

Incorporating GRADE in Cochrane Reviews: Feedback from the CEU screening programme

Toby Lasserson, Nancy Santesso, Miranda Cumpston, Rachel Marshall & Orla Ní Ógáin

Assessing the quality of the evidence is an integral part of undertaking a Cochrane Review. GRADE is an established method to help authors rate the quality of evidence and to communicate the key results of systematic reviews to users. It is a mandatory MECIR conduct standard to base the interpretation of the evidence on the five GRADE considerations (C76): risk of bias, imprecision, inconsistency, indirectness and publication bias.

Since the start of the CEU review screening programme in September 2013 we have been able to see how GRADE is used to assess and communicate the quality of a body of evidence. We also recognise some important challenges Cochrane Review authors and editorial groups have encountered with implementing GRADE.

As a result of this experience we are sharing our thoughts on some of the reviews which highlight four key aspects of bringing GRADE methods and ratings into the text of Cochrane Reviews:

1. Describing methods for assessing the quality of the evidence under the '*Data collection & analysis*' section of protocols and full reviews.
2. Explaining decisions about the quality of the evidence in reporting of results.
3. Incorporating information about the quality of evidence in the Discussion.
4. Drawing on quality of evidence ratings when summarising and interpreting the results e.g. abstracts, plain language summaries and implications for practice sections.

We recognise that not all of the following examples will represent a 'Gold Standard'. However, they are undoubtedly good enough to illustrate the general approaches that we believe will help to integrate GRADE into Cochrane Reviews.

1. Describing methods for assessing quality of the evidence

There is often only limited information presented in reviews about the implementation of GRADE. Given that GRADE is a method, it should be acknowledged as such under '*Data collection & analysis*'. Methods for rating the quality of evidence should be considered as early as possible in the review process, ideally at the protocol stage. However, even if GRADE has been adopted post-protocol it is useful to know how this method has been applied to rate the quality of evidence. The following examples present information relevant to the implementation of GRADE methods and the selection

of outcomes that are used in a Summary of Findings table (although note that The Cochrane Handbook recommends that the main outcomes for Summary of Findings table should generally be included under Types of outcome measures).

Good practice examples

'For assessments of the overall quality of evidence for each outcome that included pooled data from RCTs only, we downgraded the evidence from 'high quality' by one level for serious (or by two for very serious) study limitations (risk of bias), indirectness of evidence, serious inconsistency, imprecision of effect estimates or potential publication bias. Data from observational studies started at low quality.'

van Ginneken N, Tharyan P, Lewin S, Rao GN, Meera SM, Pian J, et al. Non-specialist health worker interventions for the care of mental, neurological and substance-abuse disorders in low- and middle-income countries. Cochrane Database of Systematic Reviews 2013, Issue 11.

<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009149.pub2/full>

'The GRADE approach was employed to interpret findings (Langendam 2013) and the GRADE profiler (GRADEPRO) allowed us to import data from Review Manager 5.2 (Review Manager) to create 'Summary of findings' tables. These tables provide outcome-specific information concerning the overall quality of evidence from studies included in the comparison, the magnitude of effect of the interventions examined, and the sum of available data on the outcomes we considered.

The following outcomes were included in the 'Summary of findings' tables.

Failure to respond at endpoint (six to 12 weeks).

Failure to respond at one to four weeks.

Failure to respond at 16 to 24 weeks.

Failure to remit at endpoint.

SMD at endpoint.

Failure to complete - any cause.

Participants with at least some SE.'

Purgato M, Papola D, Gastaldon C, Trespidi C, Magni LR, Rizzo C, et al. Paroxetine versus other anti-depressive agents for depression. Cochrane Database of Systematic Reviews 2014, Issue 4. Art. No.: CD006531. DOI: 10.1002/14651858.CD006531.pub2.

<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD006531.pub2/full>

2. Explaining decisions about the quality of the evidence in reporting results

Having implemented GRADE, authors will need to consider where to describe their decisions for rating the quality of evidence. They should look to integrate this information when presenting

results. One approach is to include the quality rating with the outcome results under 'Effects of interventions'.

Good practice example

'Health-related quality of life

Improvement on the St George's Respiratory Questionnaire (SGRQ) was greater with LABA therapy than with placebo (MD -2.32, 95% CI -3.09 to -1.54; I² = 50%, P = 0.007; Analysis 1.1), based on data from 11,397 people in 17 studies. Results were analysed using a random-effects model because heterogeneity was high, and the outcome was downgraded from high to moderate quality for this reason.'

Kew KM, Mavergames C, Walters JAE. Long-acting beta₂-agonists for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews 2013, Issue 10. Art. No.: CD010177. DOI: 10.1002/14651858.CD010177.pub2.

<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010177.pub2/full>

In addition to descriptive text in the 'Effects of intervention' section, SoF tables will need to include footnotes to explain any decisions to downgrade the quality of the evidence. When checking footnotes it is worth seeing how easy it is to identify how many levels the quality of evidence has been downgraded by, along with the considerations that were a factor in the decision (i.e. risk of bias, imprecision, indirectness, inconsistency or publication bias). Users of the review should be able to see how the authors have rated the quality of the evidence for the outcomes of interest so transparency is crucial.

It can sometimes be useful to justify decisions not to downgrade the quality of the evidence. For example, authors may decide not to downgrade for imprecision if the confidence interval for the relative effect translates to clinically small differences in absolute effects. Alternatively there may be evidence of statistical heterogeneity, but the direction of the effect is consistent across the studies. Footnotes are a useful way to record such decisions, and along with other information presented on downgrading, they will help to inform the development of the discussion and the overall interpretation of the review findings.

3. Incorporating information about the quality of evidence in the Discussion

It is common to read general statements about the overall risk of bias in the included studies under the discussion. However, rating the quality of the evidence should focus not just on the risk of bias, but also how imprecision, inconsistency, indirectness and publication bias also impact on the

credibility of the results. These considerations and the thinking behind any downgrading decisions for the GRADE ratings can be summarised and incorporated into the discussion. The optional subheading 'Quality of the evidence' is an ideal place to include this information.

Good practice examples

'Quality of the evidence

...we found the quality of evidence for most outcomes to be of moderate or low quality primarily due to risk of bias and imprecise results because of few fracture events. It could be argued that evidence for hip fractures in the community, and for pelvic fractures, could be assessed as higher quality since the incidence of events is very low and the confidence intervals narrow enough that additional research would not be required. However, the unexplained heterogeneity across studies for pelvic fractures (in particular due to the O'Halloran 2004 study) warrants additional research to determine the effects of hip protectors on pelvic fractures and the evidence was therefore assessed as low quality.'

Santesso N, Carrasco-Labra A, Brignardello-Petersen R. Hip protectors for preventing hip fractures in older people. Cochrane Database of Systematic Reviews 2014, Issue 3. Art. No.: CD001255. DOI: 10.1002/14651858.CD001255.pub5.

<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001255.pub5/full>

'Quality of the evidence

The quality of findings ranks from moderate to low across the different outcomes. The main limiting factor, which was the reason for a decrease in quality in some outcomes, was the inconsistency of results across the small number of included studies. With only three studies included, it is important to acknowledge the large potential impact if the average effect of one study differs in size or direction.'

McGregor AH, Probyn K, Cro S, Doré CJ, Burton AK, Balagué F, Pincus T, Fairbank J. Rehabilitation following surgery for lumbar spinal stenosis. Cochrane Database of Systematic Reviews 2013, Issue 12. Art. No.: CD009644. DOI: 10.1002/14651858.CD009644.pub2.

<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009644.pub2/full>

In both of these reviews the authors have summarised common reasons for downgrading the quality of the evidence, rather than simply recount each downgrading decision for the relevant outcomes.

This next example illustrates how GRADE can still be implemented and reported without always being part of a SoF table. The review included one study:

'Quality of the evidence

The individual outcomes we examined were all downgraded one level to reflect the fact that Wathen 2007 was subject to a high risk of bias due to lack of blinding. (...) Since the imprecision of the results also lowers the quality of the evidence, we downgraded a further evidence level on that basis, so overall we judged the evidence to be of low quality, which means that 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.'

Black KJL, Bevan CA, Murphy NG, Howard JJ. Nerve blocks for initial pain management of femoral fractures in children. Cochrane Database of Systematic Reviews 2013, Issue 12. Art. No.: CD009587. DOI: 10.1002/14651858.CD009587.pub2.
<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009587.pub2/full>

In other situations the quality of evidence may be more variable across the outcomes, or the downgrading decisions are more specific to individual outcomes. An appropriately detailed approach to describing the quality of the evidence may sometimes be preferable to the briefer summaries shown above.

4. Using GRADE in abstracts, plain language summaries and implications for practice

As a general rule it is important to consider the outputs of the GRADE process wherever the findings of the review are being described, summarised or interpreted. Explicitly drawing on the quality of the evidence in the summary versions of the review and in the conclusions helps to avoid inconsistent reporting of results or conclusions. Integrating the quality ratings with the results also reinforces that the quality of the evidence is a key finding of the review. The following examples show how this might be approached in the abstract and PLS.

Abstract Results

Antiretroviral therapy started at a hospital and maintained at a health centre (partial decentralisation) probably reduces attrition (RR 0.46, 95% CI 0.29 to 0.71, 4 studies, 39 090 patients, moderate quality evidence).

PLS

We found that if antiretroviral therapy was started at a hospital and continued in a health centre (partial decentralisation), there was probably less attrition and fewer patients were lost to care after one year (four studies, 39 090 patients).'

Kredo T, Ford N, Adeniyi FB, Garner P. Decentralising HIV treatment in lower- and middle-income countries. Cochrane Database of Systematic Reviews 2013, Issue 6. Art. No.: CD009987. DOI: 10.1002/14651858.CD009987.pub2.
<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009987.pub2/abstract>

Abstract

There is moderate quality evidence when pooling data from five trials in the community (5614 participants) that shows little or no effect in hip fracture risk (RR 1.15, 95% CI 0.84 to 1.58); the absolute effect is two more people (95% CI 2 fewer to 6 more) per 1000 people having a hip fracture when provided with hip protectors.

PLS

Overall, there was moderate quality evidence from these studies for the following results.

In older people living in nursing care facilities, providing a hip protector probably decreases the chance of a hip fracture slightly'

Santesso N, Carrasco-Labra A, Brignardello-Petersen R. Hip protectors for preventing hip fractures in older people. Cochrane Database of Systematic Reviews 2014, Issue 3. Art. No.: CD001255. DOI: 10.1002/14651858.CD001255.pub5.

<http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001255.pub5/abstract>

Presenting results in conjunction with the quality of the evidence focusses attention on the authors' confidence in the estimated effect, rather than just whether the results are statistically significant. A similar approach can help when formulating implications for practice. Consider the following statements:

1. *'Treatment with [intervention] leads to statistically significant reductions in [outcome].'*
2. *'There is low quality evidence that [intervention] reduces [outcome].'*

In the first statement, the emphasis on statistical significance fails to convey the quality of the evidence. Replacing 'statistically significant' with the quality rating helps to convey the authors' confidence in the effect more clearly.

Another common occurrence is to see conclusions which confuse evidence of poor quality with no evidence:

1. *'There is no evidence to determine whether [intervention] has any role in the management of [condition]'*
2. *'Due to the very low quality of the evidence for our main outcomes, the effects of [intervention] in the management of [condition] are uncertain.'*

It would only strictly be accurate to state that there is no evidence when a review is empty. The second statement emphasizes the quality of the evidence and as such is a more accurate representation of the uncertainty arising from the very low quality of the evidence, rather than implying that there it is absent.

Further Reading

This document does not capture all of the detailed aspects of implementing GRADE in a Cochrane Review. More information about GRADE methods and considerations for preparing Summary of

Findings tables (including information on presenting continuous data, upgrading quality evidence, and selecting control group risks) can be found among the following resources:

Cochrane Handbook Chapter 11 (Presenting results and 'Summary of findings' tables):

http://handbook.cochrane.org/chapter_11/11_presenting_results_and_summary_of_findings_tables.htm

A list of publications introducing GRADE: <http://www.gradeworkinggroup.org/publications/>

The 2011 series of articles about GRADE in the Journal of Clinical Epidemiology (free access):

http://www.gradeworkinggroup.org/publications/JCE_series.htm

Schedule of webinars and workshops: <http://www.gradeworkinggroup.org/news.htm>

Software for generating GRADE Evidence Profiles and Summary of Findings tables:

<http://tech.cochrane.org/revman/gradepr>