

GRADE (Grades of Recommendation Assessment, Development and Evaluation)

Holger Schünemann

Rome, May 8, 2008

Content

- Background and rationale for revisiting guideline methodology
- GRADE approach
 - Quality of evidence
 - Strength of recommendations

Content

- Background and rationale for revisiting guideline methodology

- GRADE approach
 - Quality of evidence
 - Strength of recommendations

Do evidence based guidelines make a difference?

Non-rigorous guidelines:

- Create noise & bias
- Make more aggressive recommendations
- Can harm patients and impair research efforts
- Can reduce credibility of professional societies

Evidence-based clinical practice guidelines can:

- reduce delivery of inappropriate care
- support introduction of new knowledge into clinical practice

Grimshaw et al (1992); Woolf et al (1999); Fretheim et al (2002)

“Practice guidelines ... have been demonstrated to improve patient outcomes and lower cost”

...be based on sound scientific evidence and implemented in an effective manner

S. Weingarten. Hospital
Medicine 2005

Evidence-based Medicine

The conscientious and judicious use of current best evidence from clinical care research in making health care decisions

Confidence in evidence

- There always is evidence
 - “When there is a question there is evidence”
- Evidence alone is never sufficient to make a clinical decision
- Better research \Rightarrow greater confidence in the evidence and decisions

Content

- Background and rationale for revisiting guideline methodology
- GRADE approach
 - Quality of evidence
 - Strength of recommendations

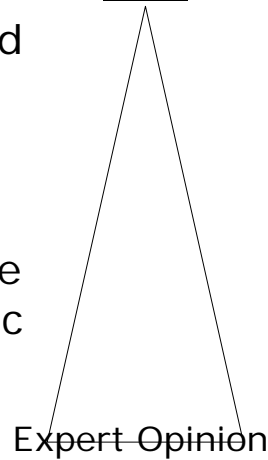
Hierarchy of evidence

STUDY DESIGN

- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations

Expert Opinion

BIAS



Expert Opinion

Why grade evidence?

- People draw conclusions about the
 - quality of evidence and strength of recommendations
- Systematic and explicit approaches can help
 - protect against errors, resolve disagreements
 - communicate information and fulfil needs
- Change practitioner behavior
- However, wide variation in approaches

Which grading system?

Recommendation for use of oral anticoagulation in patients with atrial fibrillation and rheumatic mitral valve disease

Evidence	Recommendation	Organization
■ B	Class I	➤ AHA
■ A	1	➤ ACCP
■ IV	C	➤ SIGN

The same evidence – different classification

Limitations of existing systems

- confuse quality of evidence with strength of recommendations
- lack well-articulated conceptual framework
- criteria not comprehensive or not transparent

Grades of **R**ecommendation
Assessment, **D**evelopment and
Evaluation

GRADE
Working Group

Education and debate

Grading quality of evidence and strength of recommendations

GRADE Working Group

Clinical guidelines are only as good as the evidence and judgments they are based on. The GRADE approach aims to make it easier for users to assess the judgments behind recommendations.

*Grade Working Group. CMAJ 2003, BMJ 2004, BMC 2004, BMC 2005, A-IPCCM 2006, BMJ 2008 (in press)

About GRADE

- Since 2000
- Researchers/guideline developers with interest in methodology
- Aim: to develop a common, transparent and sensible system for grading the quality of evidence and the strength of recommendations
- Evaluation of existing systems

GRADE Working Group

- David Atkins, chief medical officer^a
- Dana Best, assistant professor^b
- Martin Eccles, professor^d
- Françoise Cluzeau, lecturer^x
- Yngve Falck-Ytter, associate director^e
- Signe Flottorp, researcher^f
- Gordon H Guyatt, professor^g
- Robin T Harbour, quality and information director^h
- Margaret C Haugh, methodologistⁱ
- David Henry, professor^j
- Suzanne Hill, senior lecturer^j
- Roman Jaeschke, clinical professor^k
- Regina Kunx, Associate Professor
- Gillian Leng, guidelines programme director^l
- Alessandro Liberati, professor^m
- Nicola Magrini, directorⁿ
- James Mason, professor^d
- Philippa Middleton, honorary research fellow^o
- Jacek Mrukowicz, executive director^p
- Dianne O'Connell, senior epidemiologist^q
- Andrew D Oxman, director^r
- Bob Phillips, associate fellow^r
- Holger J Schünemann, professor^{s,5}
- Tessa Tan-Torres Edejer, medical officer^t
- David Tovey, Editor^v
- Jane Thomas, Lecturer, UK
- Helena Varonen, associate editor^u
- Gunn E Vist, researcher^f
- John W Williams Jr, professor^v
- Stephanie Zaza, project director^w
- a) Agency for Healthcare Research and Quality, **USA**
- b) Children's National Medical Center, **USA**
- c) Centers for Disease Control and Prevention, **USA**
- d) University of Newcastle upon Tyne, **UK**
- e) German Cochrane Centre, **Germany**
- f) Norwegian Centre for Health Services, **Norway**
- g) McMaster University, **Canada**
- h) Scottish Intercollegiate Guidelines Network, **UK**
- i) Fédération Nationale des Centres de Lutte Contre le Cancer, **France**
- j) University of Newcastle, **Australia**
- k) McMaster University, **Canada**
- l) National Institute for Clinical Excellence, **UK**
- m) Università di Modena e Reggio Emilia, **Italy**
- n) Centro per la Valutazione della Efficacia della Assistenza Sanitaria, **Italy**
- o) Australasian Cochrane Centre, **Australia**
- p) Polish Institute for Evidence Based Medicine, **Poland**
- q) The Cancer Council, **Australia**
- r) Centre for Evidence-based Medicine, **UK**
- s) National Cancer Institute, **Italy**
- t) World Health Organisation, **Switzerland**
- u) Finnish Medical Society Duodecim, **Finland**
- v) Duke University Medical Center, **USA**
- w) Centers for Disease Control and Prevention, **USA**
- x) University of London, **UK**
- y) BMJ Clinical Evidence, **UK**

GRADE Uptake

- World Health Organization
- National Institute Clinical Excellence (NICE)
- Agency for Health Care Research and Quality (AHRQ)
- Canadian Agency for Drugs and Technology in Health
- Allergic Rhinitis in Asthma Guidelines (ARIA)
- American Thoracic Society
- American College of Chest Physicians
- UpToDate
- British Medical Journal
- American College of Physicians
- Cochrane Collaboration
- European Society of Thoracic Surgeons
- Clinical Evidence
- Many other organizations

The GRADE approach

Clear separation of 2 issues:

1) 4 categories of quality of evidence:
very low, low, moderate, or high
quality?

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

2) Recommendation: 2 grades - weak or
strong (for or against)?

- Quality of evidence only one factor

*www.GradeWorking-Group.org

Determinants of quality

- RCTs start high
- observational studies start low
- what can lower quality? 5 factors
 - detailed design and execution
 - inconsistency
 - indirectness
 - reporting bias
 - imprecision

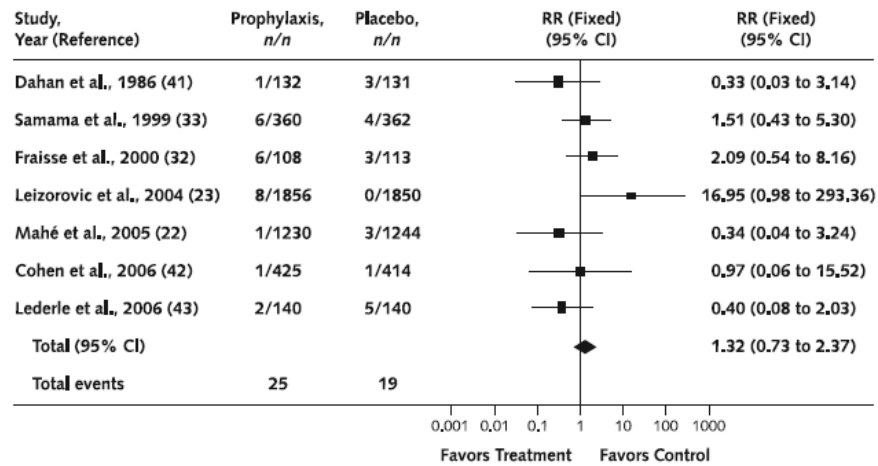
Design and Execution

- limitations
 - lack of concealment
 - intention to treat principle violated
 - inadequate blinding
 - loss to follow-up
 - early stopping for benefit
- Example: RCT suggests that danaparoid sodium is of benefit in treating HIT complicated by thrombosis
 - Key outcome: clinicians' assessment of when the thromboembolism had resolved
 - Not blinded – subjective judgement

Consistency of results

- consistency of results
- if inconsistency, look for explanation
 - patients, intervention, outcome, methods
- unexplained inconsistency downgrade quality
- Bleeding in thrombosis-prophylaxed hospitalized patients
 - seven RCTs
 - 4 lower, 3 higher risk

Example: Bleeding in the hospital



Dentali et al. Ann Int Med, 2007

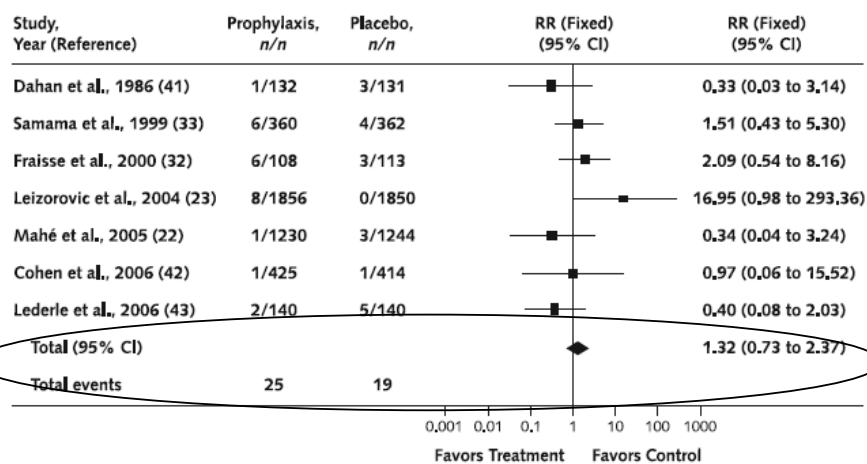
Directness of Evidence

- indirect comparisons
 - interested in A versus B
 - have A versus C and B versus C
 - formoterol versus salmeterol versus tiotropium
- differences in
 - patients (mild versus severe COPD)
 - interventions (all inhaled steroids)
 - outcomes (long-term health-related quality of life, short-term functional capacity, laboratory exercise, spirometry)

Reporting Bias & Imprecision

- reporting bias
 - reporting of studies
 - publication bias
 - number of small studies
 - reporting of outcomes
- small sample size
 - small number of events
 - wide confidence intervals
 - uncertainty about magnitude of effect

Example: Bleeding in the hospital



What can raise quality?

- large magnitude can upgrade (RRR 50%)
 - very large two levels (RRR 80%)
 - common criteria
 - everyone used to do badly
 - almost everyone does well
 - Oral anticoagulation for mechanical heart valves
- dose response relation
(higher INR – increased bleeding)

Quality assessment criteria

Quality of evidence	Study design	Lower if	Higher if
High	Randomised trial	Study quality: Serious limitations Very serious limitations Important inconsistency Directness: Some uncertainty Major uncertainty Sparse or imprecise data High probability of reporting bias	Strong association: Strong, no plausible confounders Very strong, no major threats to validity Evidence of a Dose response gradient All plausible confounders would have reduced the effect
Moderate			
Low	Observational study		
Very low			

GRADE Profiles

No of studies	Quality assessment						Summary of findings				Quality	Importance
	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	No of patients	Relative (95% CI)	Effect			
All cause mortality follow-up 4-36 months												
35	randomised trial	no serious limitations	serious ¹	no serious indirectness	serious ²	none	1000/3025	830/3093 10% 50%	HR 1.11 (1 to 1.22)	25 more per 1000 (from 0 more to 49 more) 10 more per 1,000 36 more per 1,000	⊕⊕⊕⊕ LOW	CRITICAL
Thromboembolic events												
30	randomised trial	serious ³	serious ¹	no serious indirectness	no serious imprecision	none	218/3355	112/2737 1% 8%	RR 1.69 (1.36 to 2.1)	20 more per 1000 (from 15 more to 45 more) 6 more per 1,000 66 more per 1,000	⊕⊕⊕⊕ LOW	CRITICAL
Complete response of tumor to chemotherapy												
5	randomised trial	no serious limitations	serious ¹	no serious indirectness	no serious imprecision	reporting bias ⁴	216/344	211/344	RR 1.0 (0.92 to 1.1)	0 fewer per 1000 (from 49 fewer to 61 more)	⊕⊕⊕⊕ LOW	CRITICAL
Transfusion rates (follow-up 4-26 weeks)												
34	randomised trial	no serious limitations	serious ⁵	no serious indirectness	no serious imprecision	none	864/2659	1110/2351 26% 75%	RR 0.63 (0.59 to 0.67)	175 fewer per 1000 (from 156 fewer to 194 fewer) 52 fewer per 1,000 277 fewer per 1,000	⊕⊕⊕⊕ MODERATE	CRITICAL
Increase > 2 mg/dL in Hb (mg/dL) (follow-up 4-20 weeks)												
15	randomised trial	no serious limitations	serious ⁶	no serious indirectness	no serious imprecision	strong association ⁷	1069/1844	239/1449	RR 3.42 (3.03 to 3.86)	399 more per 1000 (from 335 more to 472 more)	⊕⊕⊕⊕ HIGH	IMPORTANT

¹ Overall heterogeneity not significant, but underlying clinical heterogeneity due to risk of VTE, treatment regimens, and epo protocols (starting and stopping Hb).
² CI includes no effect and clinically important increase in mortality.
³ Criteria for determining and reporting VTE variable in studies; trials reporting varying combinations of DVT, PE, TIA, stroke, and MI.
⁴ Only 5 trials reported this outcome; does not include the largest trials powered for mortality benefit.
⁵ Tests of heterogeneity I square were significant. Reduced risk of transfusion evidence in subgroups defined by different starting Hb level, but size of benefit differs. Clinical heterogeneity in control rate transfusions, tumor type and chemo regimen, and protocols for determining transfusion need.
⁶ All trials support substantial benefit but significant heterogeneity in magnitude of benefit, clinical heterogeneity in starting Hb levels, underlying chemo regimens and tumor types, and risk of anemia.
⁷ Size of RR (3.4 pooled, range 2 to 9) would qualify as large effect.

Summary of Findings Tables

Erythropoiesis stimulants (epo) compared to placebo for anemia from cancer chemotherapy						
Patient or population: patients with anemia from cancer chemotherapy						
Settings: Outpatient Cancer treatment						
Intervention: Erythropoiesis stimulants (epo)						
Comparison: placebo						
Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	placebo	Erythropoiesis stimulants (epo)				
All cause mortality (follow-up: 4 - 36 months)	Population	268 per 1000	293 per 1000 (259 to 317)	RR 1.11 (1 to 1.22)	6918 (36)	⊕⊕⊕⊕ low ²
	Low risk population	100 per 1000	110 per 1000 (100 to 121)			
	High risk population	500 per 1000	537 per 1000 (500 to 571)			
	Population	41 per 1000	69 per 1000 (50 to 86)			
	Low risk population	10 per 1000	17 per 1000 (14 to 21)			
Thromboembolic events	High risk population	80 per 1000	135 per 1000 (109 to 160)	RR 1.69 (1.36 to 2.1)	6092 (30)	⊕⊕⊕⊕ low ³
	Population	613 per 1000	613 per 1000 (664 to 674)			
	Low risk population	10 per 1000	17 per 1000 (14 to 21)			
Complete response of tumor to chemotherapy	613 per 1000	613 per 1000 (664 to 674)	RR 1.0 (0.92 to 1.1)	600 (6)	⊕⊕⊕⊕ low ⁴	

¹ Overall heterogeneity not significant, but underlying clinical heterogeneity due to risk of VTE, treatment regimens, and epo protocols (starting and stopping Hb).
² CI includes no effect and clinically important increase in mortality.
³ Criteria for determining and reporting VTE variable in studies; trials reporting varying combinations of DVT, PE, TIA, stroke, and MI.
⁴ Only 5 trials reported this outcome; does not include the largest trials powered for mortality benefit.
⁵ Tests of heterogeneity I square were significant. Reduced risk of transfusion evidence in subgroups defined by different starting Hb level, but size of benefit differs. Clinical heterogeneity in control rate transfusions, tumor type and chemo regimen, and protocols for determining transfusion need.
⁶ All trials support substantial benefit but significant heterogeneity in magnitude of benefit, clinical heterogeneity in starting Hb levels, underlying chemo regimens and tumor types, and risk of anemia.
⁷ Size of RR (3.4 pooled, range 2 to 9) would qualify as large effect.

Strength of recommendation

- "The strength of a recommendation reflects the extent to which we can, across the range of patients for whom the recommendations are intended, be confident that desirable effects of a management strategy outweigh undesirable effects."

Desirable and undesirable effects

- Desirable effects
 - Mortality
 - improvement in quality of life, fewer hospitalizations/infections
 - reduction in the burden of treatment
 - reduced resource expenditure
- Undesirable effects
 - deleterious impact on morbidity, mortality or quality of life, increased resource expenditure

Determinants of the strength of recommendation

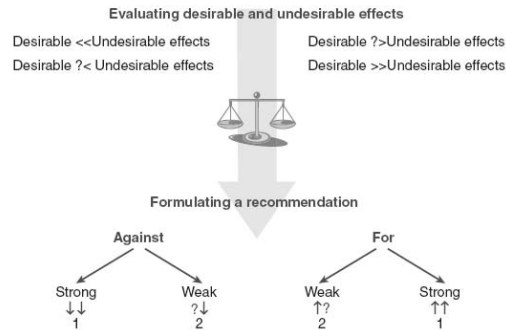
Factors that can strengthen a recommendation	Comment
Quality of the evidence	
Balance between desirable and undesirable effects	
Values and preferences	
Costs (resource allocation)	

Determinants of the strength of recommendation

Factors that can strengthen a recommendation	Comment
Quality of the evidence	The higher the quality of evidence, the more likely is a strong recommendation.
Balance between desirable and undesirable effects	The larger the difference between the desirable and undesirable consequences, the more likely a strong recommendation warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely weak recommendation warranted.
Values and preferences	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak recommendation warranted.
Costs (resource allocation)	The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted

Developing recommendations

Strength of Recommendations



The figure describes the balance between important benefits and downsides relate to a recommendation. The process begins by evaluating whether desirable effects outweigh undesirable effects or vice versa. Moving on to making a recommendation requires a decision: if the balance is clear, a strong recommendation for or against an action follows (<< and >> denote a clear balance). If the balance is not clear, a weak recommendation for or against an action follows (?< and ?> denote a balance that is not clear). Widely differing values (the importance or preference patients assign to a certain health state) can also lead to a less clear balance of benefits versus downsides.

Conclusions

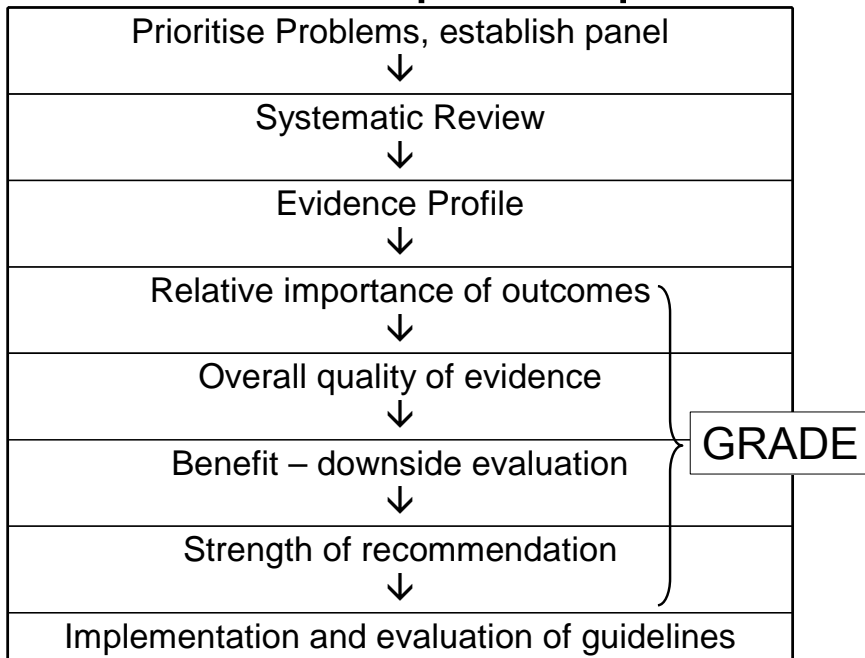
- GRADE is gaining acceptance as international standard
- Criteria for evidence assessment across questions and outcomes
- Criteria for moving from evidence to recommendations
- Simple, transparent, systematic
 - four categories of quality of evidence
 - two grades for strength of recommendations
- Transparency in decision making and judgments is key

Thank you

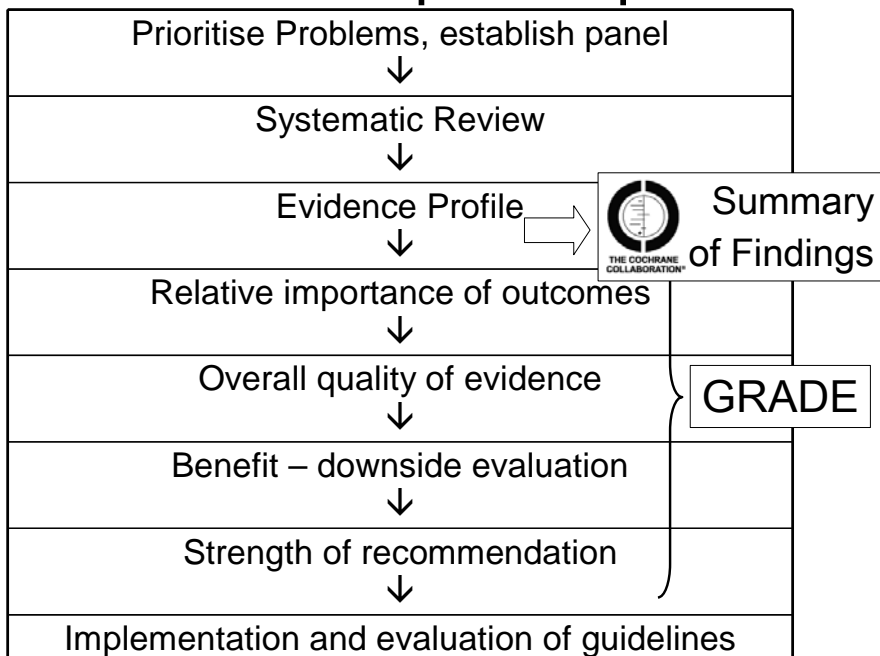
Questions for you

- Are systematic reviews for every recommendation in your guidelines a reality/possibility?
- What about cost – how do you deal with cost and how should we deal with it?

Guideline development process



Guideline development process



Implications of a *strong* recommendation

- Patients: Most people in this situation would want the recommended course of action and only a small proportion would not
- Clinicians: Most patients should receive the recommended course of action
- Policy makers: The recommendation can be adapted as a policy in most situations

Implications of a *weak* recommendation

- Patients: The majority of people in this situation would want the recommended course of action, but many would not
- Clinicians: Be prepared to help patients to make a decision that is consistent with their own values/decision aids and shared decision making
- Policy makers: There is a need for substantial debate and involvement of stakeholders