Determinants of quality: Factors that lower or increase the quality of evidence

GRADE Workshop
CBO, NHG and Dutch Cochrane Centre
CBO, April 17th, 2013

Outline

- The GRADE approach: step by step
- Factors that lower the quality of evidence
- Factors that increase the quality of evidence
- Examples + discussion
- For each factor: what information do you need and where do you find it

2

Step 1a
Specify research question

- Patients
- Intervention
- Comparison
- Outcomes

3

Step 1b
Specify outcomes for research question

		Importance of end points
Mortality Myocardial infarction Fractures	9 8 7	Critical for decision making
Pain	6 5 4	Important but not critical for decision making
Flatulence	3 2 1	Not important for decision making – of lower importance for patients

4

Step 2
Summarize all relevant evidence

- By outcome, across studies

- Use available systematic review(s)
  - Asses quality of the review (AMSTAR)
  - Needed: information on individual studies (e.g. risk of bias assessment)

Or

- Develop systematic review
  - Cochrane Handbook for Systematic Reviews of Interventions

5

Summary of Findings Table (Cochrane Review)

Patients or population: Anyone taking a long flight (lasting more than 6 hours)
Settings: International air travel
Intervention: Compression stockings[1]
Comparison: Without stockings

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Without stockings	**With stockings (95% CI)**				
Symptomatic deep vein thrombosis (DVT)	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed symptomatic DVT in these studies.
Symptomatic deep vein thrombosis –surrogate, symptomless deep vein thrombosis	**Low risk population** [2]		**RR 0.10** (0.04 to 0.26)	2637 (9 studies)	⊕⊕⊕◯ **Moderate**[3]	All of the events in the "illustrative comparative risks" column were asymptomatic and thus of questionable importance
	15 per 1000	**1.5 per 1000** (0 to 3)				
	High risk population [2]					
	25 per 1000	**2.5 per 1000** (1 to 8)				
Superficial vein thrombosis	**13 per 1000**	**6 per 1000** (2 to 15)	**RR 0.45** (0.18 to 1.13)	1804 (8 studies)	⊕⊕⊕◯ **Moderate**[4]	
Oedema Post-flight values measured on a scale from 0, no oedema, to 10, maximum oedema.	The mean oedema score ranged across control groups from **6 to 9**.	The mean oedema score in the intervention groups was on average **4.7 lower** (95% CI −4.9 to −4.5).		1246 (6 studies)	⊕⊕◯◯ **Low**[5]	

6

Step 3
Rate the evidence

- Across studies, by outcome
- RCTs: high quality evidence
- Observational studies: low quality evidence
- 5 factors can lower the quality
- 3 factors can increase the quality
- Use footnotes to explain your decisions on down/upgrading

7

Quality of evidence: 4 levels

- **High** ⊕⊕⊕⊕
  Very confident that the true effect lies close to the estimate of the effect.
- **Moderate** ⊕⊕⊕◯ Moderately confident in the effect estimate: true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- **Low** ⊕⊕◯◯ Limited confidence in the effect estimate: true effect may be substantially different from the estimate of effect.
- **Very low** ⊕◯◯◯ Very little confidence in the effect estimate: true effect is likely to be substantially different from the estimate of effect.

8

Levels of quality of evidence

Underlying methodology	Quality rating
RCTs Double-upgraded observational studies	High
Downgraded RCTs Upgraded observational studies	Moderate
Double-downgraded RCTs Observational studies	Low
Triple-downgraded RCTs Downgraded observational studies Case series/reports	Very low

9

Factors that can lower the quality of evidence

1. Study limitations
2. Inconsistency of results
3. Indirectness of evidence
4. Imprecision
5. Publication bias

If factor is present:
downgrade level of evidence by
1 ('serious') or 2 levels ('very serious')

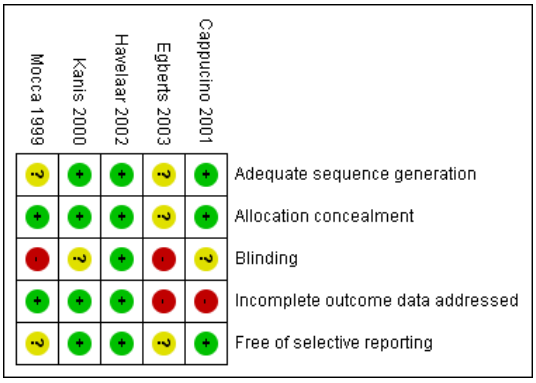
10

Downgrading: Study limitations (RCTs)

- Inadequate randomization
- Lack of allocation concealment
- Lack of blinding (patient / clinician / outcome assessor)
- Loss to follow-up / selective follow-up
- Intention to treat principle violated
- Failure to report outcomes

11

Risk of bias summary



Risk of bias	Across studies	Interpretation	Considerations	GRADE assessment
Low	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	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	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

13

Another example (fictitious)

	Concealment of allocation	Blinding of outcome assessment	Description of dropouts (%)	Analysis by ITT	Selective reporting of events
1	Adequate	Adequate	Adequate (4)	**No**	No
2	Adequate	**Unclear (NR)**	Adequate **(50)**	Yes	**Yes**
3	Adequate	Adequate	Adequate (12)	Yes	No

14

Downgrading: inconsistency

Explanations for inconsistency:

- Clinical heterogeneity
  - Populations
    different effect in sicker populations
  - Interventions
    larger effect with higher doses
  - Outcomes
    diminishing effect with time
- Methodological heterogeneity (differences study design)

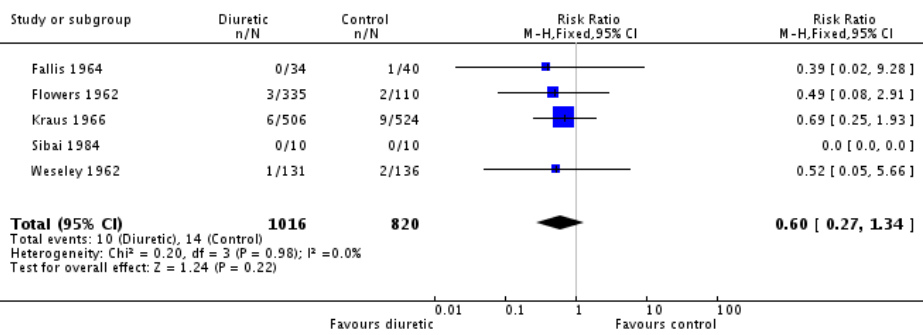
15

Inconsistency

- Unexplained heterogeneity
- Results of meta-analysis (forest plot)
  - Eye-ball test: overlap of confidence intervals?
  - I^2 statistic (>60% = substantial)

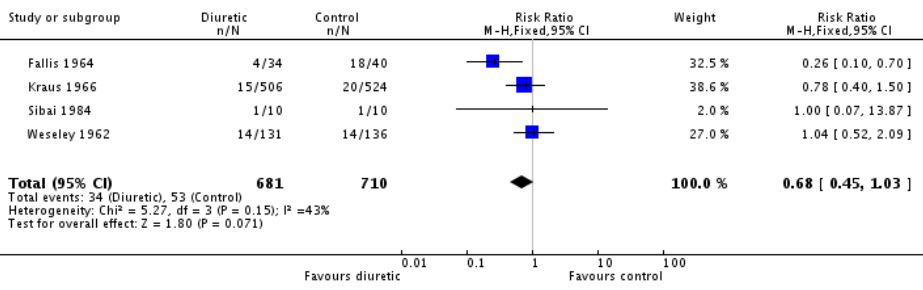
16

Review: Diuretics for preventing pre-eclampsia
Comparison: 1 Diuretic versus placebo or no treatment
Outcome: 11 Stillbirth

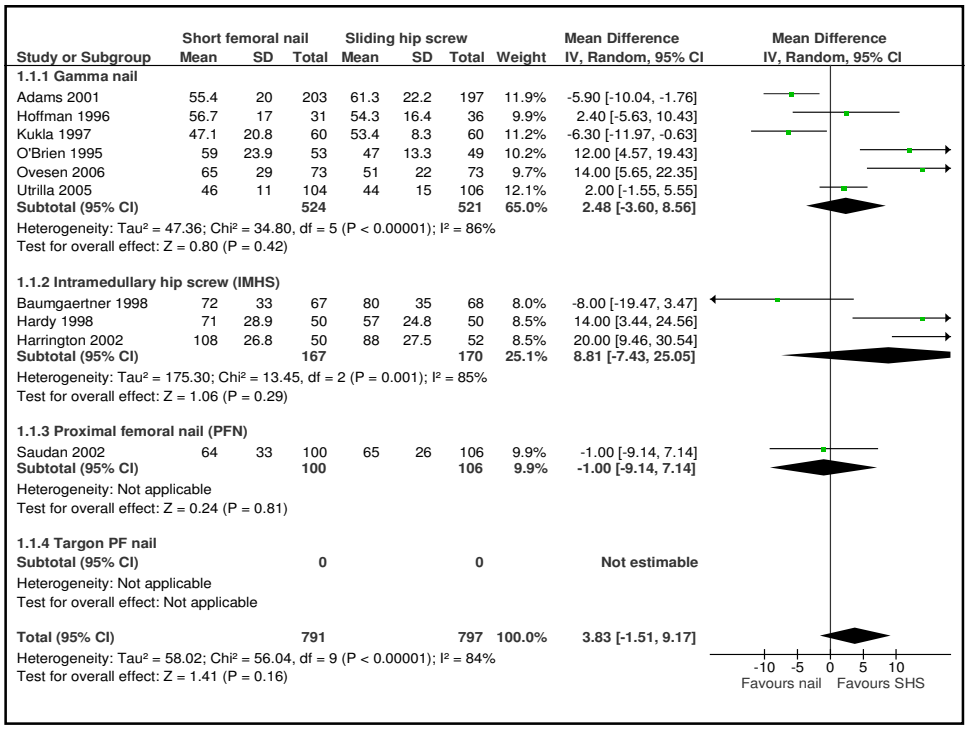


17 17

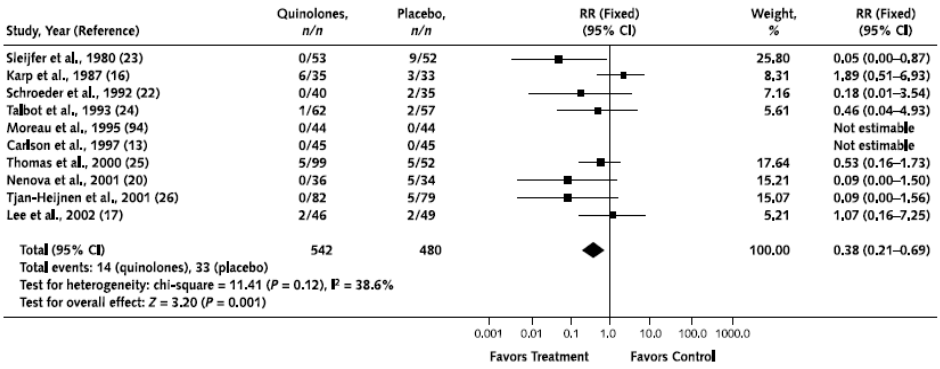
Review: Diuretics for preventing pre-eclampsia
Comparison: 1 Diuretic versus placebo or no treatment
Outcome: 1 Pre-eclampsia



18 18



Fluoroquinolone prophylaxis in neutropenia: infection-related mortality



20

Downgrading: Indirectness of evidence

Evidence comes from different research question

- Indirect comparison: drug A – drug B
  A – placebo and B – placebo
- Population: oseltamivir prophylaxis for avian flu
  seasonal influenza
- Comparator: new drug – flexible doses of haloperidol
  fixed doses of haloperidol
- Outcome (next slide)

21

Indirectness: surrogate outcomes

Condition	Patient important outcomes	Surrogate outcome
Diabetes mellitus	Symptoms, hospital admission, complications	Blood glucose, HbA1c
Dementia	Patient function, behaviour, caregiver burden	
Osteoporosis	Fractures	Bone density
Chronic respiratory disease	Quality of life, exacerbations, mortality	Pulmonary function, exercise capacity
Cardiovascular disease/risk	Vascular events, mortality	Serum lipids

22

Downgrading: Imprecision

- Basic idea: few patients and few events => wide confidence intervals => imprecise results
- SRs versus Guidelines
- Optimal information size (OIS): the number of patients required for adequate statistical power

23

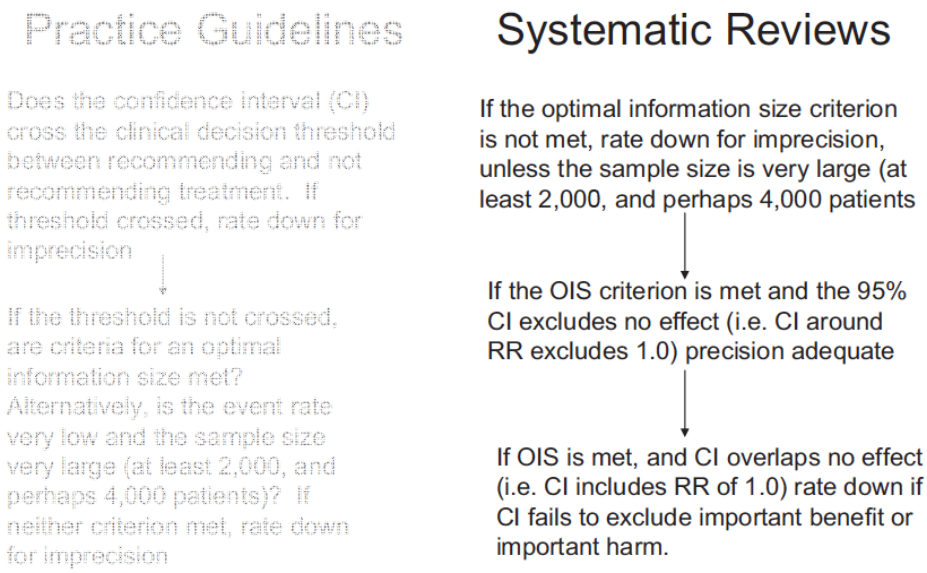


Fig. 3. Deciding whether to rate down for imprecision in guidelines and systematic reviews of binary variables.

24

Imprecision

For dichotomous outcomes consider downgrading if:

1. Sample size lower than optimal information size and/or number of events <300
2. OIS is met, but 95% confidence interval includes both no effect and appreciable benefit or harm*

* Relative Risk Reduction (RRR) >0.25% (RR<0.75 of >1.25)

25

Table 1: Optimal information size implications from Figure 5

Total Number of Events	Relative Risk Reduction	Implications for meeting OIS threshold
100 or less	≤ 30%	Will almost never meet threshold whatever control event rate
200	30%	Will meet threshold for control event rates for ~ 25% or greater
200	25%	Will meet threshold for control event rates for ~ 50% or greater
200	20%	Will meet threshold only for control event rates for ~ 80% or greater
300	≥ 30%	Will meet threshold
300	25%	Will meet threshold for control event rates ~ 25% or greater
300	20%	Will meet threshold for control event rates ~ 60% or greater
400 or more	≥ 25%	Will meet threshold for any control event rate
400 or more	20%	Will meet threshold for control event rates of ~ 40% or greater

Imprecision

For continuous outcomes consider downgrading if:

1. Sample size lower than optimal information size and/or population size <400
2. OIS is met, but 95% confidence interval includes no effect and one of the confidence limits crosses the minimal important difference (benefit or harm)*

   If SMD or effect size (ES) is calculated in the meta-analysis, consider downgrading if the upper or lower confidence limit crosses an ES of 0.5 in either direction.

* If the MID is not known, you may calculate an MID by using the rule of thumb that MID is typically 0.5 standard deviations (Norman et al., 2003).

27

Example MID

A meta-analysis found a pooled mean difference of -2.58 (-5.14 to -0.02) on a scale of 0 to 100
The pooled SD of the control groups was 17.2. The MID could be calculated as 0.5 X 17.2 = 8.6

The confidence intervals around the pooled mean difference do not include 8.6 (despite the fact that there is an effect).

Resulting point estimate is precise and the evidence is not downgraded.

28

Practice Guidelines

Does the confidence interval (CI) cross the clinical decision threshold between recommending and not recommending treatment. If threshold crossed, rate down for imprecision

↓

If the threshold is not crossed, are criteria for an optimal information size met? Alternatively, is the event rate very low and the sample size very large (at least 2,000, and perhaps 4,000 patients)? If neither criterion met, rate down for imprecision

Systematic Reviews

If the optimal information size criterion is not met, rate down for imprecision, unless the sample size is very large (at least 2,000, and perhaps 4,000 patients

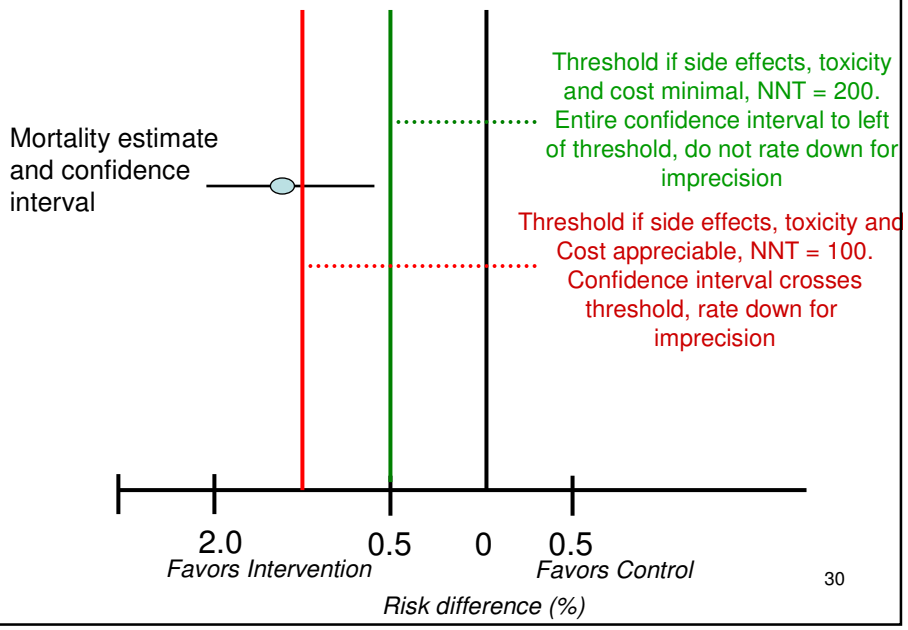
If the OIS criterion is met and the 95% CI excludes no effect (i.e. CI around RR excludes 1.0) precision adequate

If OIS is met, and CI overlaps no effect (i.e. CI includes RR of 1.0) rate down if CI fails to exclude important benefit or important harm.

Fig. 3. Deciding whether to rate down for imprecision in guidelines and systematic reviews of binary variables.

29

Figure 1, Rating down for imprecision in guidelines: Thresholds are key



30

Downgrading: Publication bias

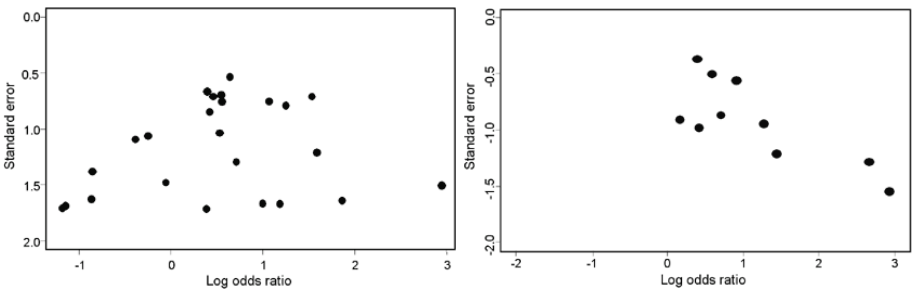
- Evidence is incomplete
- Systematic under- or overestimate of the effect due to selective publication of studies
- Investigators fail to report studies (typically those that show no effect)
- Suspicion: evidence is limited to small number of trials, all funded by industry

31

Publication bias

How to identify?

- Funnel plot



32

Publication bias

How to identify?

- Funnel plot
- Trial registers
- Abstract books
- Experts

33

Upgrading: magnitude of effect

Large effect (in the absence of plausible confounders in studies with no major threats to validity)

- Large: RR >2 or RR <0.5 (upgrade 1 level)
- Very large: RR >5 or RR <0.2 (upgrade 2 levels)

Example

Bicycle helmets to prevent head injuries in cyclists: OR 0.31, 95% CI 0.26 to 0.37 (Thompson 2000)

34

Upgrading: dose-response gradient

Example

Rofecoxib and cardiovascular events (McGettigan 2006):

RR 1.33 (95% CI 1.00 to 1.79) for doses ≤ 25 mg/day

RR 2.19 (95% CI 1.64 to 2.91) for doses >25 mg/day

35

Upgrading: residual bias

All plausible biases may be working to underestimate an apparent intervention effect

E.g. Sicker patients overrepresented in intervention group, but the intervention seems more effective than control group

Observational studies have failed to demonstrate an association, but all plausible biases would have increased an intervention effect

E.g. Exploration of apparent harmful effects

Only valid studies can be upgraded!

36

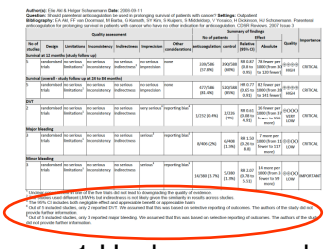
Do's and don'ts

Do

- First, consider each factor separately
- Consider all factors together to make a balanced judgment (quality of evidence is a continuum)

Do not

- Double downgrade for the same problem (e.g. inconsistency and imprecision can be related)



Footnotes: examples

1 Unclear concealment in one of five trials did not lead to downgrading the quality of evidence

2 The studies used different LMWHs but indirectness is not likely given the similarity in results across studies

3 The 95% CI includes both negligible effect and appreciable benefit or appreciable harm

4 Out of 5 included, only 2 reported DVT. We assumed that this was based in selective reporting of outcomes. The authors of the study did not provide further information.

Overall quality of evidence

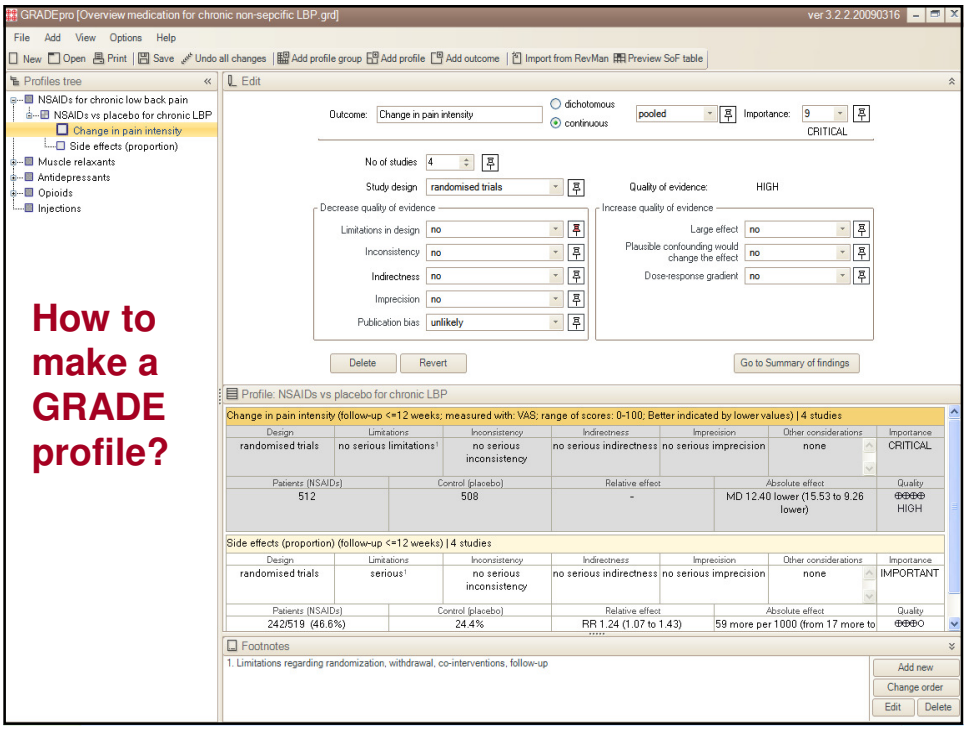
- Consider the outcomes that are critical for decision making
- Overall quality of evidence = lowest rating of the critical outcomes
- Some exceptions…

39

Using existing SRs: what information do you need?

- Review question = your research question
- Clear description of the in/exclusion criteria
- Characteristics of the included studies
- Detailed risk of bias assessment + reporting
- Estimate of effect:
  - pooled result: results of the meta-analysis, forest plot with I^2, subgroup analyses
  - no meta-analysis: effect estimates of the individual studies
- Assessment/discussion on publication bias

40



How to make a GRADE profile?