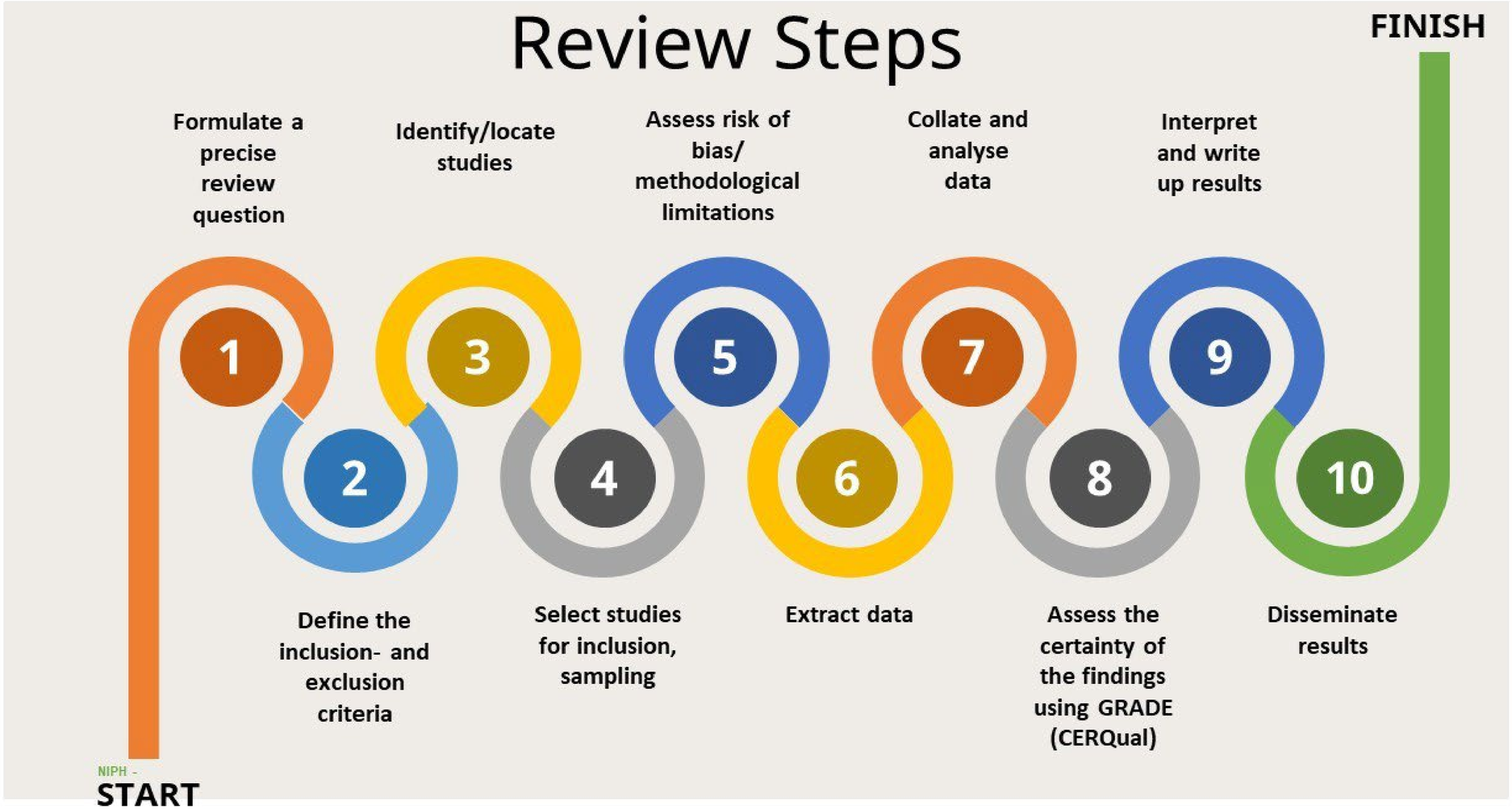
NIPH
Norwegian Institute of Public Health

GRADE

Assessing the confidence in the evidence

Eva Denison, senior researcher


Review Steps
FINISH
Formulate a precise review question
Identify/locate studies
Assess risk of bias/ methodological limitations
Collate and analyse data
Interpret and write up results
1
2
3
4
5
6
7
8
9
10
Define the inclusion- and exclusion criteria
Select studies for inclusion, sampling
Extract data
Assess the certainty of the findings using GRADE (CERQual)
Disseminate results
NIPH -
START

What is GRADE?

https://gdt.gradepro.org/app/handbook/handbook.html#h.svwngs6pm0f2

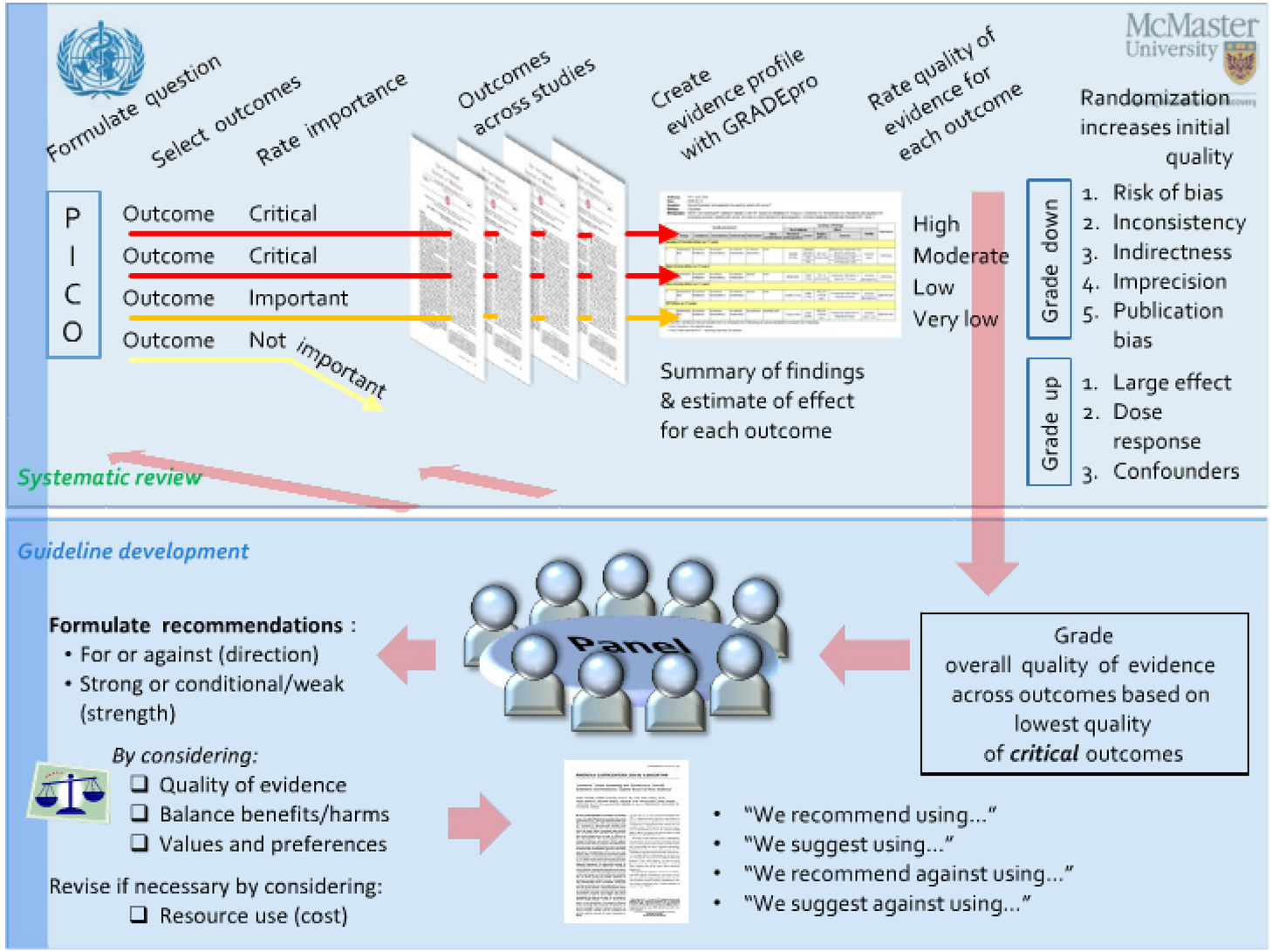
- The GRADE approach is a system for rating the quality of a body of evidence in systematic reviews and other evidence syntheses, such as health technology assessments, and guidelines and grading recommendations in health care.
- GRADE offers a transparent and structured process for developing and presenting evidence summaries and for carrying out the steps involved in developing recommendations.
- It can be used to develop clinical practice guidelines (CPG) and other health care recommendations (e.g. in public health, health policy and systems and coverage decisions).

GRADE in systematic reviews

https://gdt.gradepro.org/app/handbook/handbook.html#h.svwngs6pm0f2

- Systematic reviews should provide a comprehensive summary of the evidence but they should typically not include health care recommendations.
- Use of the GRADE approach by systematic review authors terminates after rating the quality of evidence for outcomes and clearly presenting the results in an evidence table, i.e. an GRADE Evidence Profile or a Summary of Findings table.

GRADE in context



GRADE – SoF

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Face-to-face interventions directed to parents for informing or educating parents about early childhood vaccination, as compared with control						
Patient or population: parents of preschool-aged children or expectant parents **Settings:** clinics, antenatal classes, or the mother's home **Intervention:** face-to-face information or educational interventions **Comparison:** control (no education, other education, or control not described)						
Outcomes	**Illustrative comparative risks* (95% CI)**		**Relative effect (95% CI)**	**No of Participants (studies)**	**Quality of the evidence (GRADE)**	**Comments**
	Assumed (baseline) risk	**Corresponding (intervention) risk**				
	Control (no face-to-face information or education)	**Face-to-face information or education**				
Vaccination status Final time point (3, 6, or 12 months post-intervention)	**55 per 100**[1]	**66 per 100** (57 to 75)	**RR 1.20** (1.04 to 1.37)	3004 (7 studies)	⊕⊕◯◯ **low**[2]	The results for this outcome were variable, so the true result may be substantially higher or lower than this estimate

GRADE

Grading of Recommendations
Assessment,
Development and Evaluation

http://www.gradeworkinggroup.org/

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.

Establish initial level of evidence

1.
Establish initial level of confidence

Study design	*Initial confidence in an estimate of effect*
Randomized trials →	High confidence
Observational studies →	Low confidence

Raising the level of certainty (confidence) of obesrvational studies

2.

Consider lowering or raising level of confidence

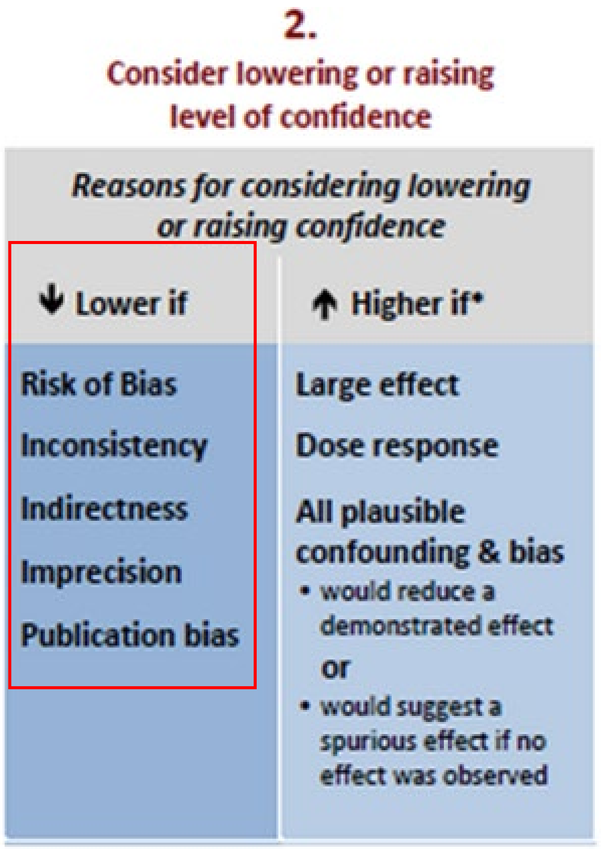
Reasons for considering lowering or raising confidence	
↓ Lower if	↑ Higher if*
Risk of Bias	Large effect
Inconsistency	Dose response
Indirectness	All plausible confounding & bias • would reduce a demonstrated effect or • would suggest a spurious effect if no effect was observed
Imprecision	
Publication bias	

- Magnitude of effect
  - Large – may increase one level
  - Very large – may increase two levels
- Dose response gradient
  - Upgrade one level
- Confounding and bias
  - Upgrade one level
  - See https://gdt.gradepro.org/app/handbook/handbook.html#h.gwd531rylwaj

Consult a statistician!

*upgrading criteria are usually applicable to observational studies only.

Lowering the level of certainty (cinfidence)



- «No serious limitations»
  - Do not downgrade
- «Serious limitations»
  - Downgrade one step
  - Give explanation
- «Very serious limitations»
  - Downgrade two steps
  - Give explanation
- Publication bias
  - Not suspected: do not downgrade
  - Suspected: downgrade one step

Determinants of quality

5 factors that can lower quality

1. limitations of detailed design and execution *(risk of bias criteria)*
2. Inconsistency *(or heterogeneity)*
3. Indirectness *(PICO and applicability)*
4. Imprecision *(number of events and confidence intervals)*
5. Publication bias

Risk of bias (part of internal validity)

Review authors' judgement:

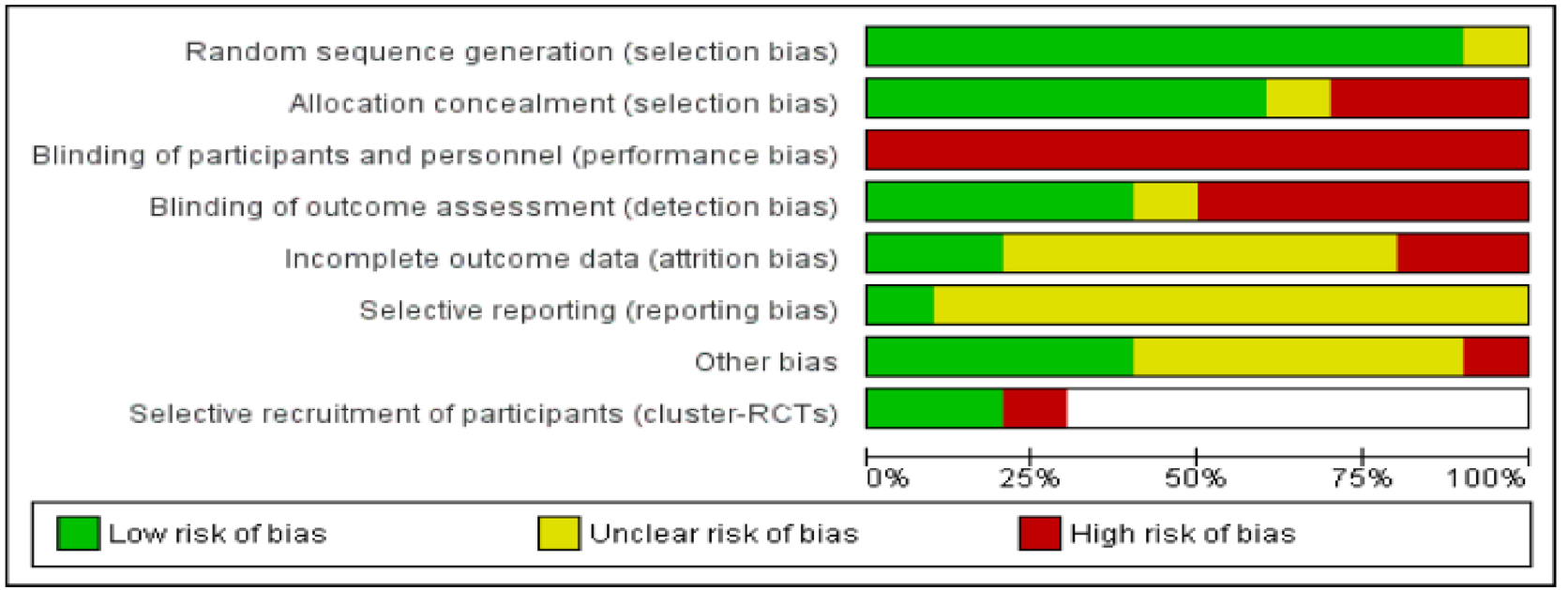
- **Was the allocation sequence adequately generated?**
- **Was allocation adequately concealed?**
- **Was knowledge of the allocated intervention adequately prevented during the study?**
  - Participants
  - Trial personnel
- **Were incomplete outcome data adequately addressed?**
- **Are reports of the study free of suggestion of selective outcome reporting?**
- **Was the study apparently free of other problems that could put it at a high risk of bias?**

17

Domains from the Cochrane Handbook: http://handbook.cochrane.org/

Risk of bias summary for a body of evidence

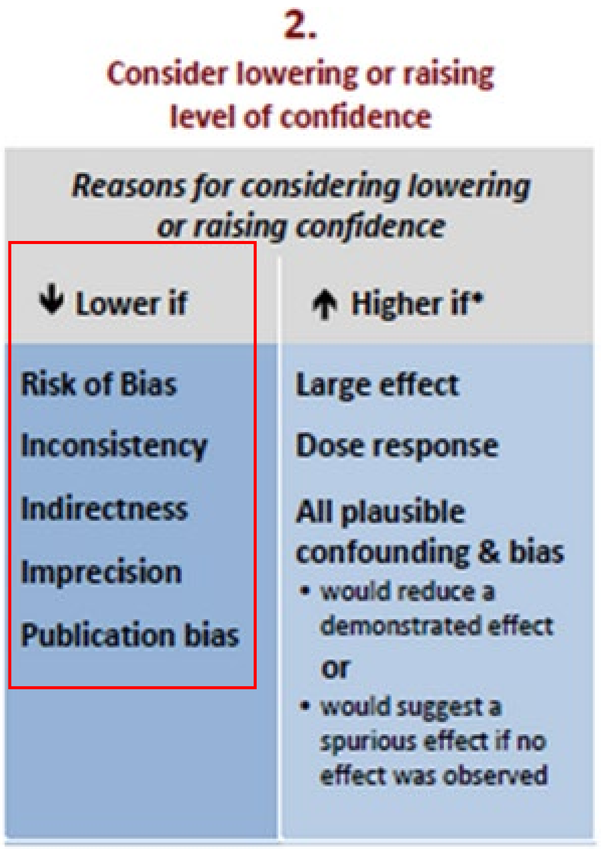
Figure 3. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies



GRADE assesment of **study limitations** across studies

Risk of bias	Across studies	Interpretation	Considerations	GRADE assessment of study limitations
Low risk of bias.	Most information is from studies at low risk of bias.	Plausible bias unlikely to seriously alter the results.	No apparent limitations.	No serious limitations, do not downgrade.
Unclear risk of bias.	Most information is from studies at low or unclear risk of bias.	Plausible bias that raises some doubt about the results.	Potential limitations are unlikely to lower confidence in the estimate of effect.	No serious limitations, do not downgrade.
			Potential limitations are likely to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.
High risk of bias.	The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results.	Plausible bias that seriously weakens confidence in the results.	Crucial limitation for one criterion, or some limitations for multiple criteria, sufficient to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.
			Crucial limitation for one or more criteria sufficient to substantially lower confidence in the estimate of effect.	Very serious limitations, downgrade two levels.

Lowering the level of certainty (cinfidence)

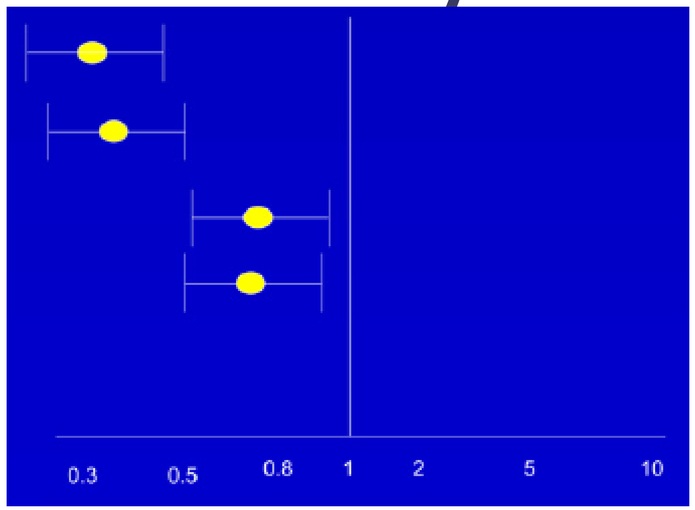


- «No serious limitations»
  - Do not downgrade
- «Serious limitations»
  - Downgrade one step
  - Give explanation
- «Very serious limitations»
  - Downgrade two steps
  - Give explanation
- Publication bias
  - Not suspected: do not downgrade
  - Suspected: downgrade one step

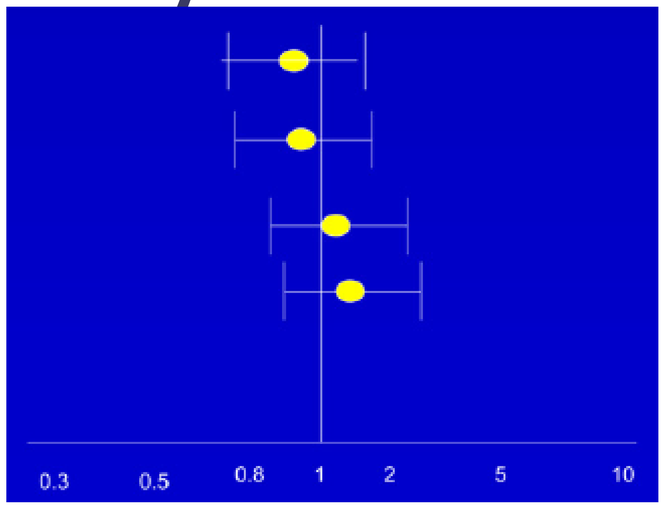
Inconsistency across the results

- Similarity in effect estimates
- Results pointing in the same directions
- The confidence intervals cover the effect estimates

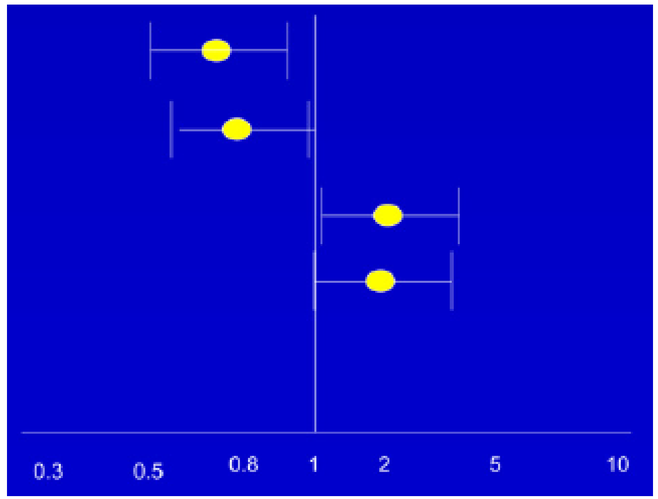
Inconsistency - Heterogeneity



No overlap,
same direction of effect

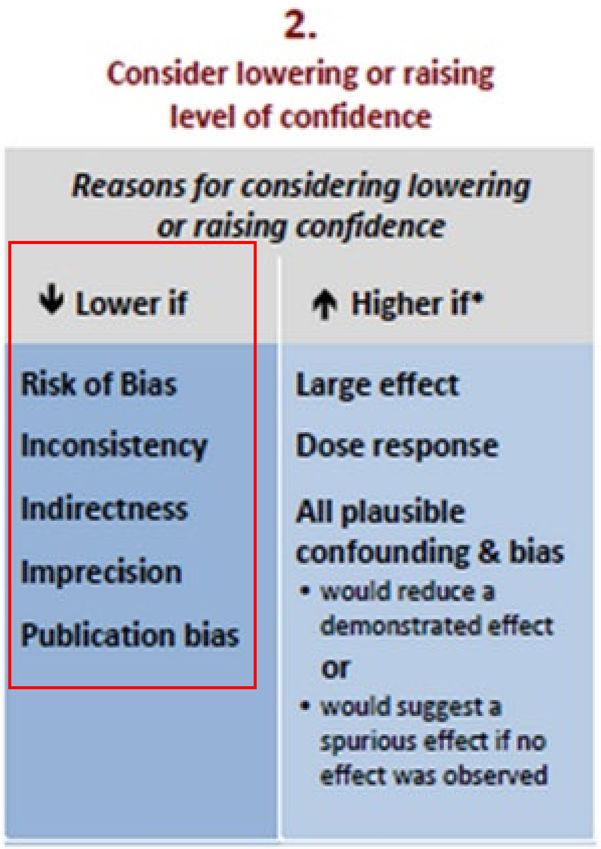


Overlap,
different directions of effect



No overlap,
different directions of effect

Lowering the level of certainty (cinfidence)



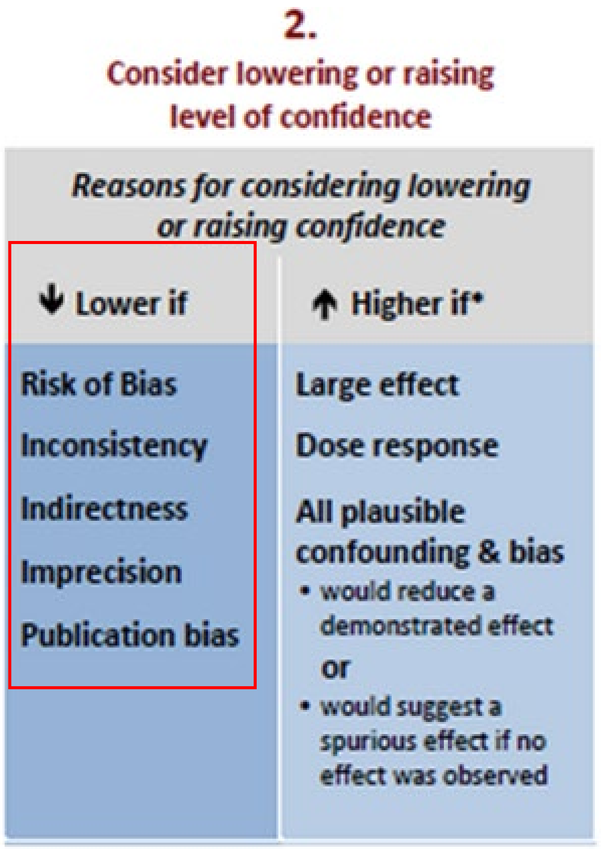
- «No serious limitations»
  - Do not downgrade
- «Serious limitations»
  - Downgrade one step
  - Give explanation
- «Very serious limitations»
  - Downgrade two steps
  - Give explanation
- Publication bias
  - Not suspected: do not downgrade
  - Suspected: downgrade one step

Indirectness of evidence

- Differences in population (applicability)
- Differences in interventions (applicability)
- Differences in outcomes measures (surrogate outcomes)
- Indirect Comparisons

https://gdt.gradepro.org/app/handbook/handbook.html#h.w6r7mtvq3mjz

Lowering the level of certainty (cinfidence)



- «No serious limitations»
  - Do not downgrade
- «Serious limitations»
  - Downgrade one step
  - Give explanation
- «Very serious limitations»
  - Downgrade two steps
  - Give explanation
- Publication bias
  - Not suspected: do not downgrade
  - Suspected: downgrade one step

Imprecision

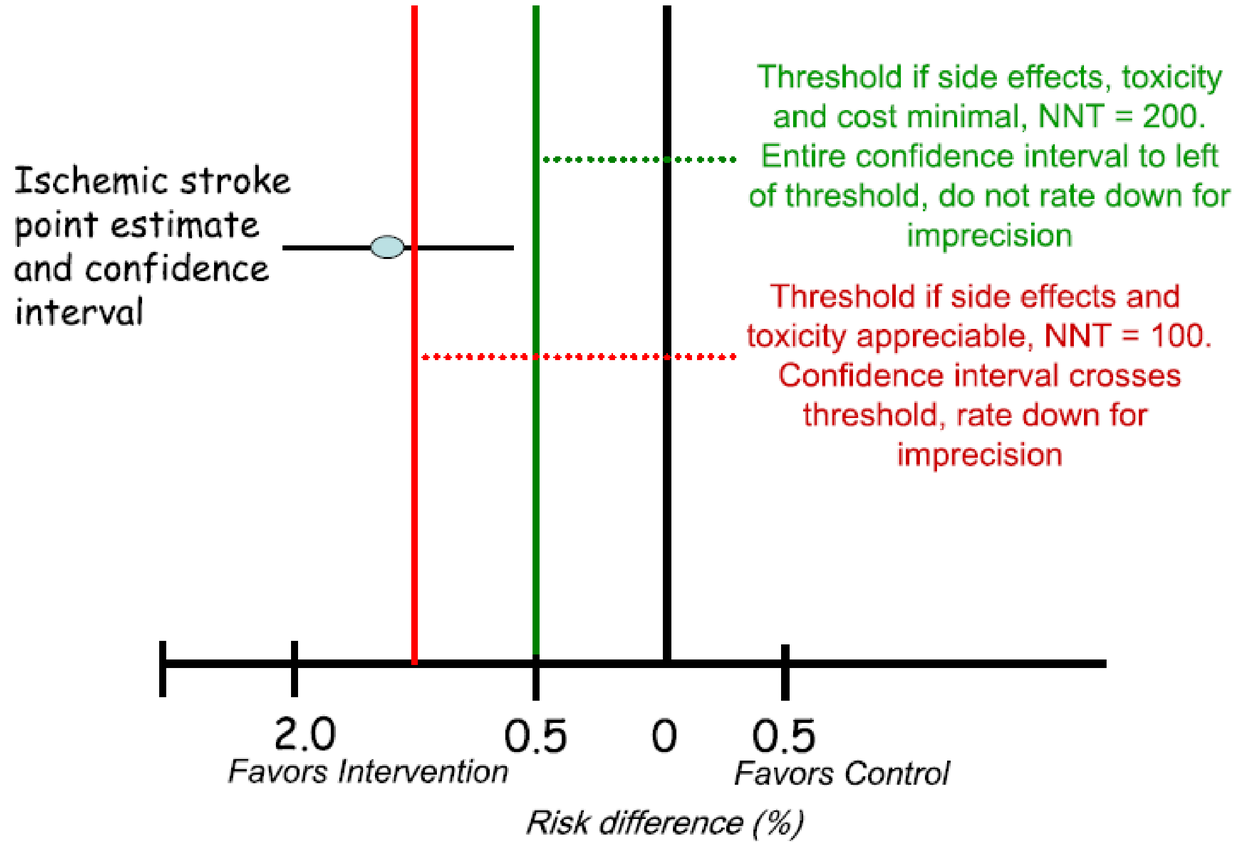


Fig. 1. Rating down for imprecision in guidelines: thresholds are key.

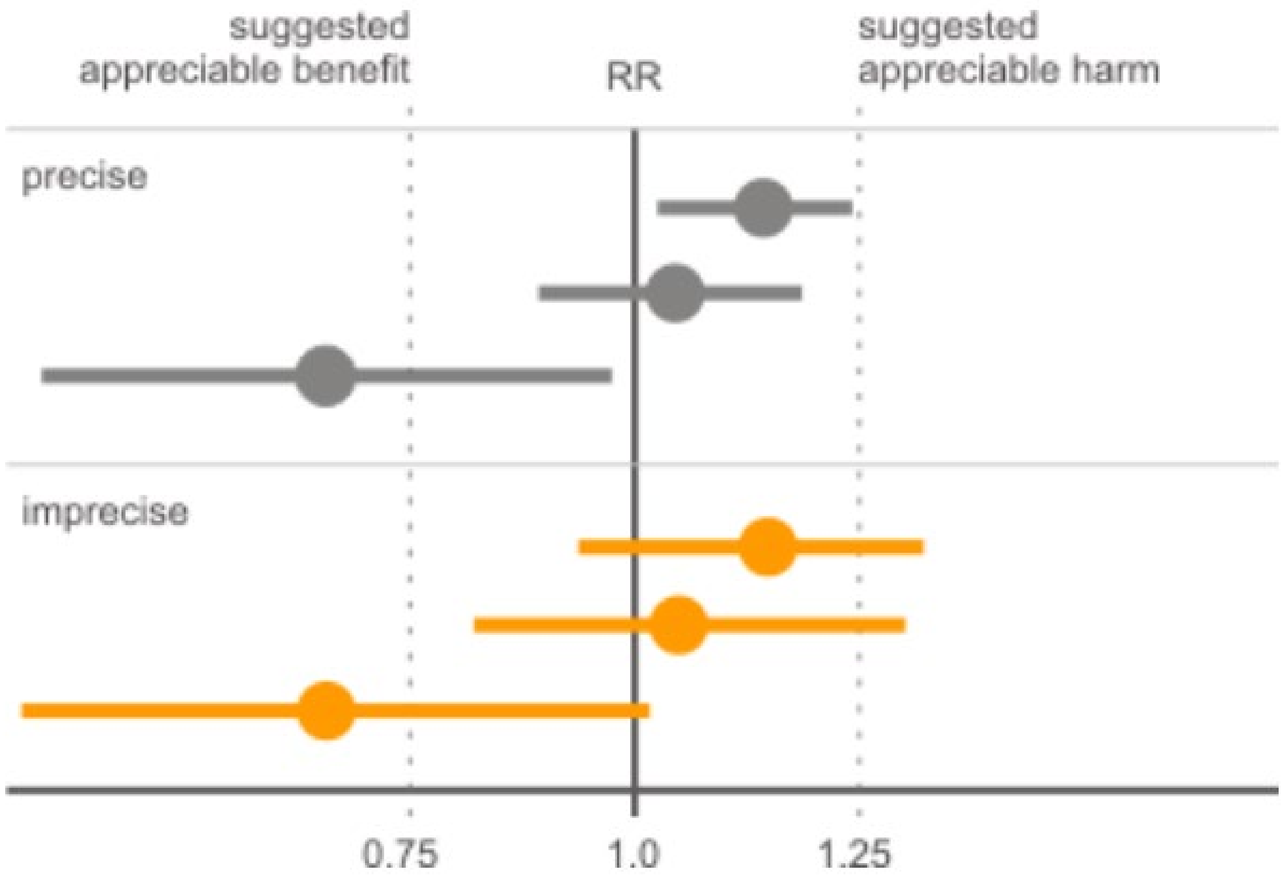
NNT: Number-Needed-to-Treat

Imprecision

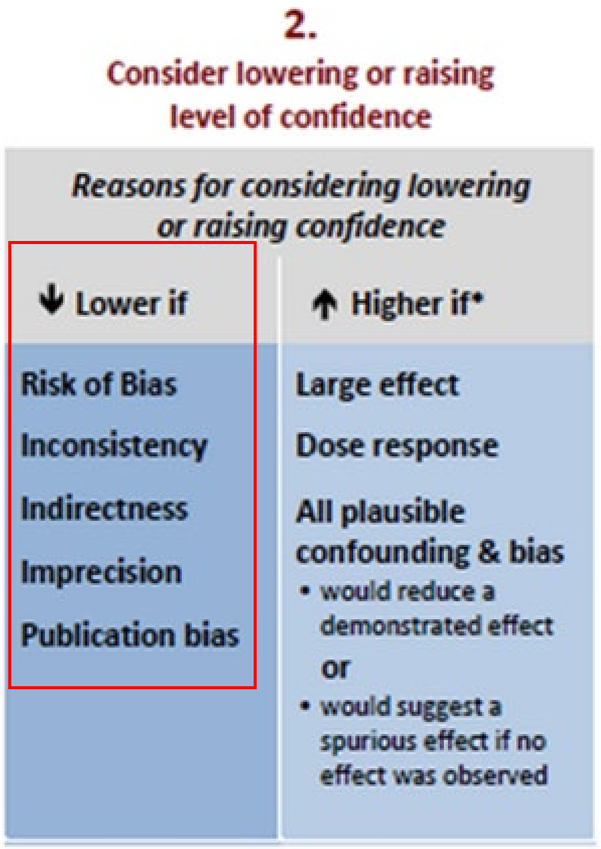
How much data do you have?

- Not a lot of data – evidence
- Broad confidence intervals
- Precision of results

Imprecision



Lowering the level of certainty (cinfidence)



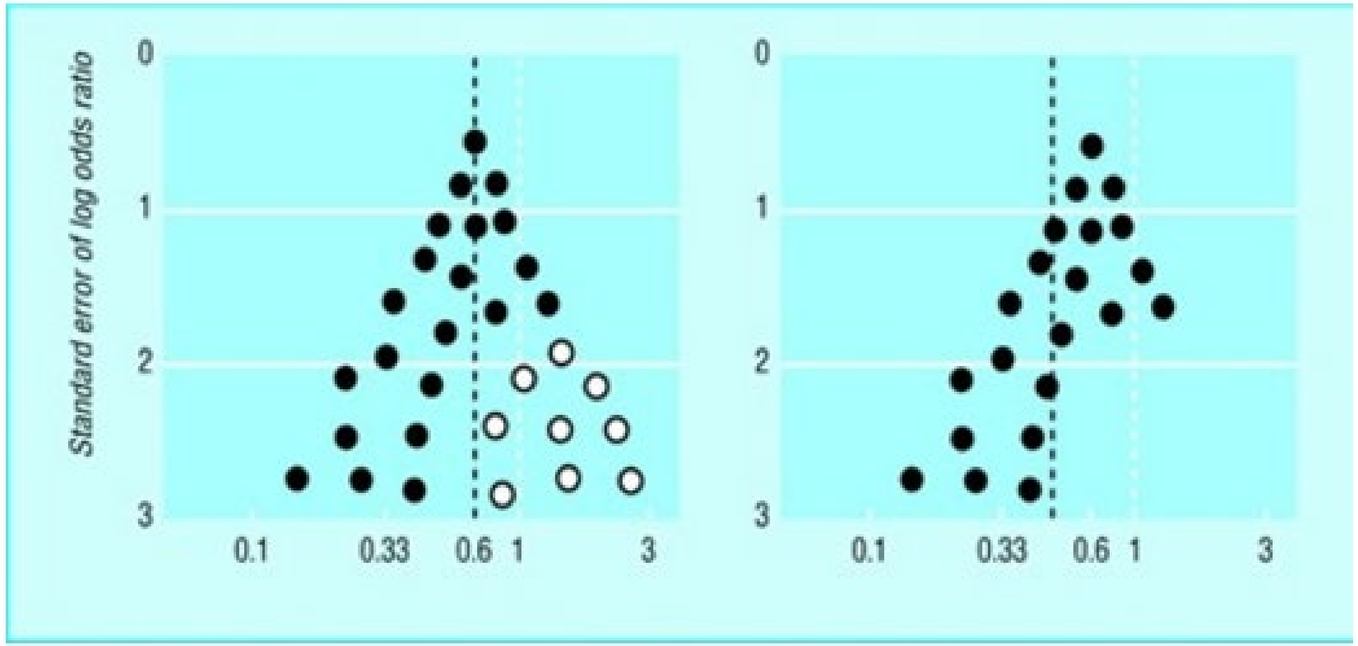
- «No serious limitations»
  - Do not downgrade
- «Serious limitations»
  - Downgrade one step
  - Give explanation
- «Very serious limitations»
  - Downgrade two steps
  - Give explanation
- Publication bias
  - Not suspected: do not downgrade
  - Suspected: downgrade one step

Publication bias

Publication bias is a systematic under-estimation or an over-estimation of the underlying beneficial or harmful effect due to the **selective publication of studies**. Confidence in the combined estimates of effects from a systematic review can be reduced when publication bias is suspected, even when the included studies themselves have a low risk of bias.

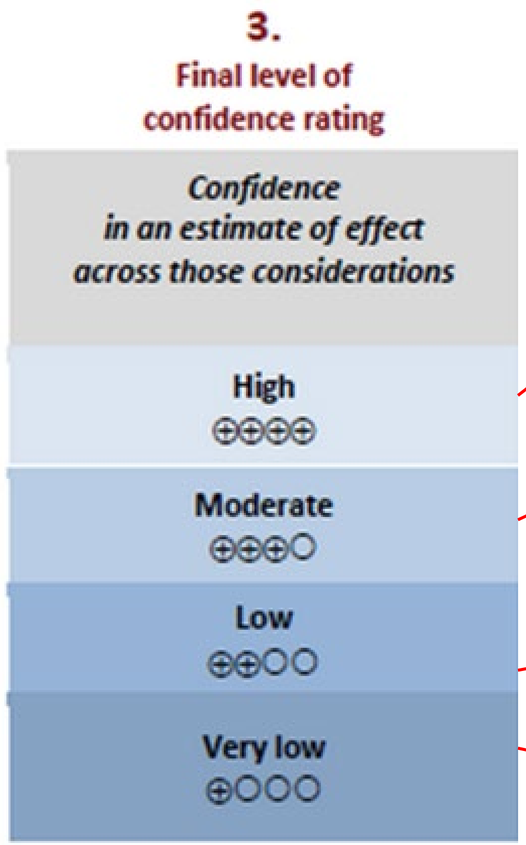
Publication bias

- Funnel plot



No Publication Bias Publication Bias

Let’s look at each step



We are very confident that the true effect lies close to that of the estimate of the effect

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

Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

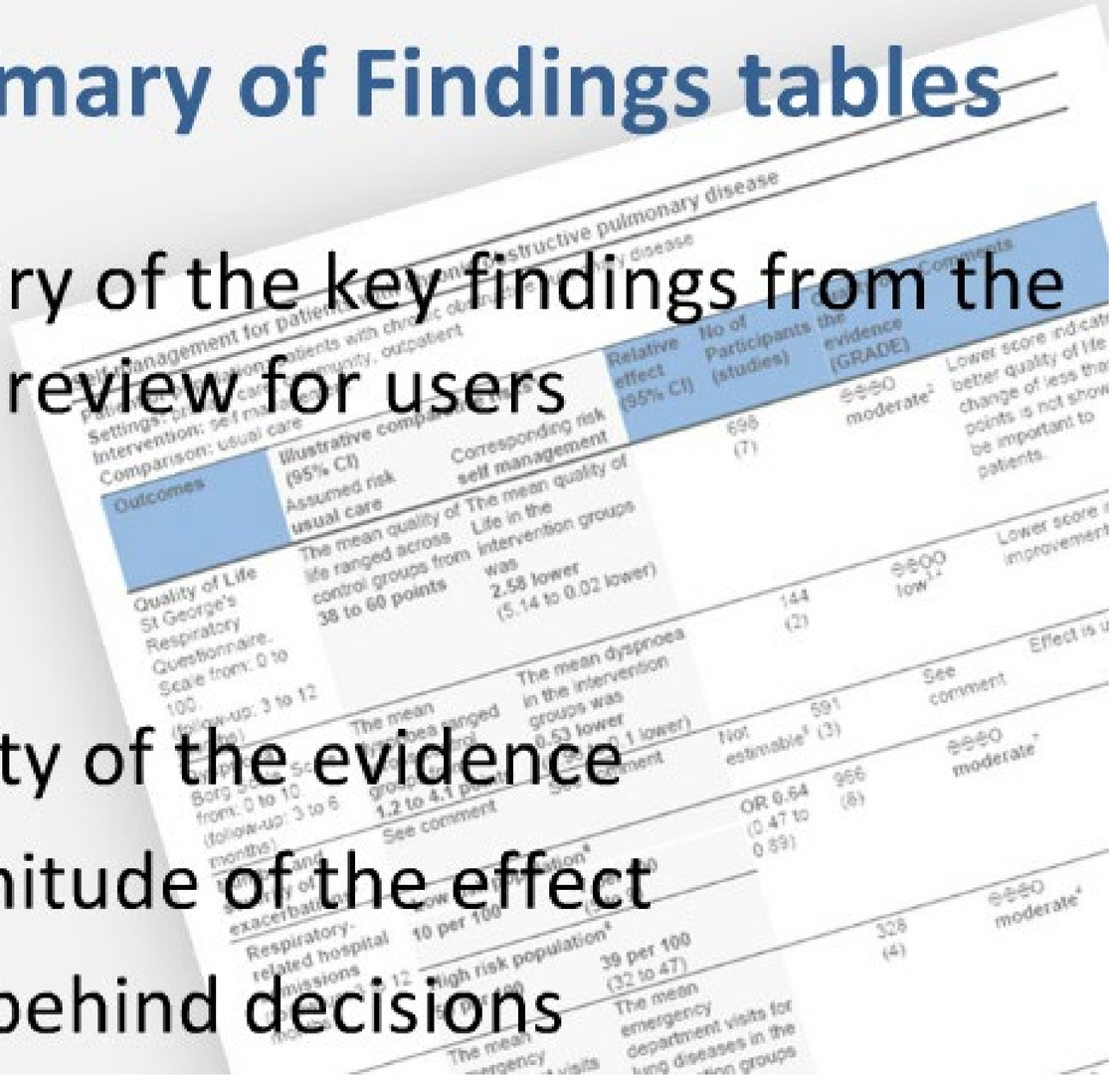
We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

GRADE categories for the quality of a body of evidence

Quality level	Symbol	Definition
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.

The Summary of Findings tables

- Is a summary of the key findings from the systematic review for users
- Presents
  - –the quality of the evidence
  - –the magnitude of the effect
  - –reasons behind decisions



Grading of Recommendations Assessment, Development and Evaluation

http://www.gradeworkinggroup.org/

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.

Exercise for osteoarthritis of the knee (Review)

Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL.

THE COCHRANE COLLABORATION®

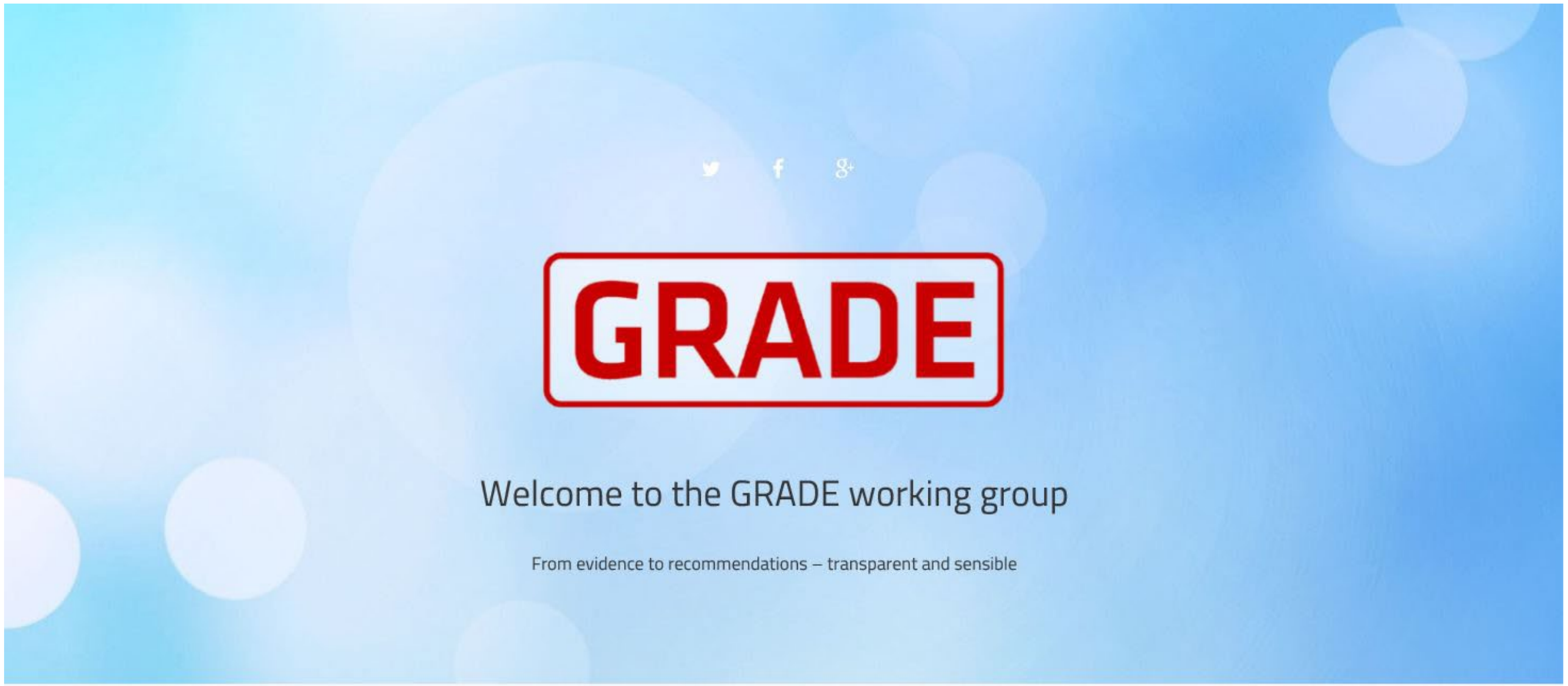
SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Immediate post-treatment effects of exercise for osteoarthritis of the knee

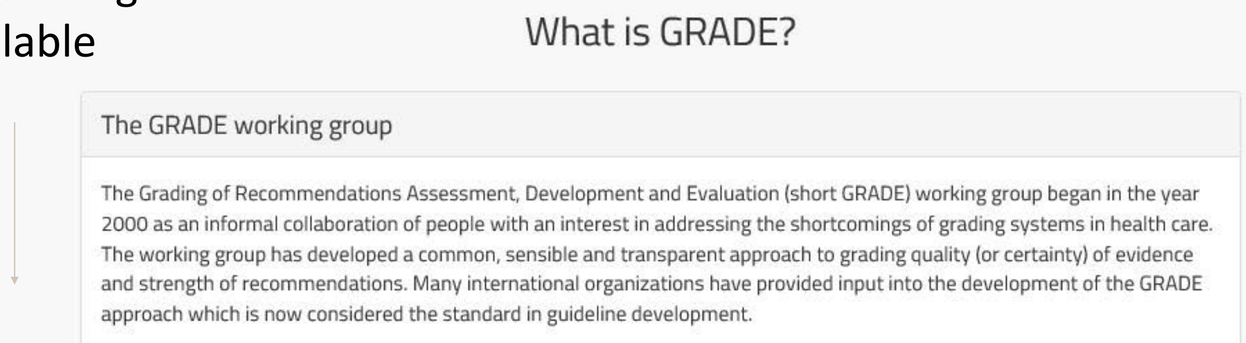
Patient or population: patients with knee OA
Settings: clinic or community
Intervention: land-based exercise
Comparison: no exercise

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	No exercise	Land-based exercise				
Pain Self-report questionnaires. Scale from 0-100 (0 represents no pain)	Mean pain in the control groups was **44 points**	Mean pain in intervention groups was **0.49 standard deviations lower** (0.39-0.59 lower) This translates to an absolute mean reduction of 12 (10-15) points compared with control group on a 0-100 scale[a]		3537 (44 studies)	⊕⊕⊕⊕ **High**	SMD -0.49 (-0.39 to -0.59) Absolute reduction in pain 12% (10%-15%); relative change 27% (21%-32%) [a] NNTB 4 (3-5)[b]
Physical function Self-report questionnaire. Scale from 0-100 (0 represents no physical disability)	Mean physical function in control groups was **38 points**	Mean physical function in intervention groups was **0.52 standard deviations lower** (0.39-0.64 lower) This translates to an absolute mean improvement of 10 (8-13) points on a 0-100 scale[c]		3913 (44 studies)	⊕⊕⊕◯ **Moderate**[d]	SMD -0.52 (-0.39 to -0.64) Absolute improvement 10% (8%-13%); relative improvement 26% (20%-32%)[c] NNTB 4 (3-5)[b]

http://www.gradeworkinggroup.org/



Many online learning resources available on web-site



Learning material - examples

- Cochrane interactive learning http://training.cochrane.org/interactivelearning
- GRADE https://cebgrade.mcmaster.ca/
- Equatornetwork http://www.equator-network.org/
- Testing treatments http://www.testingtreatments.org/