How to cite this document: Ryan R; Cochrane Consumers and Communication Review Group. 'Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Group reviews: planning the analysis at protocol stage. http://cccrg.cochrane.org, December 2016 (accessed DATE).

Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Group reviews:

Planning the analysis at protocol stage

Heterogeneity

- *Cochrane Handbook for Systematic Reviews of Interventions*, section 9.5.
- Cochrane training modules:
  - Exploring heterogeneity; https://training.cochrane.org/resources/exploring-heterogeneity).
  - Heterogeneity online learning module: https://training.cochrane.org/resources/heterogeneity-online-learning-module
- Guyatt *et al* (2011) GRADE guidelines 7. Rating the quality of evidence – inconsistency. Journal of Clinical Epidemiology, 1294-302; available at http://www.gradeworkinggroup.org/index.htm.
- Oxman (2012) Subgroup analyses: The devil is in the interpretation. BMJ 344: e2022
- Sun et al (2010) Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 340: 850-4.

Please note that the material in this document is adapted directly from the Cochrane Handbook, especially section 9.5.

Heterogeneity refers to any kind of variation across studies. No two studies will be absolutely identical, so systematic reviews need ways to assess the variability across studies in order to make sensible decisions about pooling data or making particular comparisons.

In systematic reviews, different types of variability can occur across the included studies.

- *Clinical heterogeneity* refers to differences associated with the participants, interventions or outcomes. Even though a review deliberately selects studies that may be similar in many ways based on these factors, there can still be substantial differences that mean it might not make sense to pool studies. For example, thinking through clinical heterogeneity might involve considering whether very different populations are receiving the intervention across studies, if the intervention (of forms of it) are different in important ways across studies, or whether the control or other comparison groups are very different across the included studies.

- *Methodological heterogeneity* refers to differences in the way that studies were conducted – for example, differences in study design or the study's risk of bias.

In a systematic review, a decision about whether to pool the results of studies in meta-analysis needs to consider whether there are clinical or methodological differences between studies that might affect the results. Participants, interventions, comparisons and outcomes need to be taken into consideration to determine whether they are similar enough to ensure a clinically meaningful answer. If studies are very dissimilar on some or all of these factors, it may be preferable not to pool the results.

If, on the other hand, a group of studies seem to be similar enough clinically and methodologically to pool in meta-analysis then statistical heterogeneity needs to be considered.

- *Statistical heterogeneity* is the term given to differences in the effects of interventions and comes about because of clinical and/or methodological differences between studies (ie it is a consequence of clinical and/or methodological heterogeneity). Although some variation in the effects of interventions between studies will always exist, whether this variation is greater than what is expected by chance alone needs to be determined.

It is critical to assess whether heterogeneity is present, and how much, when pooling studies using meta-analysis, as the presence of heterogeneity can affect the conclusions that can be drawn from meta- analysis.

Any *statistical* heterogeneity that is detected in results must also be taken into account when interpreting the results, as this can affect the generalisability of the conclusions that can be drawn.

There are different ways to assess this:

- Visually inspect the forest plot to look at the consistency of intervention effects across included studies. If the studies are estimating the same intervention effect there should be overlap between the confidence intervals for each effect estimate on the forest plot, but if overlap is poor, or there are outliers, then statistical heterogeneity may be likely.

- Statistically test for variation: RevMan software automatically generates statistics that test for heterogeneity when performing meta-analysis. These are the:

    - Chi^2 statistic – which is the test for heterogeneity. Heterogeneity is indicated by a Chi^2 statistic greater than the df (degrees of freedom) and a small P value (eg $P < 0.05$).

    - I^2 statistic – which is the test used to quantify heterogeneity and calculates the proportion of variation due to heterogeneity rather than due to chance. The I^2 value ranges from 0% to 100%, with higher values indicating greater heterogeneity.

      As a rough guide, the I^2 statistic can be interpreted as follows:

        - 0% to 40%: might not be important;

- 30% to 60%: may represent moderate heterogeneity*
- 50% to 90%: may represent substantial heterogeneity*
- 75% to 100%: considerable heterogeneity*.

* However, the size of the I^2 statistic should be interpreted in light of the size and direction of effects, as well as the strength of evidence for heterogeneity (eg P value from Chi^2 test). Please note that a simple 'threshold' judgement approach shouldn't be used for heterogeneity: instead it is important to try to identify possible reasons for variability (if it there is a high degree of this).

If substantial heterogeneity is found, there are different courses of action that can be taken (see the *Cochrane Handbook*, section 9.5.3):

1. Do not pool data using meta-analysis – this may produce misleading results if there is high heterogeneity, or
2. Investigate heterogeneity using subgroup analysis or meta-regression. Note that if this is a possibility, it needs to be planned and pre-specified at protocol stage. See below for more on planning subgroup analyses, or
3. Use a random-effects model for meta-analysis as this includes consideration of heterogeneity in the effect estimate. A fixed-effect model assumes that there is no statistical heterogeneity between studies (ie that the estimated effects from each study would all be the same if the studies were large enough); while the random-effects model assumes that the effects estimated within each study are not identical but do follow a specific distribution. Note, however, that even though a random-effects model helps to consider heterogeneity, it does not remove it – heterogeneity still needs to be considered in interpreting the results. Also note that a random-effects model is usually used where heterogeneity is unexplained, rather than where there are identifiable reasons for heterogeneity (eg clinical differences).

A range of different approaches for dealing with heterogeneity in reviews can be used. For example, any of the approaches outlined above can be adopted. Another possibility is to first assess clinical heterogeneity (in terms of specific study factors eg intervention types or populations) and to pool studies if judged sufficiently clinically similar. Statistical heterogeneity would then be assessed and used to interpret the results but in such a case this would be a second step in the assessment of heterogeneity.

For standard text that can be used in Cochrane Consumers and Communication Group protocols to help authors to develop their approach to considering heterogeneity; see the 'Assessment of heterogeneity' and 'Subgroup analysis and investigation of heterogeneity' sections of the Group's protocol template (available from http://cccrg.cochrane.org).

Please note that in your review you will be expected to report why you decided to pool data using meta-analysis, or not, and exactly what these decisions were based on. This means that you will need to consider and report exactly why studies were too variable to pool (if this is the decision made), or how the outcomes from different studies were similar enough to meta-analyse, so that these important decisions underpinning analyses in the review are transparent to readers.

Exploring heterogeneity

- *Cochrane Handbook for Systematic Reviews of Interventions,* section 9.5 and 9.6

In systematic reviews, authors can use different methods to examine the influence of effect modifiers - for example, to investigate whether the effects of the intervention vary based on specific features (such as type, intensity, or duration) or vary in different populations. This can be done to investigate heterogeneity, or because there are good reasons to suggest that particular features of the participants, interventions or study types will influence the effects of the intervention.

These methods include *subgroup analyses* and *meta-regression*. This quick guide will focus on issues associated with subgroup analyses. More information on meta-regression can be found in the *Cochrane Handbook*, section 9.6.4.

Some tips when planning to assess and explore heterogeneity:

Heterogeneity is dealt with in several sections within a review (and planned for at protocol stage), so it is important that there is a clear correspondence between the different sections. As a guide, the following issues should be dealt with in each of these sections:

In *Assessment of heterogeneity:*

- This section should explain that clinical and methodological heterogeneity will be assessed, and if studies are considered similar enough then statistical heterogeneity will be looked at (and how this will be done). This then leads into the possibility of meta-analysis being conducted (outlined further in the *Data synthesis* section).
- It should also include a statement that indicates that if variability (from clinical, methodological and/ or statistical sources) is too high across studies then results will not be pooled (ie a narrative synthesis will be conducted instead), and that the reasons for these decisions (ie to pool data statistically or not) will be clearly reported in the review.
- This section should also include a statement linking the assessment of heterogeneity to any planned subgroup analyses – ie that if statistical heterogeneity is apparent in pooled effect estimates that this will be explored by conducting subgroup analyses.

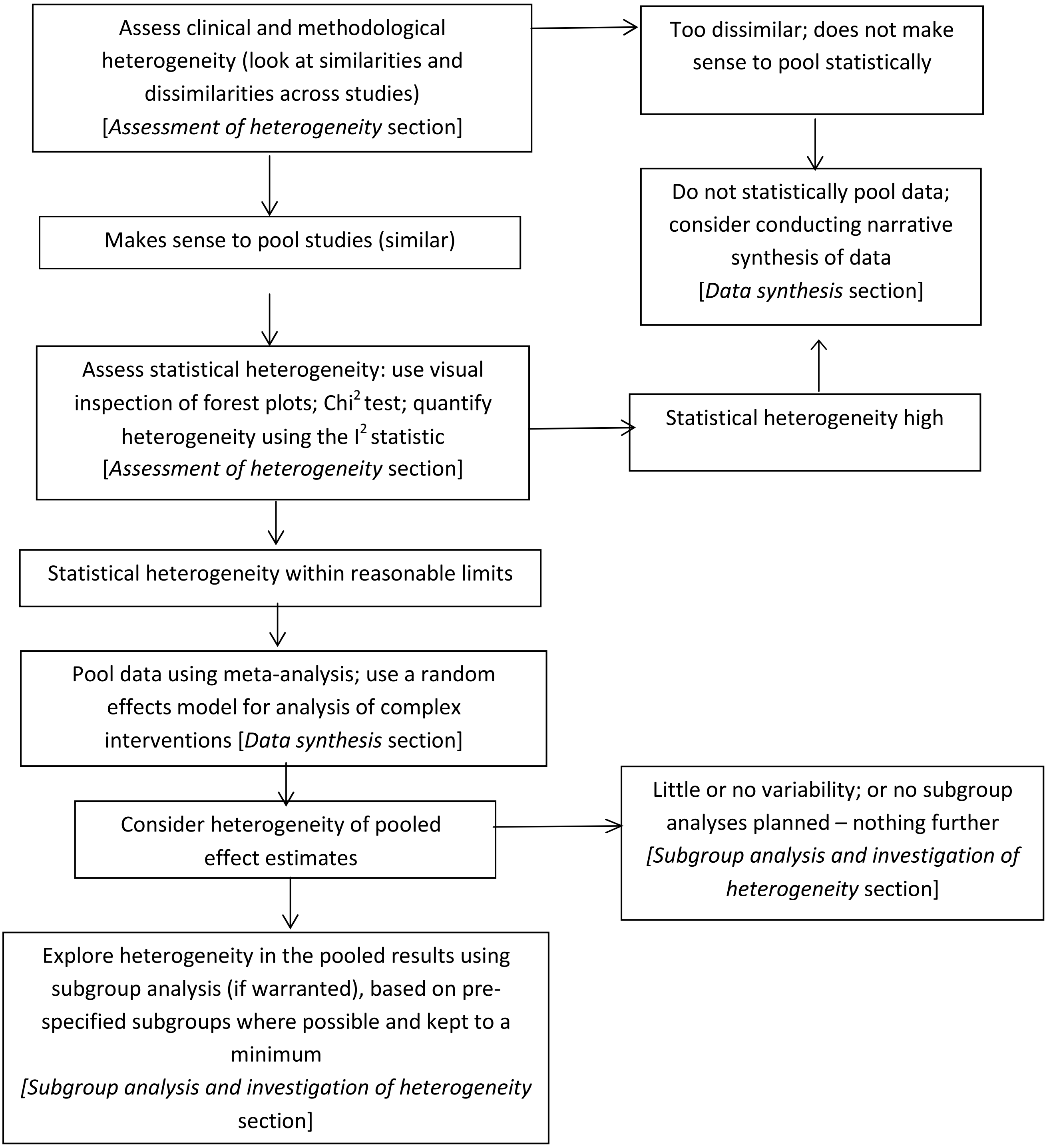
In *Data synthesis*:

- This section should outline how analysis will be conducted, planning for cases where meta-analysis is possible, as well as for the possibility that studies will be too heterogeneous to pool statistically or that meta-analysis will not be possible for all studies or outcomes (ie where narrative synthesis will need to be conducted instead).
- This section should include a statement that the decision to conduct meta-analysis or not will be made based on whether the studies make sense to pool or not (ie based on an assessment of whether participants [settings], intervention, comparison and outcomes are sufficiently similar to ensure a meaningful result). At the review stage it should also provide clear reasons as to how the decision was made (ie exactly what the decision was based on).

In *Subgroup analysis and investigation of heterogeneity:*

- This section should describe which (if any) subgroups will be investigated if variability in the pooled effect estimates is found. If heterogeneity in the pooled effect estimates is very low then it may not be necessary to perform subgroup analyses. Similarly, not all reviews are suited to including subgroup analysis – for example, if there are no strong reasons to further investigate particular variables.

Possible approach to assessing and exploring heterogeneity in reviews:



Assess clinical and methodological heterogeneity (look at similarities and dissimilarities across studies) [*Assessment of heterogeneity* section]

Too dissimilar; does not make sense to pool statistically

Do not statistically pool data; consider conducting narrative synthesis of data [*Data synthesis* section]

Makes sense to pool studies (similar)

Assess statistical heterogeneity: use visual inspection of forest plots; Chi^2 test; quantify heterogeneity using the I^2 statistic [*Assessment of heterogeneity* section]

Statistical heterogeneity high

Statistical heterogeneity within reasonable limits

Pool data using meta-analysis; use a random effects model for analysis of complex interventions [*Data synthesis* section]

Consider heterogeneity of pooled effect estimates

Little or no variability; or no subgroup analyses planned – nothing further *[Subgroup analysis and investigation of heterogeneity* section]

Explore heterogeneity in the pooled results using subgroup analysis (if warranted), based on pre-specified subgroups where possible and kept to a minimum *[Subgroup analysis and investigation of heterogeneity* section]

Subgroup analyses

Subgroup analyses split the data from all participants in order to make comparisons between subgroups in the data. These subgroups can be based on factors like participant features (eg low versus high educational levels; younger versus older people) or subsets of studies (eg geographic location).

An important aspect of splitting the data in this way is to recognise that subgroup analyses are observational and so are no longer based on randomised comparisons. As the number of subgroup analyses increases, the rate of false findings also grows. Therefore, subgroup analyses should be kept to a minimum in reviews to avoid potentially misleading results. There are no clear thresholds recommended for the numbers of subgroups which are generally acceptable in Cochrane reviews. However, advice is to keep them to a minimum, and that each subgroup must be well justified with a clear rationale.

If subgroup analyses are to be conducted this needs to be planned and pre-specified in the protocol and these planned analyses followed in the review. This helps to keep the number of subgroups investigated to a small number and also prevents knowledge of a study's results from influencing which factors are investigated.

Reasons for investigating particular subgroups must also be sound, ie, there must be a solid rationale, preferably empirical, for investigating the effects of particular factors, for example a clear theoretical basis or clinical reason. Subgroups should also be selected because they are important and/or clinically relevant to the review question; see *Cochrane Handbook* section 9.6.5.4.

Caution should also be taken when selecting factors for subgroup analysis. Ideally effect modifiers (ie factors that can affect how well the intervention works) should be chosen, whereas others like prognostic factors are less suitable for subgroup analysis unless they are also capable of modifying the intervention's effects. It is therefore often best to focus on effect modifiers such as features of the intervention (eg intensity, content, delivery), methodology (study design, quality) or study features (eg length of the study) when choosing factors to investigate in subgroup analyses.

It is also important to consider whether potential factors for investigation might be confounded as they co-occur and cannot be disentangled from the effects of others, for example more intensive follow-up may have been done in older patients and so it is not possible to separately identify the effects of these two factors on the outcome; see *Cochrane Handbook,* section 9.6.5.6.

Interpreting subgroup analyses

The results of subgroup analyses need to be interpreted with caution in all cases. See *Cochrane Handbook* 9.6.3, 9.6.3.1 and 9.6.6 for common errors in interpreting the results of subgroup analyses and suggestions on interpretation of results.

Where subgroup analyses are to be compared, and there are enough studies to do this meaningfully, authors must use a formal test to do so. Concluding that there is a difference on the basis of significantly different results can be misleading. See the *Cochrane Handbook* section 9.6.3.1.

See also the following papers which outline criteria for interpreting subgroup analyses:

- Oxman (2012) Subgroup analyses: The devil is in the interpretation. BMJ 344: e2022

- Sun et al (2010) Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 340: 850-4.

Using a structured scheme or set of criteria such as the one discussed in these papers can help authors to make a meaningful assessment about how credible the results of subgroups analysis are.

These criteria include the following considerations; but please refer to the paper for full list of criteria and explanations:

- Is the subgroup variable a characteristic measured at baseline or measured following randomisation? (those assessed post-randomisation may be less credible as they may arise due to the effects of the intervention itself)
- Was the direction of the effect pre-specified? (a subgroup effect consistent with the direction of effect that was specified up front is more credible than one that goes against the predicted direction of effect or is not specified ahead of time).
- Was the subgroup analysis hypothesis pre-specified (*a priori*)?
- Is the size of the subgroup effect large?
- Is the interaction consistent across similar/ closely related outcomes?

In plain language, these criteria can be used to assign an overall assessment of the confidence that can be placed in subgroup analyses. These judgements range from an assessment of 'very low confidence' to high confidence' depending on whether key criteria are met or not met – see Oxman 2012 for a full description of this.

Exploring relationships in the data without using subgroup analyses

Subgroup analyses can be useful for investigating statistically the effects of particular factors (effect modifiers), or questions, on the intervention's effects. As subgroup analyses should be kept to a minimum to avoid spurious statistical findings, and such analyses may often not be possible (eg too few included studies; or data not suitable for pooling), authors should also think about narratively exploring such relationships in review data.

Looking at relationships in review data is best undertaken with some clear questions in mind. As with statistical subgroup analyses, these could be undertaken to explore reasons for heterogeneity in the data (ie variability in the effects of the intervention across studies) or, where there is good reason to suggest it, to explore the influence of particular factors on the intervention's effects. This might include systematically addressing questions raised in the review's secondary objectives.

Ideally, investigating such factors narratively should be clearly justified at protocol stage, because there should be sound reasons for looking more closely at them. However, since statistical heterogeneity will not be able to be assessed, this type of narrative analysis will focus instead on clinical or methodological factors that might explain variability between studies.

Last updated: 1st December 2016