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

Study 1 → Data → Effect measure

Study 2 → Data → Effect measure

Study 3 → Data → Effect measure

Study 4 → Data → Effect measure

Study Level

Meta-analysis Level

Effect measure
To apply a meta-analysis

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

   \[
   \text{weight of study} = \frac{1}{\text{variance}}
   \]

2. Combine these using a weighted average:

   \[
   \text{pooled estimate} = \frac{\text{sum of (estimate} \times \text{weight)}}{\text{sum of weights}}
   \]

   \[
   \text{with variance} = \frac{1}{\text{sum of weights}}
   \]
Synthesis of findings - Forest plot

Does granisetron prevent nausea?

Odds ratio = 1.17 (0.53, 2.58)

Line of no effect

Estimate and confidence interval for each study. Squares $\propto$ study weights

Direction of effect

Favours Granisetron ← → Favours Placebo

Scale (effect measure)

Study ID

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Synthesis of findings - Forest plot

Study ID | OR (95% CI) | % Weight
--- | --- | ---
1 | 0.22 (0.05, 0.91) | 4.36
2 | 0.51 (0.33, 0.79) | 10.33
3 | 1.17 (0.53, 2.58) | 7.81
4 | 0.07 (0.02, 0.29) | 4.71
5 | 0.07 (0.02, 0.28) | 4.71
6 | 0.20 (0.02, 2.02) | 2.08
7 | 0.18 (0.01, 4.04) | 1.28
8 | 0.65 (0.33, 1.30) | 8.53
9 | 0.28 (0.16, 0.48) | 9.68
10 | 1.00 (0.06, 16.69) | 1.50
11 | 0.06 (0.01, 0.31) | 3.68
12 | 1.00 (0.06, 16.76) | 1.50
13 | 0.30 (0.03, 3.15) | 2.04
14 | 0.38 (0.09, 1.54) | 4.36
15 | 0.66 (0.18, 2.36) | 4.93
16 | 0.18 (0.05, 0.70) | 4.55
17 | 0.36 (0.10, 1.33) | 4.79
18 | 0.62 (0.22, 1.71) | 6.27
19 | 0.18 (0.02, 1.63) | 2.31
20 | 0.19 (0.05, 0.78) | 4.40
21 | 1.09 (0.39, 3.08) | 6.19

Overall (I-squared = 49.4%, p = 0.006)

NOTE: Weights are from random effects analysis

Estimate and confidence for the meta-analysis

Favours Granisetron  →  Favours Placebo
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
  
  o The risk of nausea with and without granisetron

<table>
<thead>
<tr>
<th></th>
<th>Nausea</th>
<th>Non-nausea</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granisetron</td>
<td>3</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>Placebo</td>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>13</td>
<td>47</td>
<td>200</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th></th>
<th>Nausea</th>
<th>Non-nausea</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Granisetron</strong></td>
<td>3</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td><strong>Placebo</strong></td>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>13</td>
<td>47</td>
<td>200</td>
</tr>
</tbody>
</table>

Relative measures
- Odds Ratio
- Risk Ratio

Absolute measure
- Risk Difference
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
  - ...

When the outcome for every participant is one of two possibilities or events.

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

When the outcome for every participant is one of two possibilities or events.
Dichotomous data

Consider a single study:

<table>
<thead>
<tr>
<th></th>
<th>Event</th>
<th>Non-Event</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td>a</td>
<td>b</td>
<td>m₁</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>c</td>
<td>d</td>
<td>m₂</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>N₁</td>
<td>N₂</td>
<td>N</td>
</tr>
</tbody>
</table>
Dichotomous data

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

<table>
<thead>
<tr>
<th></th>
<th>Dead</th>
<th>Alive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>10</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>Control</td>
<td>14</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>176</td>
<td>200</td>
</tr>
</tbody>
</table>
Effect measures for dichotomous data

• We can compare the two groups in several ways:
  o Odds ratio (OR)
  o Risk ratio (RR) = Relative Risk
  o 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?
  
  \[
  \frac{1 \text{ day of the week is Monday}}{7 \text{ days of the week}} = \frac{1}{7}
  \]

• 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?
  
  \[
  \frac{1}{6}
  \]
Risk

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

- **Risk** of a headache:
  
  \[
  \text{Risk} = \frac{\text{number of events of interest}}{\text{total number of observations}}
  \]

  \[
  = \frac{5}{20} = 0.25 = 25\%
  \]

- The risk of having a headache is 25% if you drink vodka.
• 20 people drank vodka, 5 developed a headache

• **Odds** of a headache
  
  \[ \text{Odds} = \frac{\text{number of events}}{\text{number of non events}} \]
  
  \[ = \frac{5}{15} = \frac{1}{3} \]

• 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 \textit{small} when the event is \textit{rare} but can be \textit{large} for \textit{common} events

\[
Risk = \frac{Odds}{1+Odds}
\]

\[
Odds = \frac{Risk}{1-Risk}
\]

<table>
<thead>
<tr>
<th>Event</th>
<th>Total</th>
<th>Risk</th>
<th>Odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>100</td>
<td>0.05</td>
<td>0.0526</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>95</td>
<td>100</td>
<td>0.95</td>
<td>19</td>
</tr>
</tbody>
</table>
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

Risk Ratio = \frac{\text{Risk in treatment group}}{\text{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

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

Odds Ratio allows us to compare between two groups
Risk ratio and odds ratio

<table>
<thead>
<tr>
<th></th>
<th>Event</th>
<th>Non-Event</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>Control</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>N</td>
</tr>
</tbody>
</table>

Risk ratio = \( \frac{\text{risk in treatment group}}{\text{risk in control group}} \)

Odds ratio = \( \frac{\text{odds in treatment group}}{\text{odds in control group}} \)

\[
\text{Risk Ratio} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}
\]

\[
\text{Odds Ratio} = \frac{\frac{a}{b}}{\frac{c}{d}}
\]
## Risk ratio

<table>
<thead>
<tr>
<th></th>
<th>Dead</th>
<th>Alive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td>10</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>14</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>24</td>
<td>176</td>
<td>200</td>
</tr>
</tbody>
</table>

Risk of event in **treatment** = $\frac{10}{100} = 0.10$

Risk of event in **control** = $\frac{14}{100} = 0.14$

Risk Ratio = $\frac{\frac{10}{100}}{\frac{14}{100}} = \frac{0.10}{0.14} = 0.71$
### Odds ratio

<table>
<thead>
<tr>
<th></th>
<th>Dead</th>
<th>Alive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td>10</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>14</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>24</td>
<td>176</td>
<td>200</td>
</tr>
</tbody>
</table>

Odds of event in **treatment**

\[
\text{Odds ratio} = \frac{10}{90} = \frac{14}{86} = 0.69
\]
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:

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

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

$$100 \times (1 - RR)\% = 75\%$$
Treatment effects on the log-scale: Why?

- RR, OR are *not* symmetric (take values from 0 to $+\infty$, 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 \((-\infty, \infty)\)
    - Log(RR) takes values in \((-\infty, \log(1/\text{CGR}))\)

✓ Typically the natural log transformation (log base e, written ‘ln’) is used
Log-risk ratio (LogRR)

\[ \log RR = \log \frac{a}{c} = \log \left( \frac{a}{c} \cdot \frac{c + d}{c(a + b)} \right) \]

\[ \text{var}(\log RR) = \frac{1}{a} + \frac{1}{a + b} + \frac{1}{c} + \frac{1}{c + d} \]

When \( \log RR = 0 \), there is no difference between the groups
Log-Risk Ratio (LogRR)

<table>
<thead>
<tr>
<th></th>
<th>Dead</th>
<th>Alive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>10</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>Control</td>
<td>14</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>176</td>
<td>200</td>
</tr>
</tbody>
</table>

Calculate Risk Ratio

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

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

Include in meta-analysis

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

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

Calculate a 95% C.I. for LogRR

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

Back-calculate to the original scale

95% CI for RR: $$(e^{-1.10}, e^{0.42}) = (0.33, 1.53)$$
Log-odds ratio (LogOR)

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

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

*When $logOR = 0$, there is no difference between the groups*
Risk difference (RD)

• The **difference in the risk** between the treated and control groups

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

• A measure **easy to understand** but clinical interpretation **depends on context**
  o 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
  o RD of -10% applied to a population with 7% Control Group Risk gives –3% Treatment Group Risk
Risk difference (RD)

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

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

When \( RD = 0 \), there is no difference between the groups
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:

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

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>RR</th>
<th>RD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>✗</td>
<td>✓</td>
<td>✓✓</td>
</tr>
<tr>
<td>Consistency</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Mathematics</td>
<td>✓✓</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

- 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 \textbf{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 \(\Theta_{IV}\) and \(SE(\Theta_{IV})\)
Inverse Variance Method

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

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

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

- $y_i$ can be $logOR$, $logRR$ or $RD$ 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
Let’s say we have the logOR and the variance for each of the 3 studies we identified in a systematic review.

<table>
<thead>
<tr>
<th>Study</th>
<th>( y_i ) (logOR)</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Study 2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Study 3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
Group Exercise: Inverse Variance Method Example

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

<table>
<thead>
<tr>
<th>study</th>
<th>$y_i$ (logOR)</th>
<th>variance</th>
<th>$w_i$</th>
<th>$w_iy_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\theta_{IV} = \frac{\sum w_i y_i}{\sum w_i}
\]

$w_i$ is the inverse of the variance
Try to calculate by hand the IV meta-analysis pooled result

<table>
<thead>
<tr>
<th>(y_i) (logOR)</th>
<th>variance</th>
<th>(w_i)</th>
<th>(w_iy_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>study 1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>study 2</td>
<td>2</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>study 3</td>
<td>4</td>
<td>0.5</td>
<td>2</td>
</tr>
</tbody>
</table>

\[\Sigma w_i = 2 \quad \Sigma (w_iy_i) = 5\]

\[SE(\Theta_{IV}) = \frac{1}{\sqrt{\Sigma w_i}}\]

\[\Sigma (w_iy_i) / \Sigma w_i = 5/2 = 2.5\]

\[1/\text{Sqrt}(\Sigma w_i) = 0.7\]
Group Exercise: Inverse Variance Method Example

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

<table>
<thead>
<tr>
<th>$y_i$ (logOR)</th>
<th>variance</th>
<th>$w_i$</th>
<th>$w_iy_i$</th>
<th>% weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>study 1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1/2=50%</td>
</tr>
<tr>
<td>study 2</td>
<td>2</td>
<td>0.5</td>
<td>1</td>
<td>0.5/2=25%</td>
</tr>
<tr>
<td>study 3</td>
<td>4</td>
<td>0.5</td>
<td>2</td>
<td>0.5/2=25%</td>
</tr>
</tbody>
</table>

$$\Sigma w_i = 2 \quad \Sigma (w_iy_i) = 5 \quad 100\%$$

$$\Sigma (w_iy_i) / \Sigma w_i = 5/2 = 2.5$$

$$1/\text{Sqrt}(\Sigma w_i) = 0.7$$
Group Exercise: Inverse Variance Method Example

<table>
<thead>
<tr>
<th>logOR (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.00 (0.04, 3.96)</td>
<td>50.00</td>
</tr>
<tr>
<td>2.00 (-0.77, 4.77)</td>
<td>25.00</td>
</tr>
<tr>
<td>4.00 (1.23, 6.77)</td>
<td>25.00</td>
</tr>
<tr>
<td>2.50 (1.11, 3.89)</td>
<td>100.00</td>
</tr>
</tbody>
</table>
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**
Perform a meta-analysis using
  • Inverse Variance
  • Mantel-Haenszel methods
  • Peto

It needs:
1. The $2 \times 2$ table $(a,b,c,d)$
   [data entry as dichotomous outcome]
2. A treatment effect (e.g. $\log OR$) and its standard error ($SE$)
   [data entry as generic inverse variance outcome]
Continuous data
Study level $\downarrow$

- **Study B**
  - Outcome data
  - Effect measure

- **Study C**
  - Outcome data
  - Effect measure

- **Study D**
  - Outcome data
  - Effect measure

Review level $\downarrow$

Source: Jo McKenzie & Miranda Cumpston
Types of continuous data

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

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

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

• Time-to-event
  o Time until an event occurs
  o E.g. time to relapse of cancer, time to walk 100m
How common are continuous outcomes?

# 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.

<table>
<thead>
<tr>
<th></th>
<th>Fortified group</th>
<th>Unfortified group</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>271</td>
<td>269</td>
</tr>
<tr>
<td>Mean Hb (g/L)</td>
<td>121.0</td>
<td>120.5</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>10.1</td>
<td>9.5</td>
</tr>
</tbody>
</table>
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

Graphs by Group

control intervention

post intervention haemoglobin (g/L)

Density
normal p_hb
Comparing different SDs
Effect measures: Mean difference (MD)

• The MD is calculated as:
  \[
  \text{MD} = \text{mean outcome in intervention group} - \text{mean outcome in control group}
  \]

<table>
<thead>
<tr>
<th></th>
<th>Fortified group</th>
<th>Unfortified group</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>271</td>
<td>269</td>
</tr>
<tr>
<td>Mean Hb</td>
<td>121.0</td>
<td>120.5</td>
</tr>
<tr>
<td>SD</td>
<td>10.1</td>
<td>9.5</td>
</tr>
<tr>
<td>Mean Difference (95% CI)</td>
<td>0.5 (-1.1, 2.2)</td>
<td></td>
</tr>
</tbody>
</table>

• 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
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:

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

<table>
<thead>
<tr>
<th></th>
<th>Fortified group</th>
<th>Unfortified group</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>271</td>
<td>269</td>
</tr>
<tr>
<td>Mean Hb</td>
<td>121.0</td>
<td>120.5</td>
</tr>
<tr>
<td>SD</td>
<td>10.1</td>
<td>9.5</td>
</tr>
<tr>
<td>Ratio of means (95% CI)</td>
<td>1.005 (0.991, 1.018)</td>
<td></td>
</tr>
</tbody>
</table>

- 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).
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)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Domperidone</td>
<td>Placebo</td>
<td>Domperidone</td>
<td>Placebo</td>
</tr>
<tr>
<td></td>
<td>Stats</td>
<td>Stats</td>
<td>Stats</td>
<td>Stats</td>
</tr>
<tr>
<td>Baseline daily milk volume (mls)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD or SE, as indicated)</td>
<td>21</td>
<td>184.4 (SD 167.0)</td>
<td>24</td>
<td>217.7 (SD 154.5)</td>
</tr>
<tr>
<td>Median (lower quartile, upper quartile)</td>
<td>7</td>
<td>112.8 (SD 128.7)</td>
<td>9</td>
<td>48.2 (SD 63.3)</td>
</tr>
<tr>
<td>Follow-up daily milk volume (mls)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD or SE, as indicated)</td>
<td>21</td>
<td>380.2 (SD 201.6)</td>
<td>24</td>
<td>250.8 (SD 171.6)</td>
</tr>
<tr>
<td>Median (lower quartile, upper quartile)</td>
<td>6</td>
<td>183.5 (NR)</td>
<td>8</td>
<td>66.1 (NR)</td>
</tr>
<tr>
<td>Difference in mean change between groups (mls)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (95%CI)</td>
<td></td>
<td></td>
<td></td>
<td>88.6 (95%CI 32.5 to 144.8)</td>
</tr>
</tbody>
</table>

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

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

*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:

\[
\text{Weighted average} = \frac{\text{sum of (estimate } \times \text{ weight)}}{\text{sum of weights}}
\]

• 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.
  
  o For studies of the same size, those studies with smaller SDs will be given relatively more weight compared with studies with larger SDs.

  o 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)

<table>
<thead>
<tr>
<th>Study</th>
<th>Corticosteroid</th>
<th>Placebo</th>
<th>Weighted Mean Difference (Fixed)</th>
<th>Weight</th>
<th>Weighted Mean Difference (Fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (SD)</td>
<td>N</td>
<td>Mean (SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td>Albert 1980</td>
<td>22</td>
<td>0.96 (0.24)</td>
<td>22</td>
<td>0.86 (0.31)</td>
<td></td>
</tr>
<tr>
<td>Bullard 1996</td>
<td>60</td>
<td>0.66 (0.59)</td>
<td>53</td>
<td>0.57 (0.74)</td>
<td></td>
</tr>
<tr>
<td>Davies 1999</td>
<td>29</td>
<td>0.96 (0.48)</td>
<td>25</td>
<td>0.77 (0.41)</td>
<td></td>
</tr>
<tr>
<td>Maltas 2002</td>
<td>57</td>
<td>1.15 (0.49)</td>
<td>58</td>
<td>0.90 (0.41)</td>
<td></td>
</tr>
<tr>
<td>Niewoehner 1999</td>
<td>160</td>
<td>0.97 (0.39)</td>
<td>111</td>
<td>0.87 (0.39)</td>
<td></td>
</tr>
<tr>
<td>Thompson 1996</td>
<td>13</td>
<td>1.27 (0.52)</td>
<td>14</td>
<td>0.91 (0.31)</td>
<td></td>
</tr>
<tr>
<td>Wood-Baker 1997</td>
<td>18</td>
<td>0.82 (0.38)</td>
<td>10</td>
<td>0.71 (0.27)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>359</td>
<td></td>
<td>293</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity chi-square = 4.75 df = 6 p = 0.58 P = 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)

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Deworming</th>
<th>Control</th>
<th>Mean Difference</th>
<th>Weight</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N Mean(SD)</td>
<td>N Mean(SD)</td>
<td>IVFixed,95% CI</td>
<td></td>
<td>IVFixed,95% CI</td>
</tr>
<tr>
<td>Alendazole</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adams 1994</td>
<td>28 21.1 (2.43)</td>
<td>27 21.6 (3.9)</td>
<td></td>
<td>1.3 %</td>
<td>-0.50 [-2.22, 1.22]</td>
</tr>
<tr>
<td>Awasti 2000</td>
<td>592 10.7 (2.1)</td>
<td>395 10.8 (1.9)</td>
<td></td>
<td>62.2 %</td>
<td>-0.10 [-0.35, 0.15]</td>
</tr>
<tr>
<td>Stephenson 1989</td>
<td>78 23.4 (3.8)</td>
<td>72 22.5 (4.41)</td>
<td></td>
<td>2.3 %</td>
<td>0.90 [-0.42, 2.22]</td>
</tr>
<tr>
<td>Stephenson 1993</td>
<td>96 31.3 (7.45)</td>
<td>93 31.3 (7.71)</td>
<td></td>
<td>0.8 %</td>
<td>0.0 [-2.16, 2.16]</td>
</tr>
<tr>
<td>Sur 2005</td>
<td>342 12.6 (5.73)</td>
<td>340 12.1 (4.61)</td>
<td></td>
<td>6.5 %</td>
<td>0.50 [-0.28, 1.28]</td>
</tr>
<tr>
<td>Watkins 1996</td>
<td>116 24.36 (3.99)</td>
<td>110 24.93 (4.2)</td>
<td></td>
<td>3.5 %</td>
<td>-0.57 [-1.64, 0.50]</td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td><strong>1252</strong></td>
<td><strong>1037</strong></td>
<td><strong>76.6 %</strong></td>
<td><strong>-0.05 [-0.27, 0.18]</strong></td>
<td></td>
</tr>
</tbody>
</table>
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
  o this can be problematic if differences in SDs across trials is due to variability among study populations.

• Examples include:
  o Trials with restricted eligibility criteria may be given more weight compared with trials with loose criteria (e.g. pragmatic vs explanatory trials).
  o 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
## Trials comparing mental state at 12 months between ACT and standard care

<table>
<thead>
<tr>
<th>Trial</th>
<th>ACT</th>
<th>Standard Care</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>mean (SD)</td>
</tr>
<tr>
<td>Audini</td>
<td>30</td>
<td>41.4 (14.0)</td>
</tr>
<tr>
<td>Morse</td>
<td>37</td>
<td>0.95 (0.76)</td>
</tr>
<tr>
<td>Lehman</td>
<td>67</td>
<td>4.10 (0.83)</td>
</tr>
</tbody>
</table>
Measures of mental state at 12 months between ACT and standard care

<table>
<thead>
<tr>
<th>Trial</th>
<th>ACT</th>
<th>Standard Care</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N mean (SD)</td>
<td>N mean (SD)</td>
</tr>
<tr>
<td>Audini</td>
<td>30 41.4 (14.0)</td>
<td>28 42.3 (12.4)</td>
</tr>
<tr>
<td>Morse</td>
<td>37 0.95 (0.76)</td>
<td>35 0.89 (0.65)</td>
</tr>
<tr>
<td>Lehman</td>
<td>67 4.10 (0.83)</td>
<td>58 3.80 (0.87)</td>
</tr>
</tbody>
</table>

Brief psychiatric rating scale
Brief symptom inventory
Colorado symptom index
### Trials comparing mental state at 12 months between ACT and standard care

<table>
<thead>
<tr>
<th>Trial</th>
<th>ACT N</th>
<th>mean (SD)</th>
<th>Standard Care N</th>
<th>mean (SD)</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audini</td>
<td>30</td>
<td>41.4 (14.0)</td>
<td>28</td>
<td>42.3 (12.4)</td>
<td>Brief psychiatric rating scale</td>
</tr>
<tr>
<td>Morse</td>
<td>37</td>
<td>0.95 (0.76)</td>
<td>35</td>
<td>0.89 (0.65)</td>
<td>Brief symptom inventory</td>
</tr>
<tr>
<td>Lehman</td>
<td>67</td>
<td>4.10 (0.83)</td>
<td>58</td>
<td>3.80 (0.87)</td>
<td>Colorado symptom index</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>TRIAL</th>
<th>T: n</th>
<th>mean (SD)</th>
<th>SMD (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audini</td>
<td>30</td>
<td>41.4 (14.0)</td>
<td>-0.07 (-0.58, 0.45)</td>
<td>22.9%</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>42.3 (12.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morse</td>
<td>37</td>
<td>0.95 (0.76)</td>
<td>0.08 (-0.38, 0.55)</td>
<td>28.5%</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.89 (0.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lehman</td>
<td>67</td>
<td>-4.10 (0.83)</td>
<td>-0.35 (-0.71, 0.00)</td>
<td>48.5%</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>-3.80 (0.87)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall (95% CI) -0.16 (-0.41, 0.08)

Standardised Mean Difference

Favours ACT

Favours standard
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

T: n  mean (SD)
30  41.4 (14.0)
28  42.3 (12.4)

C: n  mean (SD)
37  0.95 (0.76)
35  0.89 (0.65)

Audini

Morse

Lehman

Overall (95% CI)

-0.16 (-0.41, 0.08)

Standardised Mean Difference

Favours ACT

Favours standard

87
Comparison: ACT care versus Standard Care

Outcome: Measures of Mental State at 12 months

<table>
<thead>
<tr>
<th>Trial</th>
<th>T: n</th>
<th>mean (SD)</th>
<th>C: n</th>
<th>mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audini</td>
<td>30</td>
<td>41.4 (14.0)</td>
<td>28</td>
<td>42.3 (12.4)</td>
</tr>
<tr>
<td>Morse</td>
<td>37</td>
<td>0.95 (0.76)</td>
<td>35</td>
<td>0.89 (0.65)</td>
</tr>
<tr>
<td>Lehman</td>
<td>67</td>
<td>-4.10 (0.83)</td>
<td>58</td>
<td>-3.80 (0.87)</td>
</tr>
</tbody>
</table>

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

Overall (95% CI)

-0.16 (-0.41, 0.08)

Standardised Mean Difference
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

Standardised Mean Difference

<table>
<thead>
<tr>
<th>Trial</th>
<th>T: n</th>
<th>mean (SD)</th>
<th>C: n</th>
<th>mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audini</td>
<td>30</td>
<td>41.4 (14.0)</td>
<td>28</td>
<td>42.3 (12.4)</td>
</tr>
<tr>
<td>Morse</td>
<td>37</td>
<td>0.95 (0.76)</td>
<td>35</td>
<td>0.89 (0.65)</td>
</tr>
<tr>
<td>Lehman</td>
<td>67</td>
<td>-4.10 (0.83)</td>
<td>58</td>
<td>-3.80 (0.87)</td>
</tr>
<tr>
<td>Overall (95% CI)</td>
<td>-0.16 (-0.41,0.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison: ACT care versus Standard Care

Outcome: Measures of Mental State at 12 months

### TRIAL

<table>
<thead>
<tr>
<th></th>
<th>T: n</th>
<th>mean (SD)</th>
<th>C: n</th>
<th>mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audini</td>
<td>30</td>
<td>41.4 (14.0)</td>
<td>28</td>
<td>42.3 (12.4)</td>
</tr>
<tr>
<td>Morse</td>
<td>37</td>
<td>0.95 (0.76)</td>
<td>35</td>
<td>0.89 (0.65)</td>
</tr>
<tr>
<td>Lehman</td>
<td>67</td>
<td>-4.10 (0.83)</td>
<td>58</td>
<td>-3.80 (0.87)</td>
</tr>
</tbody>
</table>

Overall (95% CI)

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
  o *but can transform back to different scales (Section 12.6 of the CHSRI).*

• The SD may not always be a good scaling factor.
  o 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.
  o 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.
Data extraction

• Be careful to extract all reported statistics.
• While SDs may not be directly reported, they can be computed from:
  o standard errors
  o confidence intervals
  o t-tests
  o 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.
  o Contact the publication authors.
  o Use information on SDs from other trials.
  o Carry out sensitivity analyses investigating the effect of imputation.
  o Be careful to note in the review which SDs are imputed.
  o 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.

\[
SE = \frac{SD}{\sqrt{n}} \quad SD = \sqrt{n} \times SE
\]

- 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)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Oral misoprostol</th>
<th>Injectable uterotonic</th>
<th>Mean Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1 600 mcg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong 2001</td>
<td>296</td>
<td>160</td>
<td>254</td>
<td>157</td>
</tr>
<tr>
<td>Nigeria 2003</td>
<td>341</td>
<td>19.3</td>
<td>339</td>
<td>18.9</td>
</tr>
<tr>
<td>Turkey 2003</td>
<td>328</td>
<td>152</td>
<td>312</td>
<td>176</td>
</tr>
<tr>
<td>WHO 1999</td>
<td>340.9</td>
<td>295.08</td>
<td>352.6</td>
<td>309.59</td>
</tr>
<tr>
<td>WHO 2001</td>
<td>332.8</td>
<td>274.6</td>
<td>289.7</td>
<td>262.1</td>
</tr>
</tbody>
</table>

Subtotal (95% CI) 11073 11092 100.0% 10.17 [7.19, 13.15]

Heterogeneity: $\chi^2 = 113.57$, df (P < 0.00001); $I^2 = 96$

Test for overall effect: $Z = 6.69$ (P < 0.00001)

Total (95% CI) 11073 11092 100.0% 10.17 [7.19, 13.15]

Heterogeneity: $\chi^2 = 113.57$, df (P < 0.00001); $I^2 = 96$

Test for overall effect: $Z = 6.69$ (P < 0.00001)

Test for subgroup differences: Not applicable

Oral misoprostol group: $19.3 \times \sqrt{247} = 303$

Injectable uterotonic: $18.9 \times \sqrt{249} = 298$
## Prostaglandins for prevention of postpartum haemorrhage

**Outcome = blood loss (ml)**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Oral misoprostol</th>
<th>Injectable uterotonics</th>
<th>Mean Difference</th>
<th>IV, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
<td>Mean</td>
</tr>
<tr>
<td>1.2.1 600 mcg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong 2001</td>
<td>296</td>
<td>160</td>
<td>1026</td>
<td>254</td>
</tr>
<tr>
<td>Nigeria 2003</td>
<td>341</td>
<td>303.323</td>
<td>247</td>
<td>339</td>
</tr>
<tr>
<td>Turkey 2003</td>
<td>328</td>
<td>152</td>
<td>388</td>
<td>312</td>
</tr>
<tr>
<td>WHO 1999</td>
<td>340.9</td>
<td>295.08</td>
<td>199</td>
<td>352.6</td>
</tr>
<tr>
<td>WHO 2001</td>
<td>332.8</td>
<td>274.6</td>
<td>9213</td>
<td>289.7</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>11073</td>
<td></td>
<td></td>
<td>11092</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11073</td>
<td></td>
<td></td>
<td>11092</td>
</tr>
</tbody>
</table>

Heterogeneity: $\chi^2 = 9.68$, df = 4 (P = 0.05); $I^2 = 59\%$

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 III) to week IV. Restricted-calorie group; † unrestrictedor-calorie group. *P < 0.05, restricted-calorie group 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 $\sqrt{n}$ 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:

\[ \hat{\theta}_{FV} = \bar{y}_{int} - \bar{y}_{ctrl} \]

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

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

where

\[ \beta = \rho \frac{\sigma_y}{\sigma_x} \]
Other issues

Fortification randomised trial

<table>
<thead>
<tr>
<th></th>
<th>Intervention (n = 271)</th>
<th>Control (n = 269)</th>
<th>Diff in means (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>118.2 (10.5)</td>
<td>120.0 (9.0)</td>
<td>-1.8</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td>121.0 (10.1)</td>
<td>120.5 (9.5)</td>
<td>0.5 (-1.1, 2.2)</td>
<td>0.540</td>
</tr>
<tr>
<td>Change score</td>
<td>2.8 (8.3)</td>
<td>0.5 (8.4)</td>
<td>2.3 (0.9, 3.7)</td>
<td>0.002</td>
</tr>
<tr>
<td>ANCOVA</td>
<td></td>
<td></td>
<td>1.6 (0.4, 2.9)</td>
<td>0.012</td>
</tr>
</tbody>
</table>
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.
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.
    - \(\frac{(\text{mean} - \text{lowest possible score})}{SD} < 2\) indicates skew
    - \(\frac{(\text{highest possible score} - \text{mean})}{SD} < 2\) 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=l6gqY5GkwMs

• 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.
  o 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](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](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 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.
Thank you for your attention!

Questions?