GRADE- an introduction

October 2012

Esther van Zuuren

Dermatologist

Leiden University Medical Centre

The Netherlands

GRADE

GRADE =

Grading of Recommendations Assessment, Development and Evaluation

Content

- Background and rationale for revisiting guideline methodology
- GRADE approach
  - Quality of the Evidence
  - Strength of Recommendations
- GRADE working group

Why grading system?

- Medical evidence, or the recommendations that are based on the evidence, can be of different quality
- People draw conclusions about the quality of evidence and strength of recommendations
- Insufficiënt attention to quality of evidence risks inappropriate guidelines
- Different grading systems

Different grading systems

Evidence	Recommendations	Organisation
1^{++}- 4	A-D	SIGN
A-C	I-V	American Heart Association® Learn and Live
1a - 5	A-D	Centre for evidence based medicine
1-3	A-C	AAFP

Hierarchy of evidence

The GRADE approach

- Method of grading quality of evidence and strength of recommendations => evidence-based recommendations
- Developed by the GRADE working group (www.gradeworkinggroup.org)
- Methodologists, guideline developers, clinicians
- To overcome shortcomings of present grading systems

The GRADE approach

systematic reviews authors and guideline developers:

1. Specific question
2. Identify all patient important outcomes
3. Judge relative importance of outcomes
4. Summarize all relevant evidence
5. Grade quality of evidence for each outcome

The quality of the evidence reflects the extent of our confidence that the estimates of the effect are correct

The GRADE approach

guideline developers:

6. Overall quality of evidence across outcomes
7. Include judgments on values and preferences
8. Balance of desirable and undesirable effects
9. Balance of net benefits and costs
10. Grade strength of the recommendation
11. Formulate recommendation
12. Implement and evaluate

Grade the evidence

High

⨁⨁⨁⨁: true effect lies close to the estimate of the effect

Moderate

⨁⨁⨁O: true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low

⨁⨁OO: true effect may be substantially different from the estimate of effect

Very low

⨁OOO: true effect is likely to be substantially different from the estimate of effect

GRADEing the evidence

- RCTs: start high
- Observational studies: start low
- 5 factors can reduce the quality
- 3 factors can increase the quality
- Across studies, by outcome
- Explain your decisions on down/upgrading

Factors that lower the quality

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’)

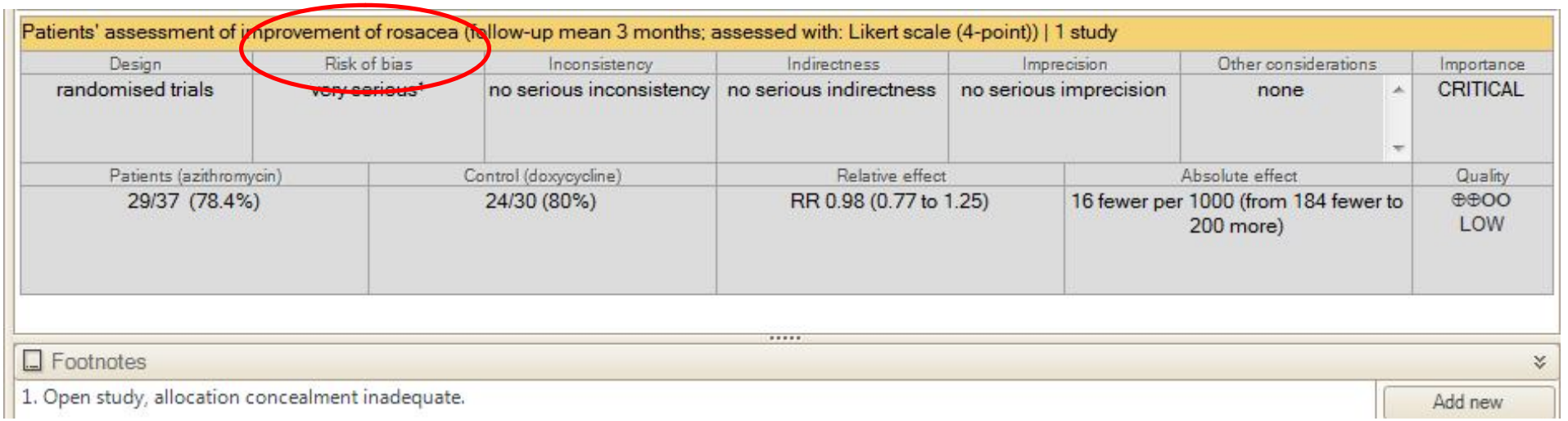
Study limitations (RCTs)

Limitations in design and execution;

Risk of Bias

- Inadequate randomisation
- Lack of allocation concealment
- Lack of blinding
- Large losses to follow-up
- Failure to adhere to an ITT analysis
- Failure to report outcomes

Example of risk of bias



Patients' assessment of improvement of rosacea (follow-up mean 3 months; assessed with: Likert scale (4-point)) | 1 study

Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Importance
randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	CRITICAL

Patients (azithromycin)	Control (doxycycline)	Relative effect	Absolute effect	Quality
29/37 (78.4%)	24/30 (80%)	RR 0.98 (0.77 to 1.25)	16 fewer per 1000 (from 184 fewer to 200 more)	⊕⊕OO LOW

Footnotes

1. Open study, allocation concealment inadequate.

Add new

Down graded 2 levels

1) Allocation concealment inadequate: the physicians had access to the computer-generated randomisation list.
2) Not blinded.

Inconsistency of results

Unexplained heterogeneity of results across studies

1) Clinical heterogeneity may arise from differences in:

- Populations

  different effect in sicker populations

- Interventions

  larger effect with higher doses

- Outcomes

  diminishing effect with time

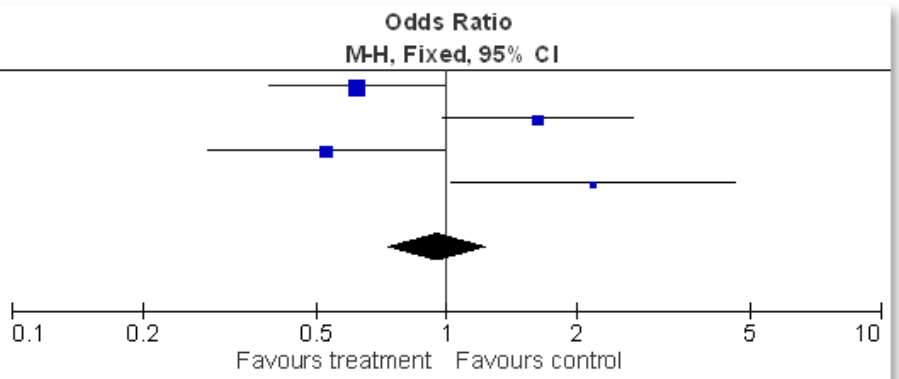
2) Methodological heterogeneity (differences in study design)

Example of inconsistency

Study or Subgroup	Treatment Events	Treatment Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Blake 1990	36	234	58	256	43.7%	0.62 [0.39, 0.98]
Coultas 2005a	60	149	37	126	22.3%	1.62 [0.98, 2.68]
Coultas 2005b	19	151	30	140	25.4%	0.53 [0.28, 0.99]
Rea 2004	28	52	22	63	8.6%	2.17 [1.02, 4.61]
Total (95% CI)		**586**		**585**	**100.0%**	**0.95 [0.73, 1.25]**
Total events	143		147			

Heterogeneity: Chi² = 15.63, df = 3 (P = 0.001); I² = 81%

Test for overall effect: Z = 0.34 (P = 0.73)



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: diabetic complications

  development of biochemical diabetes

Imprecision

- Small sample size
- Small number of events
- Wide confidence intervals
- Uncertainty about the magnitude of effect

Publication bias

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 showing benefit and funded by industry
- Check: funnel plot, trial registers books

Factors that can increase quality

1. Large magnitude of effect
   - Large: RR >2 or RR <0.5 (1 level)
   - Very large: RR >5 or RR <0.2 (2 levels)

2. Plausible biases underestimate the true effect or increase the effect if no effect was observed

3. Dose-gradient response

Summary of Findings table

Table 4 Summary of Findings of minoxidil versus placebo[21-26,28]

Minoxidil compared to Placebo for Female Pattern Hair Loss

Patient or population: Patients with Female Pattern Hair Loss
Settings: Multi centre hospital setting
Intervention: Minoxidil
Comparison: Placebo

Outcomes	Illustrative comparative risks* (95% CI) Assumed risk Placebo	Corresponding risk Minoxidil	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
The proportion of participants with self-rated clinically significant hair regrowth at the end of the study 3 and 5 point scales [1] Follow-up: 24-32 weeks	**Study population**		**RR 1.86** (1.42 to 2.43)	964 (4 studies)	⊕⊕⊖⊖ **low**[2,3]	
	134 per 1000	**250 per 1000** (191 to 327)				
	Moderate					
Change in 'Quality of life' using any validated and recognized generic or disease-specific instrument[4] - not reported	See comment	See comment	Not estimable[4]	-	See comment	
Adverse effects: safety and tolerability and any reported adverse events 1% minoxidil Laboratory values, blood pressure, participant-reported a/e at recall Follow-up: 24 weeks	**Study population**		**RR 1.12** (0.61 to 2.06)	280 (1 study)	⊕⊕⊕⊖ **moderate**[5]	
	121 per 1000	**136 per 1000** (74 to 250)				
	Moderate					
Adverse effects: safety and tolerability and any reported adverse events 2% minoxidil Structured interview, physical examination and laboratory investigations Follow-up: 32-48 weeks	**Study population**		**RR 1.4** (0.6 to 3.27)	604 (3 studies)	⊕⊕⊕⊖ **moderate**[5]	
	27 per 1000	**38 per 1000** (16 to 88)				
	Moderate					
Adverse effects: safety and tolerability and any reported adverse events 5% minoxidil Structured interview, physical examination and laboratory investigations Follow-up: 48 weeks	**Study population**		**RR 3.55** (1.1 to 11.47)	227 (1 study)	⊕⊕⊕⊖ **moderate**[5]	
	41 per 1000	**144 per 1000** (45 to 465)				
	Moderate					
Change in total hair count from baseline to the end of the study Hair counts in 1-1.5 cm2 area. Scale from: -3.25 to 50.14.	The mean change in total hair count from baseline to the end of the study ranged across control groups from **-3.25 to 20.64**	The mean change in total hair count from baseline to the end of the study in the intervention groups was **13.28 higher**		1166 (7 studies[6])	⊕⊕⊖⊖ **low**[7]	

Strength of recommendation

"The extent to which one can be confident that adherence to a recommendation will do more good than harm"

GRADE specifies only two categories of the strength of a recommendation

1. Strong recommendation:

   Do it or don't do it

2. Weak recommendation:

   Probably do it, or probably don't

Strength of recommendation

Determinants of the strength of recommendation

- Quality of the evidence
- Balance between desirable and undesirable effects
- Values and preferences
- Costs

Strength of recommendations

- Strong recommendations
  - strong methods
  - large precise effect
  - few down sides of therapy
- Weak recommendations
  - weak methods
  - imprecise estimate
  - small effect
  - substantial down sides

GRADE working group

GRADE Working Group

- www.gradeworkinggroup.org
- 2x a year meeting “GRADE working group
- Publications BMJ (2008), JCE (2010-12)

GRADEprofiler (support@gradepro.org)

- www.cc-ims.net/gradepro

Conclusions

- GRADE is gaining acceptance as international standard
- It provides criteria for evidence assessment across questions and outcomes
- It provides criteria for moving from evidence to recommendations
- It is simple, transparant and systematic
  - 4 categories of quality of evidence
  - 2 grades for strength of recommendation
- Transparency in decision making and judgements is key

Thank you very much for your attention

Questions?