Audit of planned methods for using GRADE and preparing Summary of Findings tables in protocols of Cochrane Reviews: Target 1.3 for 2015

*Newton Opiyo, Liz Bickerdike, Toby Lasserson*

# Summary points

* We compared planned implementation of methods for GRADE and Summary of Findings (SoF) tables in protocols published from two issues of The Cochrane Library in 2013 and 2015.
* The number of protocols describing the intention to use GRADE increased from 35% to 91% between the two cohorts.
* Recent protocols show improvement in the specification of GRADE considerations, methods for preparing SoF tables and SoF table outcomes.
* Cochrane author teams and editors are encouraged to make greater efforts to prioritise comparisons for SoF tables in protocols and to undertake GRADE assessments on a consensual basis.
* We encourage piloting the submission of SoF tables and/or GRADE evidence profiles for evaluation before write-up of the main results and the summary versions to ensure optimal implementation of GRADE in full reviews.

# Background

Goal 1 of the Strategy to 2020 commits Cochrane to produce high-quality systematic reviews. To support this long term goal, the 2015 target for review quality comprised an audit of two cohorts of published protocols. This was to evaluate progress made in planning the assessment of the quality of evidence using GRADE and preparation of Summary of Findings tables.

Since the initiation of pre–publication screening of new reviews by a team of editors within the CEU, we have seen improvements in the presentation of Summary of Findings tables in reviews.1 However, we continue to see variation in the implementation of GRADE and its integration in interpreting evidence in the body the review and summary versions. This may in part be attributable to variability in the way that GRADE is planned in review protocols.

Methods for using GRADE and preparing SoF tables should be considered early on during the planning of review methods in the protocol. The development of the review protocol represents an opportunity to elaborate on how GRADE will be used to summarise and interpret the review findings.

The aim of this audit was to evaluate how GRADE and SoF table methods are included in protocols of Cochrane Reviews. We were interested in assessing progress in relation to planning of GRADE and SoF table methods since the initiation of CEU pre-publication review screening in September 2013.

# Methods

We identified 40 protocols for Cochrane intervention reviews published in the *Cochrane Database of Systematic Reviews* in August 2013 and 33 protocols published in August 2015. We developed an audit tool based on the following eight items from MECIR conduct and reporting standards that we considered relevant for planning GRADE and SoF table methods in protocols of Cochrane reviews. The questions were pilot-tested on five protocols and further refined for relevance following discussions among a team of CEU editors. The audit tool is available in Appendix 1.

1. Reference to GRADE as a method for assessing quality of evidence
2. Description of GRADE considerations for assessing quality of evidence
3. Description of GRADE levels of evidence
4. Whether methods for preparing SoF tables were specified
5. Consideration for comparisons to be covered in the SoF table
6. Whether outcomes to be included in the SoF table were specified
7. Whether the number of reviewers to be involved in GRADE assessment was specified
8. Whether GRADE and SoF table methods were specified in an appropriate section (heading) in the protocol

Two CEU editors (NO and LB) independently read the protocols and rated each of the eight items as having been met (e.g. clear description of GRADE methods), not met, partially met or not applicable. Data analysis involved a comparison of the proportion of protocols meeting the specified audit standards across the two cohorts and a detailed iterative exploration to identify important issues related to implementation of GRADE and SoF table methods in Cochrane reviews.

# Results

We included 73 protocols from 28 different Cochrane review groups in the audit. The characteristics of the two cohorts are summarised below.

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | 2013 | 2015 | Total |
| Protocols (N) | 40 | 33 | 73 |
| N Cochrane Review Groups | 25 | 21 | 28 |
| N Protocols per group (median, range) | 1 (1 to 5) | 2 (1 to 5) | 2 (1 to 10) |
| N GRADE cited (%) | 14 (35%) | 30 (91%) | 44 (60%) |

The number of protocols that reference GRADE increased from 35% to 91% between the two cohorts. In addition, a higher proportion of GRADE and SoF table audit items were met by protocols published in August 2015 (59.3%, 147 out of 248 items) compared to those published in August 2013 (22.4%, 67 out of 299 items). We judged that none of the protocols from both cohorts fully met all eight audit standards. An overview of the audit findings is presented in Figure 1.

Cont./



## 

## 1. GRADE methods

A greater proportion of protocols in 2015 referenced GRADE compared with those from 2013. Reference to GRADE methods for assessing quality of evidence was made in 91% (n=30) of the protocols published in 2015 compared to 35% (n=14) of protocols published in 2013. Most of these involved citation of Section 8.5 and Chapter 12 of the Cochrane Handbook.2 However, in some protocols reference was made to specific GRADE publications.3,4

## 2. GRADE quality criteria

There were more protocols defining GRADE criteria in protocols from 2015 compared with 2013. A description of GRADE considerations for assessing quality of evidence was provided in 67% (n=22) protocols published in 2015 compared to 18% (n=7) of protocols published in 2013. This often involved reporting of the four criteria for downgrading quality of evidence (risk of bias, inconsistency, indirectness, imprecision, publication bias). Overall, only three protocols reported specific considerations related to upgrading quality of evidence (large effect, dose-response gradient, plausible confounding effect). These latter considerations were however only considered relevant where authors planned to include observational studies in the review.

## 3. GRADE levels of evidence

A description of the four categories of quality of evidence (high, moderate, low, very low) was provided in 24% (n=8) of protocols published in 2015 compared to 13% (n=5) of protocols published in 2013. However, an additional category “No evidence” (defined as “no RCTs were identified that addressed this outcome”) was added in three protocols.

## 4. Summary of Findings table methods

The methods for preparing SoF tables were more frequently presented in 2015 protocols. Seventy six percent (n=25) of protocols published in 2015 specified the methods for preparing SoF tables, compared to 33% (n=13) of protocols published in 2013. Most protocols planned to use Cochrane methods (described in Section 8.5 and Chapter 12 of the Cochrane Handbook) and GRADEpro software.5

Authors planned to use “Checklist to aid consistency and reproducibility of GRADE assessments” (Appendix 2) and attach “evidence profile tables” in two protocols. Presentation of results using a narrative SoF table format, drawing on the example from one published review6, was mentioned in two protocols. Template text for GRADE and SoF table methods was used in two protocols.

## 5. Comparisons covered in Summary of Findings tables

We considered this criteria to be relevant only where authors planned to include at least one comparison in the review. These were handled similarly between the two cohorts. Details about comparisons to be covered in the SoF tables were specified in 4 out of 17 eligible protocols published in 2015 and 3 out of 19 eligible protocols published in 2013. Most protocols where this criteria was met reported preparation of SoF tables for all the comparisons specified in the ‘Types of interventions’ section of the protocol.

## 6. Summary of Findings table outcomes

The audit showed better prioritization of outcomes for SoF tables in more recent protocols. The outcomes to be considered in the Summary of Findings table were specified in 82% (n=27) of protocols published in 2015, compared to 25% (n=10) of protocols published in 2013. Most protocols where this was done specified seven or fewer outcomes, including both harms and benefits.

## 7. Number of reviewers involved in GRADE assessment

Very few protocols reported on the number of raters involved in assessing the quality of evidence. Three protocols published in 2015 reported that two reviewers will independently undertake planned GRADE quality assessments. However, only one protocol in the 2013 cohort mentioned independent GRADE quality rating.

## 8. GRADE and Summary of Findings table heading (section in protocol)

Eighty five percent (n=28) of the protocols published in 2015 described GRADE and SoF table methods in an appropriate section (heading) compared to 20% (n=14) of protocols published in 2013. Most protocols where this was met considered GRADE and SoF table methods in the Data synthesis section or under separate sub-headings (e.g. ‘Assessment of the quality of evidence’, ‘Summary of Findings tables’). The protocols where this criteria was not met addressed GRADE and SoF table methods under the sections ‘Types of outcomes’ and ‘Sensitivity analyses’.

# Discussion

Our audit findings show that GRADE has become a more routine part of protocol development. We believe that there are also areas where greater prespecification would help to optimise implementation of GRADE and development of Summary of Findings tables.

We observed improvements in the following aspects of GRADE and SoF tables: reference to GRADE methods, description of GRADE quality criteria, definition of methods for preparing SoF tables and description of outcomes to be included in SoF tables. GRADE and SoF tables were also considered under appropriate sections in most protocols published in 2015. However, GRADE levels of evidence, comparisons to be covered in SoF tables and specification of reviewers to be involved in GRADE assessments were not well addressed in either cohort.

Examples of protocols that illustrate good practices in planning of different aspects of GRADE and SoF tables are presented in Appendix 3.

### Specifying comparisons and outcomes

Selection of outcomes for Summary of Findings tables at the protocol stage is now more common. Prespecifying outcomes (methods of measurement, time-points of reporting) helps avoid bias in the choice of outcomes7,8, and serves as a reminder to review teams to include results for the most relevant outcomes in SoF tables irrespective of the amount and quality of evidence. However, some modifications may be justified; for example, as noted in one protocol “the importance of an outcome may only become known after the protocol was written and analysis carried out”. This may be the case for adverse events. Changes to outcomes may also be necessitated by peer reviewers. In these situations authors should explain and appropriately document any change in the specified outcomes (e.g. under “Differences between protocol and review” in the full review).

Specification of comparisons to be covered in SoF tables remains an area where improvement is needed. As with biases associated with post-protocol modification or switching of outcomes, preferential reporting of results for a subset of review comparisons may introduce bias. Thirty six protocols included in this audit planned multiple comparisons, but only seven of them mentioned the comparisons to be covered in the SoF tables. Thus, it appears most reviewers are less aware of the relation between comparisons addressed in the review and SoF tables. Selective reporting of comparisons may introduce bias, for example, if reviewers only present comparisons that show treatments to be effective or less harmful, or for comparisons where there is most amount of data available. The current findings may also indicate that review teams plan to prepare SoF tables for all the comparisons specified under ‘Types of interventions’ by default. For broad review questions the full extent of relevant comparisons may only become apparent after studies are identified, and may vary depending on the analysis undertaken. Prioritizing comparisons at the protocol stage will help focus the review on addressing questions that users are likely to value most.

### Specifying GRADE considerations and number of raters

We found improvements in the number of protocols reporting the five considerations that can lead to downgrading the quality of evidence (risk of bias, inconsistency, indirectness, imprecision, publication bias). However, most protocols did not specify the number of raters involved in the assessment of these factors; only four protocols mentioned involvement of at least two reviewers. Duplicate or consensual processes for rating GRADE merit greater attention. Studies assessing the consistency of GRADE ratings support the need for more than one rater.9,10 With this in mind consideration should be given to the use of checklists10 in GRADE assessments as planned in two of the protocols studied. These checklists should help improve transparency and consistency of GRADE assessments. They may be particularly helpful for those with limited experience using GRADE.11

### Standardizing the location of reporting GRADE methods

GRADE and SoF table methods were reported in different sections (headings) in most of the protocols (e.g. under ‘Data synthesis’, or separate sub-headings ‘Assessment of the quality of evidence’, ‘Summary of findings tables’, etc.). This inconsistency in reporting reflects current lack of clear guidance on this aspect; a section about GRADE and Summary of Findings tables should be incorporated in future versions of Review Manager to aid uniformity in reporting across protocols and reviews.

### Implications for preparing Cochrane Reviews

We believe that our audit has implications for GRADE assessments and development of SoF tables in reviews. Progress made in planning implementation of GRADE in protocols is largely attributable to improved referencing of GRADE and prespecification of SoF table outcomes. We believe that efforts should now include prioritization of comparisons and encouragement to adopt methods to rate the quality of evidence in duplicate by review authors. We would like to encourage piloting early, independent evaluation of the implementation of GRADE and SoF tables once the analysis of data has been undertaken but before the review has been written up. We think that this would encourage greater investment of effort in the preparation of SoF tables and help incorporate GRADE methods and ratings into the text of the review.

### Audit limitations

This audit has a number of limitations. First, we focused on stated intentions and we are unable to draw conclusions about whether the protocols led to better implementation of GRADE. This aspect warrants further investigation with a separate audit. Secondly, we identified protocols from just over half of the registered CRGs. This might limit the applicability of these findings to all Cochrane groups. Third, we derived the eight audit items from discussions by a CEU editors involved in the quality screening programme, and from available guidance in the Cochrane Handbook and GRADE working group resources. Some of our items have not been formally agreed upon (e.g. the subheading or section for reporting GRADE methods in protocols and reviews); however, this audit represents an important step towards defining acceptable standards for implementing GRADE and SoF tables in Cochrane reviews. Lastly we acknowledge that we have not been able to address the challenge of presenting results in SoF tables when there is sparse data, where meta-analysis is not possible or where continuous data have been analysed as standardised mean differences. We believe these scenarios should not be seen isolation in the context of SoF tables and they represent challenges in the overall interpretation and reporting of review findings. Despite these shortcomings, we believe our findings have important implications for promoting best practices in developing and executing plans for implementing GRADE and SoF tables in Cochrane reviews.

# Conclusion

Our audit has found a welcome improvement in the quality of reporting of GRADE methods and plans for preparing SoF tables in published protocols of Cochrane reviews between two cohorts of protocols published in August 2013 and August 2015. We have highlighted key areas for improving the methodological rigor and consistency in reporting of GRADE and SoF table methods in protocols. Specific attention should be given to comparisons addressed in Summary of Findings tables and methods for promoting consensual processes in GRADE assessments. We encourage review authors to address the issues highlighted to improve their implementation of GRADE and preparation of SoF tables.

# References

1. Cochrane Editorial Unit. Audit of published new Cochrane Reviews of interventions: 2014 Target 1.2, available from <http://editorial-unit.cochrane.org/mecir> [date accessed: 16th December 2015].
2. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated March 2011]. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
3. Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: Presenting results and ‘Summary of findings' tables. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
4. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924-6.
5. GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. McMaster University, 2015 (developed by Evidence Prime, Inc.). Available from [www.gradepro.org](http://www.gradepro.org).
6. Chan RJ, Webster J, Marquart L. Information interventions for orienting patients and their carers to cancer care facilities. Cochrane Database of Systematic Reviews 2011, Issue 12. Art. No.: CD008273 DOI: 10.1002/14651858.CD008273.pub2.
7. Smith V, Clarke M, Williamson P, Gargon E. Survey of new 2007 and 2011 Cochrane reviews found 37% of prespecified outcomes not reported. J Clin Epidemiol. 2015 Mar;68(3):237-45.
8. Kirkham JJ, Altman DG, Williamson PR. Bias due to changes in specified outcomes during the systematic review process. PLoS One. 2010 Mar 22;5(3):e9810.
9. Hartling L, Fernandes RM, Seida J, Vandermeer B, Dryden DM: From the trenches: a cross-sectional study applying the GRADE tool in systematic reviews of healthcare interventions. PLoS One 2012, 7(4):e34697.
10. Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, et al: The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. J Clin Epidemiol 2013, 66:736–742.
11. Meader N, King K, Llewellyn A, Norman G, Brown J, Rodgers M, Moe-Byrne T, Higgins JP, Sowden A, Stewart G. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. Syst Rev. 2014 Jul 24;3:82.

# Appendix 1

Checklist for auditing GRADE and SoF tables in protocols of intervention reviews

|  |  |
| --- | --- |
| **1. Was GRADE cited as a method for assessing the quality of the body of evidence in the review?** | 🞏 Yes  🞏 Partially  🞏 No |
| *Comments* | |
| **2. Were GRADE considerations for assessing quality reported?**  Did the authors report:  🞏 Criteria for downgrading quality of evidence (risk of bias, inconsistency, indirectness, imprecision, publication bias)  🞏 Criteria for upgrading quality of evidence if appropriate (large effect, dose-response gradient, plausible confounding effect) | 🞏 Yes  🞏 Partially  🞏 No |
| *Comments* | |
| **3. Were GRADE levels of evidence specified?**  🞏 Four categories of quality of evidence (high, moderate, low, very low) | 🞏 Yes  🞏 Partially  🞏 No |
| *Comments* | |
| **4. Were the methods for preparing SoF table specified?**  🞏 GRADEpro (GRADEpro GDT)  🞏 Cochrane methods and recommendations (Section 8.5 and Chapter 12 of the Cochrane Handbook) | 🞏 Yes  🞏 Partially  🞏 No |
| *Comments* | |
| **5. Were the comparisons to be covered in SoF tables specified (if appropriate)?**  🞏 Comparisons to be covered specified if multiple comparisons addressed | 🞏 Yes  🞏 Partially  🞏 No  🞏 Not applicable |
| *Comments* | |
| **6. Were the outcomes to be included in the SoF table specified?**  🞏 Seven or fewer outcomes (both benefits and harms, patient or guideline relevant outcomes)  🞏 Time period to be covered if outcomes measured at multiple times | 🞏 Yes  🞏 Partially  🞏 No |
| *Comments* | |
| **7. Were the number of reviewers involved in GRADE assessment specified?**  🞏 Independent quality rating by at least two authors | 🞏 Yes  🞏 Partially  🞏 No |
| *Comments* | |
| **8. Were GRADE and SoF table methods specified in an appropriate section (heading)?**  🞏 GRADE and SoF table considered under a separate heading (e.g. ‘GRADE and Summary of Findings table’) | 🞏 Yes  🞏 Partially  🞏 No |
| *Comments* | |
| *Any other comments* | |
|  | |

**Appendix 2**

Checklist for GRADE quality assessment (reproduced from Meader et al10)

|  |
| --- |
| **Study limitations (Risk of Bias)**  1) Was random sequence generation used (i.e. no potential for selection bias)?   * Yes * No * Unclear   2) Was allocation concealment used (i.e. no potential for selection bias)?   * Yes * No * Unclear   3) Was there blinding of participants and personnel (i.e. no potential for performance bias)?   * Yes * No * Unclear   4) Was there blinding of outcome assessment (i.e. no potential for detection bias)?   * Yes * No * Unclear   5) Was an objective outcome used?   * Yes * No   6) Were more than 80%[[1]](#footnote-1) of participants enrolled in trials included in the analysis (i.e. no potential reporting bias)?   * Yes * No * Unclear   7) Were data reported consistently for the outcome of interest (i.e. no potential selective reporting)?   * Yes * No * Unclear   8) No other biases reported? (i.e. no potential of other bias)   * Yes * No   9) Did the trials end as scheduled (i.e. not stopped early)?   * Yes * No   **Inconsistency[[2]](#footnote-2)**  1) Point estimates did not vary widely?   * Yes * No   2) To what extent did confidence intervals overlap?   * Substantial overlap   (all confidence intervals overlap at least one of the included studies point estimate)   * Some overlap   (confidence intervals overlap but not all overlap at least one point estimate)   * No overlap   (At least one outlier: where the confidence interval of some of the studies do not overlap with those of most included studies)  3) Was the direction of effect consistent?   * Yes * No   4) What was the magnitude of statistical heterogeneity (as measured by I2)?   * Low (e.g. I2 <40%) * Moderate (e.g. I2 40-60%) * High (e.g. I2 >60%)   5) Was the test for heterogeneity statistically significant (p<0.1)?   * Not statistically significant * Statistically significant   **Indirectness**  1) Were the populations in included studies applicable to the decision context?   * Highly applicable * Applicable * Poorly applicable   2) Were the interventions in the included studies applicable to the decision context?   * Highly applicable * Applicable * Poorly applicable   3) Was the included outcome not a surrogate outcome?   * Yes * No   4) Was the outcome timeframe sufficient?   * Sufficient * Insufficient   5) Were the conclusions based on direct comparisons?   * Yes * No   **Imprecision**  1) Was the confidence interval for the pooled estimate not consistent with benefit and harm?   * Yes * No   2) What is the magnitude of the median sample size?   * High (e.g. 300 participants) * Intermediate (e.g. 100-300 participants) * Low (e.g. <100 participants)   3) What was the magnitude of the number of included studies?   * Large (e.g. >10 studies) * Moderate (e.g. 5-10 studies) * Small (e.g. <5 studies)   4) Was the outcome a common event (e.g. occurs more than 1/100)?   * Yes * No * Not applicable (i.e. not a dichotomous outcome)   *Further optional question for those engaged in guideline development[[3]](#footnote-3)*  5) Was there no evidence of serious harm associated with treatment?   * Yes * No   **Publication Bias (other considerations)**  1) Did the authors conduct a comprehensive search?   * Yes * No   2) Did the authors search for grey literature?   * Yes * No   3) Authors did not apply restrictions to study selection on the basis of language?   * Yes * No   4) There was no industry influence on studies included in the review?   * Yes * No   5) There was no evidence of funnel plot asymmetry?   * Yes * No * Unclear   6) There was no discrepancy in findings between published and unpublished trials?   * Yes * No * Unclear |

# Appendix 3

# We present in this appendix text taken from four review protocols where we identified certain aspects of the planned implementation of GRADE or preparation of SoF tables as examples of good reporting or practice.

This first example provides a fair overview of the key aspects of the process. The last sentence showed that the authors are thinking about how to include relevant results that they might not be able to incorporate fully into the meta-analysis. This will help to provide a more complete summary of the body of evidence:

*We will create a 'Summary of findings' table including the following outcomes - mortality, health-related quality of life and major complications. We will use the five GRADE considerations (study limitations, consistency of effect, imprecision, indirectness and publication bias) to assess the quality of a body of evidence as it relates to the studies that contribute data to the meta-analyses for the prespecified outcomes. We will use the methods and recommendations described in Section 8.5 and Chapter 12 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011) and using GRADEpro software. We will justify all decisions to down- or upgrade the quality of studies using footnotes and make comments to aid the reader's understanding of the review where necessary. We will consider whether there is any additional outcome information that we were unable to incorporate into the meta-analyses, note this in the comments and state if it supports or contradicts the information from the meta-analyses.*

From: Vallance AE, Wilson CH, Dennison A, Manas DM, White SA. Total pancreatectomy and islet autotransplantation for chronic pancreatitis (Protocol). Cochrane Database of Systematic Reviews 2015, Issue 8. Art. No.: CD011799. DOI: 10.1002/14651858.CD011799.

In this second example the authors provide information about how they will identify the source of assumed risk. This provides a clear basis for translating and communicating analysis results into absolute effects. Decision-makers reading the review will find this information helpful in contextualizing the results of the review. The authors also use the protocol to provide a commitment to narrative summary of the results in the absence of a meta-analysis:

*Based on the methods described in Chapter 11 of the Cochrane Handbook for Systematic Reviews of Interventions (Schünemann 2011), we will prepare a ‘Summary of findings’ table to present the meta-analysis results. Results of the meta-analysis will be presented for the main comparisons of the review, the primary outcome child pain and the following secondary outcomes: child satisfaction with virtual reality simulation, child pain-related distress and parent anxiety, as outlined in the section on Types of outcome measures. For each assumed risk cited in the table(s), we will provide a source and rationale, and the GRADE system will be used to rank the quality of the evidence using GRADEprofiler (GRADEpro) software (Schünemann 2011). If meta-analysis is not possible, we will present results in a narrative ‘Summary of findings’ table format (drawing on Chan 2011 as an example).*

From: Lambert V, Matthews A, Hicks P, Boran L, Devane D. Virtual reality simulation for reducing pain in children (Protocol). Cochrane Database of Systematic Reviews 2013, Issue 8. Art. No.: CD010686. DOI: 10.1002/14651858.CD010686.

In this third example the authors outline a process for rating the quality of evidence in duplicate. Their intended approach explicitly draws on the checklist as reproduced in Appendix 2 and originally published in Meader et al10. Given the subjective nature of downgrading decisions we think that this is likely to help embed a structured, consensual approach to rating the quality of evidence:

*We will present the overall quality of the evidence for each outcome according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, which takes into account issues not only related to internal validity (risk of bias, inconsistency, imprecision, publication bias), but also to external validity, such as directness of results.*

*Two review authors (MDH, KNK) will independently rate the quality for each outcome. We will present a summary of the evidence in a 'Summary of findings' table, which provides key information about the best estimate of the magnitude of the effect, in relative terms and absolute differences for each relevant comparison of alternative management strategies, numbers of participants and studies addressing each important outcome, and the rating of the overall confidence in effect estimates for each outcome. We will create the 'Summary of findings' table based on the methods described in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011). We will present results on the outcomes as described in the Types of outcome measures section. If meta-analysis is not possible, we will present results in a narrative 'Summary of findings' table.*

*In addition, we will establish an appendix 'Checklist to aid consistency and reproducibility of GRADE assessments' to help with standardisation of 'Summary of findings' tables (Meader 2014).*

From: Huffman MD, Karmali KN, Berendsen MA, Andrei AC, Kruse J, McCarthy PM, Malaisrie SC. Concomitant atrial fibrillation surgery for people undergoing cardiac surgery (Protocol). Cochrane Database of Systematic Reviews 2015, Issue 8. Art. No.: CD011814. DOI: 10.1002/14651858.CD011814.

For this last example we thought that the authors had prepared a reasonably comprehensive overview of the necessary details for developing GRADE ratings and preparing SoF table. However, we also noted that whilst the authors had clearly prioritized a certain number of outcomes, they were also mindful of the need to be transparent about changes to outcome selection once the process of data collection was underway. We recognize that anticipating changes to outcomes is not really a minimum reporting requirement, but we think that this demonstrates a commitment to transparent reporting of changes to methods in the full review:

*We will summarise the results for the main comparisons specified in the Types of interventions section of this protocol in 'Summary of findings' tables. The quality of evidence in relation to each outcome included in these tables will be assessed using the evidence grading system developed by the GRADE collaboration as described in Section 12.2 of the Cochrane Handbook (Schünemann 2011b).*

*In the Types of outcome measures section of this protocol we have listed the outcomes (primary and secondary) in terms of perceived order of importance for decision-making and we will include in the 'Summary of findings' tables the first seven outcomes listed. However, as noted in section 11.5.6.2 of the Cochrane Handbook, the importance of an outcome "may only become known after the protocol was written or the analysis was carried out and [review authors] should take appropriate action to include these in the 'Summary of findings' table" (Schünemann 2011a). In the event that during the review process: a) we become aware of an important outcome that we have omitted to include in our protocol or b) we become aware that we have failed to accord sufficient priority to a specific outcome(s) listed in our protocol, then we will include the relevant outcome(s) in the 'Summary of findings' tables. If it is necessary to include outcomes in the 'Summary of findings' tables that were not pre-specified in our protocol, then we will clearly explain the reasons for this in our review, as recommended by Kirkham 2010.*

From: Beirne PV, Shiely F, Hennessy S, Fitzgerald T, MacLeod F. Needle size for vaccination procedures in children and adolescents (Protocol). Cochrane Database of Systematic Reviews 2013, Issue 8. Art. No.: CD010720. DOI: 10.1002/14651858.CD010720.

1. 80% drop out is given as an example here a different proportion can be used depending on the context of the systematic review area. [↑](#footnote-ref-1)
2. Reviewers may choose to use estimates from a subgroup analysis which may explain the inconsistency but should be cautious that such an explanation of heterogeneity may be due to the play of chance. [↑](#footnote-ref-2)
3. This reflects GRADE guidance that guideline developers may use a less stringent threshold for judging imprecision of an intervention’s benefits when there is no evidence of harm compared with when judging the benefits of an intervention where there is strong evidence of harm [↑](#footnote-ref-3)