology 2014, 14(1):135.
- Weir CJ, Butcher I, Assi V, Lewis SC, Murray GD, Langhorne P, et al. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. BMC Medical Research Methodology. 2018;18(1):25.
- Wiebe N, Vandermeer B, Platt RW, Klassen TP, Moher D, Barrowman NJ: A systematic review identifies a lack of standardization in methods for handling missing variance data. J Clin Epidemiol 2006, 59:342-353.

Thank you for your attention!

Questions?