Medical Journalism

Shiraz University of Medical Science

Paramedical School

Journal Club

Presented by:
Dr. Leila Ghahramani

Fall 2019

Prepared by: Dr. Kimia Bazargan Lari

Outline for Journal Club Presentation

Contents

How to GRADE the quality of the evidence

What is GRADE?

The most prominent framework for evaluating the effectiveness of systematic reviews is GRADE (**Grading** of Recommendations, Assessment, Development and Evaluations). GRADE offers a system for rating quality of evidence in systematic reviews and guidelines and grading strength of recommendations in guidelines. GRADE is used to rate the certainty of evidence for a treatment efficacy from high to very low.

The GRADE system takes in two types of studies:

- Randomized Controlled Trials (RCTs)
- Observational studies (non RCT)

Getting started with GRADE - an overview of the process

Step 1- Selection of Outcome:

Select one of the outcomes that represent one of the key outcomes for decision-making irrespective of how many studies contribute data. Predefining the summary of findings table outcomes will help to guard against emphasizing results on the basis of the amount of evidence or the size and direction of the effect. This table lists all the data that influence the quality of evidence. The final result of GRADE has four levels of evidence: very low, low, moderate, and high which is expressed by ⊕, where the higher number of ⊕ is, the more valuable it is. Table1 is an example of evidence profile.

Table GRADE ratings and their interpretation

Symbol	Quality	Interpretation
⊕⊕⊕⊕	**High**	We are very confident that the true effect lies close to that of the estimate of the effect.
⊕⊕⊕O	**Moderate**	We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
⊕⊕OO	**Low**	Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
⊕OOO	**Very low**	We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

**Table taken from the GRADE Handbook

Step 2- Determining the quality of evidence based on the type of studies

For each outcome:

a. Identify whether the evidence for that outcome comes from:
    i. RCTs (where the rating starts at HIGH quality) or
    ii. Non-RCTs (where the rating starts at LOW quality).
b. Following the guidelines below, systematically work through each of the GRADE criteria, deciding whether to downgrade and/or upgrade the quality of the evidence and by how much. Upgrading decisions will depend on the nature of the study designs you are basing your assessments on.

Step 3- Downgrade the evidence:

At this stage, it is necessary to evaluate the effect of five quality factors for each outcome. These quality factors are:

- Risk of Bias
- Inconsistency
- Indirectness
- Imprecision
- Publication Bias

Now let's verify these five quality factors.

Reasons to downgrade the evidence

1- Risk of Bias: what are the limitations?

Generally speaking, your assessments of studies using the risk of bias tool forms the basis for identifying limitations at the outcome level. You can then determine how the risk of bias for each of the studies might influence the size, direction, consistency and precision of the overall effect.

How to GRADE for Risk of Bias

1. Select the outcome to be assessed. Summarising the risk of bias must be done by outcome, rather than by study. This is so that the risk of bias for a given result (or outcome), and therefore the confidence we can have in that result, can be determined.

2. Systematically assess the outcome against the following criteria (most are elements of the RCT risk of bias tool) for each of the studies that contribute to it to determine whether the quality of the evidence is affected:
    - Inadequate methods of sequence generation.
    - Lack of allocation concealment.
    - Lack of blinding of each of:
        - participants,
        - providers,
        - outcome assessors.

      The more subjective an outcome is, the more important effective blinding becomes. For example, symptom improvement is a more subjective outcome than mortality, and is therefore more likely to be biased if unblinded.
    - Loss to follow up.

      There is no simple rule of thumb on which to base judgements about this item. The seriousness of losses from a study must be judged based on both the numbers of participants lost and the reasons for these losses, looking particularly at whether these are unbalanced across the study groups.
    - Failure to follow intention to treat principles in analyses.
    - Selective outcome reporting of outcomes and/or analyses.
    - Other sources of bias such as stopping the trial for benefit, design specific issues relating to non-standard trial designs, such as cluster or crossover studies.

    Note: please also refer to the Risk of Bias section in the CCCG Data Extraction Template (available at: http://cccrg.cochrane.org/author-resources) for more information about making these decisions.

3. Based on the limitations of the studies, come to an overall judgement about whether to downgrade the evidence, and if so, by how much. Table 1 (below) can be used to help make this decision. The following principles may also help to make this judgement.
    - Consider how much each trial contributes to the estimated size of the effect. Usually the larger the trial, with more events, the greater the contribution to the overall effect size.
    - Be conservative when downgrading: you should be fairly confident that most of the information from studies has a substantial risk of bias before you downgrade, bearing in mind that those studies rated as at unclear risk of bias will have a substantial risk of bias associated with them.
    - Make sure you are transparent and clear about why you have reached the decisions that you have, especially if it was a close call.
    - If you have conducted a meta-analysis, consider conducting a sensitivity analysis with only low risk of bias studies. If the effect estimates are unchanged then you may be confident that the risk of bias of the studies does not alter the results.

Note: if the evidence comes from a single RCT, do not automatically downgrade it. You should carefully evaluate the RCT against the GRADE criteria, but it is possible for a very large, well-designed RCT to be rated as high quality. Therefore you need to consider the risk of bias for that one study in making your decision.

2- Inconsistency: how consistent are the results?

Heterogeneity refers to any kind of variation across studies. Statistical tests that suggest heterogeneity is present, and in systematic reviews different types of heterogeneity can occur:

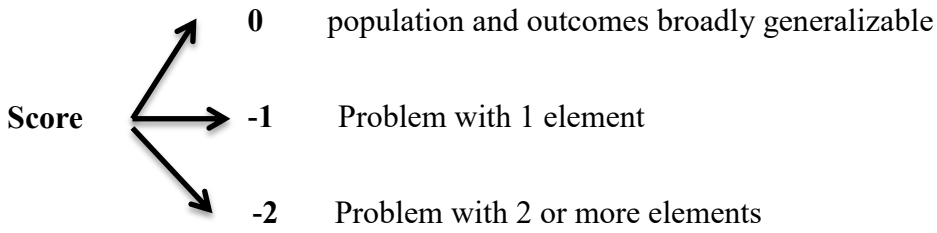
- Clinical heterogeneity: differences associated with the participants, interventions or outcomes.
- Methodological heterogeneity refers to differences in the way that studies were conducted – for example, differences in study design or risk of bias.

How to GRADE for inconsistency

1. Consider how much variability there is in the results of studies contributing to the outcome you are assessing.
   - For narrative data, consider whether there is a high degree of inconsistency in the results, such as effects in opposite directions (i.e. benefit and harm), or large variations in the degree to which the outcome is affected (i.e. very large and very small effects).
   - For meta-analysed data, consider whether:
     - there is wide variation in the effect estimates across studies
     - there is little or no overlap of confidence intervals associated with the effect estimates
     - statistical tests that suggest heterogeneity is present, for example:
       - Chi^2 test (testing the null hypothesis that the studies in the meta-analysis have the same underlying effect size) has a low p value
       - I^2 statistic (which quantifies the degree of variability between studies) is large -but please note the I^2 statistic is only one of several things to be considered when assessing heterogeneity, and the thresholds below are only a rough guide.
         As an approximate guide, an I^2 of:
         - 0% to 40% might not be important (low heterogeneity)
         - 30% to 60% might represent moderate heterogeneity
         - 50% to 90% might represent substantial heterogeneity
         - 75% to 100% might represent considerable heterogeneity.
     - whether any heterogeneity has been adequately explained.
2. Decide whether to downgrade on the basis of variability in the results:
   - not at all (inconsistency does not seem to be an issue);
   - one point (some inconsistency exists); or
   - two points (severe inconsistency is present).

3- Indirectness: how do these results apply to my review question?

Indirectness refers to how well the evidence included in the review answers the review question. It is based on the generalizability of population and outcomes from each study to our population of interest.



How to GRADE for indirectness

1. Consider again the question your review set out to address. Did the included studies provide broad answers to the question? Are there restrictions based on what was found, and that might affect applicability of the findings, in terms of:
   - population?
   - intervention?
   - comparator?
   - outcomes?

2. Decide whether the evidence that was found is more restrictive than the review question. If so, then the evidence may not directly answer the review question and you may downgrade for indirectness:
   - not at all (indirectness does not appear to be an issue)
   - one point (some indirectness exists), or
   - two points (indirectness is severe, or there is indirectness from several sources).

When considering the degree of indirectness, bear in mind that these judgements are often not clear cut, and not simply additive. A problem with indirectness of outcomes will often trigger downgrading, but all judgements need careful consideration.

4- Imprecision: how precise is the effect size?

Results are imprecise when studies include only relatively few patients or for dichotomous outcomes, there are few events, or when there is a lot of variation in the effects among the participants in continuous measures. As a result, there may be wide confidence intervals (CIs) around the effect estimate. When assessing imprecision, you should look at two things in particular:

- The number of people analyzed: is there enough information to detect a precise estimate of the effect?

- The CI around the effect estimate: does the CI (i.e. the range of values that the effect estimate might take) include meaningful benefit and harm, or a meaningful effect and no effect (consistent or inconsistent effects)?

How to GRADE for imprecision

1. Assess whether there is enough information (large enough sample size, or large enough number of events) to calculate a precise effect estimate.

    a. For dichotomous outcomes, unless events rates are very low (see also point 2 below) information is likely to be insufficient if:
        - total number of events is less than 300 (a "rule of thumb")
        - total (cumulative) sample size is lower than the calculated optimal information size: i.e. if the total number of participants in the review is less than the number of participants required for a single adequately powered trial

    b. For continuous outcomes information is likely to be insufficient if:
        - total number of participants is less than 400 (a "rule of thumb").

2. Look at the precision of the effect estimate.

    - Do the upper and lower limits include both meaningful benefit and harm (consistent or inconsistent messages) about the effect of the intervention? If the limits of confidence intervals represented the true effect, would they give the same message about the intervention, or not (e.g. does one end indicate a meaningful benefit, and the other no effect or even a harm)?
    - Does the 95% CI (or alternative estimate of precision) around the pooled or best estimate of effect include both little or no effect **and** appreciable benefit or appreciable harm?
        - For dichotomous outcomes, GRADE suggests that the threshold for 'appreciable benefit' or 'appreciable harm' that warrants downgrading is a relative risk reduction (RRR) or relative risk increase (RRI) greater than 25%. Do not overlook absolute effects since for rare events wide confidence intervals of a risk ratio can correspond to small differences in absolute terms.
        - For continuous outcomes, GRADE suggests that the thresholds are the minimal important difference (MID), either for benefit or harm. If the MID is not known, we suggest downgrading if the upper or lower confidence limit crosses the effect size (eg SMD) of 0.5 in either direction.

3. Decide whether there is imprecision in the results, based on your assessments of points 1 and 2 above, and if so, to what extent. Make a decision about whether to downgrade:
    - not at all (imprecision does not appear to be an issue)
    - one point (some imprecision exists), or
    - two points (very serious imprecision exists).

5- Publication Bias: are these all of the relevant studies?

Publication bias is a systematic under or over estimation of the underlying beneficial or harmful effect of the intervention, due to the selective publication of studies or availability of their data.

How to GRADE for Publication bias

1. Consider the size of the included studies (and number of events they include). If all results come from small studies, publication bias may be present.
2. Consider constructing a funnel plot, which graphs precision against the size of the effect. If the plot is asymmetrical (skewed) then publication bias may be present. Note, however, that asymmetry of the plot does not always indicate publication bias.
3. As it is difficult to entirely rule out the presence of publication bias, and ways of assessing it are uncertain, the GRADE recommendation is to only downgrade one level at a maximum (not two) on the basis of suspected publication bias. If publication bias is:
    a. undetected, do not downgrade
    b. strongly suspected, downgrade one level.

Reasons to upgrade the evidence

There are three major possible reasons to upgrade the quality of evidence.

Large magnitude of effect

- When methodologically strong observational studies yield large or very large and consistent estimates of the magnitude of an intervention effect, we may be more confident about the results.
- In situations like this, even though the study design is weak and is likely to overestimate the effects of the intervention, it is unlikely to explain all of the apparent benefit or harm.

Dose Response

- The presence of a dose-response gradient (relationship) may increase our confidence in the findings of observational studies and thereby increase the quality of evidence.

Effect of all plausible confounding factors

- On occasion, all plausible confounding from observational studies or randomized trials may be working against the direction of the observed effect, either to:
    - reduce the effect seen, or
    - increase the effect if no effect was observed.

How to upgrade the quality of the evidence:

Consider whether each of the following apply to your outcome:

1. Is there a large magnitude of effect?
   - Large RR >2 or <0.5 (based on consistent evidence from at least 2 studies, with no plausible confounders): upgrade 1 level
   - Very large RR >5 or <0.2 (based on direct evidence with no major threats to validity): upgrade 2 levels
2. Is there a dose-response gradient in the findings?
3. Have all plausible confounding factors been accounted for?
   - Did you find an effect, even though all of the confounding you can think of would have reduced the effect size?
   - Did you fail to find an effect, even though all of the confounding you can think of would have increased the effect size?

Note: only observational studies with no major threats to validity (i.e. that have not been downgraded) can be upgraded.

Incorporating GRADE into the review

1- Describing the methods used for assessing the quality of the evidence
2- Integrating the ratings of quality with the reporting of results: having used GRADE to assess the quality of the evidence, these ratings need to be reported together with the results.
3- Including information about the quality of the evidence in the Discussion section
4- Brief references to the ratings should also be included in the descriptions of your findings in all summary versions of the results, including the Conclusions, Abstract, and Plain Language Summary.

Table 2 can be used to work through a GRADE assessment for each outcome.

Table 2 -Table for assessing the GRADE criteria

GRADE criteria	Rating (circle one)	Footnotes (explain reasons for down- or upgrading)	Quality of the evidence (Circle one)
Outcome:			
Study design	RCT (starts as high quality) Non-RCT (starts as low quality)		⊕⊕⊕⊕ High ⊕⊕⊕◯ Moderate ⊕⊕◯◯ Low ⊕◯◯◯ Very Low
Risk of Bias *(use the Cochrane Risk of Bias tables and figures)*	No serious (-1) very serious (-2)		
Inconsistency	No serious (-1) very serious (-2)		
Indirectness	No serious (-1) very serious (-2)		
Imprecision	No serious (-1) very serious (-2)		
Publication Bias	Undetected Strongly suspected (-1)		
Other (upgrading factors, circle all that apply)	Large effect (+1 or +2) Dose response (+1 or +2) No Plausible confounding (+1 or +2)		

Reference

Ryan, R., & Hill, S. (2019).” How to GRADE the quality of the evidence.” Cochrane Consumers and Communication Group. 2016.