Grading of recommendations, assessment, development, and evaluations (GRADE)
By Dr Linda Long (Pen-TAG)
Understanding GRADE: An Introduction

Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) is currently emerging as the dominant method for appraising controlled studies and making recommendations for systematic reviews and guidelines

Used by: Cochran for use in systematic reviews
World Health Organization (WHO) guideline developers and other guideline developers
GRADE and the process from results to conclusions in systematic reviews

Cochrane Handbook 2002

More or less explicit approaches to grading the strength of evidence underlying a conclusion are available (CTFPHE 1979, Cook 1992, Gyorkos 1994 Guyatt 1995, US PSTF 1996), although there is no single approach that is universally accepted as being appropriate for the wide range of reviews included in the Cochrane Database of Systematic Reviews. A Collaborative Review Group (CRG) may elect to use a standard approach to grading the strength of evidence across its reviews. Over time, it may be possible for the Cochrane Collaboration as a whole to develop a more consistent and explicit approach to drawing conclusions about the overall strength of evidence for the main conclusions of a review. However, it is currently up to individual reviewers, in consultation with others in their CRG, to select an approach to summarising the strength of evidence that is appropriate for the question being reviewed.
What is GRADE

- GRADE is a method used by systematic reviewers and guideline developers to assess the quality of the evidence and decide whether to recommend an intervention.

- GRADE differs from other appraisal systems for three reasons:
  1. Because it separates quality of evidence and strength of recommendation.
  2. The quality of evidence is assessed for each outcome.
  3. Observational studies can be “upgraded” if they meet certain criteria.

Ensures:
- Systematic process
- Transparency
Using GRADE

The GRADE method involves **five** distinct steps:

- **STEP 1**
  Assign an a-priori ranking of “high” to randomized controlled trials and “low” to observational studies
  Randomized controlled trials are initially assigned a higher grade because they are usually less prone to bias than observational studies

- **STEP 2**
  “Downgrade” or “upgrade” initial ranking
  It is common for randomized controlled trials and observational studies to be downgraded because they suffer from identifiable bias. Also, observational studies can be upgraded when multiple high-quality studies show consistent results
Using GRADE

• **Reasons to “downgrade”**

• **Risk of bias**
  – Lack of clearly randomized allocation sequence
  – Lack of blinding
  – Lack of allocation concealment
  – Failure to adhere to intention-to-treat analysis
  – Trial is cut short
  – Large losses to follow-up

• **Inconsistency**
  When there is significant and unexplained variability in results from different trials
Using GRADE

• Reasons to “downgrade”

• Indirectness of evidence
  can refer to several things:
  – An indirect comparison of two drugs.
  – An indirect comparison of population, outcome or intervention

• Imprecision
  when wide confidence intervals mar the quality of the data

• Publication bias
  when studies with “negative” findings remain unpublished
Using GRADE

- **Reasons to “upgrade”**

- **Large effect**
  When the effect is so large that bias common to observational studies cannot possibly account for the result

- **Dose-response relationship**
  When the result is proportional to the degree of exposure

- **All plausible confounders would have reduced the treatment effect**
  When all possible confounders would only diminish the observed effect and it is thus likely that the actual effect is larger than the data suggests
Using GRADE

**STEP 3**

- Assign final grade for the quality of evidence as “high”, “moderate”, “low” or “very low” for all the critically important outcomes

<table>
<thead>
<tr>
<th>Final GRADE ranking</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>⊕⊕⊕⊕⊕ We are very confident that the effect of the study reflects the actual effect</td>
</tr>
<tr>
<td>Moderate</td>
<td>⊕⊕⊕ We are quite confident that the effect in the study is close to the true effect, but it is also possible it is substantially different</td>
</tr>
<tr>
<td>Low</td>
<td>⊕⊕ The true effect may differ significantly from the estimate</td>
</tr>
<tr>
<td>Very low</td>
<td>⊕ The true effect is likely to be substantially different from the estimated effect</td>
</tr>
</tbody>
</table>
Using GRADE

STEP 4

• Consider other factors that impact on the strength of recommendation for a course of action

• High-quality evidence does not always imply a strong recommendation. Recommendations must consider factors besides the quality of evidence

• First factor the balance between desirable and undesirable effects.

• Uncontroversial recommendation e.g. antibiotics

• Controversial recommendation: where the benefit to harm ratio is less clear. Patient values and preferences, as well as costs, need to be considered carefully
STEP 1: a priori ranking

Randomised controlled trial: HIGH
Observational study: LOW

STEP 2: Upgrade/ downgrade

Downgrade for:
- Risk of bias
- Inconsistency
- Indirectness
- Imprecision
- Publication bias

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

STEP 3: Assign final grade

High
Moderate
Low
Very low

STEP 4: consider factors affecting recommendation

Balance of desirable and undesirable effects
Cost-effectiveness
Preference of patients

STEP 5: make recommendation

Strong for using
Weak for using
Strong against using
Weak against using
Using GRADE
A practical overview

Basic steps:

• (1) Choose outcomes of interest
• (2) GRADE the evidence
• (3) Present statistical results
• (4) Use GRADE-pro to create Summary of Findings tables
What is the effectiveness and cost-effectiveness of conservative interventions for tendinopathy: an overview of systematic reviews of clinical effectiveness and systematic review of economic evaluations

Long L¹, Briscoe S¹, Cooper C¹, Hyde C¹, Crathorne L¹
¹ Peninsular Technology Assessment Group, Exeter University Medical School, Veysey Building, Salmon Pool Lane, Exeter EX2 4SG

Objectives

This systematic review aims to summarise the evidence concerning the clinical and cost effectiveness of conservative interventions for lateral elbow tendinopathy

Outcomes (pre-defined; maximum of 7): Pain, Function, Quality of Life, Remain / return to work, Sport activity, Recurrence, Adverse events
GRADE-pro

www.ims.cochrane.org/revman/gradepro
The Summary of Findings Table

• Is a summary of the key findings from the systematic review

• 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)
## Summary of findings table

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>No of Participants (studies) Follow up</th>
<th>Quality of the evidence (GRADE)</th>
<th>Relative effect (95% CI)</th>
<th>Overall results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain (short term) VAS (100 mm)</td>
<td>446 (3 studies) 4-6 w eeks</td>
<td>⊕⊕⊕⊕ MODERATE due to inconsistency</td>
<td></td>
<td>The mean pain (short term) in the intervention groups w as 9.42 lower (20.7 low er to 1.86 higher)</td>
</tr>
<tr>
<td>Pain (intermediate term) resisted wrist extension (Thomsen test)</td>
<td>455 (3 studies) 12 w eeks</td>
<td>⊕⊕⊕⊕ MODERATE due to inconsistency</td>
<td></td>
<td>The mean pain (intermediate term) in the intervention groups w as 9.04 lower (19.37 low er to 1.28 higher)</td>
</tr>
<tr>
<td>Function (intermediate term) Mean grip strength</td>
<td>448 (3 studies) 12 w eeks</td>
<td>⊕⊕⊕⊕ MODERATE due to inconsistency</td>
<td></td>
<td>The mean function (intermediate term) in the intervention groups w as 0.05 standard deviations higher (0.13 low er to 0.24 higher)</td>
</tr>
<tr>
<td>Quality of life</td>
<td>O/C NR</td>
<td>O/C NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remain/return to work</td>
<td>O/C NR</td>
<td>O/C NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sport activity</td>
<td>O/C NR</td>
<td>O/C NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recurrence</td>
<td>O/C NR</td>
<td>O/C NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events (mild)</td>
<td>60 (1 study) 5 w eeks</td>
<td>⊕⊕⊕⊕ MODERATE due to inconsistency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events (general)</td>
<td>542 (1 study) 52 w eeks</td>
<td>⊕⊕⊕⊕ MODERATE due to inconsistency</td>
<td>OR 4.3 (2.9 to 6.3)</td>
<td></td>
</tr>
</tbody>
</table>

*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; NR, not reported; O/C, outcome; OR, odds ratio; RR, risk ratio; VAS, visual analogue scale

### 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.

1. conflicting results for pain relief compared to other placebo controlled trials of ESWT; 2. No explanation w as provided; 3. Tingling during therapy (5 in placebo group), aching after therapy (1 in placebo group), soreness after therapy (4 in placebo group) and increased pain symptoms after therapy (3 in placebo group); 4. Conflicting results, w ith 4 other RCTs reporting no significant adverse events; 5. Significantly more side effects were reported in ESWT group. The most frequent side effects in ESWT group were transitory redening of the skin (21.1%), pain (4.5%) and small hematomas (3%). Migraine occurred in 4 participants and syncope in 3 participants following ESWT. 5 other RCTs reported adverse events in ESWT group including increased pain, localized redness, tingling, and nausea during treatment, and aching, soreness and increased pain symptoms after therapy. Treatment discontinuation due to nausea and pain (slight tremor) in treatment arm w as reported in 1 RCT. Other adverse events included localised swelling, bruising or petechiae (1 RCT). Most observed side effects resolved by final follow-up. 6. 4 RCTs reported no significant adverse events in any treatment groups.
**GRADE Evidence Profile**

**Question:** Should shock wave therapy (ESWT) vs placebo be used for lateral epicondylitis?


<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>No of patients</th>
<th>Effect</th>
<th>Quality</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of studies</td>
<td>Design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
</tr>
<tr>
<td>Pain (short term) (follow-up 4-6 weeks; measured with: VAS (100 mm); range of scores: 3.6--19; Better indicated by lower values)</td>
<td>3</td>
<td>randomised trials</td>
<td>no serious risk of bias</td>
<td>serious¹</td>
</tr>
<tr>
<td>Pain (intermediate term) (follow-up 12 weeks; measured with: resisted wrist extension (Thomsen test); Better indicated by lower values)</td>
<td>3</td>
<td>randomised trials</td>
<td>no serious risk of bias</td>
<td>serious¹</td>
</tr>
<tr>
<td>Function (intermediate term) (follow-up 12 weeks; measured with: Mean grip strength; Better indicated by lower values)</td>
<td>3</td>
<td>randomised trials</td>
<td>no serious risk of bias</td>
<td>serious²</td>
</tr>
<tr>
<td>Quality of life - not reported</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Remain/return to work - not reported</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sport activity - not reported</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
**“Review of reviews”: Quality of evidence across studies for the outcome**

<table>
<thead>
<tr>
<th>Level of Quality of Evidence*</th>
<th>Based on:</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-quality evidence</td>
<td>One or more updated, high-quality systematic reviews that are based on at least 2 high-quality primary studies with consistent results</td>
</tr>
</tbody>
</table>
| Moderate-quality evidence     | One or more updated systematic reviews of high or moderate quality  
• Based on at least 1 high-quality primary study  
• Based on at least 2 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  
• Based on inconsistent results in the reviews  
• Based on inconsistent results in primary studies |
| No evidence from systematic reviews | There is no systematic review identified on this topic |

* Based on principles from Grading of Recommendations Assessment, Development, and Evaluation (GRADE)
## Overall evidence summary

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Comparison</th>
<th>Results (combined)</th>
<th>Quality of evidence (based on GRADE principles)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESWT</td>
<td>PBO</td>
<td>No difference in pain</td>
<td>low</td>
<td>Evidence from one high quality review in need of updating. Inconsistent results in primary studies for pain and function.</td>
</tr>
<tr>
<td></td>
<td>GCI</td>
<td>No difference in function</td>
<td>low</td>
<td></td>
</tr>
<tr>
<td>Laser therapy</td>
<td>PBO</td>
<td>Unclear</td>
<td>low</td>
<td>Evidence from one high quality review in need of updating. Inconsistent results in primary studies for pain and function.</td>
</tr>
<tr>
<td></td>
<td>Other PT modalities</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Resources

The Cochrane Handbook

Information on how to create Summary of Findings Tables using the information from Cochrane systematic reviews and GRADEing the evidence.

**Chapter 11**: Presenting results and ‘Summary of findings’ tables

**Chapter 12**: Interpreting results and drawing conclusions

Webinars and online modules; [http://cebgrade.mcmaster.ca/](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.
Workshop materials and presentations about Summary of Findings Tables and the GRADE approach

Presentation 1: Background to Summary of Findings Tables and the GRADE approach
Presentation 2: An introduction to Summary of Findings Tables and GRADEpro and calculating and presenting results
Handout: Calculations in Summary of Findings Tables
Workshop exercise materials: Workshop for Developing Summary of Findings Tables from Cochrane Reviews

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)**

Other articles

*Guyatt et al.* What is “quality of evidence” and why is it important to clinicians? BMJ 2008; 336: 995-998