Cochrane Consumers and Communication

Describing results

Authors face many challenges when summarising results in reviews.

This document aims to help authors to develop clear, consistent messages about the effects of interventions in reviews, and provide information to help you to:

- identify which sections of the review to focus on to ensure that messages about the review's findings are consistent
- avoid some of the most common errors identified in Cochrane reviews in relation to reporting of results
- describe the results so that size of effect and quality (certainty) of the evidence are clear
- use consistent wording to describe the results.

This document contains 3 sections:

1. Instructions to authors
2. Rationale and background material
3. Additional supporting material

This material is intended as a practical supplement to the advice in the Cochrane Handbook. It is intended to assist authors to apply the advice to CCCG reviews in a step-by-step way.

We will be seeking ongoing advice from the Cochrane Central Editorial Unit to ensure that this document stays up to date with developing methods.

If you use this resource in preparing your review, please cite it as a reference.

A suggested citation is:
Ryan R, Synnot A, Hill S (2016) Describing results. Cochrane Consumers and Communication Group, available at: http://cccrg.cochrane.org/author-resources. Version 2.0 December 2016.

Acknowledgements:
We gratefully acknowledge the input of Nancy Santesso, Claire Glenton, Sue Cole, Miranda Cumpston and Toby Lasserson to this resource, and to Bronwen Merner for assistance in finalising the document.

This is a new resource designed specifically for CCCG authors. If you have any comments or suggestions to improve our author resources, please contact Dr Rebecca Ryan, Deputy Coordinating Editor, CCCG at r.ryan@latrobe.edu.au.

Table of Contents

1. Instructions to authors

Before you begin

Describing the effects of interventions (results) relies on more than considering only whether a result is statistically significant or not.

Describing and interpreting the results **must** take into account a number of key factors, including:

- the size of the effect (magnitude), including clinical significance, and
- the quality (or certainty[1]) of the evidence on which the result is based (including the precision of the effect).

Since the results are described in different sections of the review it is critical that these all give the same messages about the effects of the interventions. Inconsistencies can arise in different ways, and these can create confusion and detract from the review's key messages.

In addition, language used to describe the results of the review needs to be clear and unambiguous.

For additional information on specific background issues, please see the following sections:

- **Section 2.1** for examples of inconsistencies in reporting of results across different sections of the review;
- **Section 2.2** for common mistakes in presenting or describing the results in reviews.
- **Section 2.3** for why the reporting of results in terms of statistical significance alone should be avoided
- **Section 2.4** for a rationale on why GRADE should be used to assess the quality of the evidence.

Sections of the review most affected

In particular, you should take care to ensure that the following sections give consistent messages about the review's results:

- Abstract
- Summary of findings tables (if used)
- Plain language summary
- Effects of interventions
- Discussion
- Implications for practice

The summary of the main results must be consistent across these three sections in particular, as most readers will access these first

[1] Please note that either 'quality' or 'certainty' can be used as terms. There may be less confusion with the use of 'certainty' to describe the outputs of GRADE, as this separates the GRADE assessment from the Risk of bias assessment more clearly, but it is up to you which term you choose to use.

Main steps for summarising the results

In this section we provide guidance to help you to describe the results clearly and accurately, including how to:

- Assess the quality and size of effects
- Use consistent language/ wording to describe the effects of interventions.

As mentioned above, describing the effects of interventions (results) means considering more than the statistical significance of the result.

There are two key things to look at when writing the results:

1. the **quality** of the evidence, and
2. the **size** (magnitude, or importance) of the effect.

Considering both of these aspects to describe the results can be used throughout the review, including the sections listed above.

1. Quality of the evidence

It is now **mandatory** that the quality of the evidence is assessed in all reviews. GRADE is the most widely-used system for Cochrane reviews, and links well with risk of bias assessments.

For updated reviews alternative approaches may be acceptable but authors should contact the CCCG editorial base for advice.

Please note that we strongly encourage authors to apply GRADE to all outcomes reported in a review –not just those that will be presented in Summary of Findings (SoF) tables. Using GRADE helps to ensure that consistent descriptions of the results and quality of the evidence are used. This is a key step in synthesising the results and in describing the findings in different sections of the review.

Please see the accompanying document 'How to GRADE the quality of the evidence' available at http://cccrg.cochrane.org/author-resources for detailed instructions on how to work through this assessment.

2. Size of the effect

Look at the size of the effect: is it a large or important effect (benefit or harm)? Or a smaller (or negligible) effect?

This decision is easier to make in some cases than others.

For some outcomes, there are well-established thresholds or levels at which the effect is likely to be important (clinically important and important to patients). For example, 4 points on the Health Related Quality of Life scale is of a size that is important to patients.

In other cases, it may be more difficult to make this decision – for example, if the outcomes and scales used to measure them are less widely used, not well validated, or used in a population for which they were not originally intended.

In such cases, there are a number of different approaches that can be used:

- Use a particular cut-off point for the effect size, such as that suggested by GRADE:
    - For dichotomous outcomes, an RR < 0.75 or RR > 1.25 is considered important.
    - For continuous outcomes, a SMD of 0.5 can be used as a rule of thumb for an important difference

- Use your judgement, based on your knowledge of the field and the outcomes used to measure effects.
    - If you are very familiar with the area and outcomes reported and have a degree of confidence in your judgements about what might be an important versus less important effect size then you can use this as a basis for making decisions in the review. This should be reported with supporting evidence in the review.

      For example:
      'Considering that Mayo-Wilson 2011 have previously deemed a reduction in all-cause mortality among children of 24% to be clinically meaningful, we judged the reduction from UCTs found in this review by 74% as clinically meaningful and very large in size.'

    - If you are not very confident in your ability to make such judgements, you should acknowledge this in the review and make the best judgement possible. Some examples of how this might be described are given below:

'We could not judge the clinical meaningfulness of this level of change due to the lack of international standards. The mean proportion in the control group at baseline was also unclear, but we judged this treatment effect to be of probably moderate size.'

Judging the clinical meaningfulness and effect size of standard mean differences (SMDs) was not straightforward, as there are no internationally agreed standards on which level of change is clinically meaningful or even which can be considered 'small' or 'large' in size.

Consequently we could not judge the clinical meaningfulness of the change in this outcome. However, considering the high mean proportion in the control group at baseline (75%), we considered the treatment effect to be very small in size.'

All three examples above are taken from: Pega F, Liu SY, Walter S, Lhachimi SK. Unconditional cash transfers for assistance in humanitarian disasters: effect on use of health services and health outcomes in low-and middle-income countries. Cochrane Database of Systematic Reviews 2015, Issue 9. Art.No.:CD011247.DOI:10.1002/14651858.CD011247.pub2.

Please note that any judgements made about how important the size of the effect is should be explained in the review, preferably with supporting research.

Using standardised wording to describe the results

As mentioned above, one advantage of using GRADE to rate the quality of the evidence is that it presents the opportunity to use standardised wording, or statements, that reflect the certainty of the evidence.

Table 1 presents standardised wording (statements) which can be used to describe the results and which take into account both the quality and the importance (size) of the effect. Authors may choose to use the wording in the table below, or devise their own set of statements.

In this scheme, the quality of the evidence has been assessed as High, Moderate, Low or Very low. Ideally GRADE would be used to make this assessment, but this matrix could be adapted (see section below for more on this).

Because this standardised wording distinguishes between results of greater or lesser quality, and those or more or lesser importance, it gives a matrix of options to allow consistent description of the results across the review.

Table 1: How to decide on standard statements to describe the results

Level (quality) of evidence	Important benefit or harm	Less important benefit or harm	No important benefit/harm or null effect
High	improves*	improves slightly	little or no difference in [outcome]
Moderate	probably improves	probably improves slightly	probably little or no difference in [outcome]
Low	may improve	may improve slightly	may have little or no difference in [outcome]**
Very low	We are uncertain whether [intervention] improves [outcome]***		
No events or rare events	Use comments in SoF table in a plainer language or summarise the results		
No studies	No studies were found that looked at [outcome]		

*** Substitute the appropriate verb for 'improves' throughout the table, depending on the results: for example, 'increases', 'reduces', 'leads to', 'prevents'**
**** This can also be worded as 'may lead to similar [outcome]'**
***** There is a debate about whether results which are rated as 'very low quality' should present numbers or not. Both approaches are currently used.**

The standardised statements above are based on the following paper:

Glenton C., Santesso N., Rosenbaum S., Stromme Nilsen E., Radar T., Ciapponi A., Dilkes H, (2010). Presenting the results of Cochrane systematic reviews to a consumer audience: a qualitative study. Med Decision Making DOI: 10.1177/0272989X10375853

How to use this table to decide which standardised statement to adopt:

1. Decide whether the size of the effect is:
    - An **important** benefit or harm (i.e. large enough to be clinically meaningful)
    - A **less important** benefit or harm (i.e. not large enough to be clinically meaningful), or
    - **No important** benefit or harm

    Select the relevant column in the table.

2. Decide whether the quality of the evidence for the outcome is High, Moderate, Low or Very low.
    Select the relevant row in the table.

3. See where the column and row you have selected meet, and use the standardised wording in that cell to describe the results.

4. Note that you may need to adapt the wording, depending on the result (outcome) you are reporting on (e.g. using 'reduces' instead of 'improves').

Some examples using these standardised statements:

- A result for an effect size that shows that a reminder intervention:
    - has an 'important' benefit for medicines adherence, and
    - is based on high quality evidence, could be described as:

    'A reminder improves adherence'

- A result for an effect size that shows that a decision aid intervention:
    - has a small ('less important') benefit for the outcome of knowledge, and
    - is based on high quality evidence, could be described as:

    'The decision aid increases knowledge slightly'

- A result for an effect size that shows that an educational intervention:
    - has no important benefit for self-efficacy
    - is based on low quality evidence, could be described as:

    'Education may lead to little or no difference in self-efficacy'

This standardised wording is based on work to develop consistent language for Plain Language Summaries.

However it can be usefully used throughout the review, and we encourage authors to adopt standardised language such as this, particularly for the following sections:

- Abstract
- Summary of findings table(s)
- Plain language summary
- Discussion - summary of the main results.

Please note that it is not mandatory to use this wording exactly as outlined above; however if you are not using this scheme it is important that you develop your own matrix for consistently describing the results that distinguishes between different rankings of quality and size of effects.

Wherever the findings of the review are being described or presented, the GRADE ratings should be considered. This includes those sections which present summarised findings, as considering the GRADE rating here helps to avoid inconsistency in reporting of findings across the review. It also highlights that the assessment of quality of the evidence is a major output of the review.

The standardised language can also be used to describe the results in the 'Effects of interventions' section of the review, particularly to give an overview of the findings for a particular outcome, once the specifics have been described.

Alternatively, you can use different ways to describe the results in the 'Effects of interventions' section, such as those used in the following examples which incorporate GRADE ratings with the results.

Examples of reporting in the results section when GRADE has been used to rate the quality of the evidence

Example 1:

GRADE ratings can be incorporated into the text by simply describing both the result and the quality of the evidence.

'Skill acquisition: There is moderate quality evidence that multimedia education was more effective than usual care or no education (MD of inhaler technique score 18.32%, 95% CI 11.92 to 24.73, two studies with 94 participants) and written education (risk ratio (RR) of improved inhaler technique 2.14, 95% CI 1.33 to 3.44, two studies with 164 participants). There is very low quality evidence that multimedia education was equally effective as education by a health professional (MD of inhaler technique score -1.01%, 95% CI -15.75 to 13.72, three studies with 130 participants).'

Ciciriello et al. Multimedia educational interventions for consumers about prescribed and over-the-counter medications. Cochrane Database of Systematic Reviews 2013, Issue 4. Art. No.: CD008416.
http://www.ncbi.nlm.nih.gov/pubmed/23633355

Please note, relating to the example above, there is a debate about whether results which are rated as 'very low quality' should present numbers or not. Both approaches are currently used.

Example 2:

Alternatively, you may incorporate the quality of evidence rating into the results and also use the standardised GRADE language to construct the wording.

'Pandey 2007 found that the intervention probably increases the number of children who received one or more vaccinations, compared to the control group (RR 1.67 (95% CI 1.21 to 2.31), moderate certainty evidence). For Andersson 2009, the results indicate that the intervention probably increase the uptake of both measles and the full course of DPT vaccines (RR 1.63 (95% CI 1.03 to 2.58) for

measles; (RR 2.17 (95% CI 1.43 to 3.29) for DPT). For both, the evidence was of moderate certainty. The intervention may make little or no difference to the number of children who received polio vaccination in the last 12 months (RR 1.01 (95% CI 0.97 to 1.05), low certainty evidence)'.

In the SoF table, these results were described using standardised language as follows:

'One study showed that the intervention probably increases the number of children who received one or more vaccinations. A second study showed that the intervention probably increases the uptake of both measles and DPT vaccines but makes little or no difference to the number of children who received polio vaccine.'

Saeterdalet al., Interventions aimed at communities to inform and/or educate about early childhood vaccination. *Cochrane Database of Systematic Reviews* 2014, Issue 11. Art.No.: CD010232.

2. Rationale and background material

2.1 Examples of inconsistencies in reporting of results

Examples of inconsistencies across different sections of the review include the following:

- inconsistent reporting of the main results e.g. reporting on a selection of different outcomes in each of the abstract, SoF tables and PLS
- inconsistent descriptions of the results e.g. highlighting selected results in the abstract, which do not correspond to the main results of the review, and/or to those reported in the SoF tables.
- if GRADE has been used to assess the quality of the evidence: not integrating the quality of the evidence into the language to summarise results, or only doing so in the SoF tables and not elsewhere in the review where results are described.
- if GRADE was not used: not reporting the results in a standardised way that takes into account the quality of the evidence (based on risk of bias assessments and other elements of quality such as inconsistency and imprecision).

In addition, language used to describe the results of the review needs to be clear and unambiguous.

So much effort goes into assessing studies, synthesising data, summarising the results and quality of the evidence in SoF tables and elsewhere, assessing risk of bias and applying GRADE that it is a shame if the final review does not reflect these key steps, which were performed with the aim of enabling clear and unambiguous language in the review.

2.2 Common mistakes in presenting or describing the results

Since Cochrane reviews are often large and complex pieces of research, there are many errors that can be introduced when describing and/or interpreting the effects of interventions.

Some of the most common are included in the table below, alongside suggested approaches for good practice, or examples.

Table 2: Common problems in reporting results in Cochrane reviews*

Common problem	Suggested good practice or examples
Inconsistent main messages across sections of the review - particularly the Abstract, SoF tables, PLS, Effects of interventions, Discussion & Implications sections	Use the GRADE ratings as a basis for describing the findings throughout the review
Under-reporting of the primary outcomes and harms, often with emphasis on positive secondary endpoints - particularly within the Abstract	Report the main (primary) outcomes, irrespective of the findings and the strength of evidence. In general, outcomes important enough to have been selected for the SoF tables should be included in the abstract, and vice versa.

Little or no use of the information presented in the SoF table in the Abstract, leading to inconsistent messages about effects	Describe the quality of evidence according to GRADE ratings, and ensure consistency with the SoF table(s).
Describing results that are imprecise as being the same as 'no effect' or 'no difference' or 'equally effective'	Highlight the uncertainty in the effects rather than making a judgement about whether the effects are 'present' or 'absent'. For example: 'We cannot tell from our results whether the intervention has an important effect on [outcome] because the sample size was small/the results were too imprecise to rule out a small or no effect'
Too much emphasis on statistical significance: • A failure to detect a statistically significant effect is misinterpreted as a lack (absence) of an effect • Where a statistically significant result is found, too much emphasis is placed on the presence of an effect	Emphasise the size (magnitude), the precision (confidence intervals) and the importance of the effect estimate. Integrating the GRADE ratings into the language used to describe results can help to provide a context for the results and to avoid reporting results simply as statistically significant or not (or present and absent).
Wording that associates the quality of evidence with statistical significance For example: 'moderate quality evidence of no statistical significance'	Emphasis on the quality of the evidence and the estimate of effect. For example: 'The effect of the intervention was uncertain due to imprecision (moderate quality evidence).'
Discussion of the quality of the evidence restricted to considering the risk of bias criteria only, without considering how other factors might impact on quality of evidence (such as imprecision, indirectness, inconsistency and publication bias).	Emphasis on how the GRADE ratings (domains) may influence the findings of key outcome results. Use information about the possible impacts on the quality of evidence than risk of assessments alone.
Very little use of the quality of evidence ratings from SoF tables, information on the decisions about downgrading the evidence, or information about the GRADE methods used.	Refer to, and explain the reasons for downgrading the quality of evidence contained in the GRADE or SoF tables, as needed. Describe the methods used to GRADE the quality of the evidence.
Confusing 'evidence of no effect' with 'no evidence of effect'	Where evidence is inconclusive about the effects of an intervention on an outcome, this represents 'no evidence of effect' (rather than 'evidence of no effect'), i.e. the result suggests that either an increase or decrease in the

	outcome is possible as a result of the intervention – we are uncertain about the result.
Lack of consistency in the way results are interpreted and reported from one outcome to another	Sometimes similar results are obtained from meta-analysis for different outcomes – e.g. finding uncertain results for both outcome of interest. These results must then be described in consistent ways – e.g. 'we are uncertain about the effects of the intervention on outcome x'; rather than emphasising one finding over another (e.g. stating that effects are uncertain for one outcome, but that there was a small effect for the other outcome); or describing the results with a different emphasis that might no longer be an objective reporting of the findings.
Describing uncertain results as 'no evidence of effect'	Stating 'no evidence of effect' can be misleading as it does not consider the quality of the evidence as an input to deciding how certain we can be about that result, and relies heavily simply on a test of statistical significance. It is preferable to report the result in terms of both the size and quality of the evidence.
Confusing evidence that is poor quality with no evidence	Sometimes evidence that is poor quality can be confused with no evidence, for example: 'There is no evidence to decide whether the intervention improves knowledge.' It is really only accurate to state that there is no evidence when no studies were found to measure an outcome; and this statement does not refer to the quality of the evidence (and hence our level of certainty about it). This could be more accurately stated to emphasise that the quality of the evidence is very low and so leads to uncertainty about the effects of the intervention, for example: 'As the evidence for our main outcomes is of very low quality, the effects of the intervention on knowledge are uncertain.'

*adapted from 'Incorporating GRADE in Cochrane reviews: feedback from the CEU screening programme' Lasserson T., Santesso N., Cumpston M., Marshall R., NíÓgáin O. Available at: http://editorial-unit.cochrane.org/mecir

2.3 Avoiding the use of 'statistically significant' to explain the results

It is recommended that Cochrane authors avoid dichotomising results into statistically 'significant' and 'statistically non-significant' and avoiding using these and related terms, as they are commonly misinterpreted.

Trying to make a yes or no judgement about whether an intervention works on not, on the basis of a test of statistical significance alone (e.g. P<0.05, or whether the confidence interval crosses the line of no effect), is misleading. This is because the precision of the effect estimate (e.g. the confidence intervals around it) is only one of several different factors to consider when assessing how much confidence to place in the result.

For example, authors may make a judgement that an intervention works based on the finding of a statistically significant difference between intervention and control groups. This is misleading as it does not take into account the size of the effect (i.e.is it important?), the precision of the effect estimate, or the quality of the evidence on which it is based.

Similarly, failing to detect a statistically significant result does not necessarily mean that there was no effect. It may be that the result was too imprecise (i.e. too few participants, leading to wide confidence intervals that are consistent with either an important effect or no effect), and measures of statistical significance only give an indication of the likelihood of the result occurring by chance (rather than being due to a real effect).

Statistical significance (or lack thereof) should therefore not be used in place of carefully interpreting the size or importance of the effect.

Use of the term ‘trend’ when a result approaches statistical significance should also be avoided (see further reading at the end of this document) - for example, a ‘promising trend’ is a misleading way to describe a statistically non-significant result in favour of the intervention.

Instead, authors should focus on:

- the size of the effect (magnitude), including clinical significance, and
- the quality of the evidence on which the result is based (including the precision of the effect).

2.4 Why GRADE?

Incorporating a comment about evidence quality into the description of the result is key to accurately describing the results. The use of GRADE to do this is **mandatory** for all new reviews. GRADE is one of the most commonly used and accepted schemes for assessing the quality of the evidence.

Since GRADE systematically rates the quality of evidence based on a range of important factors (risk of bias, inconsistency, imprecision, indirectness and publication bias), this is a comprehensive way of rating how much confidence we can place in the results.

We strongly encourage all authors, particularly those undertaking new reviews, to:

- use GRADE to assess the quality of evidence for **all outcomes** in reviews (not only those reported in SoF tables), and to
- use these ratings to describe the results **throughout the review.**

Presently, many CCCG reviews only include GRADE ratings for evidence that is included in Summary of Findings tables (SoF), leaving the remaining outcomes without an accompanying GRADE rating, and often failing to use the GRADE ratings elsewhere in the review to describe the results.

Similarly, updates of reviews often do not include GRADE ratings at all, as the original review may have pre-dated the use of GRADE. When updating however, it is possible to apply GRADE to both the existing and newly added body of evidence (although this may take a very substantial amount of work, especially if the review is very large). Authors considering updates to their reviews should contact the CCCG editorial base for advice.

More information about how to use GRADE to assess the quality of the evidence is available at:

- How to GRADE the quality of the evidence, available at http://cccrg.cochrane.org/author-resources

3. Additional supporting material

More information on how to interpret and write results is available at:

- Chapters 11 and 12 of the Cochrane Handbook

More information about interpreting p values is available at:

- Chapter 12 of the Cochrane Handbook
- Effective Practice of Care (EPOC) Author resources. Results should not be reported as statistically significant or statistically non-significant http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/22 Interpreting statistical significance 2013 08 12_2.pdf
- The Cochrane Training website has an interactive training module on interpreting imprecision at https://training.cochrane.org/resource/interpreting-results-and-drawing-conclusions-online-learning-module, see section 4 on 'Interpreting results of statistical outputs', and an introduction to meta-analysis and interpreting imprecision at https://training.cochrane.org/resource/introduction-meta-analysis.
- Wood et al. 2014. Trap of trends to statistical significance: likelihood of near significant P value becoming more significant with extra data. *BMJ* 348: g2215 http://www.bmj.com/content/348/bmj.g2215
- McCormack et al. 2013. How confidence intervals become confusion intervals. *BMC Medical Research Methodology*, 13:134 http://www.biomedcentral.com/1471-2288/13/134