Cochrane Training

Assessing the certainty of evidence

Trusted evidence.
Informed decisions.
Better health.

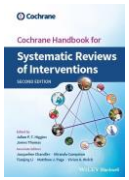
Cochrane Training

Steps of a Cochrane Review

1. define the question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. **interpret results and draw conclusions**
10. improve and update review

Session outline

- **a reminder on how to read and understand meta-analysis results**
- GRADE approach to assessing certainty of evidence
- factors affecting certainty



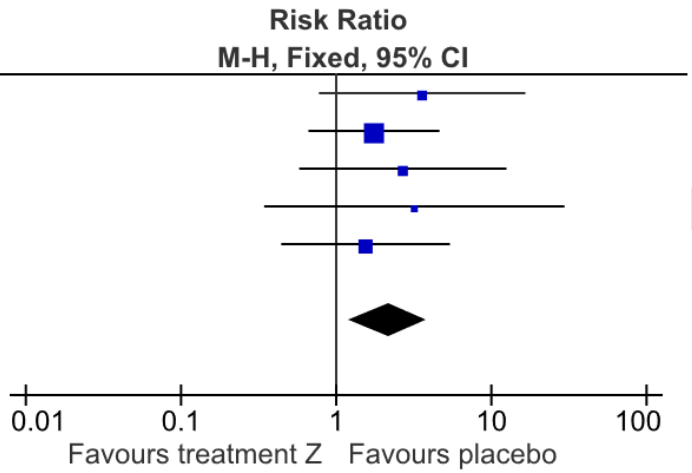
See Chapters 14 & 15 of the Handbook

Understanding dichotomous outcomes

Study or Subgroup	Treatment Z Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Study 1	8	76	2	68	13.5%	3.58 [0.79, 16.27]
Study 3	12	169	6	148	40.8%	1.75 [0.67, 4.55]
Study 4	7	56	2	43	14.4%	2.69 [0.59, 12.29]
Study 6	3	32	1	34	6.2%	3.19 [0.35, 29.09]
Study 7	6	98	4	101	25.1%	1.55 [0.45, 5.31]
Total (95% CI)		**431**		**394**	**100.0%**	**2.17 [1.20, 3.91]**
Total events	36		15			

Heterogeneity: Chi² = 1.09, df = 4 (P = 0.90); I² = 0%

Test for overall effect: Z = 2.58 (P = 0.010)



Outcome: nausea

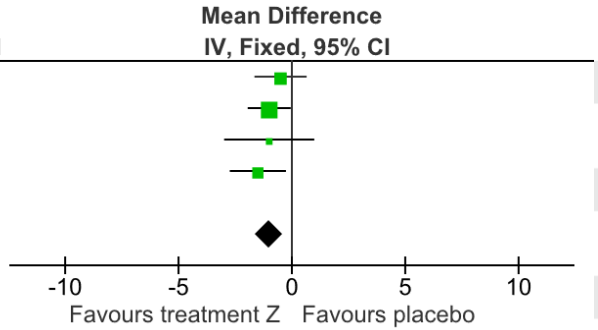
What does this Risk Ratio (RR) mean?

The risk of nausea with treatment Z is 2.17 times the risk when taking placebo. The effect could range from 1.2 to 3.91 times.

Understading continuous outcomes

Study or Subgroup	Treatment Z Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Study 1	-1	3.3	73	-0.5	3.5	72	27.2%	-0.50 [-1.61, 0.61]
Study 3	-2	3.7	167	-1	4.4	147	40.6%	-1.00 [-1.91, -0.09]
Study 6	4	4	32	5	4	33	8.8%	-1.00 [-2.95, 0.95]
Study 7	5.5	4.4	98	7	4.2	101	23.3%	-1.50 [-2.70, -0.30]
Total (95% CI)			**370**			**353**	**100.0%**	**-0.98 [-1.56, -0.40]**

Heterogeneity: Chi² = 1.45, df = 3 (P = 0.69); I² = 0%
Test for overall effect: Z = 3.33 (P = 0.0009)



Outcome: pain, on a scale from 1 (no pain) to 10 (severe pain)

What does this Mean Difference (MD) mean?

Pain may be reduced by 0.98 points when taking treatment Z in comparison to taking placebo. The effect could range from reduction by 1.56 points to 0.40 points.

NARRATIVE SYNTHESIS

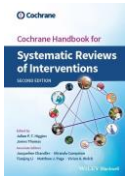
- when synthesising results is appropriate
- but statistical analysis cannot be performed

Example:

Three studies found that the risk of headaches is increased when taking the intervention compared to no intervention.

Session outline

- a reminder on how to read and understand meta-analysis results
- **GRADE approach to assessing certainty of evidence**
- factors affecting certainty



See Chapters 14 & 15 of the Handbook

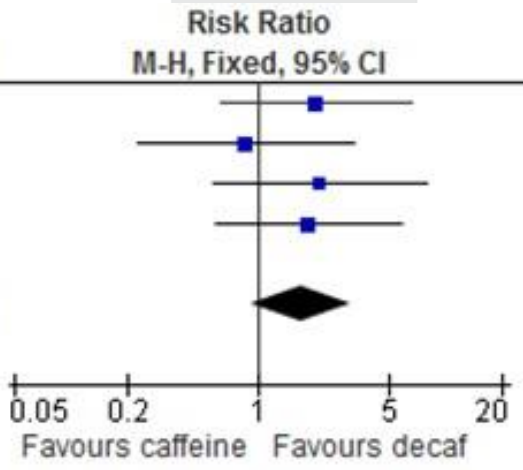
An example: Caffeine review

Outcome: sleep disruption

Study or Subgroup	Caffeinated coffee Events	Total	Decaffeinated coffee Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mama-Kaffa 1999	7	53	4	61	26.1%	2.01 [0.62, 6.50]
Morrocona 1998	3	15	4	17	26.3%	0.85 [0.23, 3.20]
Oohlahlazza 1998	6	35	3	37	20.4%	2.11 [0.57, 7.81]
Piazza-Allerta 2003	7	35	4	37	27.2%	1.85 [0.59, 5.77]
Total (95% CI)		**138**		**152**	**100.0%**	**1.68 [0.92, 3.09]**
Total events	23		15			

Heterogeneity: $Chi^2 = 1.25$, $df = 3$ ($P = 0.74$); $I^2 = 0\%$

Test for overall effect: $Z = 1.68$ ($P = 0.09$)



What was the effect size? What does it mean?

How certain you are that this is the true result?

What would you consider when deciding whether you are confident in the result of this meta analysis?

Cochrane Training

People typically use some criteria to assess the certainty of results

…but often used different criteria

…and it was not systematic

Cochrane Training

GRADE approach

Ensures:

- systematic process
- transparency

Considers:

- effect estimate: direction and size
- certainty of evidence: confidence in the effect estimate
- each outcome separately

www.gradeworkinggroup.org

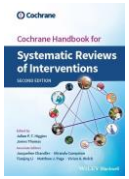
GRADE

HOW DO WE EXPRESS HOW CERTAIN WE ARE IN THE RESULTS?
GRADE: Four levels of certainty

Certainty level	Current definition	Grade
High	**We are very confident that the true effect lies close to that of the estimate of the effect**	⨁⨁⨁⨁
Moderate	**We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different**	⨁⨁⨁⊖
Low	**Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect**	⨁⨁⊖⊖
Very low	**We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect**	⨁⊖⊖⊖

Session outline

- a reminder on how to read and understand meta-analysis results
- GRADE approach to assessing certainty of evidence
- **factors affecting certainty**



See Chapters 14 & 15 of the Handbook

Where the certainty of evidence starts depends on study design

BODY OF EVIDENCE FROM RCT → **4 – high**

3 – moderate

2 – low ← **BODY OF EVIDENCE FROM NON-RANDOMIZED STUDIES***

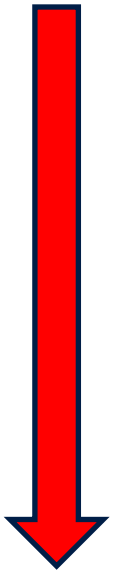
1 – very low

*Different for ROBINS-I tool

Certainty of evidence: rating down

- risk of bias
- imprecision
- inconsistency
- indirectness
- publication bias

BODY OF EVIDENCE FROM RCT



4 – high

3 – moderate

2 – low

1 – very low

Certainty of evidence: rating down

- rating down 1 level for serious concerns
- rating down 2 levels for very serious concerns
- some (minor) concerns on more than one domain can amount to one full level rated down
- reducton for individual domains add up to generate the final rating for the outcome
- not possible to go below ‘very low’

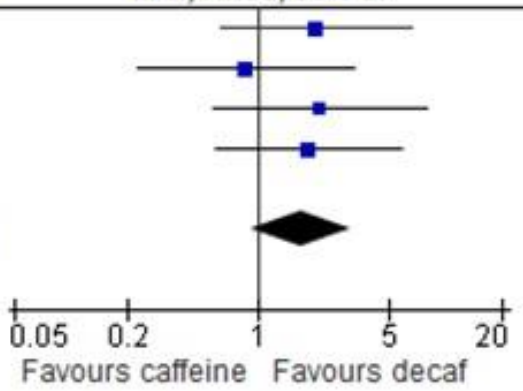
An example: Caffeine review

Outcome: sleep disruption

Study or Subgroup	Caffeinated coffee Events	Total	Decaffeinated coffee Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mama-Kaffa 1999	7	53	4	61	26.1%	2.01 [0.62, 6.50]
Morrocona 1998	3	15	4	17	26.3%	0.85 [0.23, 3.20]
Oohlahlazza 1998	6	35	3	37	20.4%	2.11 [0.57, 7.81]
Piazza-Allerta 2003	7	35	4	37	27.2%	1.85 [0.59, 5.77]
Total (95% CI)		**138**		**152**	**100.0%**	**1.68 [0.92, 3.09]**
Total events	23		15			

Heterogeneity: $Chi^2 = 1.25$, $df = 3$ ($P = 0.74$); $I^2 = 0\%$

Test for overall effect: $Z = 1.68$ ($P = 0.09$)



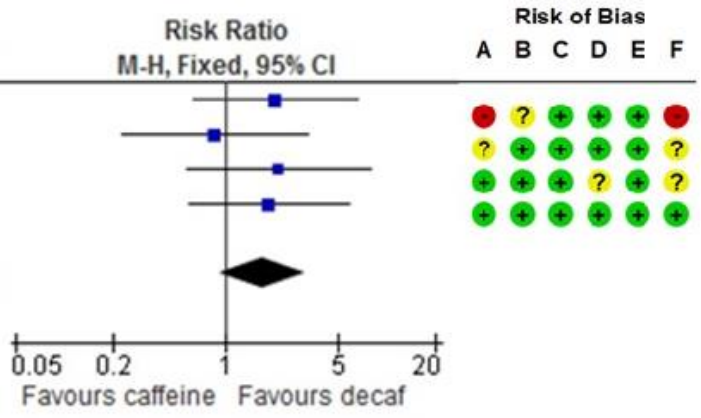
Risk of bias

for studies that provided data for each outcome:

- what is the overall risk of bias of these studies
- does the risk of bias reduce our confidence in the effect?
- consider the relative importance of each domain for the particular outcome

Study or Subgroup	Caffeinated coffee Events	Total	Decaffeinated coffee Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mama-Kaffa 1999	7	53	4	61	26.1%	2.01 [0.62, 6.50]
Morrocona 1998	3	15	4	17	26.3%	0.85 [0.23, 3.20]
Oohlahlazza 1998	6	35	3	37	20.4%	2.11 [0.57, 7.81]
Piazza-Allerta 2003	7	35	4	37	27.2%	1.85 [0.59, 5.77]
Total (95% CI)		**138**		**152**	**100.0%**	**1.68 [0.92, 3.09]**
Total events	23		15			

Heterogeneity: Chi² = 1.25, df = 3 (P = 0.74); I² = 0%
Test for overall effect: Z = 1.68 (P = 0.09)



Is the risk of bias…

- Not serious
- Serious
- Very serious

…enough that my confidence in the result is reduced?

Imprecision

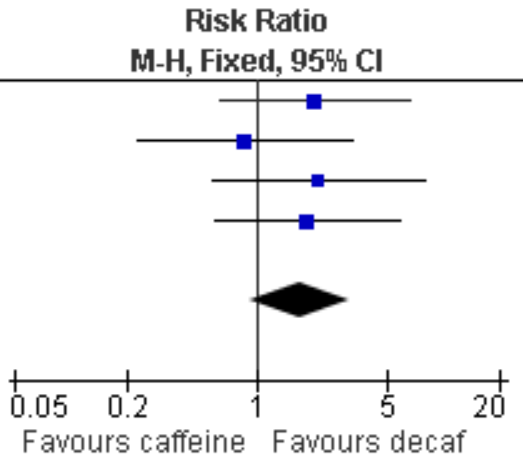
A) sample size and number of events

- Optimal Information Size - a sample size sufficient for one adequately powered study
- rule of thumb – min. 400 events (dichotomous), 400 people (continuous)

	Caffeinated coffee		Decaffeinated coffee			Risk Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**M-H, Fixed, 95% CI**
Mama-Kaffa 1999	7	53	4	61	26.1%	2.01 [0.62, 6.50]
Morrocona 1998	3	15	4	17	26.3%	0.85 [0.23, 3.20]
Oohlahlazza 1998	6	35	3	37	20.4%	2.11 [0.57, 7.81]
Piazza-Allerta 2003	7	35	4	37	27.2%	1.85 [0.59, 5.77]
Total (95% CI)		**138**		**152**	**100.0%**	**1.68 [0.92, 3.09]**
Total events	23		15			

Heterogeneity: Chi² = 1.25, df = 3 (P = 0.74); I² = 0%

Test for overall effect: Z = 1.68 (P = 0.09)



Imprecision

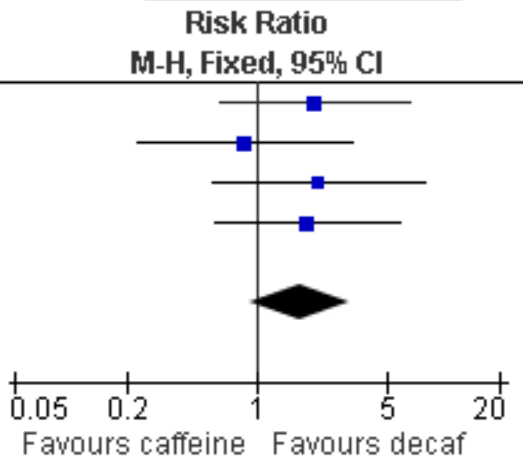
B) confidence intervals

- do they include contradictory conclusions?
  - i.e. no effect **AND** an appreciable benefit or harm

Study or Subgroup	Caffeinated coffee Events	Total	Decaffeinated coffee Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mama-Kaffa 1999	7	53	4	61	26.1%	2.01 [0.62, 6.50]
Morrocona 1998	3	15	4	17	26.3%	0.85 [0.23, 3.20]
Oohlahlazza 1998	6	35	3	37	20.4%	2.11 [0.57, 7.81]
Piazza-Allerta 2003	7	35	4	37	27.2%	1.85 [0.59, 5.77]
Total (95% CI)		**138**		**152**	**100.0%**	**1.68 [0.92, 3.09]**
Total events	23		15			

Heterogeneity: Chi² = 1.25, df = 3 (P = 0.74); I² = 0%

Test for overall effect: Z = 1.68 (P = 0.09)



But what is an appreciable harm or benefit?

Determine appreciable benefit or harm

Relative effects

- use rules of thumb, risk ratio <0.75 and >1.25
- means 25% reduction or increase in an outcome

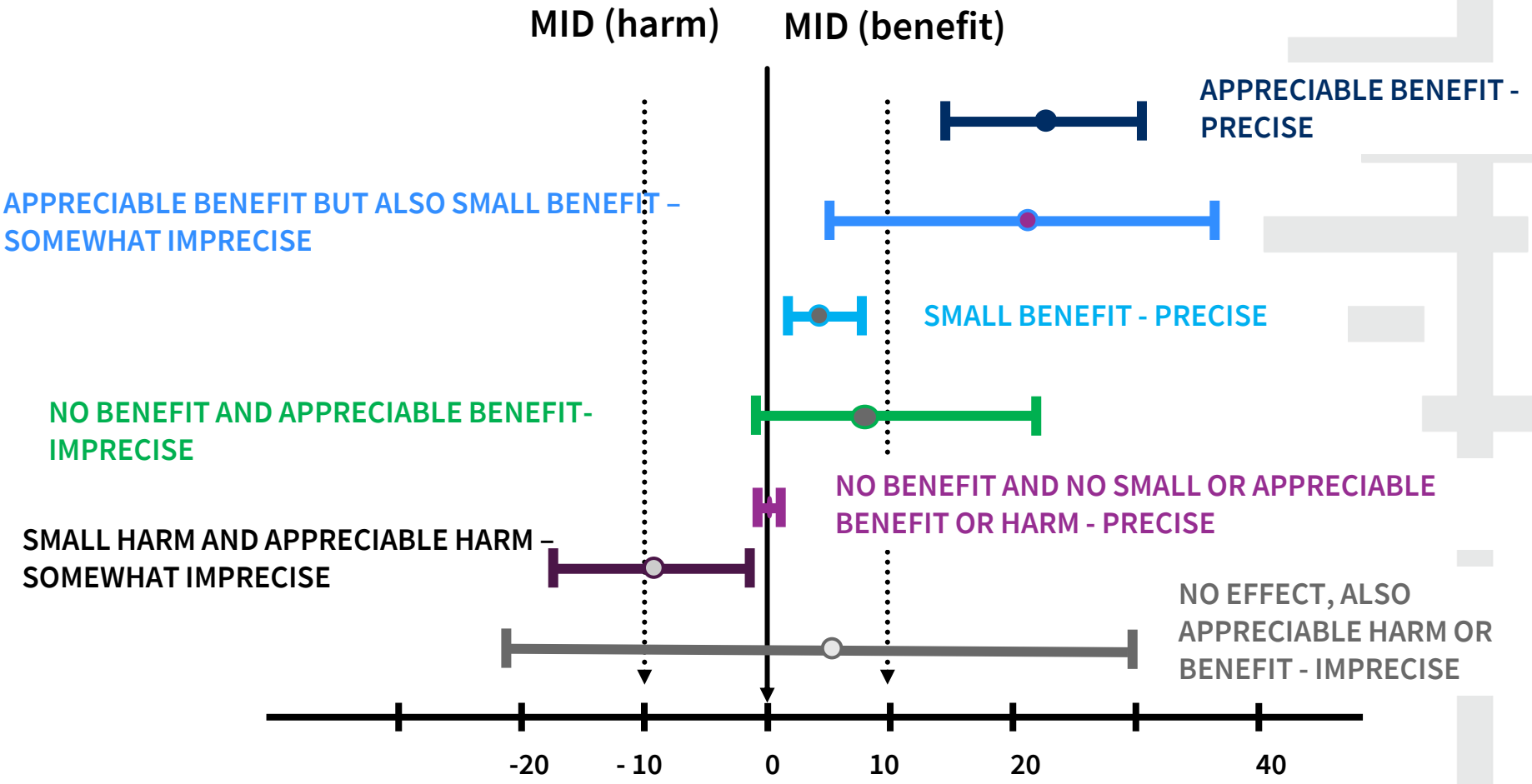
Absolute effects

- calculate the effects using the pooled estimate and a common baseline risk

Example

- using baseline risk of 10 per 100, the RR 1.68 (0.92 to 3.09) translates to 7 more people (1 fewer person to 21 more people)
- Do these CIs include important harm or benefit and no effect?

A meaningful difference: your cut-off is 10 more people



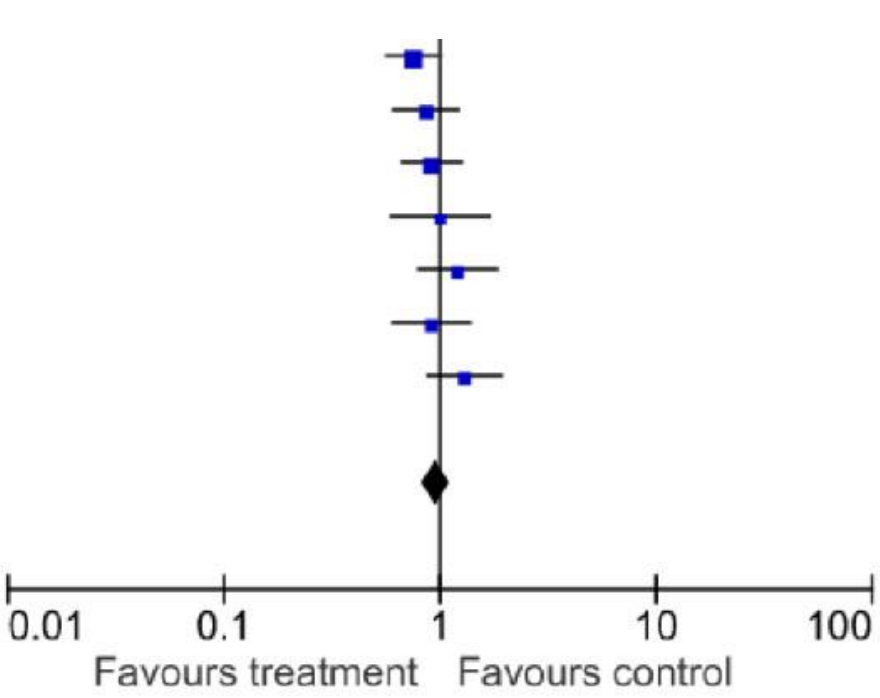
Is the imprecision…

- Not serious
- Serious
- Very serious

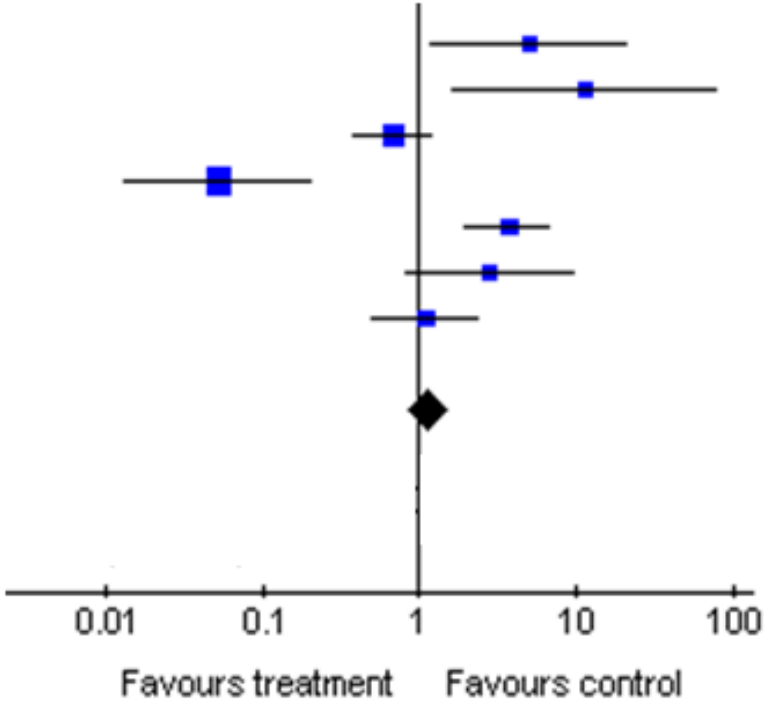
…enough that my confidence in the result is reduced?

Inconsistency (heterogeneity)

Forest plot A

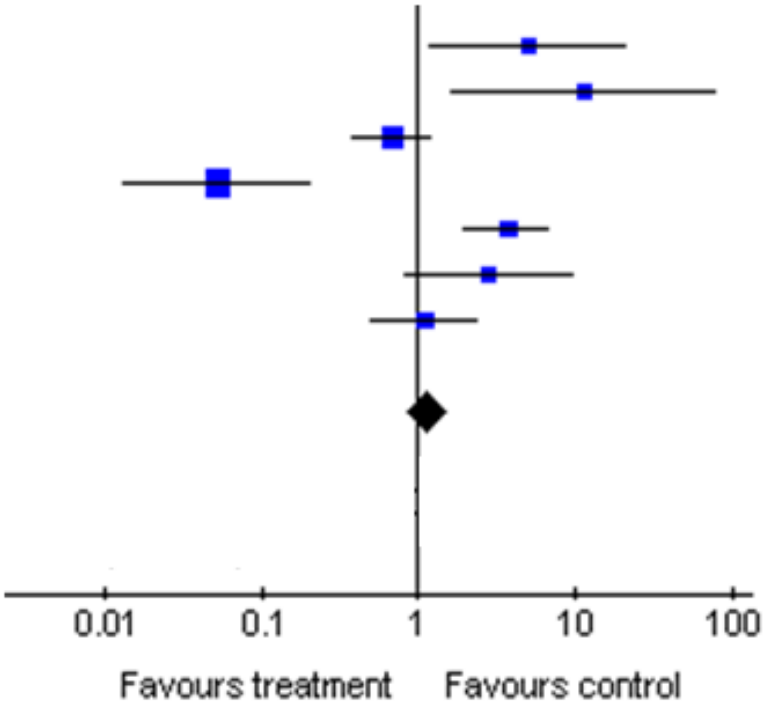


Forest plot B

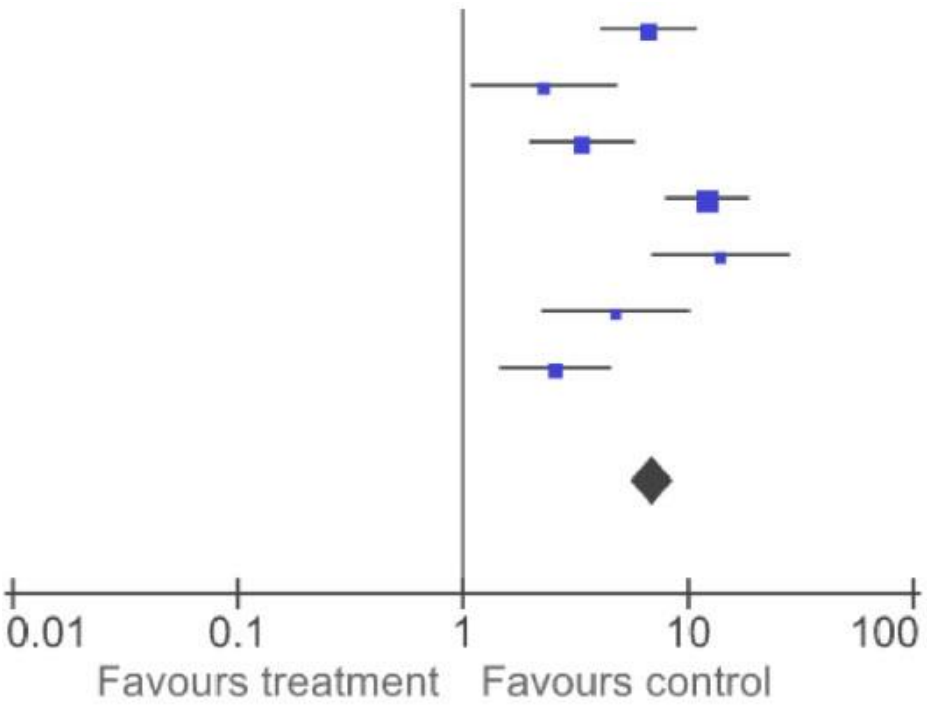


Inconsistency (heterogeneity)

Forest plot B



Forest plot C



Inconsistency (heterogeneity)

Other tools for identifying heterogeneity:

- Chi-squared (χ^2, or Chi^2) test, also known as the Q test
- I^2 statistic
  - <30-40% - low or unimportant heterogeneity
  - 30-60% - moderate heterogeneity
  - 50-90% - substantial heterogeneity
  - 75-100% - considerable heterogeneity

Only a guide, not a definitive proof of heterogeneity!

Inconsistency (heterogeneity)

Cautious exploration of possible causes:

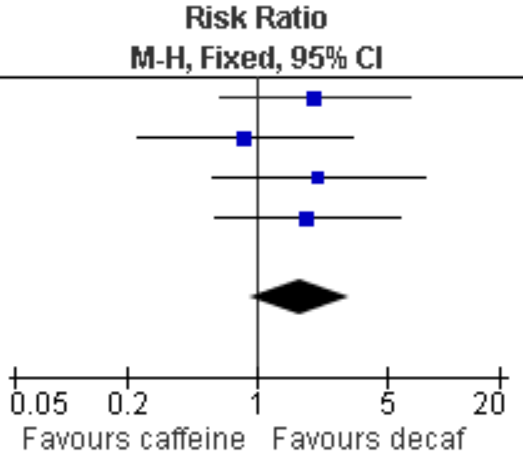
- differences in patients?
- interventions?
- outcome definitions?
- methodological diversity?

Inconsistency (heterogeneity)

Study or Subgroup	Caffeinated coffee Events	Total	Decaffeinated coffee Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mama-Kaffa 1999	7	53	4	61	26.1%	2.01 [0.62, 6.50]
Morrocona 1998	3	15	4	17	26.3%	0.85 [0.23, 3.20]
Oohlahlazza 1998	6	35	3	37	20.4%	2.11 [0.57, 7.81]
Piazza-Allerta 2003	7	35	4	37	27.2%	1.85 [0.59, 5.77]
Total (95% CI)		**138**		**152**	**100.0%**	**1.68 [0.92, 3.09]**
Total events	23		15			

Heterogeneity: $Chi^2 = 1.25$, $df = 3$ ($P = 0.74$); $I^2 = 0\%$

Test for overall effect: $Z = 1.68$ ($P = 0.09$)



Is the inconsistency…

- Not serious
- Serious
- Very serious

…enough that my confidence in the result is reduced?

Indirectness

- how well do the included studies address your review question?
- compare your eligibility criteria to the ‘Characteristics of included studies’ tables
- NOT generalisability to other populations

Indirectness

- EVIDENCE IS INDIRECT IF:
  - indirect comparisons
  - studies in your review only addressed a restricted aspect of your review question (PICC

    e.g.:
    - special populations
    - specific versions of the intervention
    - less applicable comparisons
    - surrogate outcomes

Is the indirectness…

- Not serious
- Serious
- Very serious

…enough that my confidence in the result is reduced?

Cochrane Training

Publication bias

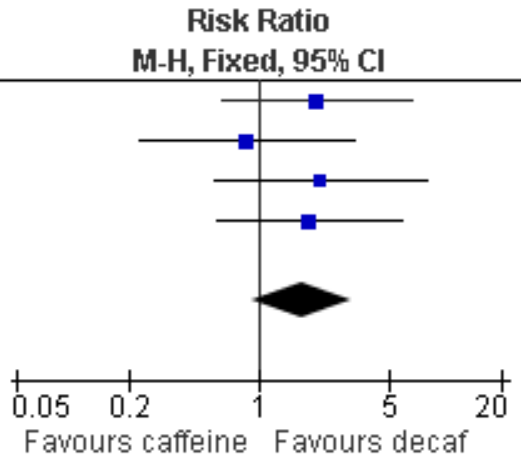
- is there a high probability of publication bias?
  - small studies with mostly positive results, especially with industry sponsorship
  - many of your included studies fail to report this outcome
- Did you conduct a comprehensive search of multiple sources? Were all published studies found?
- Did you search trial registries and it appears that you are not missing results from some studies due to bias?
- Did you contact experts in the field and seem to have the data available?

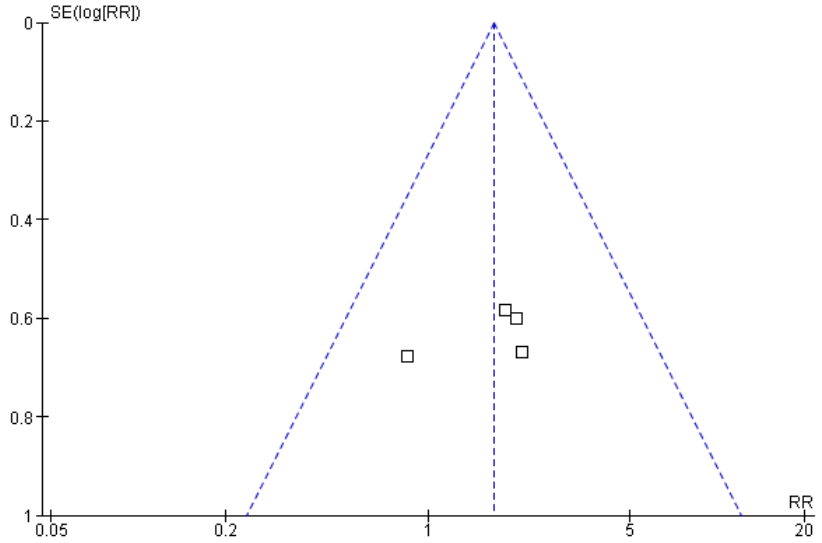
Publication bias

Study or Subgroup	Caffeinated coffee Events	Total	Decaffeinated coffee Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Mama-Kaffa 1999	7	53	4	61	26.1%	2.01 [0.62, 6.50]
Morrocona 1998	3	15	4	17	26.3%	0.85 [0.23, 3.20]
Oohlahlazza 1998	6	35	3	37	20.4%	2.11 [0.57, 7.81]
Piazza-Allerta 2003	7	35	4	37	27.2%	1.85 [0.59, 5.77]
Total (95% CI)		**138**		**152**	**100.0%**	**1.68 [0.92, 3.09]**
Total events	23		15			

Heterogeneity: Chi² = 1.25, df = 3 (P = 0.74); I² = 0%

Test for overall effect: Z = 1.68 (P = 0.09)





- need at least 10 studies
- interpret with caution
- downgrade by maximum of one level

Cochrane Training

Is the publication bias…

- Undetected
- Strongly suspected

…that my confidence in the result is reduced?

Caffeine example: overall judgement

- risk of bias
- imprecision
- inconsistency
- indirectness
- publication bias

4 – high

3 – moderate

2 – low

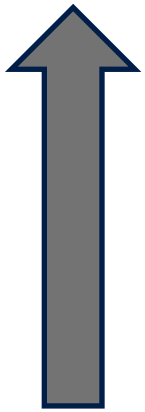
1 – very low

What if there was no meta-analysis?

The GRADE assessment shoud still be done!

Certainty of evidence: upgrading

- large effect
- dose response effect
- all the plausible biases affecting the result would tend to underestimate an apparent intervention effect or vice versa



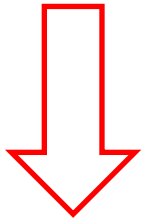
4 – high

3 – moderate

Non-randomized

2 – low

risk of bias, indirectness, inconsistency, imprecision, publication bias



1 – very low

Upgrading example: large effect

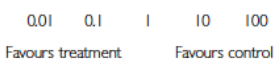
Analysis 1.3. Comparison 1 Motorcycle helmet versus no helmet, Outcome 3 Head Injury (adjusted).

Review: Helmets for preventing injury in motorcycle riders

Comparison: 1 Motorcycle helmet versus no helmet

Outcome: 3 Head Injury (adjusted)

Study or subgroup	log [Adjusted Odds Ratio] (SE)	Adjusted Odds Ratio IV,Random,95% CI	Weight	Adjusted Odds Ratio IV,Random,95% CI
1 Case-control studies				
Gabella 1995	-0.8796 (0.342)		8.9 %	0.41 [0.21, 0.81]
Tsai 1995	-1.3471 (0.3089)		10.9 %	0.26 [0.14, 0.48]
Subtotal (95% CI)			**19.8 %**	**0.32 [0.20, 0.51]**
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 1.03$, df = 1 (P = 0.31); $I^2 = 3\%$				
Test for overall effect: Z = 4.89 (P < 0.00001)				
2 Cross-sectional studies				
Christian 2003	-1.4697 (0.2547)		15.9 %	0.23 [0.14, 0.38]
Romano 1991	-1.335 (0.2057)		23.9 %	0.26 [0.18, 0.39]
Rowland 1996	-1.1314 (0.2233)		20.4 %	0.32 [0.21, 0.50]
Sauter 2005	-0.8439 (0.2263)		19.9 %	0.43 [0.28, 0.67]
Subtotal (95% CI)			**80.2 %**	**0.30 [0.24, 0.39]**
Heterogeneity: $Tau^2 = 0.02$; $Chi^2 = 4.11$, df = 3 (P = 0.25); $I^2 = 27\%$				
Test for overall effect: Z = 8.98 (P < 0.00001)				
Total (95% CI)			**100.0 %**	**0.31 [0.25, 0.38]**
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 5.18$, df = 5 (P = 0.39); $I^2 = 3\%$				
Test for overall effect: Z = 11.41 (P < 0.00001)				



Upgrading example: dose response

Example: childhood lymphoblastic leukemia

- risk for CNS malignancies 15 years after cranial irradiation
- no radiation: 1% (95% CI 0% to 2.1%)
- 12 Gy: 1.6% (95% CI 0% to 3.4%)
- 18 Gy: 3.3% (95% CI 0.9% to 5.6%)

Upgrading example: effect of confounding

Example: Metformin

- Hypoglycaemic drug phenformin causes lactic acidosis
- The **related** agent metformin is under suspicion for the same toxicity.
- Large non-randomized studies have failed to demonstrate an association even though clinicians would be more alert to lactic acidosis in the presence of the agent

What to include in your protocol

- add a subheading in RevMan Methods section
  - e.g. ‘Assessment of the certainty of the evidence’
- that you will use the GRADE approach to assess the important outcomes in your review
- that two authors will independently conduct GRADE assessments, and how you will resolve disagreements
- that you will report the GRADE assessment in the ‘Results’ and ‘Discussion’ section, and include in the ‘Summary of findings’ table

Take home message

- consider your results in the context of the overall certainty of the evidence
- GRADE is a structured system to help you consider certainty

References

- Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, Guyatt GH. **Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence**.
- Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, Akl EA, Guyatt GH**. Chapter 15: Interpreting results and drawing conclusions**.

In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane, 2020. Available from www.training.cochrane.org/handbook.

- www.tech.cochrane.org/revman/gradepro

Acknowledgements

- Compiled by Dario Sambunjak, Nancy Santesso, Miranda Cumpston and Sue Brennan
- Based on materials by Nancy Santesso, the Cochrane GRADEing Methods Group and Cochrane Australia