Grading of Recommendations Assessment, Development, and Evaluation (GRADE)

Mary Lyn Gaffield

16 June 2011

www.gradeworkinggroup.org

Training Course in Sexual and Reproductive Health Research
Geneva Workshop 2011

What is considered a WHO Guideline ?

YES:

- Systematic statements/ recommendations to aid decision making about health interventions, clinical, public health and health system interventions
- Compilations of recommendations ('package of interventions')

- **NO**
  - standards (e.g., pharmacopoeia, food),
  - standard operating procedures (e.g., lab test manuals)
  - Research protocols
  - Reports of EXPERT COMMITTEES

- **MAYBE**
  - compilations of clinical information without clear recommendations
    - Implementation/training guides
    - Journal articles with recommendations

Recommendations versus evidence

- Recommendations are judgements
- Quality of evidence
- Trade off between benefits and harms
- Costs
- Values and preferences

Minimum standards for reporting in WHO guidelines

- Who was involved and their declaration of interests
- How the guideline was developed, including
  - how the evidence was identified
  - how the recommendations were made
- Use by date (review by date)

Why bother grading?

- **People always draw conclusions about:**
  - Quality of evidence
  - Strength of a recommendation
- **Systematic and explicit approaches can help:**
  - Protect against errors
  - Resolve disagreements
  - Facilitate critical appraisal
  - Communicate information

The GRADE approach

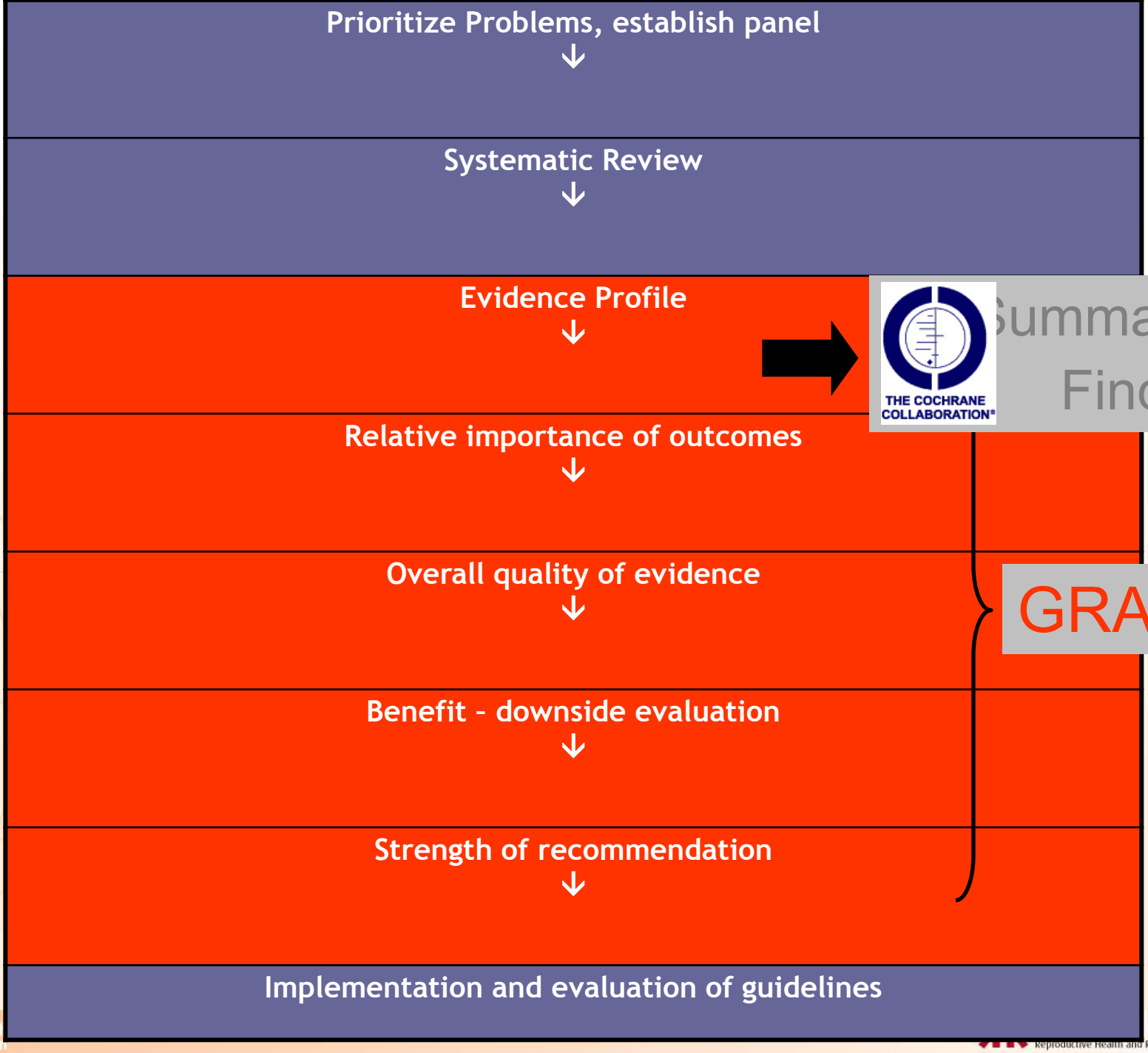
Clear separation of the two issues:

1) Quality of the evidence (High, moderate, low, very low)

- methodological quality of evidence
- likelihood of bias
- by outcome

2) Two grades of recommendation: Strong or Weak (for or against)

- Quality of evidence only one factor


Prioritize Problems, establish panel
Systematic Review
Evidence Profile
Relative importance of outcomes
Overall quality of evidence
Benefit - downside evaluation
Strength of recommendation
Implementation and evaluation of guidelines
THE COCHRANE COLLABORATION
Summary of Findings
GRADE

GRADE and Summary of findings (SoF) table

- **The extent to which one can be confident that an estimate of effect or association is correct. Although the degree of confidence is a continuum, four categories are suggested:**
  - High
  - Moderate
  - Low
  - Very low

- **The quality of the evidence for each of the** critical outcomes **(across studies) is shown in the SoF table**

Quality of evidence - four categories

High	**Further research is very unlikely to change our confidence in the estimate of effect**	⊕⊕⊕⊕
Moderate	**Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.**	⊕⊕⊕◯
Low	**Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate**	⊕⊕◯◯
Very low	**Any estimate of effect is very uncertain**	⊕◯◯◯

Criteria for assessing the quality of the evidence

- Study design

  - RCTs start high

  - Observational studies start low
    - Based on potential for risk of bias

  Both can be downgraded and upgraded

Criteria for assessing the quality of the evidence

- **What lowers quality of evidence? 5 factors:**
  - Study limitations
  - Inconsistency
  - Indirectness
  - Imprecision
  - Publication bias

Criteria for assessing the quality of evidence

- **Study limitations, Randomized controlled trials**
  - No random sequence generation
  - Lack of allocation concealment
  - No true intention to treat principle
  - Inadequate blinding
  - Loss to follow-up
  - Early stopping for benefit

Criteria for assessing the quality of evidence

- **Study limitations, cohorts**
  - Selection of participants to groups
  - Lack of important differences between groups
  - Adjustment for potential confounding factors
    - Intervention group composition
  - Measurement of outcome
  - Loss to follow-up
  - Appropriate time to follow-up

Criteria for assessing the quality of evidence

- ***Consistency*** **(similarity of estimates of effects across studies)**
  - If the estimates are inconsistent and we can not explain the inconsistency, then our confidence in the estimate of effect for that outcome decreases.

  - Arbitrary decisions but need to look at:
    - Size of effect
    - Confidence Interval overlap
    - Statistical difference and heterogeneity measure

Criteria for assessing the quality of evidence

- ***Directness of evidence (the extent of similarity to those of interest)***
  - **Population** (age, sex, diagnosis)
  - **Intervention** (dose, treatment regimen)
  - **Outcome measure** (importance, surrogate outcome, method of measurement, time of measurement)
  - **Comparison**(A vs. B but have to rely on A vs. C and B vs. C)

Criteria for assessing the quality of evidence

- ***Precision* (small sample size)**
  - Small number of events
  - Wide confidence intervals
  - Uncertainty about the magnitude of effect

- ***Publication biases /reporting bias***
  - Outcome bias
  - Publication bias
    - Funnel plots

Criteria for assessing the quality of evidence

- *Criteria that increase the quality of evidence:*
  - Strong evidence of association
  - Very strong evidence of association
  - Evidence of a dose-response gradient
  - All plausible confounders would have reduced the effect

Criteria for assessing the quality of the evidence

Quality of evidence	Study design
High	Randomized trial
Moderate	
Low	Observational study
Very low	

Lower if…
Study limitations
Inconsistency
Indirectness
Imprecision
Publication bias

Higher if…
Large effect (e.g., RR 0.5) Very large effect (e.g., RR 0.2)
Evidence of dose-response gradient
All plausible confounding would reduce a demonstrated effect

GRADE Evidence profile

GRADE Evidence Profile

Author(s): Alonso-Coello P, Mills E, Lopez-Yarto M, Zhou Q, Johanson JF, Guyatt GH.
Date: 20/03/2005
Question: Should laxatives be used for symptomatic hemorrhoids?
Patient or population: Adults with symptomatic hemorrhoids
Settings: Ambulatory care
Systematic review: Alonso-Coello P, Mills E, Lopez-Yarto M, Zhou Q, Johanson JF, Guyatt GH. Laxatives for symptomatic hemorrhoids.

Quality assessment						Summary of findings					
						No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Consistency	Directness	Other considerations	laxatives		Relative (95% CI)	Absolute (95% CI)		
Overall improvement (non validated scale Follow up: 3 months[2])											
4	Randomised trials	Serious limitations (-1)[1]	No important inconsistency	No uncertainty	None	37/148 (25%)	69/146 (47,3%)	RR 0.53 (0.38 to 0.73)	240/1 000 (370 to 120)	⨁⨁⨁◯ Moderate	8
Bleeding (non validated scale Follow up: three months[2])											
5	Randomised trials	Serious limitations (-1)[1]	No important inconsistency	No uncertainty	None	32/128 (25%)	56/123 (45,5%)	RR 0.50 (0.28 to 0.89)	260/1 000 (440 to 70)	⨁⨁⨁◯ Moderate	6
Prolapse (non validated scale Follow up: Three months)											
3	Randomised trials	Serious limitations (-1)[1]	No important inconsistency	No uncertainty	None	29/113 (25,7%)	34/110 (30,9%)	RR 0.79 (0.37 to 1.67)	/1 000 (to)	⨁⨁⨁◯ Moderate	7
Adverse events (Follow up: 30 weeks average follow-up[2])											
3	Randomised trials	Serious limitations (-1)[1]	No important inconsistency	No uncertainty	Imprecise or sparse data (-1)[3]	40/131 (30,5%)	8/135 (5,9%)	RR 6.0 (0.57 to 64.84)	/1 000 (to)	⨁⨁◯◯ Low	6

Footnotes:

1. Quality rated down from high to moderate because of general concerns about methods of individual studies, validity of outcome measures, possibility of publication bias, and some variability in effects, rather than a limitation in one category.
2. Different time point analysis in the studies (6, 12 weeks and 18 months).
3. Wide confidence intervals. Minor gastrointestinal complaints that do not stop patients continuing taking the treatment.

Why SoF table?

- Easier to get an overview of the main findings
- Consideration about importance of outcomes
- Helps identify ‘missing information’ such as lack of adverse events reporting
- An easy to understand SoF table may encourage use of the evidence

What is a SoF table?

- A table that show the main results only
- Based on the GRADE approach to evaluating the quality of evidence
- Show the quality for each of the most important outcomes

SoF Table

Summary of findings:

Compression stockings compared with no compression stockings for people taking long flights

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 Without stockings	Corresponding risk With stockings				
Symptomatic deep vein thrombosis (DVT)	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed symptomatic DVT in these studies.
Symptom-less deep vein thrombosis	**Low risk population** [2]		**RR 0.10** (0.05 to 0.25)	2637 (9 studies)	⨁⨁⨁⨁ **High**	
	10 per 1000	**1 per 1000** (0 to 3)				
	High risk population [2]					
	30 per 1000	**3 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**[3]	
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 −4.5 to −4.9).		1246 (5 studies)	⨁◯◯◯ **Low**[4]	
Pulmonary embolus	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants developed pulmonary embolus in these studies.
Death	See comment	See comment	Not estimable	2821 (9 studies)	See comment	0 participants died in these studies.
Adverse effects	See comment	See comment	Not estimable	1182 (4 studies)	See comment	The tolerability of the stockings was described as very good with no complaints of side effects in 4 studies.[5]

*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; RR: Risk ratio GRADE: GRADE Working Group grades of evidence (see explanations)

Strength of a recommendation

Although the degree of confidence is a continuum, two categories are used: strong and weak.

A **strong recommendation** is one for which the panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects.

A **weak recommendation** is one for which the panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but the panel is not confident about these trade-offs. Reasons for not being confident can include:

- absence of high quality evidence;
- presence of imprecise estimates of benefits or harms;
- uncertainty or variation in how different individuals value the outcomes;
- small benefits;
- the benefits may not be worth the costs (including the costs of implementing the recommendation).

Formulate question

Select outcomes

Rate importance

Outcomes across studies

Create evidence profile with GRADEpro

Rate quality of evidence for each outcome

P
I
C
O

Outcome Critical

Outcome Critical

Outcome Important

Outcome Not important

Summary of findings & estimate of effect for each outcome

High
Moderate
Low
Very low

Randomization increases initial quality

Grade down
1. Risk of bias
2. Inconsistency
3. Indirectness
4. Imprecision
5. Publication bias

Grade up
1. Large effect
2. Dose response
3. Confounders

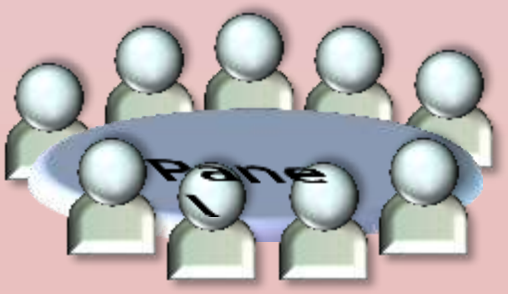
Systematic review

Guideline development

Formulate recommendations:
- For or against (direction)
- Strong or weak (strength)

By considering:
- ❑Quality of evidence
- ❑Balance benefits/harms
- ❑Values and preferences

Revise if necessary by considering:
- ❑Resource use (cost)



Grade
overall quality of evidence
across outcomes based on
lowest quality
of ***critical*** outcomes

- "We recommend using…"
- "We suggest using…"
- "We recommend against using…"
- "We suggest against using…"