Does the evidence directly answer my question?

Outcome: Mortality		
Domain (original question asked)	**Description (evidence found and included, including evidence from other studies) - consider the domains of study design and study execution, inconsistency, imprecision and publication bias**	**Judgment - Is the evidence is sufficiently direct?**
Population: All patients with advanced cancer	A total of 8 randomized trials included patients with various types of cancer, 2 trials included only patients with small cell lung cancer, others included predominantly breast cancer. The studies were well executed and enrolled patients that were similar to those seen in practice. There was some degree of inconsistency in the baseline risk and related imprecision. Publication bias was not of concern.	Yes ☐ Probably yes ☒ Probably no ☐ No ☐
Intervention: Heparins	Trials included both low molecular heparin and unfractionated heparin. The observational studies do not suggest differential effects for the heparins.	Yes ☐ Probably yes ☒ Probably no ☐ No ☐
Comparator: No anticoagulation	Trials used placebo injections	Yes ☒ Probably yes ☐ Probably no ☐ No ☐
Direct comparison	Studies directly compared the intervention against the comparator of interest (default)	Yes ☒ Probably yes ☐ Probably no ☐ No ☐
Outcome: Mortality	Mortality was determined through follow-up of patients in the trial (e.g. telephone)	Yes ☒ Probably yes ☐ Probably no ☐ No ☐
Final judgment about indirectness across domains for the outcome mortality:	**The identified evidence is directly relevant to the question. NRS will not provide strong complimentary data for the effects of the intervention. NRS suggest that the baseline risk for the population is similar in the trials compared to the population not included in trials.**	**☒ No indirectness ☐ Serious indirectness ☐ Very serious indirectness** **Footnote: The degree of indirectness does not lower our confidence that the estimates of effect would be similar for healthcare decision making. It is not useful to look for NRS evidence**

Outcome: Non-fatal extracranial bleeding		
Domain (original question asked)	**Description (evidence found and included, including evidence from other studies) - consider the domains of study design and study execution, inconsistency, imprecision and publication bias**	**Judgment - Is the evidence is sufficiently direct?**
Population: All patients with atrial fibrillation	A total of 11 randomized trials included patients with atrial fibrillation. The quality of evidence was rated down to imprecision (pooled risk ratio bleeding 1.42 (95% CI 0.89-2.29) but not inconsistency in the baseline risk and related imprecision. There were no issues of bias, inconsistency, or publication bias. In general patients enrolled were younger and healthier than those seen in clinical practice.	Yes ☐ Probably yes ☐ Probably no ☒ No ☐
Intervention: Warfarin	Trials included only adjusted dose warfarin	Yes ☒ Probably yes ☐ Probably no ☐ No ☐
Comparator: Aspirin	Trials used aspirin 75-325 mg	Yes ☒ Probably yes ☐ Probably no ☐ No ☐
Direct comparison	Studies directly compared the intervention against the comparator of interest (default)	Yes ☒ Probably yes ☐ Probably no ☐ No ☐
Outcome: Non-fatal major extracranial bleed	Bleeding was determined through follow-up of patients in the trial (e.g. in-person and telephone)	Yes ☒ Probably yes ☐ Probably no ☐ No ☐
Final judgment about indirectness across domains for the outcome non-fatal extracranial bleeding:	**The identified evidence is indirectly relevant to the question. NRS can provide strong complementary data for the effects of the intervention. NRS suggest that the baseline risk bleeding in the population as well as relative risk is lower in the trials compared to the population not included in trials. In addition, baseline risk of bleeding can be stratified by CHADS2 scoring using NRS data.**	**☐ No indirectness ☒ Serious indirectness ☐ Very serious indirectness** **Footnote: The degree of indirectness lowers our confidence that the estimates of effect would be similar for healthcare decision making. It is useful to look for NRS evidence**

5. Publication Bias

Should always be suspected

- Only small “positive” studies
- For profit interest
- Various methods to evaluate – none perfect, but clearly a problem

Publication bias

Do you strongly suspect publication bias and your certainty in the result is lowered?

- small studies with mostly positive results
- asymmetry in forest plots
- proof that studies have been withheld or not shared
- limited search for studies

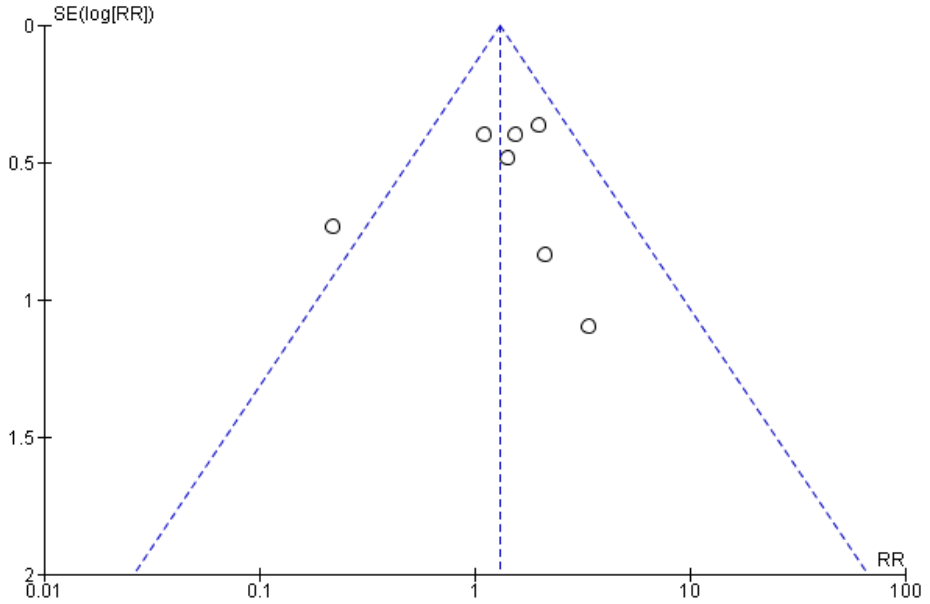
Electronic searches

To identify exercise trials, we searched the following five electronic databases:

..., the Cochrane Central Register of Controlled Trials (CENTRAL) (*The Cochrane Library*); MEDLINE; EMBASE; and Current Contents from 1966 to January 2000 with no language restrictions, according to the methods suggested by Dickersin 1994

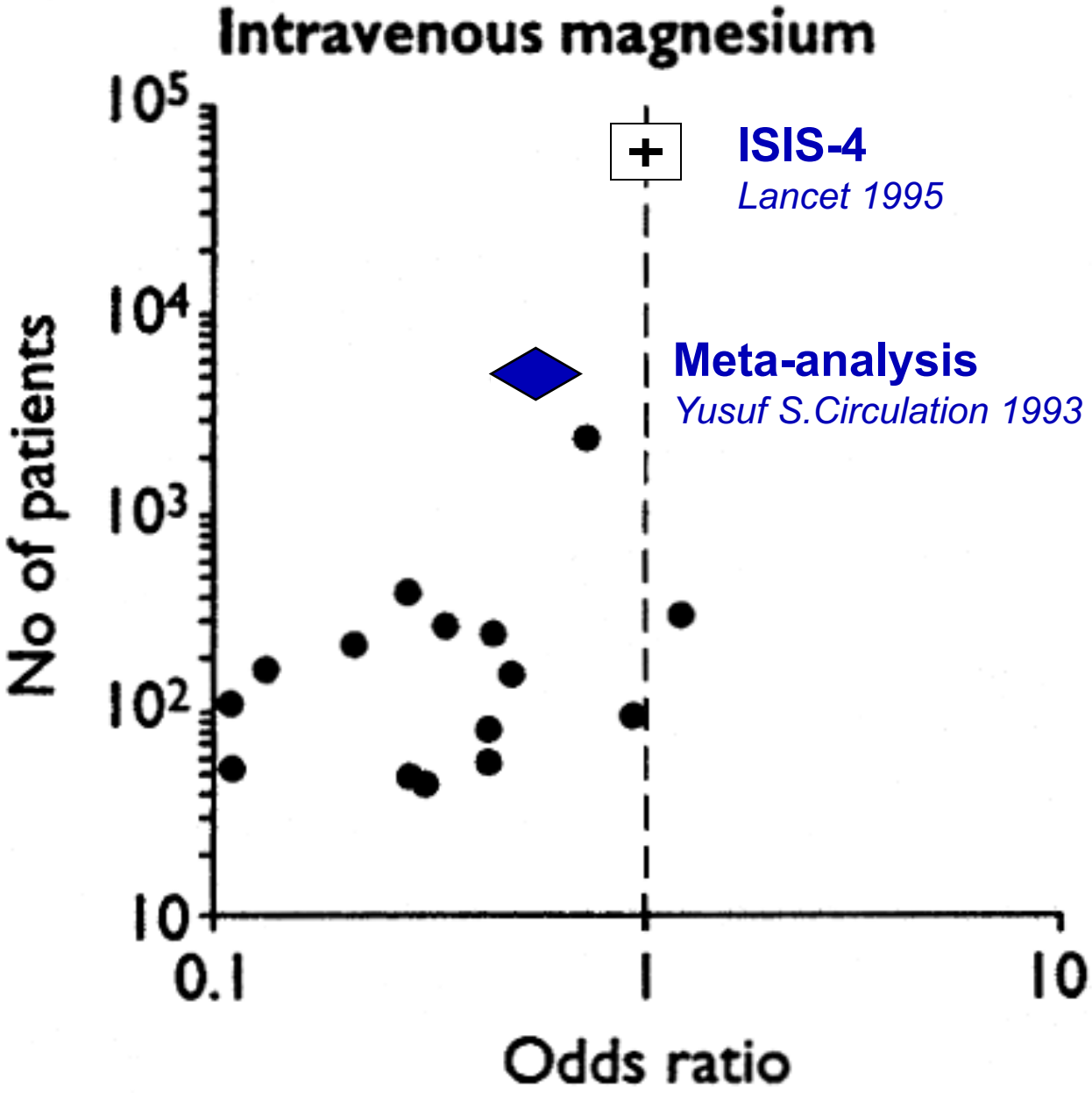
Searching other resources

In addition, we searched the reference lists of included trials and trials registers, and contacted content experts for additional studies and data.



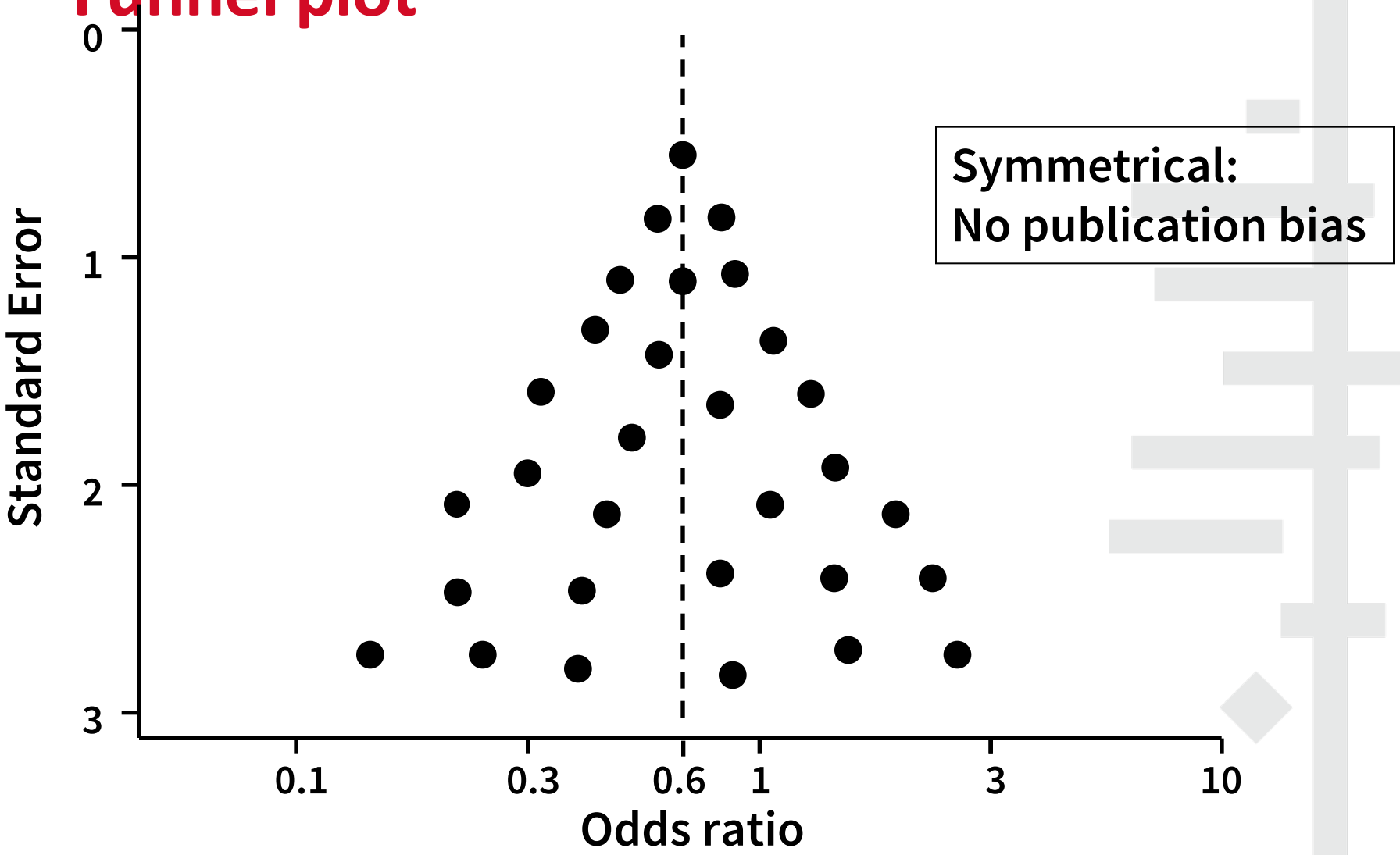
I.V. Mg in acute myocardial infarction

Publication bias



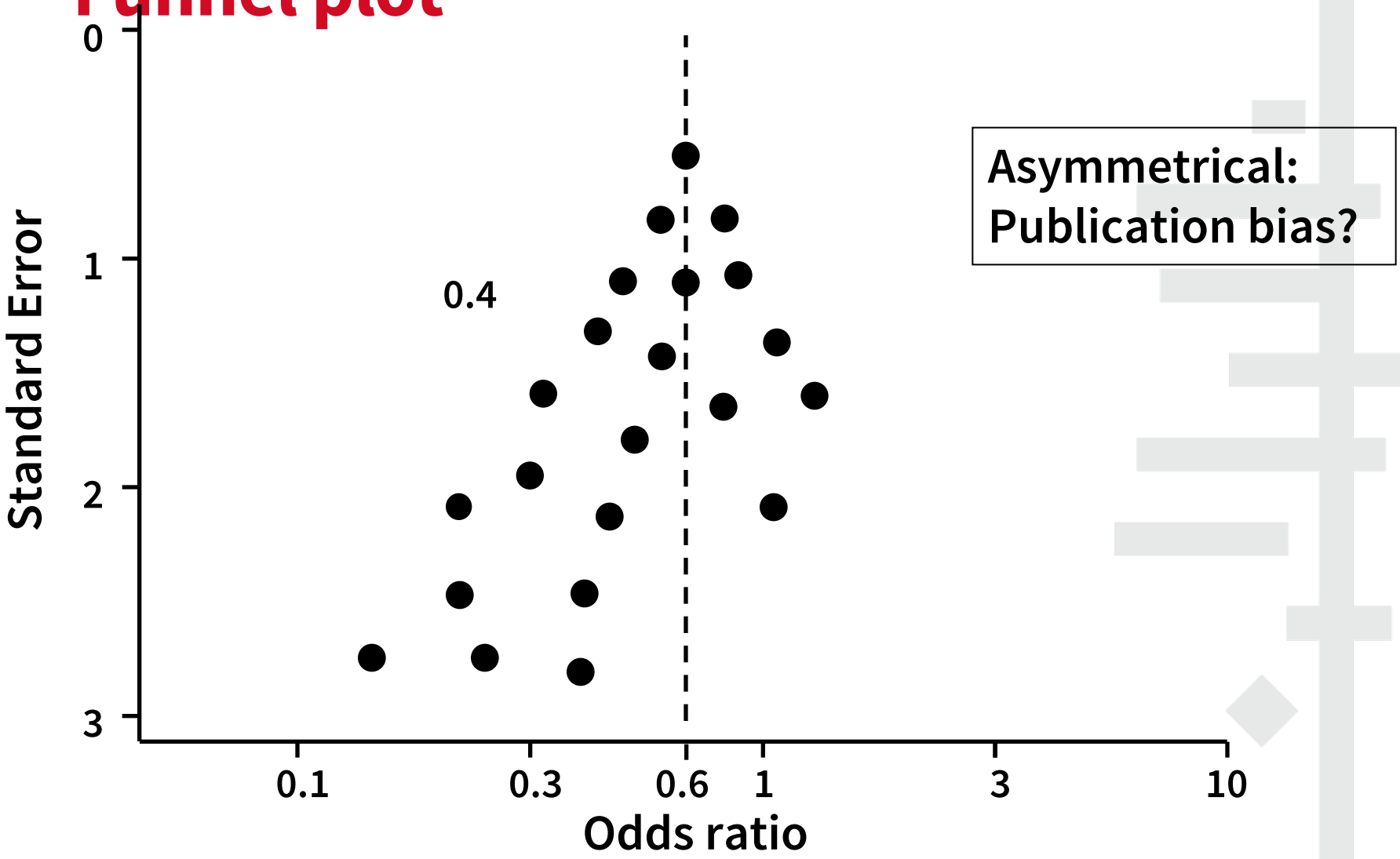
Egger M, Smith DS. BMJ 1995;310:752-54

Funnel plot



Egger M, Cochrane Colloquium Lyon 2001

Funnel plot



Asymmetrical:
Publication bias?

Egger M, Cochrane Colloquium Lyon 2001

What can raise quality?

1. large magnitude can upgrade (RRR 50%/RR 2)

- very large two levels (RRR 80%/RR 5)
- criteria
  - everyone used to do badly
  - almost everyone does well
- parachutes to prevent death when jumping from airplanes

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

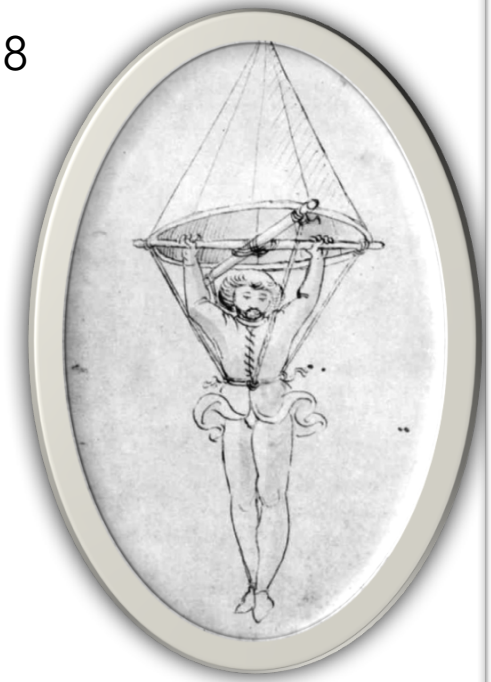
Gordon C S Smith, Jill P Pell

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

Relative risk reduction:
….> 99.9 % (1/100,000)

U.S. Parachute Association reported 821 injuries and 18 deaths out of 2.2 million jumps in 2007



Reminders for immunization uptake

Review: Patient reminder and recall systems to improve immunization rates

Comparison: 7 Patient Reminders (summary) vs. control

Outcome: 1 Immunized

	n/N	n/N	M-H,Random,95% CI	M-H,Random,95% CI
4 Other-adult				
Hogg1998T101	21/866	4/458		2.82 [0.96, 8.27]
Sansom2003T514	242/279	197/245		1.59 [1.00, 2.55]
Siebers1985T36	20/72	3/39		4.62 [1.28, 16.70]
Subtotal (95% CI)	**1217**	**742**		**2.19 [1.21, 3.99]**

Total events: 283 (Patient Reminder Sum), 204 (Control)

Heterogeneity: $Tau^2 = 0.10$; $Chi^2 = 2.93$, df = 2 (P = 0.23); $I^2 = 32\%$

Test for overall effect: Z = 2.57 (P = 0.010)

Citation: Jacobson Vann JC, Szilagyi P. Patient reminder and recall systems to improve immunization rates. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD003941. DOI: 10.1002/14651858.CD003941.pub2.

Non-randomised studies

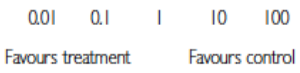
Analysis 1.3. Comparison 1 Motorcycle helmet versus no helmet, Outcome 3 Head Injury (adjusted).

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)				



What can raise quality?

2. dose response relation

- (higher INR – increased bleeding)
- 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%)

3. all plausible residual confounding may be working to reduce the demonstrated effect or increase the effect if no effect was observed

All plausible residual bias and confounding would result in an overestimate of effect

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

Vaccine – adverse effects

Practical example – bringing it all together

Flavanoids for Hemorrhoids

venotonic agents

- mechanism unclear, increase venous return

popularity

- 90 venotonics commercialized in France
- none in Sweden and Norway
- France 70% of world market

possibilities

- French misguided
- rest of world missing out

Systematic Review

14 trials, 1432 patients

key outcome

- risk not improving/persistent symptoms
- 11 studies, 1002 patients, 375 events
- RR 0.4, 95% CI 0.29 to 0.57

minimal side effects

is France right?

what is the quality of evidence?

What can lower quality?

Study limitations/risk of bias

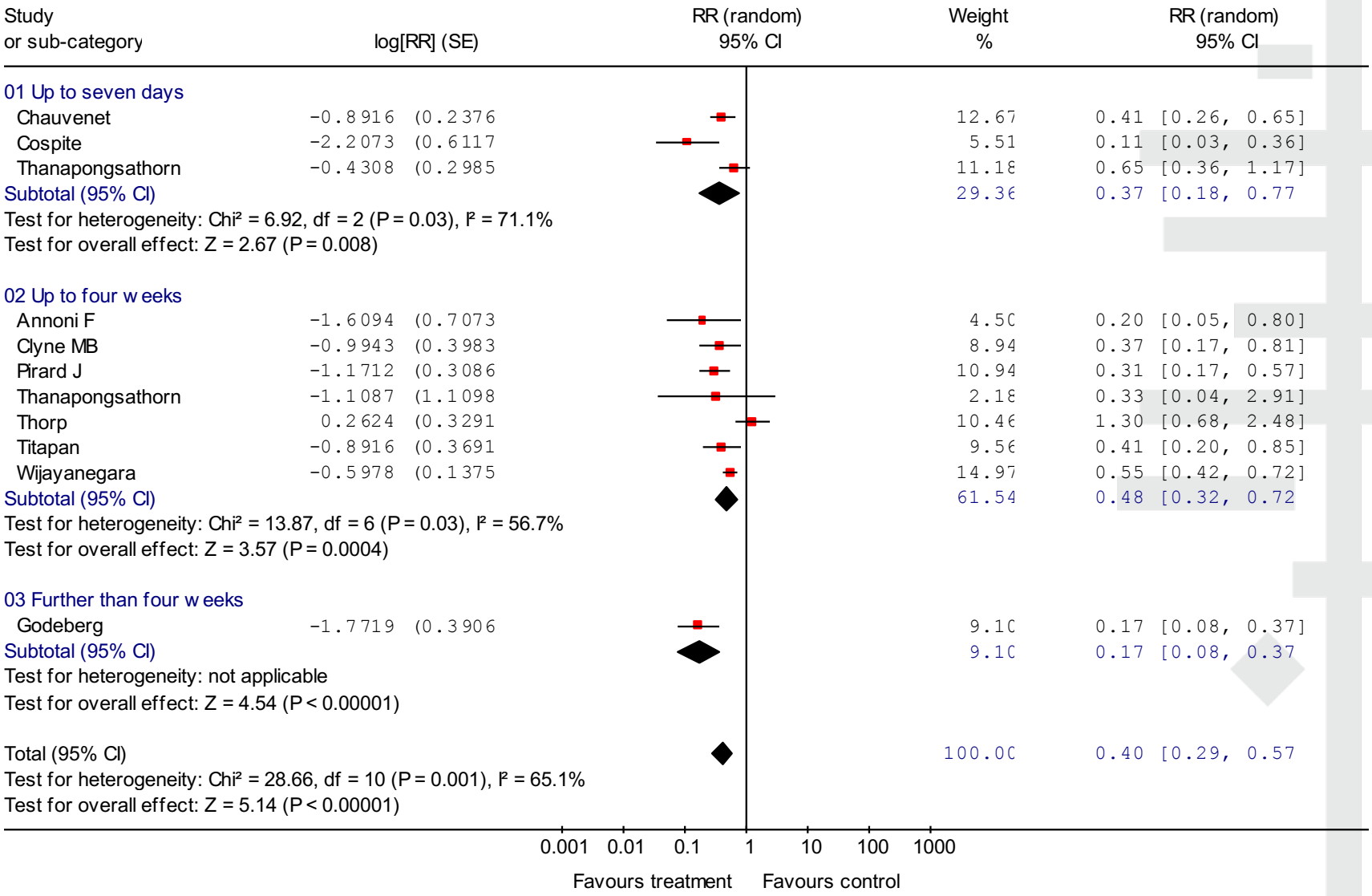
- lack of detail re concealment
- questionnaires not validated

rate down quality for study limitations/RoB?

indirectness – no problem

inconsistency, need to look at the results

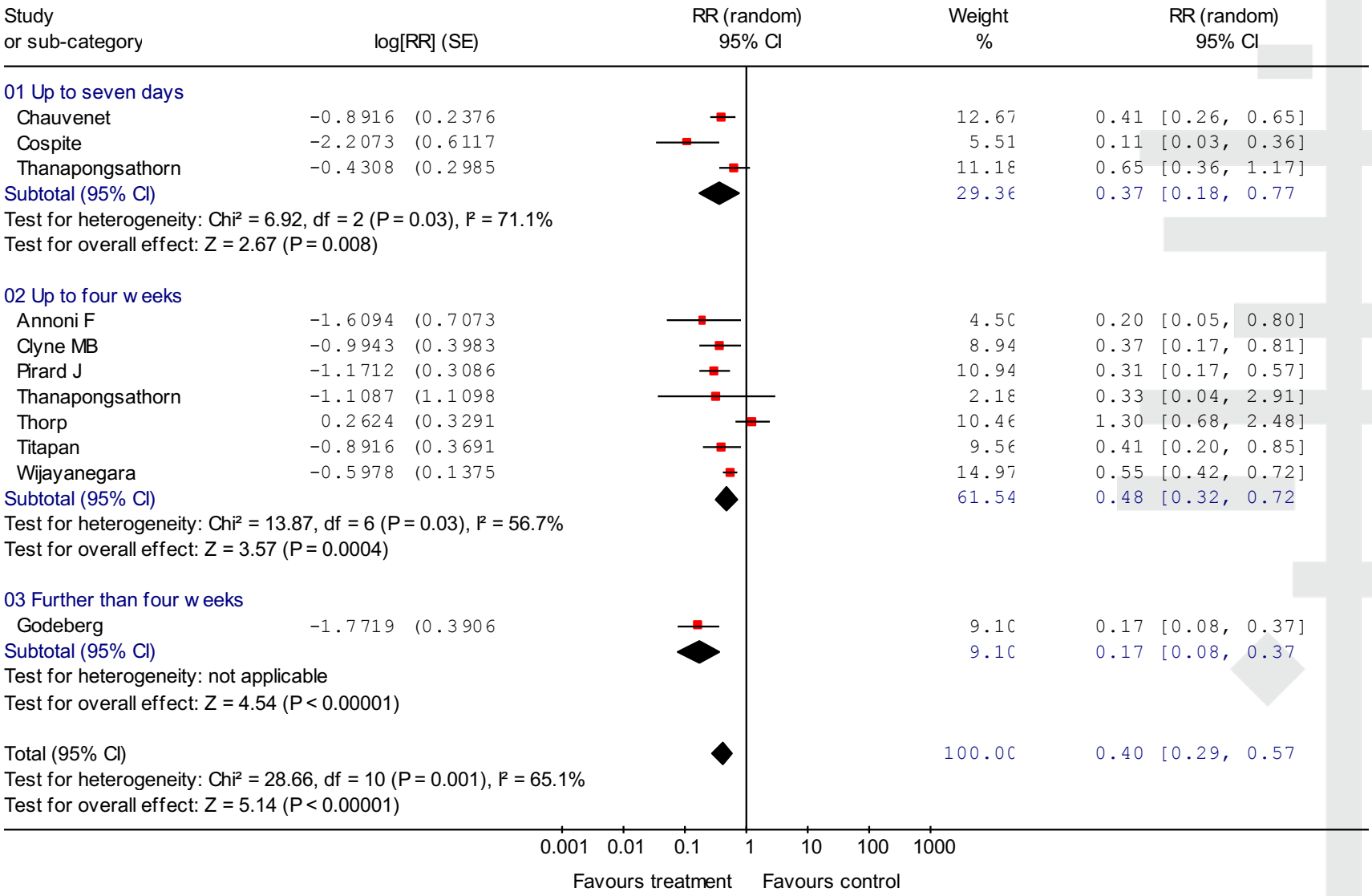
Review : Phlebotonics for hemorrhoids
Comparison: 01 Venotonics vs placebp
Outcome: 08 Overall improvement: no improvement/some improvement



Would you downgrade for inconsistency?

- [ ] No, there **is no serious inconsistency**
- [ ] Yes, there is serious inconsistency
- [ ] Yes, there is very serious inconsistency

Review : Phlebotonics for hemorrhoids
Comparison: 01 Venotonics vs placebp
Outcome: 08 Overall improvement: no improvement/some improvement



Is the imprecision…

- Not serious
- Serious
- Very Serious

…that is confidence/certainty in the result is reduced?

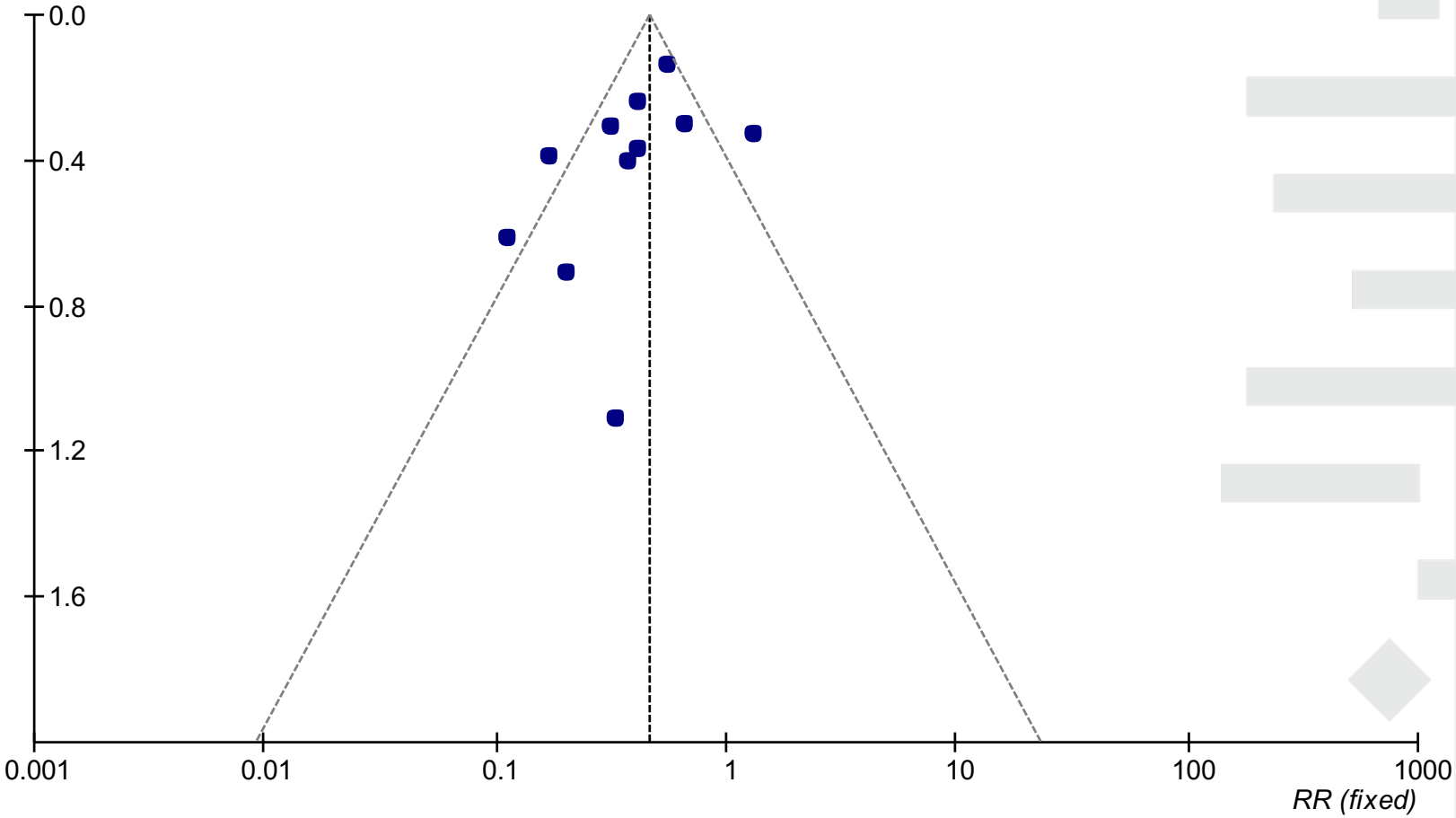
Publication bias?

size of studies

- 40 to 234 patients, most around 100

all industry sponsored

Review: Phlebotonics for hemorrhoids
Comparison: 01 Venotonics vs placebp
Outcome: 08 Overall improvement: no improvement/some improvement



Would you downgrade for publication bias?

- [ ] No, there is no publication bias
- [ ] Yes, there is publication bias
- [ ] Yes, there is very serious publication bias

What can lower quality?

risk of bias

- lack of detail re concealment
- questionnaires not validated

Inconsistency

- heterogeneity $p < 0.001$; I^2 65.1%

indirectness

imprecision

- RR 0.4, 95% CI 0.29 to 0.57

Publication bias

- 40 to 234 patients, most around 100

Your final judgment

Ratings
⨁⨁⨁⨁ High certainty
⨁⨁⨁◯ Moderate certainty
⨁⨁◯◯ Low certainty
⨁◯◯◯ Very low certainty

Interpreting the certainty in or quality of evidence

Ratings	Definitions
⨁⨁⨁⨁ High certainty	The panel is very confident that the true effect lies close to that of the estimate of the effect
⨁⨁⨁◯ Moderate certainty	The panel is 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 certainty	The panel's confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
⨁◯◯◯ Very low certainty	The panel has very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Self management for patients with chronic obstructive pulmonary disease

Patient or population: patients with chronic obstructive pulmonary disease
Settings: primary care, community, outpatient
Intervention: self management[1]
Comparison: usual care

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk **usual care**	Corresponding risk **self management**				
Quality of Life St George's Respiratory Questionnaire. Scale from: 0 to 100. (follow-up: 3 to 12 months)	The mean quality of life ranged across control groups from **38 to 60 points**	The mean quality of Life in the intervention groups was **2.58 lower** (5.14 to 0.02 lower)		698 (7)	⊕⊕⊕O **moderate**[2]	Lower score indicates better quality of life. A change of less than 4 points is not shown to be important to patients.
Dyspnoea Borg Scale. Scale from: 0 to 10. (follow-up: 3 to 6 months)	The mean dyspnoea ranged across control groups from **1.2 to 4.1 points**	The mean dyspnoea in the intervention groups was **0.53 lower** (0.96 to 0.1 lower)		144 (2)	⊕⊕OO **low**[3,4]	Lower score indicates improvement
Number and severity of exacerbations[5]	See comment	See comment	Not estimable[5]	591 (3)	See comment	Effect is uncertain
Respiratory-related hospital admissions (follow-up: 3 to 12 months)	**Low risk population**[6]		OR 0.64 (0.47 to 0.89)	966 (8)	⊕⊕⊕O **moderate**[7]	
	10 per 100	**7 per 100** (5 to 9)				
	High risk population[6]					
	50 per 100	**39 per 100** (32 to 47)				
Emergency department visits for lung diseases (follow-up: 6 to 12 months)	The mean emergency department visits for lung diseases ranged across control groups from **0.2 to 0.7 visits per person per year**	The mean emergency department visits for lung diseases in the intervention groups was **0.1 higher** (0.2 lower to 0.3 higher)		328 (4)	⊕⊕⊕O **moderate**[4]	
Doctor and nurse visits (follow-up: 6 to 12 months)	The mean doctor and nurse visits ranged across control groups from **1 to 5 vists per person per year**	The mean doctor and nurse visits in the intervention groups was **0.02 higher** (1 lower to 1 higher)		629 (8)	⊕⊕⊕O **moderate**[8]	

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; OR: Odds ratio;

Assessing Certainty in Evidence by Outcome

Table: GRADE's approach to rating quality of evidence (aka confidence in effect estimates)

For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)

1. Establish initial level of confidence		2. Consider lowering or raising level of confidence		3. Final level of confidence rating
Study design	***Initial confidence in an estimate of effect***	***Reasons for considering lowering or raising confidence***		***Confidence in an estimate of effect across those considerations***
		⬇ Lower if	**⬆ Higher if***	
Randomized trials ➔	**High confidence**	**Risk of Bias** **Inconsistency** **Indirectness** **Imprecision** **Publication bias**	**Large effect** **Dose response** **All plausible confounding & bias** • would reduce a demonstrated effect **or** • would suggest a spurious effect if no effect was observed	**High** ⨁⨁⨁⨁
				Moderate ⨁⨁⨁◯
Observational studies ➔	**Low confidence**			**Low** ⨁⨁◯◯
				Very low ⨁◯◯◯

*upgrading criteria are usually applicable to observational studies only.

Lowering certainty in RCTs

Table: GRADE's approach to rating quality of evidence (aka confidence in effect estimates)

For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)

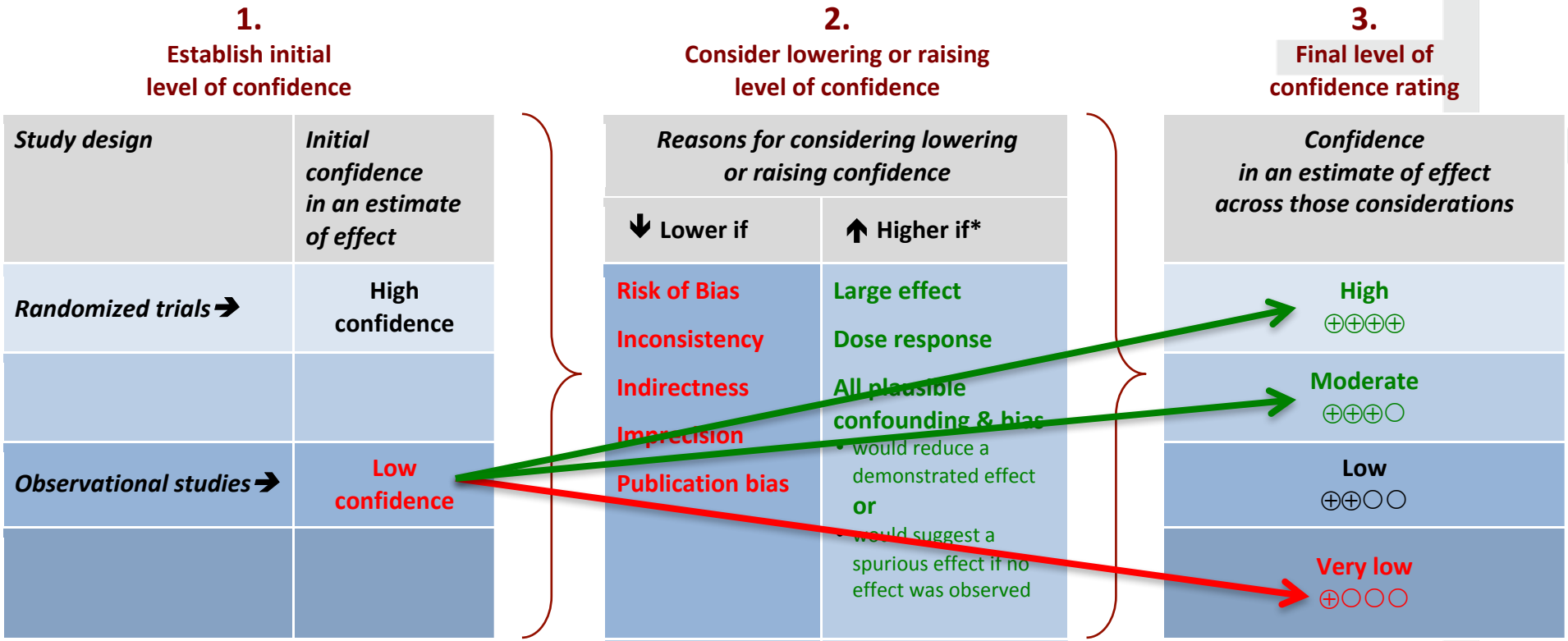
1. Establish initial level of confidence		**2. Consider lowering or raising level of confidence**		**3. Final level of confidence rating**
Study design	***Initial confidence in an estimate of effect***	***Reasons for considering lowering or raising confidence***		***Confidence in an estimate of effect across those considerations***
		⬇ Lower if	**⬆ Higher if***	
Randomized trials ➔	**High confidence**	**Risk of Bias** **Inconsistency** **Indirectness** **Imprecision** **Publication bias**	**Large effect** **Dose response** **All plausible confounding & bias** • would reduce a demonstrated effect **or** • would suggest a spurious effect if no effect was observed	**High** ⨁⨁⨁⨁
				Moderate ⨁⨁⨁◯
Observational studies ➔	**Low confidence**			**Low** ⨁⨁◯◯
				Very low ⨁◯◯◯

*upgrading criteria are usually applicable to observational studies only.

Altering certainty in observational studies

Table: GRADE's approach to rating quality of evidence (aka confidence in effect estimates)

For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)

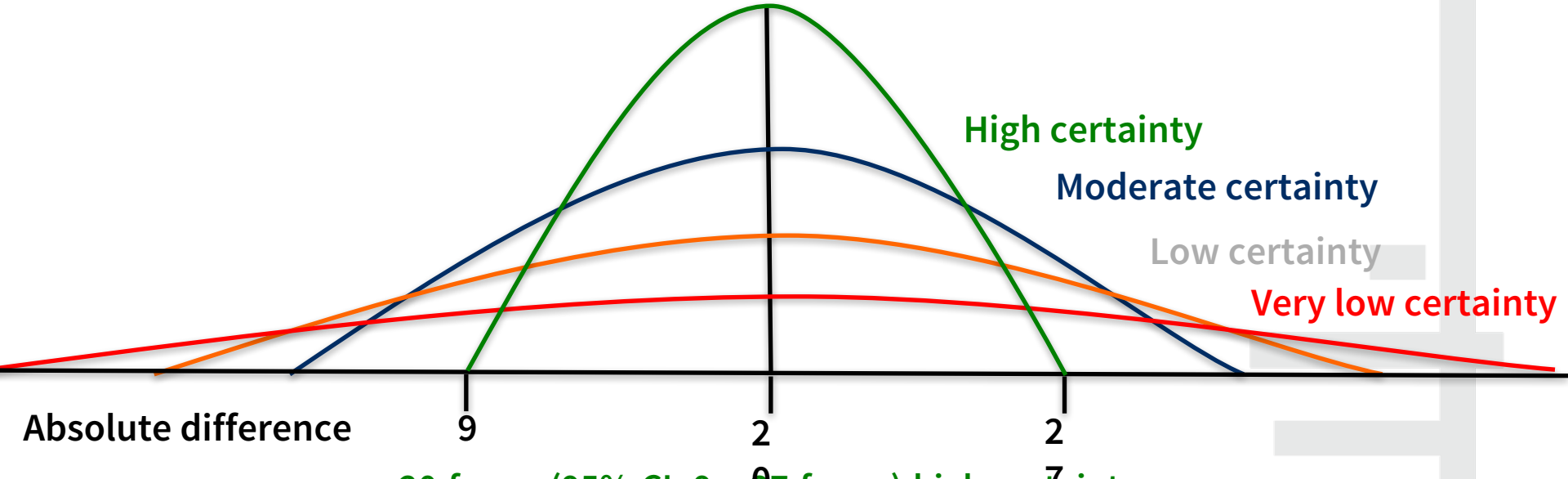


*upgrading criteria are usually applicable to observational studies only.

Magnitude of Effect

Likelihood of and certainty in the evidence or effect

Certainty or Quality of evidence Confidence in effect



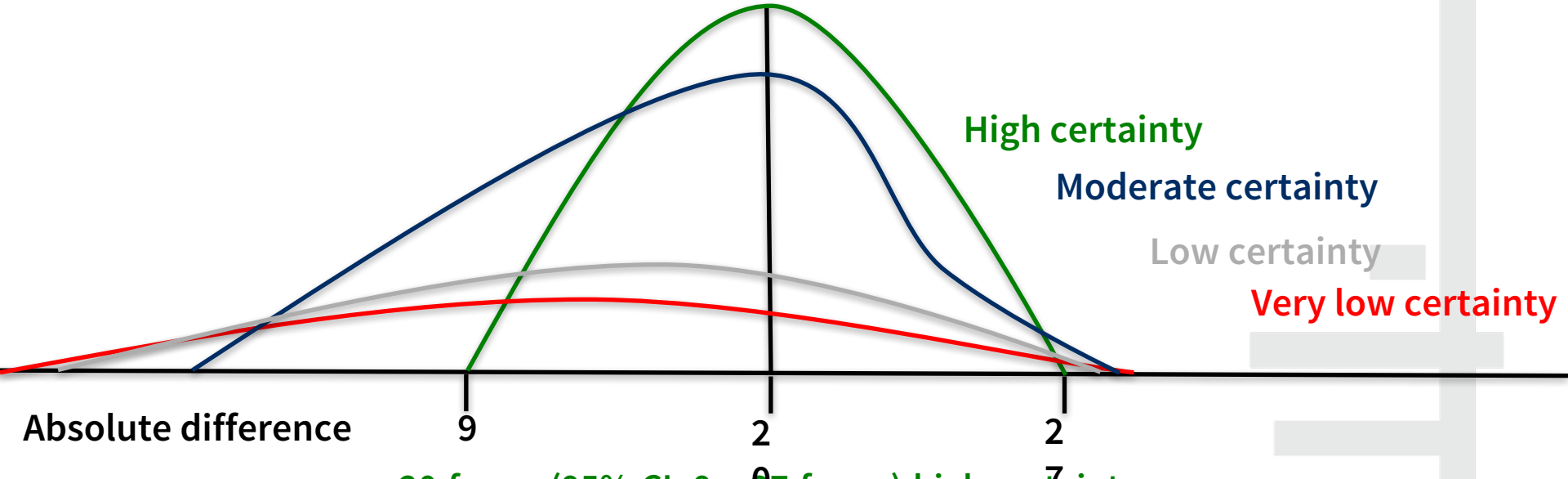
20 fewer (95% CI: 9 – 27 fewer) high certainty with sufficiently narrow confidence range – no downgrading
Certainty range identical to CI: distribution known

20 fewer (95% CI: 9 – 27 fewer) moderate certainty due to indirectness or serious concerns about any other downgrading domain including imprecision – wide certainty range despite narrow confidence intervals – certainty range wider than 9 – 27 but shape and width not exactly known

20 fewer (95% CI: 9 – 27 fewer) low certainty due to risk of bias and indirectness – very wide certainty range despite narrow confidence intervals

20 fewer (95% CI: 9 – 27 fewer) very low certainty due to risk of bias, indirectness and publication bias – extremely wide certainty range

Schünemann, JCE 2016



20 fewer (95% CI: 9 – 27 fewer) high certainty with sufficiently narrow confidence range – no downgrading
Certainty range identical to CI: distribution known

20 fewer (95% CI: 9 – 27 fewer) moderate certainty due to risk of bias or serious concerns about any other downgrading domain including imprecision – wide certainty range despite narrow confidence intervals – certainty range wider than 9 – 27 but shape and width not exactly known

20 fewer (95% CI: 9 – 27 fewer) low certainty due to risk of bias and indirectness – very wide certainty range despite narrow confidence intervals

20 fewer (95% CI: 9 – 27 fewer) very low certainty due to risk of bias, indirectness and publication bias – extremely wide certainty range

Pulling it together (results section)

Headaches

We pooled 7 studies with 567 participants and found a risk ratio of 1.38 (95% CI, 0.96 to 2.00). The certainty of the evidence is low quality due to some risk of bias for no allocation concealment, and due to imprecise results including the potential for no effect on headaches and appreciable increase. Overall, caffeinated coffee may increase the risk of headaches.

Results section

- present all your results systematically, including:
  - outcomes combined in meta-analysis
  - outcomes for which no evidence was found
  - outcomes for which meta-analysis was not possible
    - e.g. studies too different
    - e.g. data not available in comparable format

Caution when making conclusions

Do not make recommendations

- Recommendations involve consideration of setting, values and preferences of patients, resources, etc.

✖ “If people want to stay alert during the day, they should drink coffee.”

✖ “Joint lavage should be discouraged in patients with osteoarthritis.”

Instead....

Indicate level of evidence and effect on outcomes

- “Moderate quality evidence shows that alertness will probably improve by drinking coffee.”
- “There is low quality evidence that joint lavage provides little or no difference in symptoms of knee OA.”

"no evidence of effect" versus "evidence of no effect"

Results

Combining the results of six randomised clinical trials including 710 patients demonstrated no significant effects of propylthiouracil versus placebo on all-cause mortality (relative risks (RR) 0.93, 95% confidence interval (CI) 0.66 to 1.30), liver-related mortality (RR 0.80, 95% CI 0.50 to 1.29), complications of the liver disease, or liver histology.

Authors' conclusions

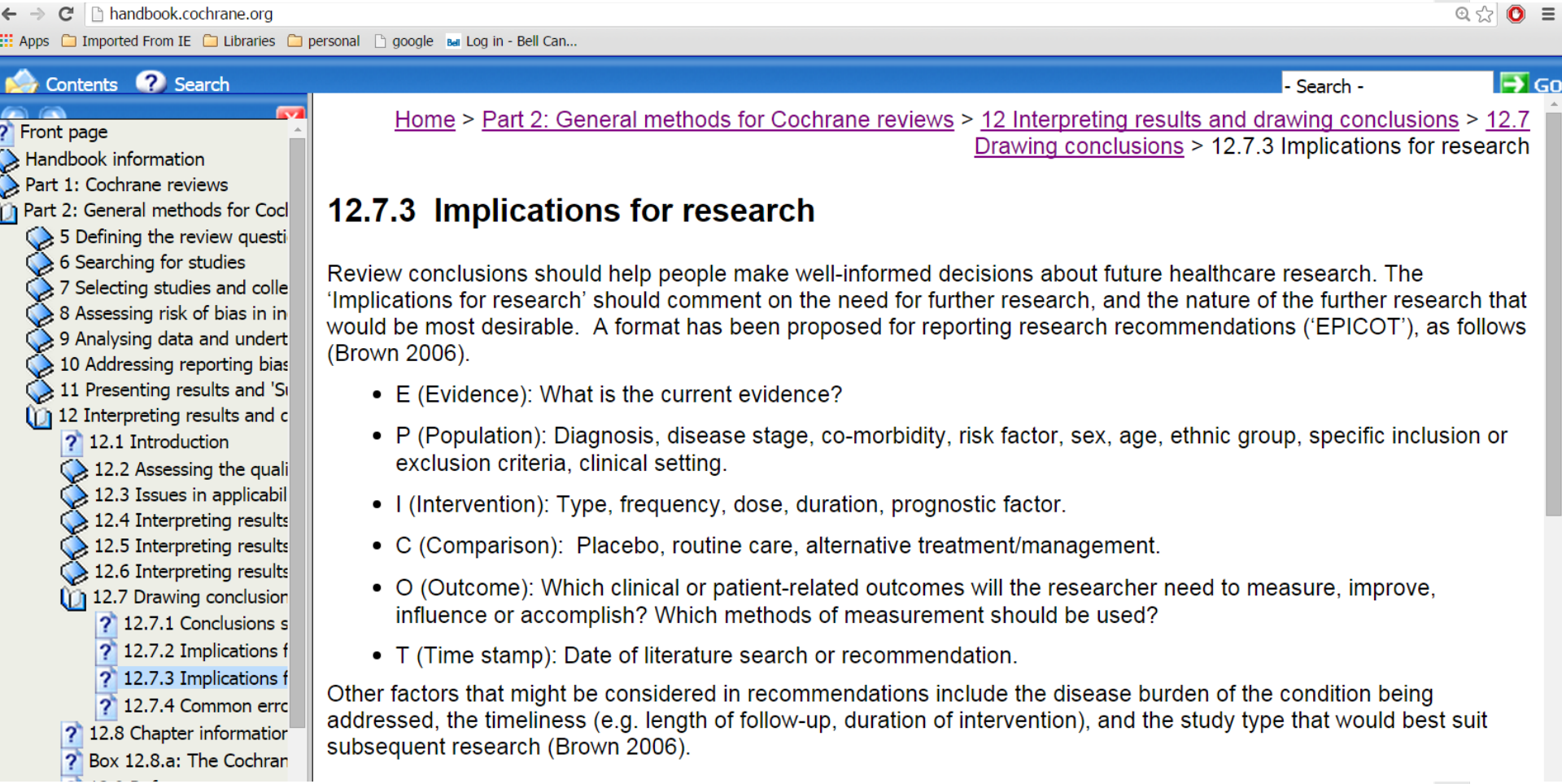
..... there is **no evidence** for using propylthiouracil for alcoholic liver disease....

Discussion section

- summary of the main results
- completeness and applicability of the results
- overall quality of the evidence
- potential biases in the review process – **NOT OF THE STUDIES**
  - Unable to translate all articles, did not conduct a sensitivity analysis when data was missing, we judged studies at high risk of bias when allocation concealment unknown, …
- agreement with other studies or reviews

Implications for research

Don't forget your GRADE assessment and how studies/evidence could be improved



Summary of Findings tables

- next

Thank you!

Today - Thursday

Before lunch

- Review of yesterday's work
  - Worked example
- Move from evidence to decisions
- Understand key criteria

Lunch to afternoon break

- Complete your own EtD framework

What you learned

Importance of proper question development

Selecting outcomes

GRADE domains for certainty

- Considered all downgrading domains
- Reviewed upgrading domains
- Use of ROBINS-I
- Specific examples of GRADEing
- Judgments and transparency – not always truth

Use of GRADEpro

Practical example – bringing it all together

Flavanoids for Hemorrhoids

venotonic agents

- mechanism unclear, increase venous return

popularity

- 90 venotonics commercialized in France
- none in Sweden and Norway
- France 70% of world market

possibilities

- French misguided
- rest of world missing out

Systematic Review

14 trials, 1432 patients

key outcome

- risk not improving/persistent symptoms
- 11 studies, 1002 patients, 375 events
- RR 0.4, 95% CI 0.29 to 0.57

minimal side effects

is France right?

what is the quality of evidence?

What can lower quality?

Study limitations/risk of bias

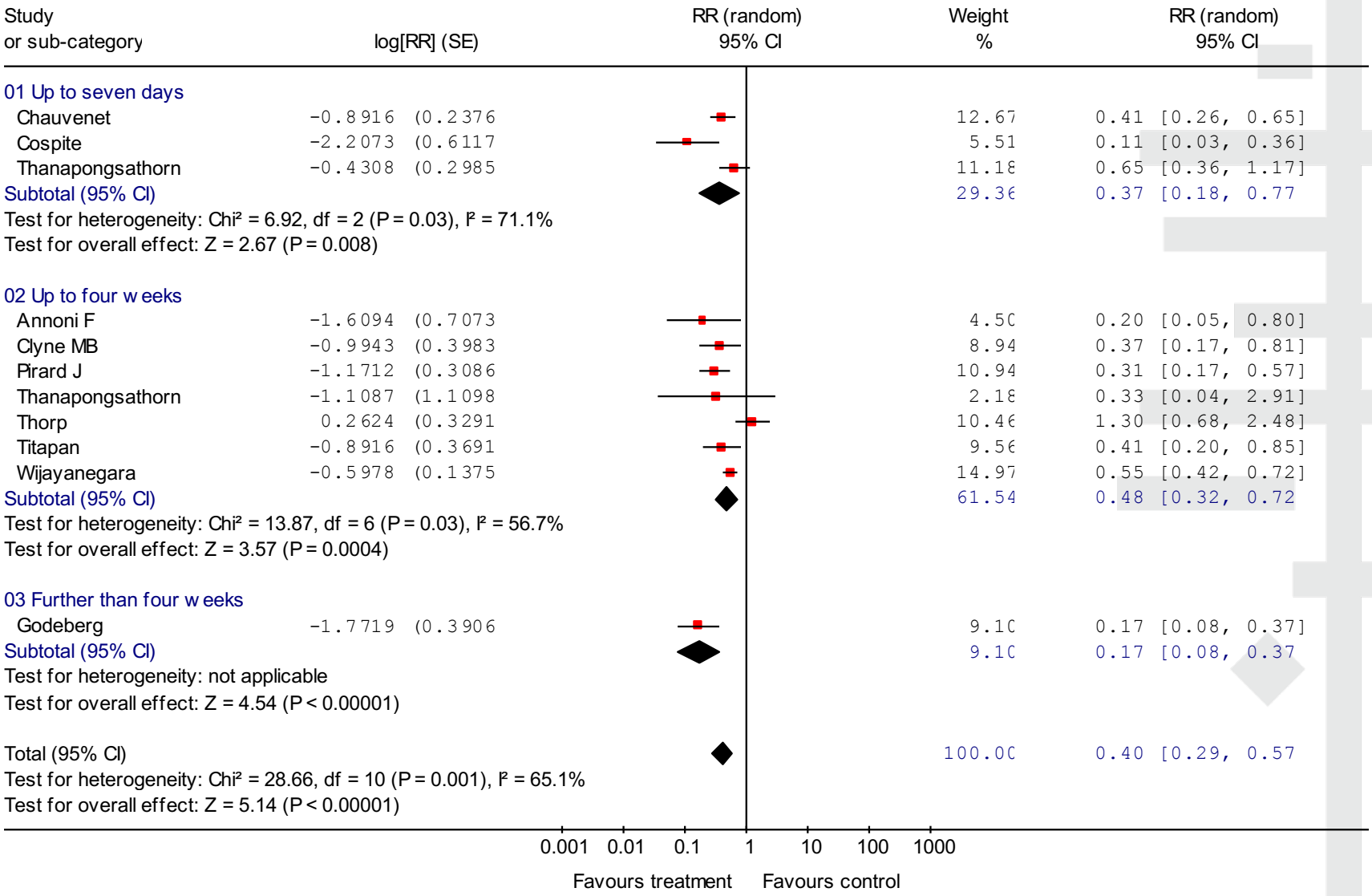
- lack of detail re concealment
- questionnaires not validated

rate down quality for study limitations/RoB?

indirectness – no problem

inconsistency, need to look at the results

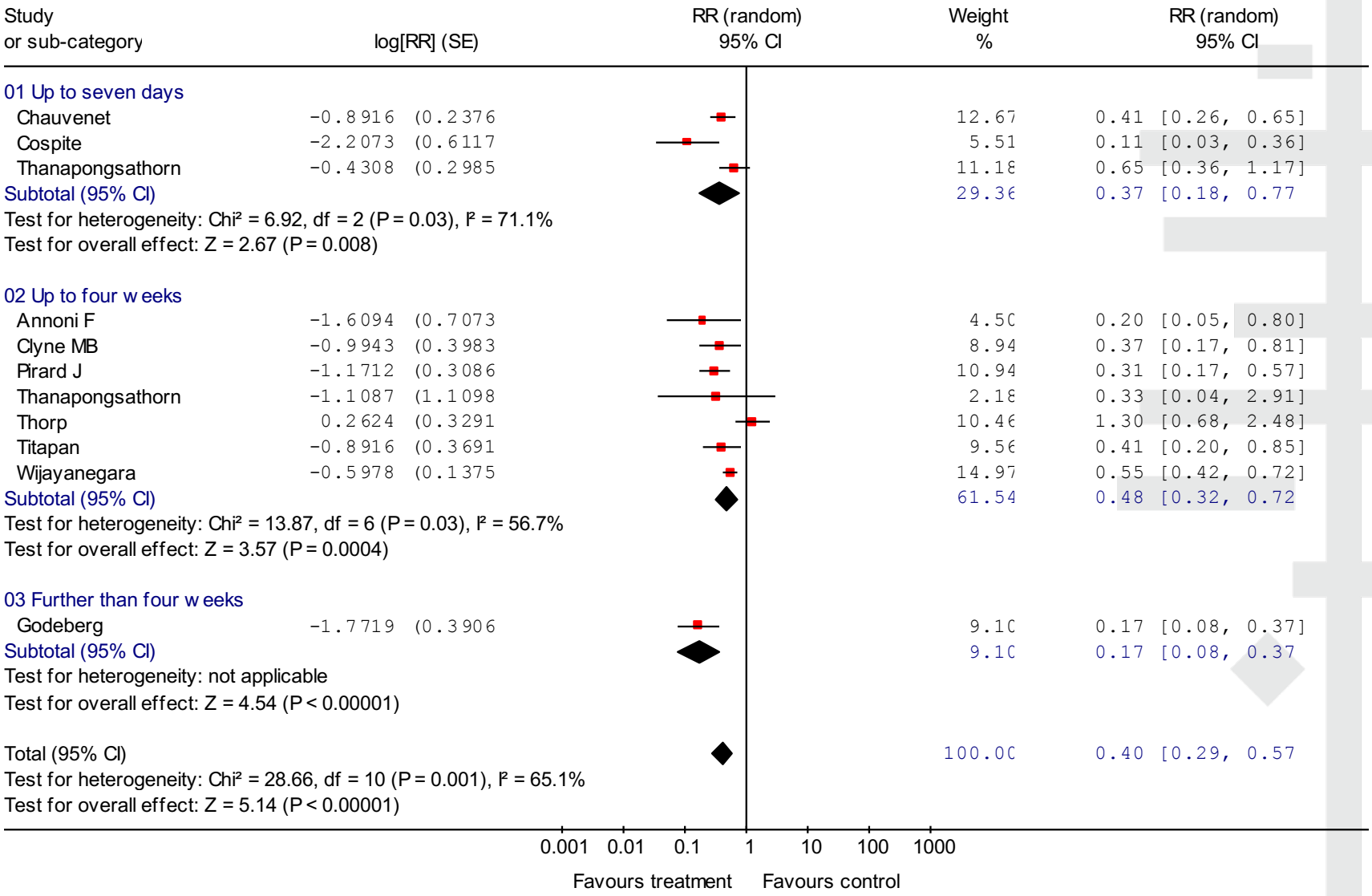
Review : Phlebotonics for hemorrhoids
Comparison: 01 Venotonics vs placebp
Outcome: 08 Overall improvement: no improvement/some improvement



Would you downgrade for inconsistency?

- [ ] No, there **is no serious inconsistency**
- [ ] Yes, there is serious inconsistency
- [ ] Yes, there is very serious inconsistency

Review : Phlebotonics for hemorrhoids
Comparison: 01 Venotonics vs placebp
Outcome: 08 Overall improvement: no improvement/some improvement



Is the imprecision…

- Not serious
- Serious
- Very Serious

…that is confidence/certainty in the result is reduced?

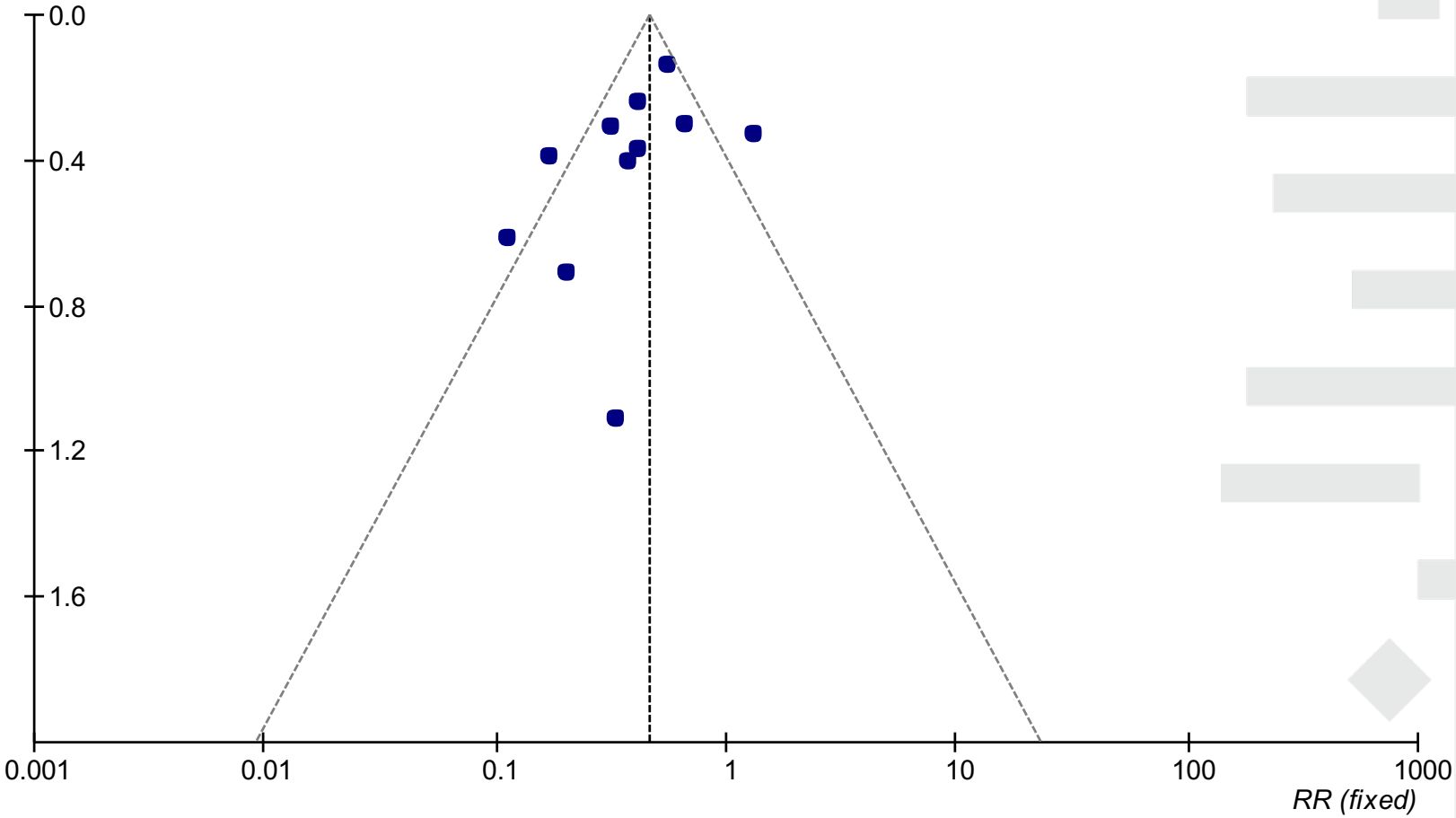
Publication bias?

size of studies

- 40 to 234 patients, most around 100

all industry sponsored

Review: Phlebotonics for hemorrhoids
Comparison: 01 Venotonics vs placebp
Outcome: 08 Overall improvement: no improvement/some improvement



Would you downgrade for publication bias?

- [ ] No, there is no publication bias
- [ ] Yes, there is publication bias
- [ ] Yes, there is very serious publication bias

Overall certainty?

risk of bias

- lack of detail re concealment
- questionnaires not validated

Inconsistency

- heterogeneity $p < 0.001$; I^2 65.1%

indirectness

imprecision

- RR 0.4, 95% CI 0.29 to 0.57
- 1002 patients, 375 events

Publication bias

- 40 to 234 patients in studies, most around 100

Your final judgment

Ratings
⨁⨁⨁⨁ High certainty
⨁⨁⨁◯ Moderate certainty
⨁⨁◯◯ Low certainty
⨁◯◯◯ Very low certainty

Interpreting the certainty in or quality of evidence

Ratings	Definitions
⨁⨁⨁⨁ High certainty	The panel is very confident that the true effect lies close to that of the estimate of the effect
⨁⨁⨁◯ Moderate certainty	The panel is 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 certainty	The panel's confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
⨁◯◯◯ Very low certainty	The panel has very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Self management for patients with chronic obstructive pulmonary disease

Patient or population: patients with chronic obstructive pulmonary disease
Settings: primary care, community, outpatient
Intervention: self management[1]
Comparison: usual care

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk **usual care**	Corresponding risk **self management**				
Quality of Life St George's Respiratory Questionnaire. Scale from: 0 to 100. (follow-up: 3 to 12 months)	The mean quality of life ranged across control groups from **38 to 60 points**	The mean quality of Life in the intervention groups was **2.58 lower** (5.14 to 0.02 lower)		698 (7)	⊕⊕⊕O **moderate**[2]	Lower score indicates better quality of life. A change of less than 4 points is not shown to be important to patients.
Dyspnoea Borg Scale. Scale from: 0 to 10. (follow-up: 3 to 6 months)	The mean dyspnoea ranged across control groups from **1.2 to 4.1 points**	The mean dyspnoea in the intervention groups was **0.53 lower** (0.96 to 0.1 lower)		144 (2)	⊕⊕OO **low**[3,4]	Lower score indicates improvement
Number and severity of exacerbations[5]	See comment	See comment	Not estimable[5]	591 (3)	See comment	Effect is uncertain
Respiratory-related hospital admissions (follow-up: 3 to 12 months)	**Low risk population**[6]		OR 0.64 (0.47 to 0.89)	966 (8)	⊕⊕⊕O **moderate**[7]	
	10 per 100	**7 per 100** (5 to 9)				
	High risk population[6]					
	50 per 100	**39 per 100** (32 to 47)				
Emergency department visits for lung diseases (follow-up: 6 to 12 months)	The mean emergency department visits for lung diseases ranged across control groups from **0.2 to 0.7 visits per person per year**	The mean emergency department visits for lung diseases in the intervention groups was **0.1 higher** (0.2 lower to 0.3 higher)		328 (4)	⊕⊕⊕O **moderate**[4]	
Doctor and nurse visits (follow-up: 6 to 12 months)	The mean doctor and nurse visits ranged across control groups from **1 to 5 vists per person per year**	The mean doctor and nurse visits in the intervention groups was **0.02 higher** (1 lower to 1 higher)		629 (8)	⊕⊕⊕O **moderate**[8]	

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; OR: Odds ratio;

Assessing Certainty in Evidence by Outcome

Table: GRADE's approach to rating quality of evidence (aka confidence in effect estimates)

For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)

1. Establish initial level of confidence		2. Consider lowering or raising level of confidence		3. Final level of confidence rating
Study design	***Initial confidence in an estimate of effect***	***Reasons for considering lowering or raising confidence***		***Confidence in an estimate of effect across those considerations***
		⬇ Lower if	**⬆ Higher if***	
Randomized trials ➔	**High confidence**	**Risk of Bias** **Inconsistency** **Indirectness** **Imprecision** **Publication bias**	**Large effect** **Dose response** **All plausible confounding & bias** • would reduce a demonstrated effect **or** • would suggest a spurious effect if no effect was observed	**High** ⨁⨁⨁⨁
				Moderate ⨁⨁⨁◯
Observational studies ➔	**Low confidence**			**Low** ⨁⨁◯◯
				Very low ⨁◯◯◯

*upgrading criteria are usually applicable to observational studies only.

Lowering certainty in RCTs

Table: GRADE's approach to rating quality of evidence (aka confidence in effect estimates)

For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)

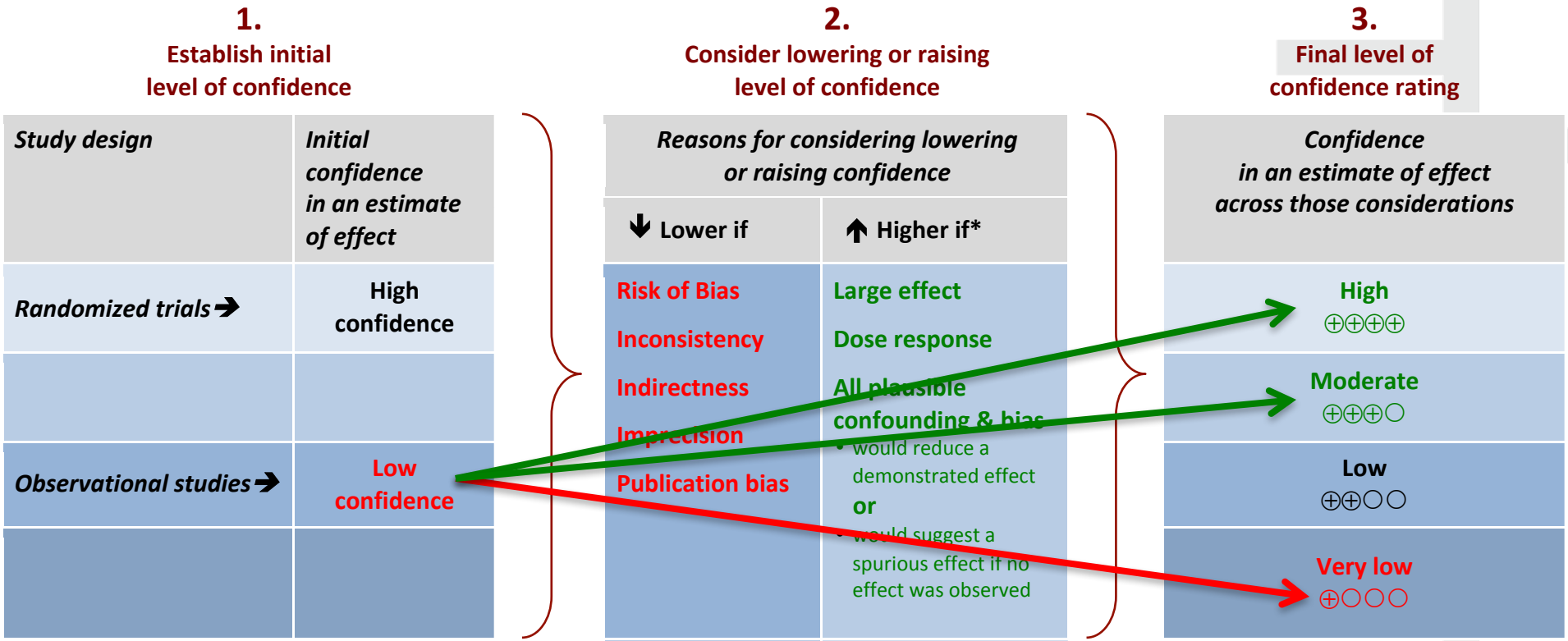
1. Establish initial level of confidence		**2. Consider lowering or raising level of confidence**		**3. Final level of confidence rating**
Study design	***Initial confidence in an estimate of effect***	***Reasons for considering lowering or raising confidence***		***Confidence in an estimate of effect across those considerations***
		⬇ Lower if	**⬆ Higher if***	
Randomized trials ➔	**High confidence**	**Risk of Bias** **Inconsistency** **Indirectness** **Imprecision** **Publication bias**	**Large effect** **Dose response** **All plausible confounding & bias** • would reduce a demonstrated effect **or** • would suggest a spurious effect if no effect was observed	**High** ⨁⨁⨁⨁
				Moderate ⨁⨁⨁◯
Observational studies ➔	**Low confidence**			**Low** ⨁⨁◯◯
				Very low ⨁◯◯◯

*upgrading criteria are usually applicable to observational studies only.

Altering certainty in observational studies

Table: GRADE's approach to rating quality of evidence (aka confidence in effect estimates)

For each outcome based on a systematic review and across outcomes (lowest quality across the outcomes critical for decision making)

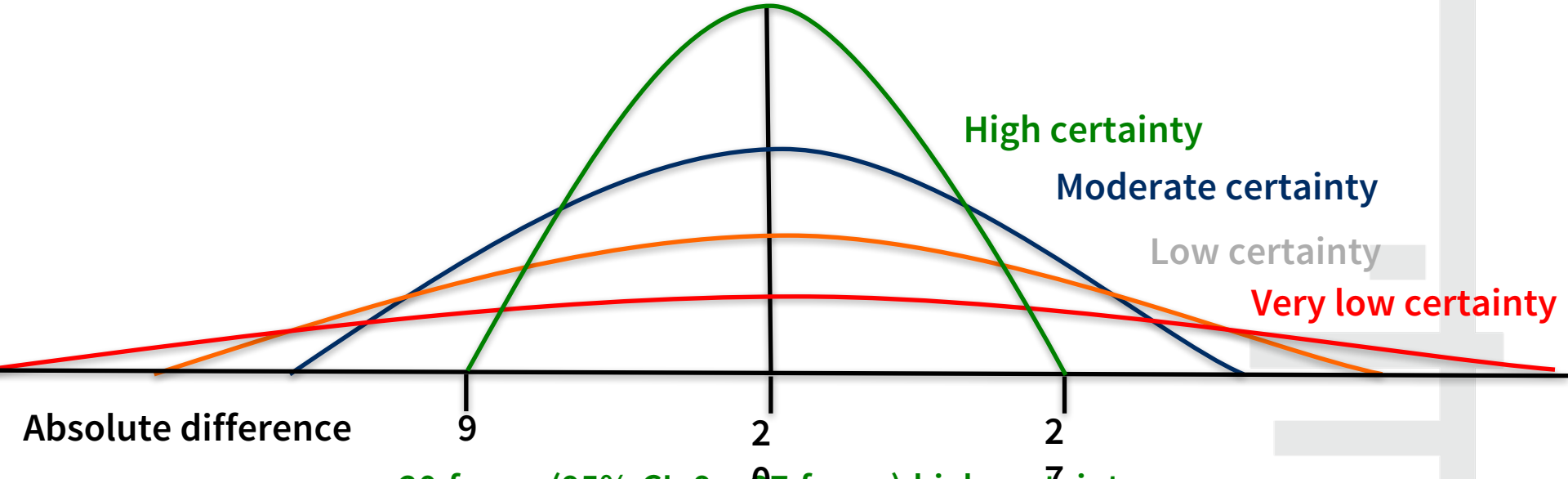


*upgrading criteria are usually applicable to observational studies only.

Magnitude of Effect

Likelihood of and certainty in the evidence or effect

Certainty or Quality of evidence Confidence in effect



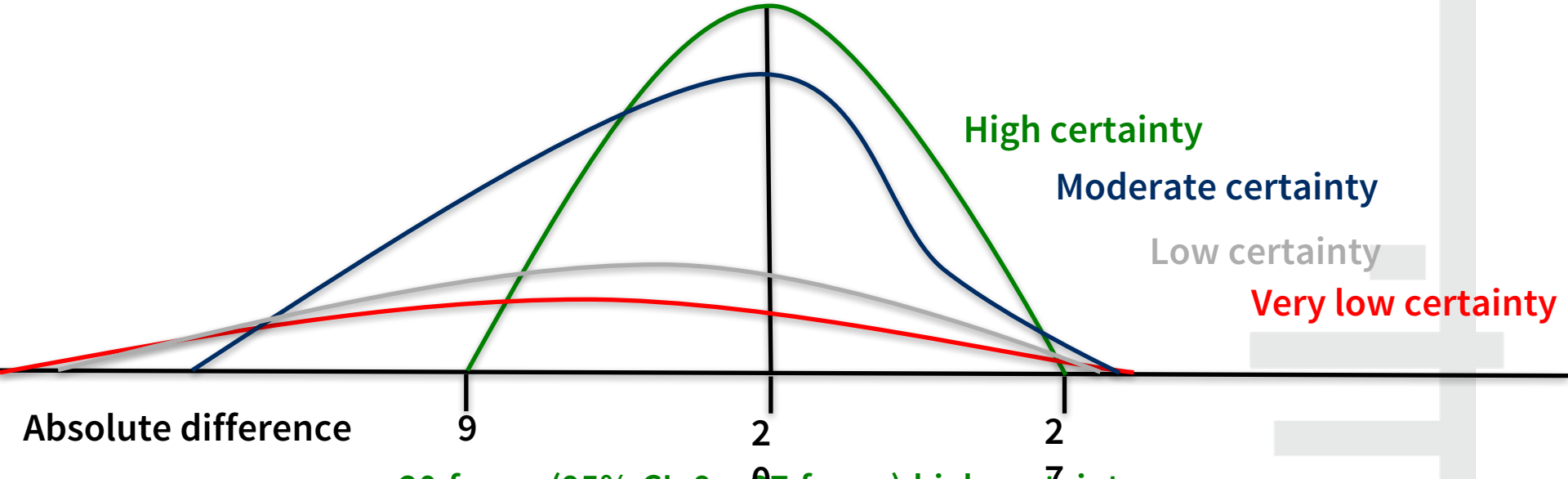
20 fewer (95% CI: 9 – 27 fewer) high certainty with sufficiently narrow confidence range – no downgrading
Certainty range identical to CI: distribution known

20 fewer (95% CI: 9 – 27 fewer) moderate certainty due to indirectness or serious concerns about any other downgrading domain including imprecision – wide certainty range despite narrow confidence intervals – certainty range wider than 9 – 27 but shape and width not exactly known

20 fewer (95% CI: 9 – 27 fewer) low certainty due to risk of bias and indirectness – very wide certainty range despite narrow confidence intervals

20 fewer (95% CI: 9 – 27 fewer) very low certainty due to risk of bias, indirectness and publication bias – extremely wide certainty range

Schünemann, JCE 2016



20 fewer (95% CI: 9 – 27 fewer) high certainty with sufficiently narrow confidence range – no downgrading
Certainty range identical to CI: distribution known

20 fewer (95% CI: 9 – 27 fewer) moderate certainty due to risk of bias or serious concerns about any other downgrading domain including imprecision – wide certainty range despite narrow confidence intervals – certainty range wider than 9 – 27 but shape and width not exactly known

20 fewer (95% CI: 9 – 27 fewer) low certainty due to risk of bias and indirectness – very wide certainty range despite narrow confidence intervals

20 fewer (95% CI: 9 – 27 fewer) very low certainty due to risk of bias, indirectness and publication bias – extremely wide certainty range

Results section

- present all your results systematically, including:
  - outcomes combined in meta-analysis
  - outcomes for which no evidence was found
  - outcomes for which meta-analysis was not possible
    - e.g. studies too different
    - e.g. data not available in comparable format

Caution when making conclusions

Do not make recommendations

- Recommendations involve consideration of setting, values and preferences of patients, resources, etc.

✖ “If people want to stay alert during the day, they should drink coffee.”

✖ “Joint lavage should be discouraged in patients with osteoarthritis.”

Instead....

Indicate level of evidence and effect on outcomes

- “Moderate quality evidence shows that alertness will probably improve by drinking coffee.”
- “There is low quality evidence that joint lavage provides little or no difference in symptoms of knee OA.”

DO NOT USE

"no evidence of effect" versus "evidence of no effect"

Results

Combining the results of six randomised clinical trials including 710 patients demonstrated no significant effects of propylthiouracil versus placebo on all-cause mortality (relative risks (RR) 0.93, 95% confidence interval (CI) 0.66 to 1.30), liver-related mortality (RR 0.80, 95% CI 0.50 to 1.29), complications of the liver disease, or liver histology.

Authors' conclusions

..... there is **no evidence** for using propylthiouracil for alcoholic liver disease....

Discussion section

- summary of the main results
- completeness and applicability of the results
- overall quality of the evidence
- potential biases in the review process – **NOT OF THE STUDIES**
  - Unable to translate all articles, did not conduct a sensitivity analysis when data was missing, we judged studies at high risk of bias when allocation concealment unknown, …
- agreement with other studies or reviews

Implications for research

Don't forget your GRADE assessment and how studies/evidence could be improved

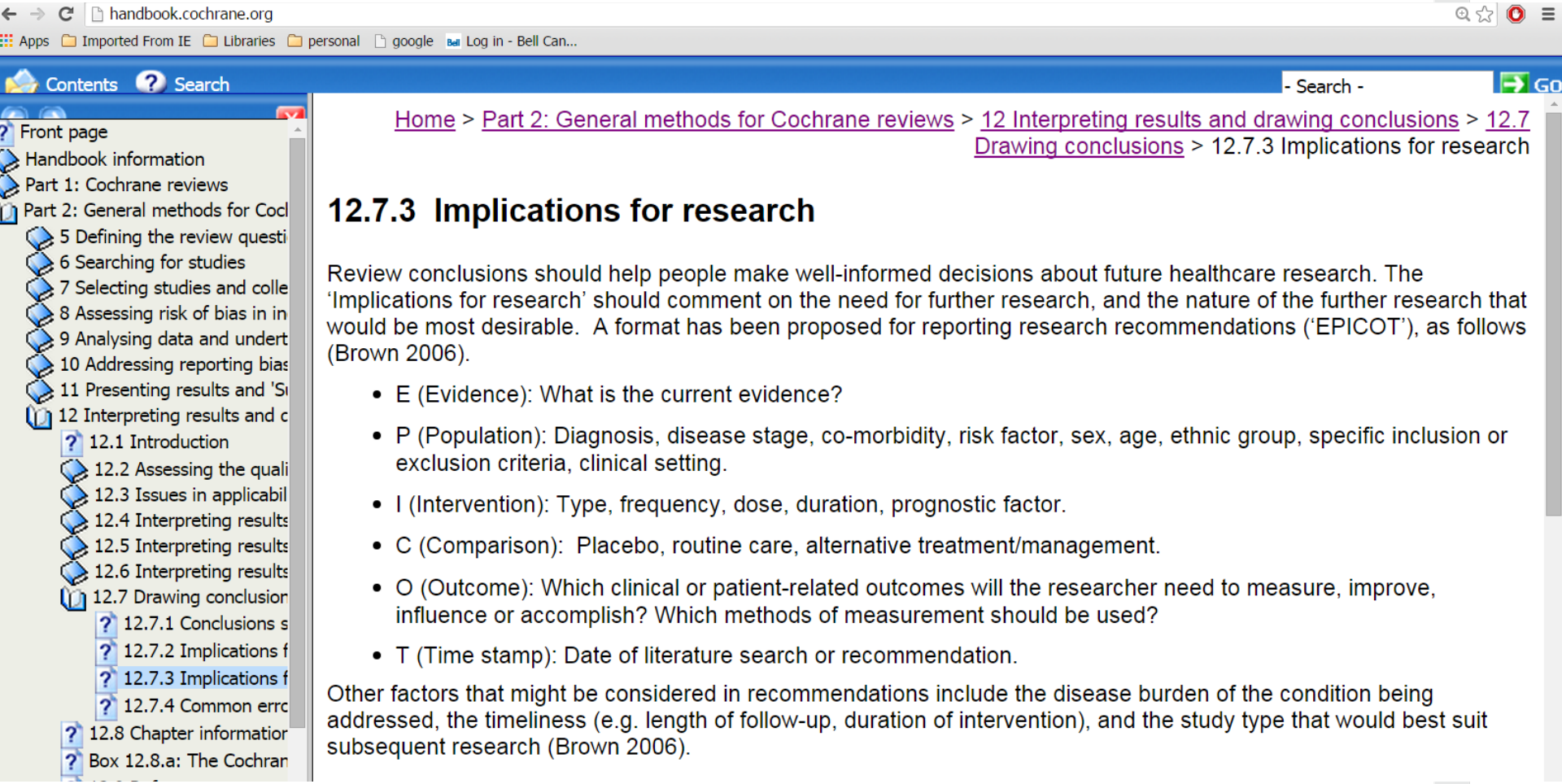


Table 8. Interpretation of the certainty in a body of evidence according to individual GRADE domains

By outcome	Implications for research	Examples	Implications for practice
Risk of bias	Need for methodologically better designed and executed studies	All studies suffered from lack of blinding of outcome assessors. Trials of this type are required.	The estimates of effect may be biased because of a lack of blinding.
Inconsistency	Unexplained inconsistency: need for individual participant data meta-analysis (IPDMA); need for studies in relevant subgroups	Studies in patients with small cell lung cancer are needed to understand if the effects differ from those in patients with pancreatic cancer.	Unexplained inconsistency: consider and interpret overall effect estimates as for the certainty in a body of evidence Explained inconsistency (if results are presented in strata): consider and interpret effects estimates by subgroup
Indirectness	Need for studies that more directly address the PICO question of interest	Studies in patients with early cancer are needed because the evidence is from studies with advanced cancer.	It is uncertain if the results directly apply to the patients or the way that the intervention is applied in your setting.
Imprecision	Need for more studies with more participants to reach optimal information size	Studies with approximately 200 more events in the treatment and control group are required.	Same as for certainty in a body of evidence
Publication bias	Need to investigate and identify unpublished data; large studies might help resolve this issue		Same as for certainty in a body of evidence

By outcome	Implications for research	Examples	Implications for practice
Large effects	No implications	No implications	The effect is large in the populations that were included in the studies. The effect is going to be in the vicinity of the observed effect.
Dose effects	No implications	No implications	The greater the reduction in the exposure the larger is the expected benefit (harm).
Opposing bias and confounding	Studies controlling for the residual bias and confounding are needed.	Studies controlling for following possible confounders are required smoking, degree of education.	The effect could be even larger than the one that is observed in the studies presented here.

Now that we have transparent evidence

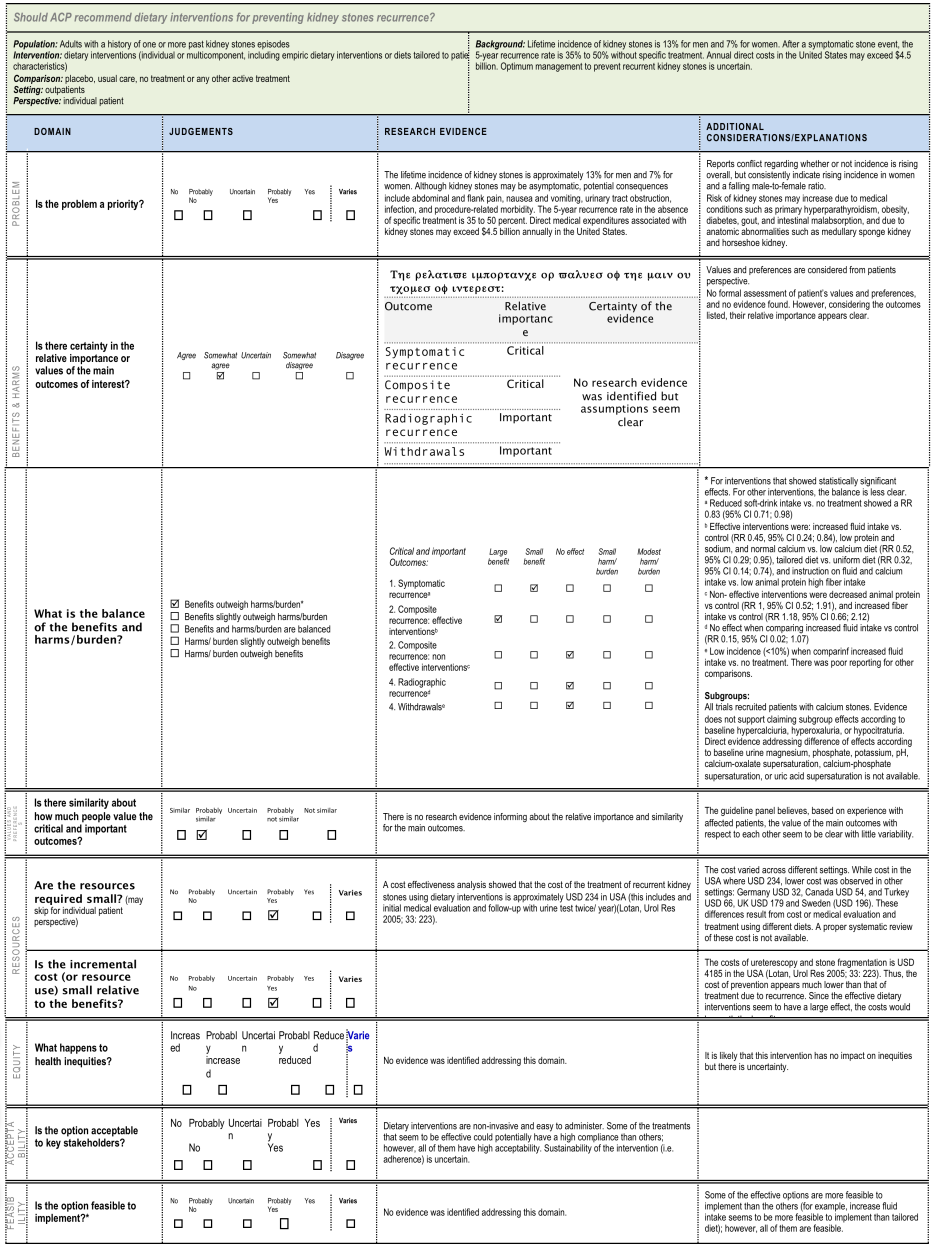
Table 1. Summary of Findings Table Showing the Relative Risks and Absolute Effects over 12 Months for Each Important Outcome after Treatment with a Low-Molecular-Weight Heparin in Patients Receiving Chemotherapy for Cancer.*

Outcome after 12 Months	Participants	Relative Risk (95% CI)	Anticipated Absolute Effect		Quality of Evidence (GRADE) and Comments†
			Risk without LMWH	Risk Difference with LMWH (95% CI)	
	no. (no. of studies)		*no. of events per 1000 patients*		
					venous thrombosis
Major bleeding	6518 (11)	1.06 (0.71–1.57)	16	1 more (5 fewer to 9 more)	Moderate-quality evidence owing to imprecision; the increase may be acceptable to patients, given that VTE, which occurs more frequently, may be equally unpleasant
Minor bleeding	6020 (9)	1.18 (0.89–1.55)	27	5 more (3 fewer to 15 more)	Moderate-quality evidence owing to imprecision; however, this outcome is unlikely to be critical for decision making

Should every cancer patient receive heparin?

Akl & Schünemann, New Engl J Med, 2012

- Question/Problem
- Benefits and harms
- Quality of evidence
- Values
- Resources
- Equity
- Acceptability
- Feasibility
- Recommendation



Should ACP recommend dietary interventions for preventing kidney stones recurrence?

Population: Adults with a history of one or more past kidney stones episodes
Intervention: dietary interventions (individual or multicomponent, including empiric dietary interventions or diets tailored to patie characteristics)
Comparison: placebo, usual care, no treatment or any other active treatment
Setting: outpatients
Perspective: individual patient

Background: Lifetime incidence of kidney stones is 13% for men and 7% for women. After a symptomatic stone event, the 5-year recurrence rate is 35% to 50% without specific treatment. Annual direct costs in the United States may exceed $4.5 billion. Optimum management to prevent recurrent kidney stones is uncertain.

	DOMAIN	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS/EXPLANATIONS
PROBLEM	**Is the problem a priority?**	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☐ · Yes ☐ · Varies ☐	The lifetime incidence of kidney stones is approximately 13% for men and 7% for women. Although kidney stones may be asymptomatic, potential consequences include abdominal and flank pain, nausea and vomiting, urinary tract obstruction, infection, and procedure-related morbidity. The 5-year recurrence rate in the absence of specific treatment is 35 to 50 percent. Direct medical expenditures associated with kidney stones may exceed $4.5 billion annually in the United States.	Reports conflict regarding whether or not incidence is rising overall, but consistently indicate rising incidence in women and a falling male-to-female ratio. Risk of kidney stones may increase due to medical conditions such as primary hyperparathyroidism, obesity, diabetes, gout, and intestinal malabsorption, and due to anatomic abnormalities such as medullary sponge kidney and horseshoe kidney.
BENEFITS & HARMS	**Is there certainty in the relative importance or values of the main outcomes of interest?**	Agree ☐ · Somewhat agree ☑ · Uncertain ☐ · Somewhat disagree ☐ · Disagree ☐	Τηε ρελατιωε ιμπορτανχε ορ ωαλυεσ οφ τηε μαιν ουτχομεσ οφ ιντερεστ: Outcome — Relative importance — Certainty of the evidence Symptomatic recurrence — Critical Composite recurrence — Critical — No research evidence was identified but assumptions seem clear Radiographic recurrence — Important Withdrawals — Important	Values and preferences are considered from patients perspective. No formal assessment of patient's values and preferences, and no evidence found. However, considering the outcomes listed, their relative importance appears clear.
BENEFITS & HARMS	**What is the balance of the benefits and harms/burden?**	☑ Benefits outweigh harms/burden* ☐ Benefits slightly outweigh harms/burden ☐ Benefits and harms/burden are balanced ☐ Harms/ burden slightly outweigh benefits ☐ Harms/ burden outweigh benefits	*Critical and important Outcomes:* — Large benefit — Small benefit — No effect — Small harm/ burden — Modest harm/ burden 1. Symptomatic recurrence[a]: ☐ ☑ ☐ ☐ ☐ 2. Composite recurrence: effective interventions[b]: ☑ ☐ ☐ ☐ ☐ 2. Composite recurrence: non effective interventions[c]: ☐ ☐ ☑ ☐ ☐ 4. Radiographic recurrence[d]: ☐ ☐ ☑ ☐ ☐ 4. Withdrawals[e]: ☐ ☐ ☑ ☐ ☐	* For interventions that showed statistically significant effects. For other interventions, the balance is less clear [a] Reduced soft-drink intake vs. no treatment showed a RR 0.83 (95% CI 0.71; 0.98) [b] Effective interventions were: increased fluid intake vs. control (RR 0.45, 95% CI 0.24; 0.84), low protein and sodium, and normal calcium vs. low calcium diet (RR 0.52, 95% CI 0.29; 0.95), tailored diet vs. uniform diet (RR 0.32, 95% CI 0.14; 0.74), and instruction on fluid and calcium intake vs. low animal protein high fiber intake [c] Non- effective interventions were decreased animal protein vs control (RR 1, 95% CI 0.52; 1.91), and increased fiber intake vs control (RR 1.18, 95% CI 0.66; 2.12) [d] No effect when comparing increased fluid intake vs control (RR 0.15, 95% CI 0.02; 1.07) [e] Low incidence (<10%) when comparinf increased fluid intake vs. no treatment. There was poor reporting for other comparisons. **Subgroups:** All trials recruited patients with calcium stones. Evidence does not support claiming subgroup effects according to baseline hypercalciuria, hyperoxaluria, or hypocitraturia. Direct evidence addressing difference of effects according to baseline urine magnesium, phosphate, potassium, pH, calcium-oxalate supersaturation, calcium-phosphate supersaturation, or uric acid supersaturation is not available.
VALUES	**Is there similarity about how much people value the critical and important outcomes?**	Similar ☐ · Probably similar ☑ · Uncertain ☐ · Probably not similar ☐ · Not similar ☐	There is no research evidence informing about the relative importance and similarity for the main outcomes.	The guideline panel believes, based on experience with affected patients, the value of the main outcomes with respect to each other seem to be clear with little variability.
RESOURCES	**Are the resources required small?** (may skip for individual patient perspective)	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☑ · Yes ☐ · Varies ☐	A cost effectiveness analysis showed that the cost of the treatment of recurrent kidney stones using dietary interventions is approximately USD 234 in USA (this includes and initial medical evaluation and follow-up with urine test twice/ year)(Lotan, Urol Res 2005; 33: 223).	The cost varied across different settings. While cost in the USA where USD 234, lower cost was observed in other settings: Germany USD 32, Canada USD 54, and Turkey USD 66, UK USD 179 and Sweden (USD 196). These differences result from cost or medical evaluation and treatment using different diets. A proper systematic review of these cost is not available.
RESOURCES	**Is the incremental cost (or resource use) small relative to the benefits?**	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☑ · Yes ☐ · Varies ☐		The costs of ureteroscopy and stone fragmentation is USD 4185 in the USA (Lotan, Urol Res 2005; 33: 223). Thus, the cost of prevention appears much lower than that of treatment due to recurrence. Since the effective dietary interventions seem to have a large effect, the costs would
EQUITY	**What happens to health inequities?**	Increased ☐ · Probably increased ☐ · Uncertain ☐ · Probably reduced ☐ · Reduced ☐ · Varies ☐	No evidence was identified addressing this domain.	It is likely that this intervention has no impact on inequities but there is uncertainty.
ACCEPTABILITY	**Is the option acceptable to key stakeholders?**	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☐ · Yes ☐ · Varies ☐	Dietary interventions are non-invasive and easy to administer. Some of the treatments that seem to be effective could potentially have a high compliance than others; however, all of them have high acceptability. Sustainability of the intervention (i.e. adherence) is uncertain.	
FEASIBILITY	**Is the option feasible to implement?***	No ☐ · Probably No ☐ · Uncertain ☐ · Probably Yes ☐ · Yes ☐ · Varies ☐	No evidence was identified addressing this domain.	Some of the effective options are more feasible to implement than the others (for example, increase fluid intake seems to be more feasible to implement than tailored diet); however, all of them are feasible.

Recommendation

Should ACP recommend any dietary intervention for preventing kidney stones recurrence?

Overall balance of consequences

Undesirable consequences clearly outweigh *desirable consequences*	*Undesirable consequences* probably outweigh *desirable consequences*	The balance between desirable and undesirable consequences is too uncertain*	The balance of desirable and undesirable consequences indicates they are very similar*	*Desirable consequences* probably outweigh *undesirable consequences*	*Desirable consequences* clearly outweigh *undesirable consequences*
☐	☐	☐	☐	☑	☐
We recommend against the option or for the alternative	*We suggest not to use the option or to use the alternative*	*No recommendation*		*We suggest using the option*	*We recommend the option*

GRADE DECIDE

Evidence to decision

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

Criteria	How the factor influences the direction and strength of a recommendation
Problem	The problem is determined by the importance and frequency of the health care issue that is addressed (burden of disease, prevalence or baseline risk). If the problem is of great importance a strong recommendation is more likely.
Values and preferences	Values and preferences or the importance of outcomes. This describes how important health outcomes are to those affected, how variable the importance is and if there is uncertainty about this.
Certainty in the evidence	The higher the certainty in the evidence the more likely is a strong recommendation.
Health benefits and harms and burden and their balance	This requires an evaluation of the absolute effects of both the benefits and harms and their importance. The greater the net benefit or net harm the more likely is a strong recommendation for or against the option.
Resource implications	This describes how resource intense an option is, if it is cost-effective and if there is incremental benefit. The more advantageous or clearly disadvantageous these resource implications are the more likely is a strong recommendation.
Equity	The greater the likelihood to reduce inequities or increase equity and the more accessible an option is, the more likely is a strong recommendation.
Acceptability	The greater the acceptability of an option to all or most stakeholders, the more likely is a strong recommendation.
Feasibility	The greater the feasibility of an option to all or most stakeholders, the more likely is a strong recommendation.

For groups making recommendations

Question

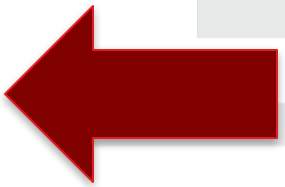
- Details
- Subgroups
- Background

Assessment

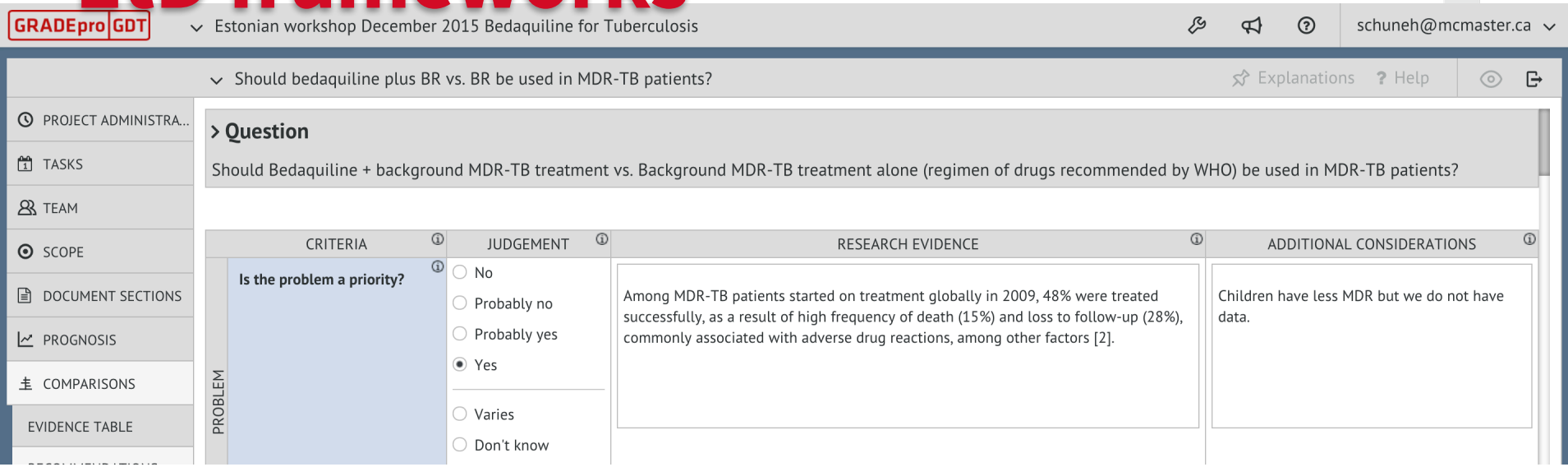
- Criteria
- Judgements
- Research evidence
- Additional considerations

Conclusions

- Type of recommendation
- Recommendation
- Justification
- Implementation considerations
- Monitoring and evaluation
- Research considerations



EtD frameworks



Criteria on which a recommendation is based

Judgements that must be made in relation to each criterion

Research evidence to inform each judgement

Additional considerations that inform or explain each judgement

GRADE Evidence to Decision (EtD) framework

Can help guideline panels (and decision makers) move from evidence to a recommendation or decision by

Informing judgements about the pros and cons of each option (intervention)

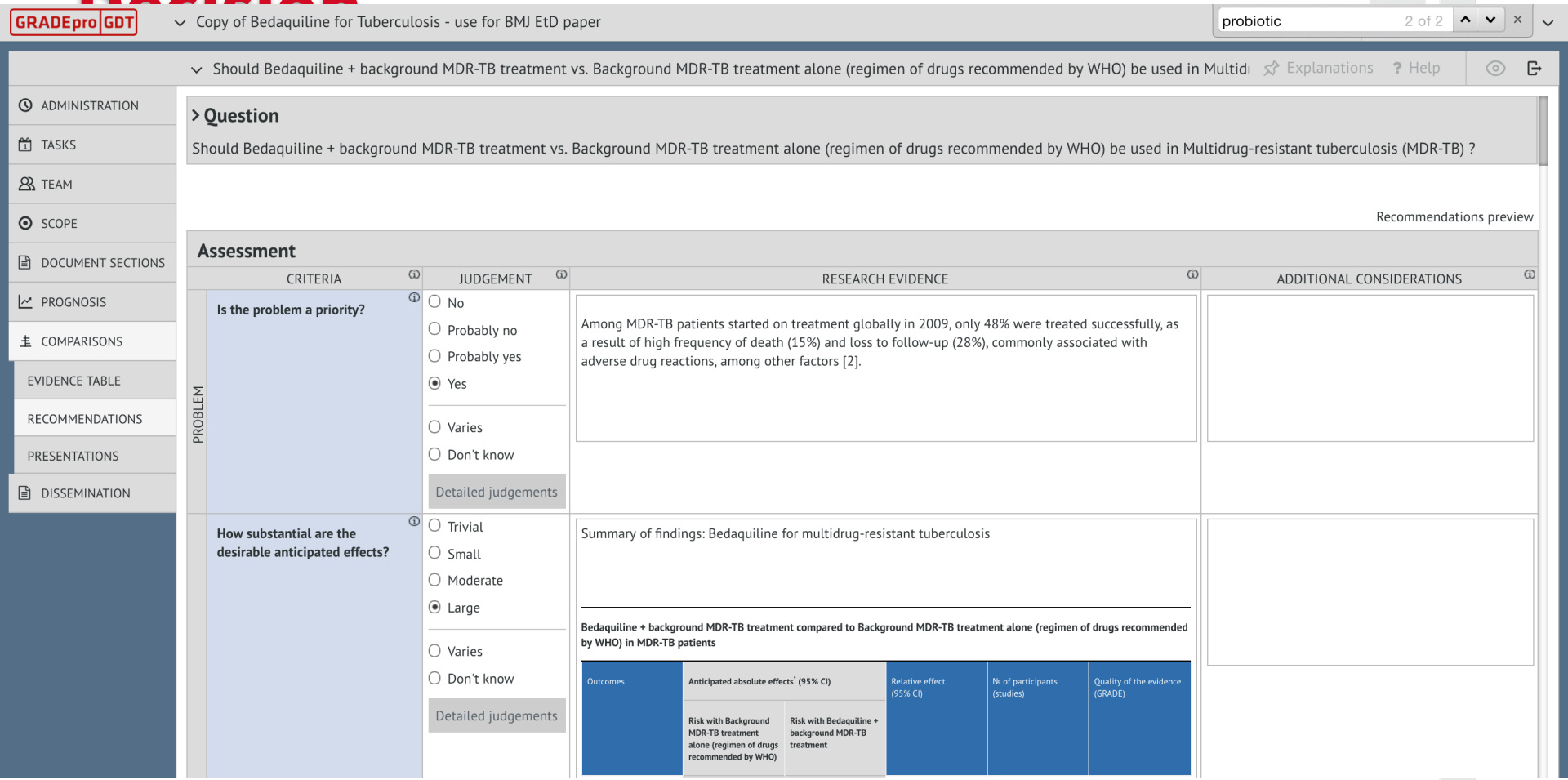
Considering each important factor that determine a decision (criteria)

Providing a concise summary of the best available research evidence to inform judgements

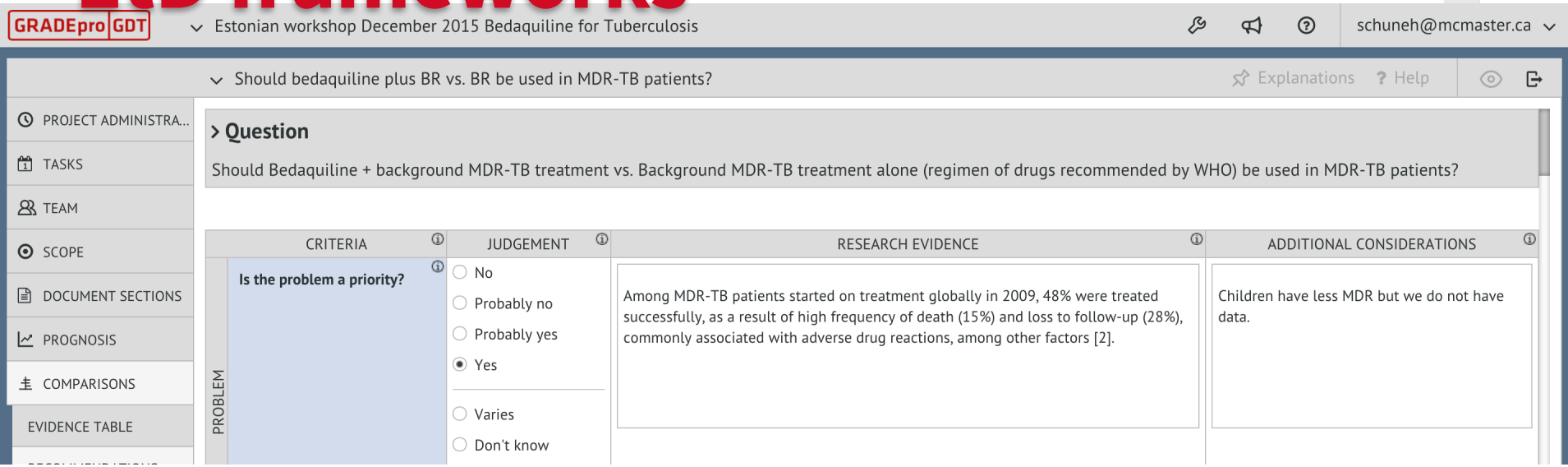
Helping to structure discussion and identify reasons for disagreements

Making the basis for decisions transparent and adaptable for target audiences

Interactive Evidence to Decision



EtD frameworks



Criteria on which a recommendation is based

Judgements that must be made in relation to each criterion

Research evidence to inform each judgement

Additional considerations that inform or explain each judgement

Live use of iEtDs

EtDs are shared with panel members before the meeting and online:

Clarify the process

During the preparation for input on the evidence (all members including conflicted members could be involved)

For initial agreement on the included evidence and additional considerations

If possible, feasible and appropriate for agreement on judgments for specific decision criteria (but may all happen at an in-person meeting)

Final draft EtDs before a final meeting

What are guideline panel members doing?

Discuss evidence

Question

Should Acyclovir vs. Placebo be used for treatment of first clinical episodes of Herpes Simplex Virus 2?

CRITERIA	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM **Is the problem a priority?**	○ No ○ Probably no ○ Probably yes ◉ Yes ○ Varies ○ Don't know Detailed judgements	Globally, it is estimated that XXXXXXX people are newly infected with HSV2 each year. When symptoms of genital herpes occur, there are generally one or more genital or anal blisters called ulcers. First-episode infections of genital herpes are more extensive and primary lesions last two to six weeks versus approximately one week for lesions in recurrent disease. Infection with HSV2 also may increase the risk of acquiring HIV infection. Moreover, HSV2 can be transmitted to neonates from an infected pregnant mother.	
DESIRABLE EFFECTS **How substantial are the desirable anticipated effects?**	○ Trivial ○ Small ○ Moderate ◉ Large ○ Varies ○ Don't know Detailed judgements	We found 5 randomised controlled trials comparing acyclovir in different doses compared to placebo. See Table below for the summary of the evidence.	

Acyclovir compared to Placebo for treatment of first clinical episodes of Herpes Simplex Virus 2

Outcomes	№ of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Placebo	Risk difference with Acyclovir
Duration of symptoms from onset of treatment assessed with: time to resolution	238 (5 RCTs) [1]	⨁⨁◯◯ LOW [2,3]	-	The mean duration of symptoms from onset of treatment was **0** days	MD **3.2 days fewer** (4.94 fewer to 1.46 fewer)
Pain	129 (3 RCTs) [4]	⨁⨁◯◯ LOW [2,3]	-	The mean pain was **0** days	MD **2.1 days fewer** (2.95 fewer to 1.25 fewer)

Add relevant considerations

Question

Should Acyclovir vs. Placebo be used for treatment of first clinical episodes of Herpes Simplex Virus 2?

CRITERIA	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM **Is the problem a priority?**	○ No ○ Probably no ○ Probably yes ◉ Yes ○ Varies ○ Don't know Detailed judgements	Globally, it is estimated that XXXXXXX people are newly infected with HSV2 each year. When symptoms of genital herpes occur, there are generally one or more genital or anal blisters called ulcers. First-episode infections of genital herpes are more extensive and primary lesions last two to six weeks versus approximately one week for lesions in recurrent disease. Infection with HSV2 also may increase the risk of acquiring HIV infection. Moreover, HSV2 can be transmitted to neonates from an infected pregnant mother.	
DESIRABLE EFFECTS **How substantial are the desirable anticipated effects?**	○ Trivial ○ Small ○ Moderate ◉ Large ○ Varies ○ Don't know Detailed judgements	We found 5 randomised controlled trials comparing acyclovir in different doses compared to placebo. See Table below for the summary of the evidence.	

Acyclovir compared to Placebo for treatment of first clinical episodes of Herpes Simplex Virus 2

Outcomes	№ of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Placebo	Risk difference with Acyclovir
Duration of symptoms from onset of treatment assessed with: time to resolution	238 (5 RCTs) [1]	⨁⨁◯◯ LOW [2,3]	-	The mean duration of symptoms from onset of treatment was **0** days	MD **3.2 days fewer** (4.94 fewer to 1.46 fewer)
Pain	129 (3 RCTs) [4]	⨁⨁◯◯ LOW [2,3]	-	The mean pain was **0** days	MD **2.1 days fewer** (2.95 fewer to 1.25 fewer)

Make judgments
(when research evidence complete) – w/o COI

› Question

Should Acyclovir vs. Placebo be used for treatment of first clinical episodes of Herpes Simplex Virus 2?

CRITERIA	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM **Is the problem a priority?**	○ No ○ Probably no ○ Probably yes ◉ Yes ○ Varies ○ Don't know Detailed judgements	Globally, it is estimated that XXXXXXX people are newly infected with HSV2 each year. When symptoms of genital herpes occur, there are generally one or more genital or anal blisters called ulcers. First-episode infections of genital herpes are more extensive and primary lesions last two to six weeks versus approximately one week for lesions in recurrent disease. Infection with HSV2 also may increase the risk of acquiring HIV infection. Moreover, HSV2 can be transmitted to neonates from an infected pregnant mother.	
DESIRABLE EFFECTS **How substantial are the desirable anticipated effects?**	○ Trivial ○ Small ○ Moderate ◉ Large ○ Varies ○ Don't know Detailed judgements	We found 5 randomised controlled trials comparing acyclovir in different doses compared to placebo. See Table below for the summary of the evidence.	

Acyclovir compared to Placebo for treatment of first clinical episodes of Herpes Simplex Virus 2

Outcomes	№ of participants (studies) Follow-up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Placebo	Risk difference with Acyclovir
Duration of symptoms from onset of treatment assessed with: time to resolution	238 (5 RCTs) [1]	⨁◯◯◯ LOW [2,3]	-	The mean duration of symptoms from onset of treatment was **0** days	MD **3.2 days fewer** (4.94 fewer to 1.46 fewer)
Pain	129 (3 RCTs) [4]	⨁◯◯◯ LOW [2,3]	-	The mean pain was **0** days	MD **2.1 days fewer** (2.95 fewer to 1.25 fewer)

EtDs

structured decision-making processes

transparent evidence syntheses that inform about the certainty in that evidence

- evidence profiles, evidence to decision frameworks with judgments

confidence in estimates of intervention effects only “a” part

accept uncertainty and be able to communicate it for better research and implementation

messages