Grading evidence and recommendations

Nicola Magrini
on behalf of the GRADE working group

GIN-WHO meeting
Budapest 11 october 2006

How can we judge the extent of our confidence that adherence to a recommendation will do more good than harm?

Why bother about grading?

- People draw conclusions about the
  - **quality of evidence**
  - **strength of recommendations**
- Systematic and explicit approaches can help
  - protect against errors
  - resolve disagreements
  - facilitate critical appraisal
  - communicate information
- However, there is wide variation in currently used approaches

Who is confused?

Evidence	Recommendation	Organization
II-2	B	USPSTF
C+	1	ACCP
Strong	Strongly recommended	GCPS

Still not confused?

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

Evidence	Recommendation	Organization
B	Class I	AHA
C+	1	ACCP
IV	C	SIGN

Weaknesses of available grading systems

- Take mostly in considerations study design
- Do not consider separately the evidence for benefits and harms
- Do not explicitly evaluate the risk-benefit ratio

GRADE

Grades of **R**ecommendation **A**ssessment, **D**evelopment and **E**valuation

GRADE Working Group

David Atkins, chief medical officer[a]
Dana Best, assistant professor[b]
Peter A Briss, chief[c]
Martin Eccles, professor[d]
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, methodologist[i]
David Henry, professor[j]
Suzanne Hill, senior lecturer[j]
Roman Jaeschke, clinical professor[k]
Gillian Leng, guidelines programme director[l]
Alessandro Liberati, professor[m]
Nicola Magrini, director[n]
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[f]
Bob Phillips, associate fellow[r]
Holger J Schünemann, associate professorg[g,s]
Tessa Tan-Torres Edejer, medical officer/scientist[t]
Helena Varonen, associate editor[u]
Gunn E Vist, researcher[f]
John W Williams Jr, associate 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) University of Buffalo, **USA**
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**

Education and debate

BMJ, 2004

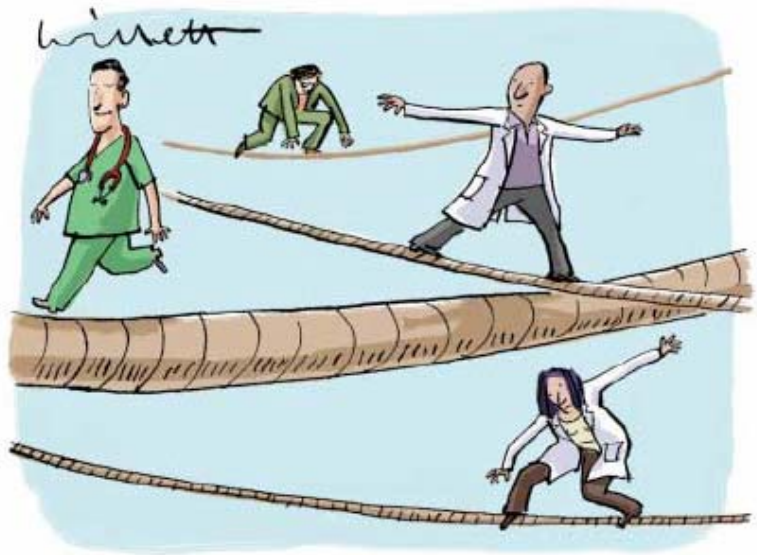
Grading quality of evidence and strength of recommendations

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

Summary

Users of clinical practice guidelines and other recommendations need to know how much confidence they can place in the recommendations. Systematic and explicit methods of making judgments can reduce errors and improve communication. We have developed a system for grading the quality of evidence and the strength of recommendations that can be applied across a wide range of interventions and contexts. In this article we present a summary of our approach from the perspective of a guideline user. Judgments about the strength of a recommendation require consideration of the balance between benefits and harms, the quality of the evidence, translation of the evidence into specific circumstances, and the certainty of the baseline risk. It is also important to consider costs (resource utilisation) before



GRADE: some references

- Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches The GRADE Working Group. *BMC Health Services Research* 2004, 4:38
- Systems for grading the quality of evidence and the strength of recommendations II: Pilot study of a new system. *BMC Health Services Research* 2005, 5:25
- Grading Strength of Recommendations and Quality of Evidence in Clinical Guidelines. Report From an American College of Chest Physicians Task Force. *Chest 2006; 129:174–181*
- Addressing Resource Allocation Issues in Recommendations From Clinical Practice Guideline Panels. Suggestions From an American College of Chest Physicians Task Force. *Chest 2006; 129:182–187*
- An Official ATS Statement: Grading the Quality of Evidence and Strength of Recommendations in ATS Guidelines and Recommendations. *Am J Respir Crit Care Med 2006;174:605–614,*

GRADE – the process

1. Choice and ratings of outcomes
2. Quality of evidence and evidence profile
3. Risk-benefit ratio, considering cost, access and feasibility

4. Strong or weak recommendation

GRADE: the various steps

Define the specific question

1. Choose and rate the importance of the outcomes (of benefit and harm) considered to be relevant

2. Produce an evidence table/balance sheet for all relevant outcomes

2. Evaluate the quality of the available evidence for each relevant outcome

2. Evaluate the quality of evidence across all relevant outcomes

3. Rate the balance of benefits and harms

4. Define the strength of the recommendation

GRADE – the process

1. **Choice and ratings of outcomes**
2. Quality of evidence and evidence profile
3. Risk-benefit ratio, considering cost, access and feasibility
4. Strong or weak recommendation

Choice of (detailed) outcomes HERCEPTIN IN BREAST CANCER

- Overall survival at 5 years
- Overall survival at 3 years
- Disease free survival at 5 years
- Disease free survival at 2 years
- Heart failure
- QoL

Rate from 1 (not important at all) to 9 (of critical importance) the relevance of each outcome

GRADE: outcome rating

Please indicate the relative importance of each outcome (by circling a number between 1 and 9).

Death

Not critical to making a decision	1 2 3 4 5 6 7 8 9	Critical to making a decision

Non-fatal MI

Not critical to making a decision	1 2 3 4 5 6 7 8 9	Critical to making a decision

Outcome X

Not critical to making a decision	1 2 3 4 5 6 7 8 9	Critical to making a decision

GRADE – the process

1. Choice and ratings of outcomes
2. **Quality of evidence and evidence profile**
3. Risk-benefit ratio, considering cost, access and feasibility
4. Strong or weak recommendation

Quality of evidence

Definition: 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, we suggest using four categories:
  - High
  - Moderate
  - Low
  - Very low

Categories of quality

- **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. ⨁◯◯◯

Quality assessment criteria

Quality of evidence	Study design	Lower if	Higher if
High	Randomised trial	**Study quality**: -1 Serious limitations -2 Very serious limitations -1 Important **inconsistency** **Directness**: -1 Some uncertainty -2 Major uncertainty -1 **Sparse or imprecise data** -1 High probability of **reporting bias**	**Strong association**: +1 Strong, no plausible confounders +2 Very strong, no major threats to validity +1 Evidence of a **Dose response gradient** +1 All plausible **confounders** would have reduced the effect
Moderate			
Low	Observational study		
Very low			

Judgements about the quality of evidence

The quality of the evidence (i.e. our confidence) depends on:

- **study design** (e.g. RCT, case-control study)
- **study quality/limitations** (protection against bias; e.g. concealment of allocation, blinding, follow-up)
- **consistency of results** (among studies)
- **Directness/generalizability of the evidence** including the
  - **populations** (those of interest versus similar; for example, older, sicker or more co-morbidity)
  - **interventions** (those of interest versus similar; for example, drugs within the same class)
  - **outcomes** (important versus surrogate outcomes)
  - **comparison** (*A - C* versus *A - B & C - B*)

GRADE: why consistency

Discussion Sections in Reports of Controlled Trials Published in General Medical Journals

Islands in Search of Continents?

Michael Clarke, DPhil; Iain Chalmers, MSc

Context.—Several journals have adopted the Consolidated Standards of Reporting Trials (CONSORT) recommendations to make assessment of the quality of randomized controlled trials (RCTs) easier. One of these recommendations is that the trial's results be discussed in light of the totality of the available evidence.

Objective.—To assess the extent to which reports of RCTs published in 5 general medical journals have discussed new results in light of all available evidence.

Design.—Assessment of the discussion sections in all 26 reports of RCTs published during May 1997 in *Annals of Internal Medicine*, *BMJ*, *JAMA*, *The Lancet*, and *The New England Journal of Medicine*.

Main Outcome Measure.—The inclusion or mention of a systematic review in the discussion section of each article.

systematic scrutiny. The typical discussion section usually addresses a number of dimensions, but, crucially, it is in this section that readers will look for an answer to Bradford Hill's "bottom line" question for any research article: "What does it mean, anyway?"[3] This was recognized in the CONSORT statement, which included the recommendation that trialists should "state general interpretation of the data in light of the totality (our emphasis) of the available evidence."[1]

GRADE: why consistency

To deserve the public's continued support and trust, researchers and journals need to ensure that reports of research end with scientifically defensible answers to Bradford Hill's question,"What does it mean, anyway?"

Those who turn to reports of trials to help guide treatment deserve nothing less than a discussion of the totality of the relevant evidence, as rightly recommended by the

CONSORT Group.

To paraphrase John Donne, No trial is an island, entire of itself; every trial is a piece of the continent, a part of the main.

Judgements about the quality of evidence

The quality of the evidence (i.e. our confidence) may also be REDUCED when there is:

- ↓ Sparse or imprecise data
- ↓ Reporting bias

The quality of the evidence (i.e. our confidence) may be INCREASED when there is:

- ↑ A strong association
- ↑ A dose response relationship

Quality assessment criteria

Quality of evidence	Study design	Lower if	Higher if
High	Randomised trial	**Study quality**: -1 Serious limitations -2 Very serious limitations -1 Important **inconsistency** **Directness**: -1 Some uncertainty -2 Major uncertainty -1 **Sparse or imprecise data** -1 High probability of **reporting bias**	**Strong association**: +1 Strong, no plausible confounders +2 Very strong, no major threats to validity +1 Evidence of a **Dose response gradient** +1 All plausible **confounders** would have reduced the effect
Moderate			
Low	Observational study		
Very low			

GRADE – evidence profile

Table 2 Quality assessment of trials comparing selective serotonin reuptake inhibitors (SSRIs) with tricyclic antidepressants for treatment of moderate depression in primary care[2]

	Quality assessment					Summary of findings					
						No of patients		Effect			
No of studies	Design	Quality	Consistency	Directness	Other modifying factors*	SSRIs	Tricyclics	Relative (95% CI)	Absolute	Quality	Importance
Depression severity (measured with Hamilton Depression Rating Scale after 4 to 12 weeks)											
Citalopram (8) Fluoxetine (38) Fluvoxamine (25) Nefazodone (2) Paroxetine (18) Sertraline (4) Venlafaxine (4)	Randomised controlled trials	No serious limitations	No important inconsistency	Some uncertainty about directness (outcome measure)†	None	5044	4510	WMD 0.034 (–0.007 to 0.075)	No difference	Moderate	Critical
Transient side effects resulting in discontinuation of treatment											
Citalopram (8) Fluoxetine (50) Fluvoxamine (27) Nefazodone (4) Paroxetine (23) Sertraline (6) Venlafaxine (5)	Randomised controlled trials	No serious limitations	No important inconsistency	Direct	None	1948/7032 (28%)	2072/6334 (33%)	RRR 13% (5% to 20%)	5/100	High	Critical
Poisoning fatalities[1]											
UK Office for National Statistics (1)	Observational data	Serious limitation‡	Only one study	Direct	Very strong association	1/100 000/ year of treatment	58/100 000/ year of treatment	RRR 98% (97% to 99%)§	6/10 000	Moderate	Critical

WMD = weighted mean difference, RRR = relative risk reduction.
*Imprecise or sparse data, a strong or very strong association, high risk of reporting bias, evidence of a dose-response gradient, effect of plausible residual confounding.
†There was uncertainty about the directness of the outcome measure because of the short duration of the trials.
‡It is possible that people at lower risk were more likely to have been given SSRIs and it is uncertain if changing antidepressant would have deterred suicide attempts.
§There is uncertainty about the baseline risk for poisoning fatalities.

GRADE – quality of evidence (1)

	Quality assessment				
No of studies	**Design**	**Quality**	**Consistency**	**Directness**	**Other modifying factors***
Depression severity (measured with Hamilton Depression Rating Scale after 4 to 12 weeks)					
Citalopram (8)	Randomised controlled trials	No serious limitations	No important inconsistency	Some uncertainty about directness (outcome measure)†	None
Fluoxetine (38)					
Fluvoxamine (25)					
Nefazodone (2)					
Paroxetine (18)					
Sertraline (4)					
Venlafaxine (4)					
Transient side effects resulting in discontinuation of treatment					
Citalopram (8)	Randomised controlled trials	No serious limitations	No important inconsistency	Direct	None
Fluoxetine (50)					
Fluvoxamine (27)					
Nefazodone (4)					
Paroxetine (23)					
Sertraline (6)					
Venlafaxine (5)					
Poisoning fatalities§					
UK Office for National Statistics (1)	Observational data	Serious limitation‡	Only one study	Direct	Very strong association

GRADE – quality of evidence (2)

	Summary of findings					
	No of patients		Effect			
No of studies	SSRIs	Tricyclics	Relative (95% CI)	Absolute	Quality	Importance
Depression severity						
Citalopram (8) Fluoxetine (38) Fluvoxamine (25) Nefazodone (2) Paroxetine (18) Sertraline (4) Venlafaxine (4)	5044	4510	WMD 0.034 (–0.007 to 0.075)	No difference	Moderate	Critical
Transient side effects						
Citalopram (8) Fluoxetine (50) Fluvoxamine (27) Nefazodone (4) Paroxetine (23) Sertraline (6) Venlafaxine (5)	1948/7032 (28%)	2072/6334 (33%)	RRR 13% (5% to 20%)	5/100	High	Critical
Poisoning fatalities[§]						
UK Office for National Statistics (1)	1/100 000/ year of treatment	58/100 000/ year of treatment	RRR 98% (97% to 99%)§	6/10 000	Moderate	Critical

GRADE – the process

1. Choice and ratings of outcomes
2. Quality of evidence and evidence profile
3. **Risk-benefit ratio,**

   Considering cost, access and feasibility

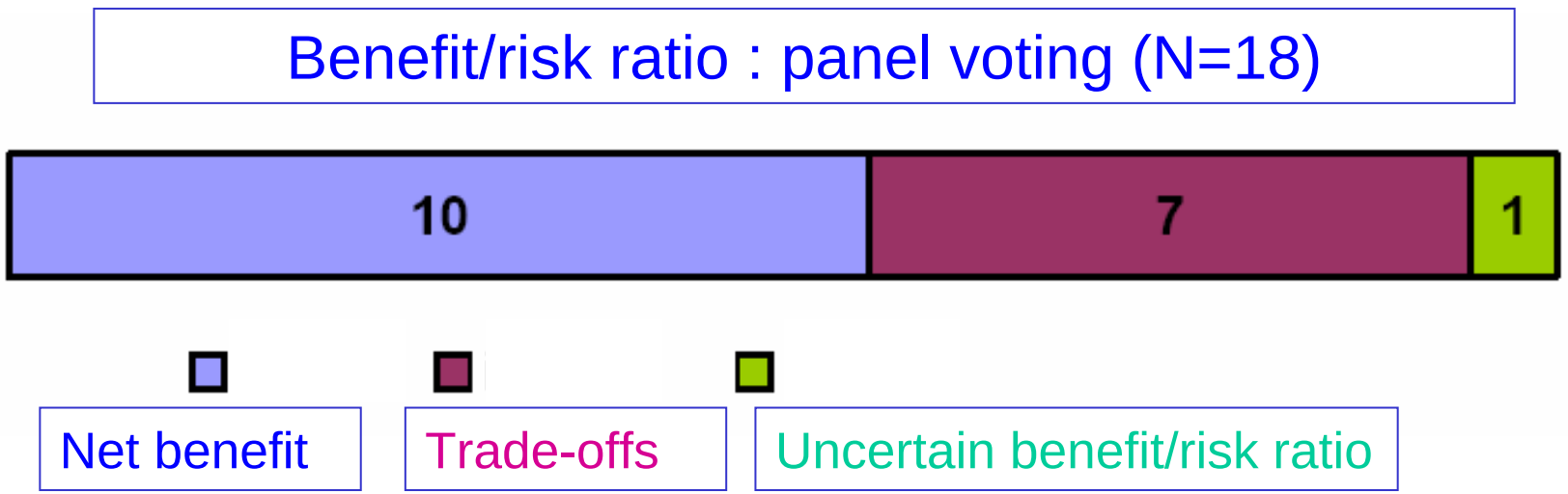
4. Strong or weak recommendation

Benefit/risk ratio

considering also cost, access, feasibility

trastuzumab in adjuvant breast cancer

the experience of Emilia Romagna NHS R&D group



GRADE – the process

1. Choice and ratings of outcomes
2. Quality of evidence and evidence profile
3. Risk-benefit ratio, considering cost, access and feasibility

4. Strong or weak recommendation

Strength of recommendation

The degree of confidence that the desirable effects of adherence to a recommendation outweigh the undesirable effects.

Desirable effects
- health benefits
- less burden
- savings

Undesirable effects
- Harms
- more burden
- costs

Judgements about the balance between benefits and harms

Although the degree of confidence is a continuum, we suggest using two categories: strong and weak.

- **Strong recommendation:** the panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects.
- **Weak recommendation:** the panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but is not confident.

Strength, R/B ratio and quality of evidence for a **strong** recommendation

TABLE 6. GRADING RECOMMENDATIONS

Grade of Recommendation	Clarity of Risk/Benefit	Quality of Supporting Evidence	Implications
Strong recommendation High-quality evidence	Benefits clearly outweigh harms and burdens, or vice versa	Consistent evidence from well-performed randomized controlled trials or exceptionally strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research is very unlikely to change our confidence in the estimate of effect.
Strong recommendation Moderate-quality evidence	Benefits clearly outweigh harms and burdens, or vice versa	Evidence from randomized controlled trials with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or unusually strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Strong recommendation Low-quality evidence	Benefits clearly outweigh harms and burdens, or vice versa	Evidence for at least one critical outcome from observational studies, from randomized controlled trials with serious flaws or indirect evidence	Recommendation may change when higher quality evidence becomes available. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Strong recommendation Very-low-quality evidence (very rarely applicable)	Benefits clearly outweigh harms and burdens , or vice versa	Evidence for at least one of the critical outcomes from unsystematic clinical observations or very indirect evidence	Recommendation may change when higher quality evidence becomes available; any estimate of effect, for at least one critical outcome, is very uncertain.

Schunemann, ATS, 2006

Strength, R/B ratio and quality of evidence for a **weak** recommendation

applicable)			
Weak recommendation High-quality evidence	Benefits closely balanced with harms and burdens	Consistent evidence from well-performed randomized controlled trials or exceptionally strong evidence from unbiased observational studies	The best action may differ depending on circumstances or patients or societal values. Further research is very unlikely to change our confidence in the estimate of effect.
Weak recommendation Moderate-quality evidence	Benefits closely balanced with harms and burdens	Evidence from randomized, controlled trials with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or unusually strong evidence from unbiased observational studies	Alternative approaches likely to be better for some patients under some circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Weak recommendation Low-quality evidence	Uncertainty in the estimates of benefits, harms, and burdens; benefits may be closely balanced with harms and burdens	Evidence for at least one critical outcome from observational studies, from randomized controlled trials with serious flaws, or indirect evidence	Other alternatives may be equally reasonable. 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.
Weak recommendation. Very low quality of evidence	Major uncertainty in the estimates of benefits, harms, and burdens; benefits may or may not be balanced with harms and burdens	Evidence for at least one critical outcome from unsystematic clinical observations or very indirect evidence	Other alternatives may be equally reasonable. Any estimate of effect, for at least one critical outcome, is very uncertain.

Schunemann, ATS, 2006

Strong recommendation: implications

TABLE 1. EXAMPLES OF IMPLICATIONS OF STRONG AND WEAK RECOMMENDATIONS FOR DIFFERENT GROUPS OF GUIDELINE USERS

Strong Recommendations

- For patients: Most individuals in this situation would want the recommended course of action and only a small proportion would not. Formal decision aids are not likely to be needed to help individuals make decisions consistent with their values and preferences.
- For clinicians: Most individuals should receive the intervention. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator.
- For policy makers: The recommendation can be adapted as policy in most situations.

Example: Early anticoagulation in patients with deep venous thrombosis for the prevention of pulmonary embolism; antibiotics for the treatment of community-acquired pneumonia.

Schunemann, ATS, 2006

Categories of recommendations

Although the degree of confidence is a continuum, we suggest using two categories: **strong and weak**.

- **Strong recommendation:** the panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects.
- **Weak recommendation:** the panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but is not confident.

Recommend ↑↑

Suggest ↑

Judgements about the strength of a recommendation

Reasons for not being confident can include:

- absence of high quality evidence
- imprecise estimates
- uncertainty or variation in how different individuals value the outcomes
- small net benefits
- uncertainty whether the net benefits are worth the costs (including the costs of implementing the recommendation)

Judgements about the strength of a recommendation

- No precise threshold for going from a strong to a weak recommendation
- The presence of important concerns about one or more of the above factors make a weak recommendation more likely.
- Panels should consider all of these factors and make the reasons for their judgements explicit.
- Recommendations should specify the perspective that is taken (e.g. individual patient, societal) and which outcomes were considered (including which, if any costs).

GRADE
a flexible system

GRADE: flexible strong or weak

BOX 7. EXAMPLE: WEAK RECOMMENDATIONS BASED ON HIGH-QUALITY EVIDENCE

Several RCTs compared the use of combination chemotherapy and radiotherapy versus radiotherapy alone in unresectable, locally advanced non–small cell lung cancer (stage IIIA) (24, 25). The overall quality rating for these trials could be considered high by a guideline panel. Compared with radiotherapy alone, the combination of chemotherapy and radiotherapy reduces the risk for death corresponding to a mean gain in life expectancy of few months (24), but increases harm and burden related to chemotherapy. Thus, considering the values and preferences patients would place on the small survival benefit in view of the harms and burdens, guideline panels may offer a weak recommendation (Table 1) despite the high quality of the available evidence.

GRADE: flexible

strong or weak

according to baseline risk

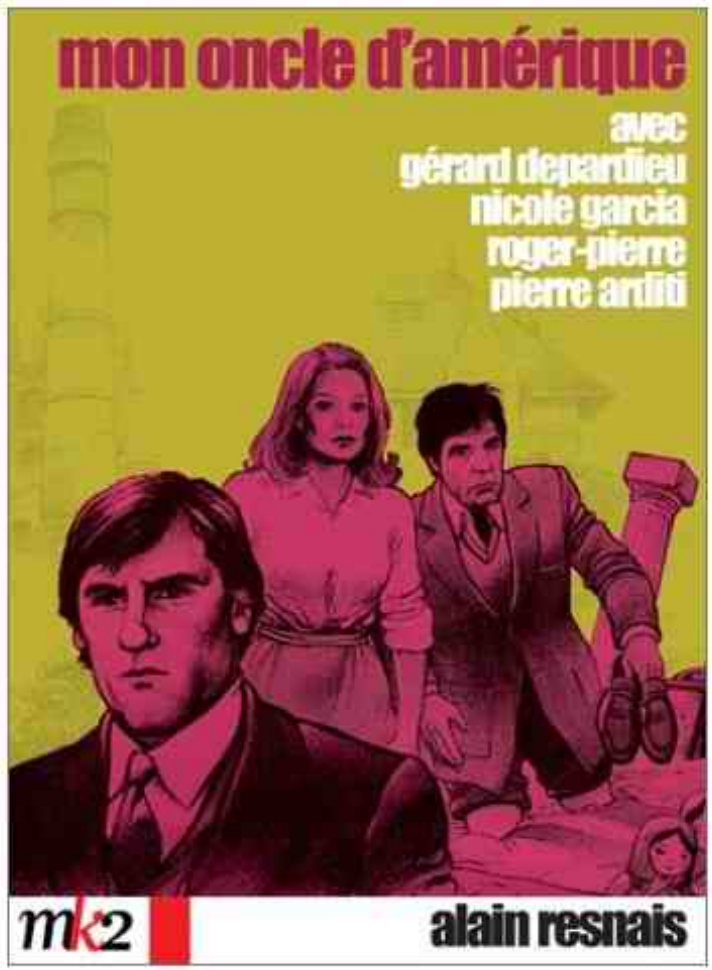
BOX 5. EXAMPLE: INFLUENCE OF BASELINE RISK ON THE STRENGTH OF RECOMMENDATIONS

Consider a 65-year-old patient with mild COPD and frequent exacerbations for whom inhaled corticosteroids are a treatment option. This individual's risk for suffering an exacerbation in the next year may be 20%. Considering the RR of inhaled corticosteroids for reducing exacerbations (RR, 0.76; 95% CI, 0.72–0.80) and this baseline risk, one can derive a simplified absolute magnitude of the effect (19). Inhaled corticosteroids, relative to placebo, will reduce the absolute risk by approximately 4.8% (= 20% − [0.76 × 20%]). Some patients who are very averse to experiencing an exacerbation may consider the downsides of inhaled corticosteroids (thrush, fracture risk, burden of inhalers) well worth it. Given the relatively narrow CI that follows from the CI around the RRR, one could make a strong recommendation for using inhaled corticosteroids if all patients were equally adverse to exacerbations. More patients are, however, likely to consider the benefit not worth the harms and burden of taking inhalers if their baseline risk is lower. For instance, if the baseline risk for an exacerbation is 5%, the absolute risk reduction is only 1.2% (= 5% − [0.76 × 5%]) but the possible harms and burden remain unchanged. Fewer patients with lower baseline risk would make the choice of taking inhaled steroids. When, across the range of patient values, fully informed patients are liable to make different choices, guideline panels should offer weak recommendations and explain the rationale for their recommendation.

Comparison of GRADE and other systems

- Explicit definitions
- Explicit, sequential judgements
- Components of quality
- Overall quality
- Relative importance of outcomes
- Balance between health benefits and harms
- Balance between incremental health benefits and costs
- Consideration of equity
- Evidence profiles
- International collaboration
- Consistent judgements?
- Communication?

America doesn’t exist, I have been there



GRADE – value judgments/subjectivity vs transparency

BMC Health Services Research

BioMed Central

Research article

Open Access

Systems for grading the quality of evidence and the strength of recommendations II: Pilot study of a new system

David Atkins[1], Peter A Briss[2], Martin Eccles[3], Signe Flottorp[4], Gordon H Guyatt[5], Robin T Harbour[6], Suzanne Hill[7], Roman Jaeschke[8], Alessandro Liberati[9], Nicola Magrini[10], James Mason[3], Dianne O'Connell[11], Andrew D Oxman[4], Bob Phillips[12], Holger Schünemann[5,13], Tessa Tan-Torres Edejer[14], Gunn E Vist*[4], John W Williams Jr[15] and The GRADE Working Group

GRADE – value judgments/subjectivity vs transparency

Table 8: Results, summary of the recommendations made the 17 evaluators for each of the 12 examples in the pilot study

Example	Do it	Probably do it	Toss up	Probably don't do it	Don't do it	Consensus
1	4/16	7/16	3/16	-	2/16	Probably do it
2	6/16	8/16	2/16	-	-	Do it
3	-	6/15	7/15	2/15	-	Need more information
4	13/15	2/15	-	-	-	Do it
5	11/16	5/16	-	-	-	Do it
6	11/17	5/17	1/17	-	-	Do it
7	1/17	7/17	2/17	6/17	1/17	Probably do it
8	2/15	7/15	4/15	2/15	-	Do it
9	1/17	4/17	8/17	4/17	-	Probably don't do it/Tossup
10	-	2/17	6/17	7/17	2/17	No consensus
11	7/17	8/17	2/17	-	-	Do it
12	-	-	-	4/13	9/13	No consensus

Judgements about confidence in evidence and recommendations are complex. The GRADE system represents our current thinking about how to reduce errors and improve communication of these complex judgements.

BMC, 2004

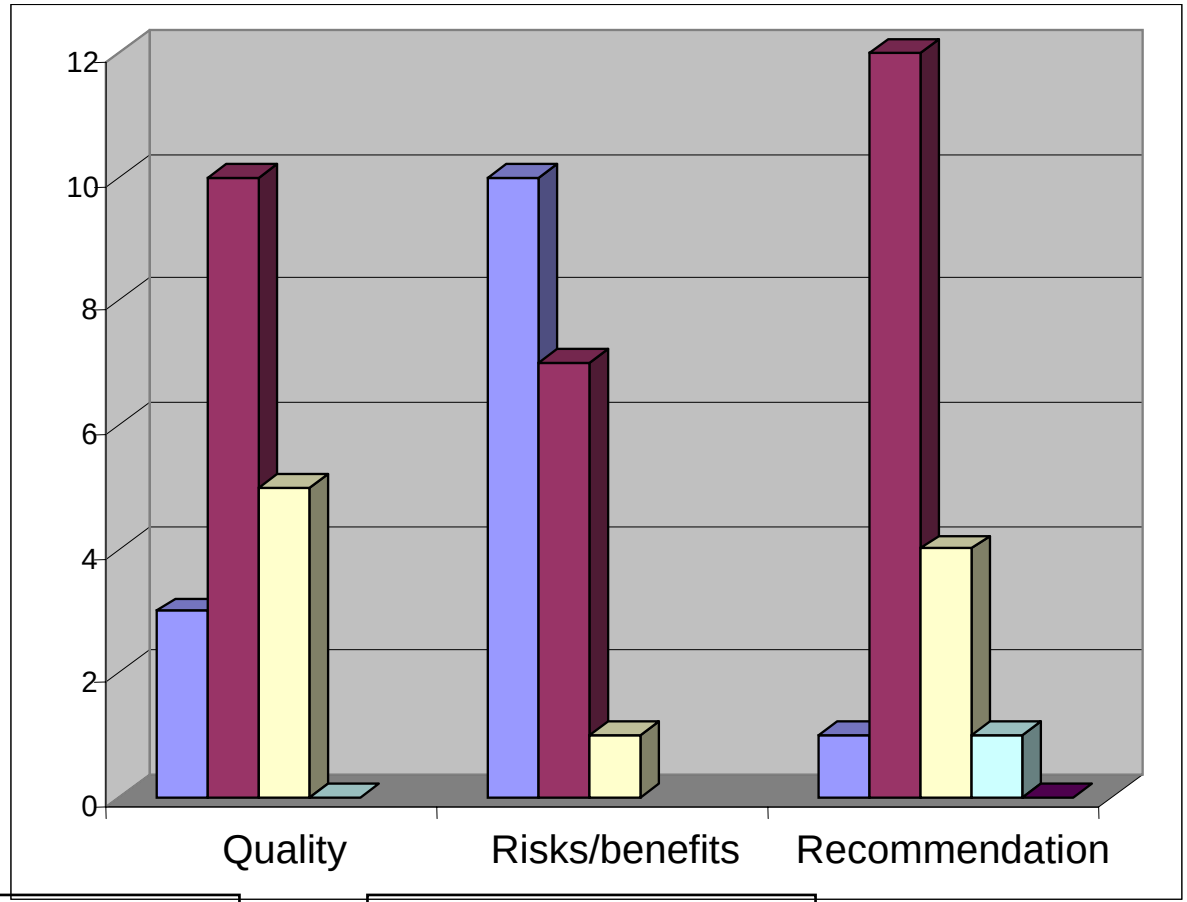
We will serve the public more responsibly and ethically when research designed to reduce the likelihood that we will be misled by bias and the play of chance has become an expected element of professional and policy making practice, not an optional add-on.

Iain Chalmers

GRADE: an example

Herceptin in breast cancer
(ASR-RER Italy, 2006)

In women with breast cancer C-erb+ is trastuzumab (herceptin) recommended as adjuvant therapy ?



quality

- high
- intermediate
- low
- Very low

risks/benefit

- Net benefits
- uncertain
- No benefits
- Very low

recommendation

- Strong +
- Weak +
- Not possible to recommend either way
- Weak -
- Strong -

Conclusions (I)

- GRADE is a highly structured approach
- GRADE allows the reader to understand the various stages of the evaluation process and is more explicit in comparison to other existing grading systems
- It is fairly demanding for the preparation of choice of relevant outcomes, evidence summaries / balance sheet
- It allows the panel to vote for the various dimensions to be considered (outcomes, quality of evidence, risk/benefit profile,

Conclusions (II)

- GRADE proved very useful for defining recommendations on the use of new drugs even in the presence of single or few RCT
- GRADE steps allowed an explicit discussion and proved a good guide for the panels for the evaluation of available evidence, risk-benefit profile and strength of recommendations
- New drugs have often limited long-term comparative data and GRADE approach

“To me, “knowledge” is about access and understanding. Action is then a choice (hopefully an informed one) for the individual.”

Chris Silagy, Australia 1999