GRADE
Judging the applicability and strength of evidence in Health Technology Assessments (HTAs)

Linda Long and Chris Hyde
STEPS in a Systematic Review

• Ask the question
• Get the stuff
  – SEARCH

• Check the stuff
  – APPRAISAL

• Synthesise the stuff
  – what does it mean?
  – meta-analysis (quantitative data)
  – meta-synthesis (qualitative data)

• Apply in the decision
Using evidence from systematic reviews

• The endpoint of a systematic review is:
  – A summary of the evidence
  – An estimate of effect for various outcomes

• We need to make a judgement on the strength of evidence across all studies for each important outcome

• Guideline developers can use this judgement to formulate their recommendation(s) and consider the direction (for or against) and strength (strong or weak) of their recommendation(s)
The GRADE approach is a transparent system for grading the quality of evidence in systematic reviews.

GRADE Working Group (www.gradeworkinggroup.org)
- International group of guideline developers, methodologists and clinicians from around the world (>100 contributors) – since 2000
What is GRADE?

• GRADE describes the extent to which you can be confident that an estimate of effect is near the true value for each outcome, across all studies.

• When judging the quality of a body of evidence, the assessment considers several factors, including *The Risk of Bias* of each included study.

• Other factors for assessing quality are:
  – imprecision, inconsistency, indirectness, or publication bias.
Hierarchy of evidence

- Sys Reviews-Metanalysis
- RCT's
- Cohort studies
- Case-Control
- Cross-sectional studies
- Case series, Case reports
- Ideas, opinions, editorials, anecdotal
Downgrade for:
Risk of bias
Inconsistency
Indirectness
Imprecision
Publication bias

Upgrade for:
Large consistent effect
Dose response
Confounders only reducing size of effect

High
Moderate
Low
Very low

STEP 1:
a priori ranking

STEP 2:
Upgrade/ downgrade

STEP 3:
Assign final grade – judgement on the quality of the body of evidence

STEP 4:
Consider factors affecting recommendation

STEP 5:
Make recommendation

Randomised controlled trial: HIGH
Observational study: LOW

Balance of desirable and undesirable effects
Cost-effectiveness
Preference of patients
Strong for using
Weak for using
Strong against using
Weak against using
Who uses GRADE?

• GRADE is becoming increasingly popular for guideline developers and systematic reviewers

• Used by:
  – Cochrane for use in systematic reviews
  – World Health Organization (WHO) guideline developers
  – Other guideline developers including NICE
Our experience of using GRADE: In HTA Systematic Reviews

• **GRADEpro** software is used to create Summary of Findings (SoF) tables

• Presents:
  – The quality of the evidence
  – The magnitude of the effect
  – Reasons behind decisions (records the judgements that are being made to evaluate the quality of the evidence)
### EPO HTA:

#### Summary of Findings table

<table>
<thead>
<tr>
<th>No. of participants (no. of studies)</th>
<th>Quality assessment</th>
<th>Overall quality of evidence</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESRAs vs pla</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td>Hb change (overall) (measured by change in Hb levels (g/dl) from baseline until the end of treatment period, better indicated by higher values)</td>
<td>Serious</td>
<td>Serious; significant heterogeneity ((I^2 = 75.9%); p &lt; 0.01)</td>
<td>No serious indirectness</td>
</tr>
<tr>
<td>3170 (16 trials reported in 26 papers([17,48,50,51,53,58–60,63–67,69–71,74,77–83,85,86])</td>
<td>Sample sizes, (n): control 1489, ESAs 1681; WMD 1.59 (95% CI 1.33 to 1.84)</td>
<td>The random-effects meta-analysis demonstrated a statistically significant difference in Hb change (increase from baseline) in favour of treatment</td>
<td></td>
</tr>
<tr>
<td>Meta-analysis: 18 trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haematological response (overall) (assessed by proportion of participants with an increase in Hb level of (\geq 2) g/dl or an increase in haematocrit of (\geq 6) percentage points, unrelated to transfusion)</td>
<td>Serious</td>
<td>No serious inconsistency</td>
<td>No serious indirectness</td>
</tr>
<tr>
<td>2228 (10 trials reported in 19 papers([17,48,50,51,53,58–60,63–67,71,74,77–83,85,86])</td>
<td>Study event rates, (n/N) (%): control 182/1015 (17.9), ESAs 759/1213 (62.6); RR 3.29 (95% CI 2.84 to 3.81)</td>
<td>The random-effects meta-analysis demonstrated a statistically significant difference in haematological response in favour of treatment</td>
<td></td>
</tr>
<tr>
<td>Meta-analysis: 12 trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBCT requirements (overall) (assessed by proportion of participants requiring RBCT)</td>
<td>Serious</td>
<td>No serious inconsistency</td>
<td>No serious indirectness</td>
</tr>
<tr>
<td>4779 (22 trials reported in 33 papers([17,48,50,51,53,58–60,63–67,71,73–86])</td>
<td>Study event rates, (n/N) (%): control 835/2299 (36.3), ESAs 554/2480 (22.3); RR 0.63 (95% CI 0.57 to 0.69)</td>
<td>The random-effects meta-analysis demonstrated a statistically significant difference in RBCT requirement in favour of treatment</td>
<td></td>
</tr>
<tr>
<td>Meta-analysis: 24 trials</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: 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 quality: We are very uncertain about the estimate.

Ref: Cochrane Handbook Chapter 12.2 Assessing the quality of a body of evidence
Tendinopathy HTA: Summary of Findings table

<table>
<thead>
<tr>
<th>Question</th>
<th>Should laser therapy vs. placebo be used for LET?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference</strong></td>
<td>Smidt et al.99</td>
</tr>
<tr>
<td><strong>Quality assessment</strong></td>
<td></td>
</tr>
<tr>
<td>Studies (n)</td>
<td>Design</td>
</tr>
<tr>
<td>Pain (0–6 weeks) (follow-up 3 weeks; measured with VAS)</td>
<td>I RCT</td>
</tr>
<tr>
<td>Pain (7 weeks) (follow-up 7 weeks; measured with VAS; range of scores = –0.27)</td>
<td>I RCT</td>
</tr>
<tr>
<td>Pain (13 weeks) (follow-up 13 weeks; measured with VAS)</td>
<td>I RCT</td>
</tr>
<tr>
<td>Function</td>
<td>NR</td>
</tr>
<tr>
<td>QoL</td>
<td>NR</td>
</tr>
<tr>
<td>Remained/return to work</td>
<td>NR</td>
</tr>
<tr>
<td>Sport activity</td>
<td>NR</td>
</tr>
<tr>
<td>Recurrence</td>
<td>NR</td>
</tr>
<tr>
<td>Adverse events</td>
<td>NR</td>
</tr>
</tbody>
</table>

*NR, not reported.

\(^a\) Low sample size and wide CIs.

\(^b\) Contradictory results for intermediate- and long-term follow-up assessment.
GRADE Classification of Evidence for “Reviews of Reviews”

- **High-quality evidence** One or more updated, high-quality systematic reviews based on at least:
  - one high-quality primary study
  - two primary studies of moderate quality with consistent results

- **Moderate-quality evidence** One or more updated systematic reviews of high or moderate quality based on
  - at least:
    - one high-quality primary study
    - two primary studies of moderate quality with consistent results

- **Low-quality evidence** One or more systematic reviews of variable quality based on:
  - primary studies of moderate quality
  - inconsistent results in the reviews
  - inconsistent results in primary studies

- **No evidence from systematic reviews** There is no systematic review identified on this topic
Our experience of using GRADE: In HTA Systematic Reviews

• Health Technology Assessments
  • What is the clinical effectiveness and cost-effectiveness of conservative interventions for tendinopathy? An overview of systematic reviews of clinical effectiveness and systematic review of economic evaluations
  • The effectiveness and cost-effectiveness of erythropoiesis-stimulating agents (epoetin and darbepoetin) for treating cancer treatment-induced anaemia

• NICE guidance
  • Not used routinely in technology appraisal (although we have used it)
  • Is used in NICE’s clinical guidelines

• Cochrane
  • Routinely used for effectiveness reviews
  • More contentious for test accuracy reviews
Concerns identified:
1. Some risks of bias in included studies.
2. Wide confidence intervals.
3. Unable to include studies described as showing "no difference" in meta-analysis. Heavy weighting towards very small studies with apparently very precise estimates of blood loss. Very marked heterogeneity.
4. Some heterogeneity.
5. Evidence of publication bias.
6. Unable to incorporate results of all studies on intracerebral haemorrhage where control of bleeding measured in different manner.
7. Small number of included studies.

Outcomes:
Deaths; no of patients with bleeding; transfusion requirements; & thromboembolic events
Down-grade factors

• Risk of bias
  – Cochrane risk for RCTs
• Inconsistency
  – Heterogeneity
• Indirectness
  – Tricky – see next slide
• Imprecision
  – 95% CI include no effect value
• Publication bias
  – Is it present
Indirectness of evidence

• Guideline developers face two types of indirectness of evidence. The first occurs when, for example, considering use of one of two active drugs. Although randomised comparisons of the drugs may be unavailable, randomised trials may have compared one of the drugs with placebo and the other with placebo. Such trials allow indirect comparisons of the magnitude of effect of both drugs. Such evidence is of lower quality than would be provided by head to head comparisons of the two drugs.

• Increasingly, recommendations must simultaneously tackle multiple interventions. For example, possible approaches to thrombolysis in myocardial infarction include streptokinase, alteplase, reteplase, and tenecteplase. Attempts to deal with multiple interventions inevitably involve indirect comparisons. A variety of recently developed statistical methods may help in generating estimates of the relative effectiveness of multiple interventions. Their confident application requires, in addition to evidence from indirect comparisons, substantial evidence from direct comparisons—evidence that is often unavailable.

• The second type of indirectness includes differences between the population, intervention, comparator to that intervention, and outcome of interest, and those included in the relevant studies.
Using GRADE in test accuracy

• GRADE has expanded from effectiveness to many other areas (www.gradeworkinggroup.org)

• Quite difficult to keep track

• Links to papers
  • [http://gradepro.org/guidelines-development#develop-publics](http://gradepro.org/guidelines-development#develop-publics)
  • J Clin Epi papers are better and more accessible
Challenges for diagnostics

• Accuracy as a starting point for conclusions on patient impact is problematic
• Limited applicability of standard down-grade factors
  • Risk of bias
  • Inconsistency
  • Indirectness
  • Imprecision
  • Publication bias
• Trying it out
CERqual (Confidence in the Evidence from Reviews of Qualitative Research)

• Assessing confidence for acceptability and feasibility

• Four components:
  – methodological limitations
  – relevance
  – coherence
  – adequacy of data
Discussion

• Is GRADE helpful?
• Should we be doing it more?
• Others in IHR who have experience who we may want to talk in future
  – Ruth Garside
  – Lyndsey Anderson
  – Someone from GRADE Working Group
Resources

- Webinars and online modules; http://cebgrade.mcmaster.ca/

- Online modules for GRADE criteria and Summary of Findings Tables
  A variety of online modules have been created to help GRADE the evidence in systematic reviews and create Summary of Findings Tables. Each module is approximately 15 to 20 minutes long and can be watched in any order. Topics include an introduction of GRADE, imprecision, risk of bias, publication bias.

- Introduction and overview of GRADE and Summary of Findings Tables
  This 40 minute webinar is a recording of an online webinar hosted by the Canadian Cochrane Network and Centre on 11 February 2010. It provides a general overview of how to interpret results of systematic reviews and draw conclusions using the GRADE approach, how to summarise and present those results in a Summary of Findings Table, and how to start with GRADE-pro to create Summary of Findings tables.

- GRADEing the evidence
  This 40 minute webinar is a recording of an online webinar hosted by the Canadian Cochrane Network and Centre on 11 February 2010. It explains the GRADE criteria used to evaluate the quality of evidence in a systematic review. It provides examples of each of the 5 main criteria: risk of bias, imprecision, inconsistency, indirectness, and publication bias, as well as 3 other criteria: magnitude of effect, dose response, confounding.

- How to create a Summary of Findings Table using GRADEpro
  This 40 minute webinar is a demonstration of how to use GRADE-pro to create a Summary of Findings Table. It explains and shows the step-by-step process from importing RevMan data into GRADE-pro, creating the table, and then importing a completed table back into RevMan. This is an online webinar which was recorded on 02 March 2010.
Additional materials

A series of articles freely available and published in the *Journal of Clinical Epidemiology* about GRADE and Summary of Findings Tables. Each article explains key issues in GRADE (e.g. choosing outcomes).

1. Introduction—GRADE evidence profiles and summary of findings tables
2. Framing the question and deciding on important outcomes
3. Rating the quality of evidence
4. Rating the quality of evidence—study limitations (risk of bias)