**Methods used in reviews**

**Search strategies**

**Access to specialised register by reviewers**

Neonatal reviewers are provided with a listing of references from the neonatal specialized register as requested. Reviewers obtain their own copy of a particular reference, but if for some reason it is not available to them through their own library, we will provide copies of references as required. Reviewers are advised that reference lists provided to them from the neonatal specialized register are not all-inclusive and that the search results provided are to be used in addition to searches they must do.

**Additional search strategies**

Reviewers will report in their review the search strategy used for detecting relevant trials. Cite the standard search method of the Cochrane Neonatal Group which is described in the Cochrane Library (see Specialized Register).

Reviewers will report any additional effort to detect relevant trials, including use of sources such as other trial registries, computerized bibliographic databases, review articles, abstracts, conference/symposia proceedings, dissertations, books, expert informants, granting agencies, industry, personal files. Unpublished trials are sought and if identified, must state how identified.

Detailed searching instructions for Cochrane neonatal review authors are outlined in a document prepared by the Neonatal editors - "[*Overview of Searching Databases for Randomised Trials in Neonatology*](https://neonatal.cochrane.org/sites/neonatal.cochrane.org/files/public/uploads/Overview%20of%20Searching%20Databases%20for%20Randomized%20Trials%20in%20Neonatology.doc)".

**Study selection**

Standard method of Cochrane for conducting a systematic review is followed as described in the Cochrane Handbook.

Inclusion criteria are based on characteristics of study design, population, intervention and outcomes. If relevant, reviewer states if contacted the investigators for additional information or clarification of patient characteristics, details of interventions, definitions of events, additional outcomes, losses to followup. Type of data retrieved and for what trials is described.

**Assessment of methodological quality**

Please refer to our *Model Teaching Review (currently being updated).*

The "*Methods*" section of our *Model Teaching Review* includes direction provided in the *Cochrane Handbook for Systematic Reviews of Interventions,* as well as specific requirements of the Neonatal Review Group.

--------------------------

The Cochrane Handbook describes four biases which characteristically can arise in the design or conduct of randomized trials. The Neonatal group bases its quality assessments on systematic assessment of the opportunity for each of these biases to arise. Thus, the reviewer should judge from the report of the trial whether each of the following criteria was met, and report the results of this assessment as part of the review.

Bias: Selection
Method of Avoidance: Blinding of randomization
Reviewer's Judgement: Yes, Can't tell, No

Bias: Performance
Method of Avoidance: Blinding of intervention
Reviewer's Judgement: Yes, Can't tell, No

Bias: Attrition
Method of Avoidance: Complete followup
Reviewer's Judgement: Yes, Can't tell, No

Bias: Detection
Method of Avoidance: Blinding of outcome measurement
Reviewer's Judgement: Yes, Can't tell, No

Thus, a trial which met each criterion would be described as:

Blinding of randomization: yes
Blinding of intervention: yes
Complete followup: yes
Blinding of outcome measurement: yes

It is not necessary to name bias which are avoided. Because there is no agreement on the relative importance of the various biases and their avoidance, we do not assign an overall methodology score.

**Data collection**

To the extent possible, outcome data on all patients randomized is extracted. It will be stated whether a second reviewer worked independently and at what stages of the review (assessment of trials for inclusion and methodologic quality, data extraction) and, if so, state level of agreement and how differences were resolved.

**Analysis**

*1. ANALYSIS*

a) Categorical data

Extract the proportion of randomized participants who experience adverse outcomes (e.g. death, not survival) in the treatment and control groups. Then the event rates will be the adverse event rate, and the relative risk will be the ratio of adverse events in the treated and control groups. A relative risk less than 1 will indicate a benefit in the treatment group as compared to controls. The point estimate will be plotted to the left of a RR of 1, labelled "Favours Treatment" on the graph. A risk difference (Treatment minus Control) which is a negative number will be plotted to the left of RD=0 and labelled "Favours Treatment" on the graph.

For measures of treatment effect use relative risk (RR), relative risk reduction (RR), risk difference (RD) and number needed to treat (1/RD). Relative risk and risk difference are computed by RevMan and should be calculated when appropriate. Relative risk reduction and number needed to treat should be calculated by hand and used in the text of the review when appropriate for discussing the findings.

b) Continuous data

Extract the mean and standard deviation in the treatment and control groups. Check that reported standard deviations (needed by RevMan) are what they purport to be. In original papers, se's are occasionally reported as SD's. If the SD looks very small, be suspicious; you may be able to check by recalculating statistical tests.

Some trials omit SD and se when reporting mean values. The SD can be imputed using the coefficient of variation (CV), based on data from another trial in the meta-analysis having the closest mean and N. This maneuver, which is only recommended as a last resort (call the original trial authors first to obtain SD), should be appropriately footnoted.

Make sure continuous data are really continuous, e.g. not ordinal or nominal. Nominal, ordinal, interval or ratio data may be collapsed into dichotomies and analyzed using categorical methods. Dichotomies should be justified a priori as being clinically relevant or biologically important.

A mean difference (Treatment minus Control) which is a negative number will be plotted to the left of MD=0. A negative number may or may not represent a clinical benefit. When it represents a clinical harm, it is necessary to reverse the meta-analysis graph labels "Favors Treatment, Favors Control" (by using the edit option and then the "graph" option).

c) If the data are too sparse, or of too low quality, or too heterogeneous to proceed with statistical aggregation, perform a narrative, qualitative summary and avoid meta-analysis.

d) Use the fixed effects "assumption free" model and specify such in statistical methods section.

e) Use 95% confidence intervals for the individual trial results and the typical estimates.

f) If the 95% confidence interval does not cross the "no effect" line (i.e. RR=1, RD=0, WMD=0), then the effect is statistically significant at p<.05. This level of statistical significance is also indicated when z for the typical effect is >1.96. The p value indicated by any z-value can be obtained from tables of the standard normal distribution.

g) It is not necessary to include a power analysis of individual trials or typical estimates. The confidence intervals are sufficient expression of the power.

h) Consider a cumulative meta-analysis to reveal the contribution of successive trials (ordered by publication date with typical estimate recalculated as each trial added - RevMan currently has no provision to do this automatically). Trials may also be ordered by study quality. It is not recommended that trials be ordered by baseline risk because of inherent biases to this approach.

i) Effect Measures for Counts and Rates (for a fuller treatment see the revised edition of the Cochrane Reviewers' Handbook 8.2.4 and 8.6.7). If a reviewer is contemplating an outcome which may occur more than once to a single patient there are three options for analysis. We use the number of transfusions that may be required by a neonate as an example.

i) The outcome may be stated as a binary one: need for transfusion 0 versus 1+ which would be analyzed using relative risk in the usual way.

ii) Depending on whether the number of transfusions is common, if what is reported is the number of transfusions per infant in which case the data are analyzed as a continuous measure. "Common" is not readily defined but you may prefer this approach if the trial had more transfusions than patients, ie. many patients had two or more transfusions.

iii) If transfusions are rare you could calculate the number of transfusions per person-day in each arm. This is equivalent to a person-years analysis. Eg If there are 30 transfusions in total in 100 participants studied for 14 days each, you have 30/ 1400 = 0.021 or 2.1 per 100 days. When this is entered for both arms of the trial the rate ratio (relative risk) methodology is used in the usual way. This method is not commonly used because it assumes the risk of events is constant across time and participants. As this is an uncertain assumption in most circumstances, the CNRG does not recommend this approach.

j) Effect measures for time-to-event (survival) outcomes (revised Cochrane Handbook 8.2.5 and 8.6.8)
If the time to death is of interest, rather than simply the occurrence of death, appropriate analysis is of the time-to-event. An example from neonatology is the time to blockage of a catheter (measured in hours or days). Five options exist for analysis.

i) The outcome may be stated as a binary one by selecting a fixed point of follow-up for analysis and counting the number of neonates with a blocked catheter. For example, after 7 days of follow-up, calculate in each treatment arm how many infants had one or more blocked catheters. This type of data is analyzed using relative risk in the usual way. This method ignores important information about the time-to-event.

ii) It is more appropriate to calculate the hazard of blockage for each treatment group, and the hazard ratio for the comparison between treatments. The hazard ratio is analyzed in meta-view using the relative risk procedure. Proportional hazards, which assume that the risk of the event is constant over the follow-up period, are typically used in this type of analysis. There is no procedure for calculating hazards in meta-view and this statistic should be sought in the original trial report.

iii) If catheter blockage is frequent, the number of blocked catheters per patient can be assessed and analyzed as a continuous measure in the usual way. (It would not be appropriate to analyze the total number of all catheters used since these may be removed and exchanged for reasons other than blockage.)

iv) If the time to event is measured as a continuous variable, it is not appropriate to exclude those not experiencing the event. The time to the end of the observation period should be substituted for those not experiencing the event.

v) If multiple blockages are common, it may be possible to average the time to blockage for all catheters used in a single patient, and to compute the mean of means for all patients in that treatment arm. This approach has the advantage of using all the data for each patient.

2. EVALUATION OF HETEROGENEITY

The following discussion is modified from the Cochrane Handbook, Section 8.

2.1 What is heterogeneity?

Inevitably, studies brought together in a systematic review will differ. Any kind of variability among studies in a systematic review may be termed heterogeneity. It can be helpful to distinguish between different types of heterogeneity. Variability in the participants, interventions and outcomes studied may be described as clinical diversity (sometimes called clinical heterogeneity), and variability in trial design and quality may be described as methodological diversity (sometimes called methodological heterogeneity). Variability in the treatment effects being evaluated in the different trials is known as statistical heterogeneity, and is a consequence of clinical and/or methodological diversity among the studies. Statistical heterogeneity manifests itself in the observed treatment effects being more different from each other than one would expect due to random error (chance) alone. We will follow convention and refer to statistical heterogeneity simply as heterogeneity.

If there is concern about heterogeneity, define what is clinically important and examine potential sources of heterogeneity (e.g. differences in study participants, treatment regimen, study quality, or in definition and measurement of treatment outcomes).

Clinical variation will lead to heterogeneity if the treatment effect is affected by the factors that vary across studies - most obviously, the specific interventions or patient characteristics. In other words, the true treatment effect will be different in different studies.

Differences between trials in terms of methodological factors, such as use of blinding and concealment of allocation, or if there are differences between trials in the way the outcomes are defined and measured, may be expected to lead to differences in the observed treatment effects. Significant statistical heterogeneity arising from methodological diversity or differences in outcome assessments suggests that the studies are not all estimating the same quantity, but does not necessarily suggest that the true treatment effect varies. In particular, heterogeneity associated solely with methodological diversity would indicate the studies suffer from different degrees of bias. Empirical evidence suggests that some aspects of design can affect the result of clinical trials, although this is not always the case.

The scope of a review will largely determine the extent to which studies included in a review are diverse. Sometimes a review will include trials addressing a variety of questions, for example when several different interventions for the same condition are of interest. Trials of each intervention should be analyzed and presented separately. Meta-analysis should only be considered when a group of trials is sufficiently homogenous in terms of participants, interventions, and the way outcomes are defined and measured, to provide a meaningful summary. This is a decision based on the reviewer's judgment and is not reliant on a statistical test of heterogeneity. It is often appropriate to take a broader perspective in a meta-analysis than in a single clinical trial. A common analogy is that systematic reviews bring together apples and oranges, and that combining these can yield a meaningless result. This is true if apples and oranges are of intrinsic interest on their own, but may not be if they are used to contribute to a wider question about fruit. For example, a meta-analysis may reasonably evaluate the average effect of a class of drugs by combining results from trials where each evaluates the effect of a different drug from the class.

There may be specific interest in a review in investigating how clinical and methodological aspects of trials relate to their results. Where possible these investigations should be specified a priori, i.e. in the systematic review protocol. It is legitimate for a systematic review to focus on examining the relationship between some clinical characteristic(s) of the studies and the size of treatment effect, rather than on obtaining a summary effect estimate across a series of trials.Meta-regression may best be used for this purpose, although it is not implemented in RevMan (see The Cochrane Handbook Section 8.8.3 "Meta-regression").

2.2 Identifying and measuring heterogeneity
It is important to consider to what extent the results of studies are consistent. If confidence intervals for the results of individual studies (generally depicted graphically using horizontal lines) have poor overlap, this generally indicates the presence of statistical heterogeneity. More formally, a statistical test for heterogeneity is available. This chi-squared test is included in the graphical output of Cochrane reviews. It assesses whether observed differences in results are compatible with chance alone. A low p-value (or a large chi-squared statistic relative to its degree of freedom) provides evidence of heterogeneity of treatment effects (variation in effect estimates beyond chance).

Care must be taken in the interpretation of the chi-squared test, since it has low power in the (common) situation of a meta-analysis when trials have small sample size or are few in number. This means that while a statistically significant result may indicate a problem with heterogeneity, a non-significant result must not be taken as evidence of no heterogeneity. This is also why a P-value of 0.10, rather than the conventional level of 0.05, is sometimes used to determine statistical significance. A further problem with the test, which seldom occurs in Cochrane reviews, is that when there are many studies in a meta-analysis, the test has high power to detect a small amount of heterogeneity that may be clinically unimportant.

Some argue that, since clinical and methodological diversity always occur in a meta-analysis, statistical heterogeneity is inevitable. Thus the test for heterogeneity is irrelevant to the choice of analysis; heterogeneity will always exist whether or not we happen to be able to detect it using a statistical test. Methods have been developed for quantifying inconsistency across studies that move the focus away from testing whether heterogeneity is present to assessing its impact on the meta-analysis. A useful statistic for quantifying inconsistency is I = [(Q - df)/Q] x 100%, where Q is the chi-squared statistic and df is its degrees of freedom (Higgins 2003, Higgins 2002). This describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance). A rough guide to the degree of heterogeneity I estimates is: low, moderate and high heterogeneity at values of >25%, >50% and >75%, respectively.

2.3 Strategies for addressing heterogeneity

A number of options are available if (statistical) heterogeneity is identified among a group of trials that would otherwise be considered suitable for a meta-analysis.

2.3.1. Check again that the data are correct
Severe heterogeneity can indicate that data have been incorrectly extracted or entered into RevMan. For example, if standard errors have mistakenly been entered as standard deviations for continuous outcomes, this could manifest itself in overly narrow confidence intervals with poor overlap and hence substantial heterogeneity. Unit of analysis errors may also be causes of heterogeneity (see Cochrane Handbook Section 8.3 "Study designs and identifying the unit of analysis").

2.3.2. Do not do a meta-analysis
A systematic review need not contain any meta-analyses (O'Rourke 1989). If there is considerable variation in results, and particularly if there is inconsistency in the direction of effect, it may be misleading to quote an average value for the treatment effect.

2.3.3. Explore heterogeneity
It is clearly of interest to determine the causes of heterogeneity among results of studies. This process is problematic since there are often many characteristics that vary across studies from which one may choose. Heterogeneity may be explored by conducting subgroup analyses (see The Cochrane Handbook Section 8.8.2 "Undertaking subgroup analyses") or meta-regression (see The Cochrane Handbook Section 8.8.3 "Meta-regression"), though this latter method is not implemented in RevMan. Ideally, investigations of characteristics of trials that may be associated with heterogeneity should be pre-specified in the protocol of a review (see The Cochrane Handbook Section 8.1.5 "Writing the analysis section of the protocol"). Reliable conclusions can only be drawn from analyses that are truly pre-specified before inspecting the trials' results, and even these conclusions should be interpreted with caution. In practice, authors will often be familiar with some trial results when writing the protocol, so true pre-specification is not possible. Explorations of heterogeneity that are devised after heterogeneity is identified can at best lead to the generation of hypotheses. They should be interpreted with even more caution and should generally not be listed among the conclusions of a review. Also, investigations of heterogeneity when there are very few studies are of questionable value.

2.3.4. Ignore heterogeneity
Fixed effect meta-analyses ignore heterogeneity. The pooled effect estimate from a fixed effect meta-analysis is normally interpreted as being the best estimate of the treatment effect. However, the existence of heterogeneity suggests that there may not be a single treatment effect but a distribution of treatment effects. Thus the pooled fixed effect estimate may be a treatment effect that does not actually exist in any population, and therefore have a confidence interval that is meaningless as well as being too narrow, (see The Cochrane Handbook Section 8.7.4 "Incorporating heterogeneity into random effects models"). The P-value obtained from a fixed effect meta-analysis does however provide a meaningful test of the null hypothesis that there is no effect in every study.

2.3.5. Perform a random effects meta-analysis
A random effects meta-analysis may be used to incorporate heterogeneity among trials. This is not a substitute for a thorough investigation of heterogeneity and is not the recommended approach of the Cochrane Neonatal Review Group. It is intended primarily for heterogeneity that cannot be explained. An extended discussion of this option appears in the Cochrane Handbook Section 8.7.4 "Incorporating heterogeneity into random effects models".

2.3.6. Change the effect measure
Heterogeneity may be an artificial consequence of an inappropriate choice of effect measure. For example, when trials collect continuous outcome data using different scales or different units, extreme heterogeneity may be apparent when using the mean difference but not when the more appropriate standardized mean difference is used. Furthermore, choice of effect measure for dichotomous outcomes (odds ratio, relative risk, or risk difference) may affect the degree of heterogeneity among results. In particular, when control group event rates vary, homogeneous odds ratios or risk ratios will necessarily lead to heterogeneous risk differences, and vice versa. However, it remains unclear whether homogeneity of treatment effect in a particular meta-analysis is a suitable criterion for choosing between these measures (see also The Cochrane Handbook Section 8.6.3.4 "Which measure for dichotomous outcomes?").

2.3.7. Exclude studies
Heterogeneity may be due to the presence of one or two outlying trials with results that conflict with the rest of the trials. In general it is unwise to exclude studies from a meta-analysis on the basis of their results as this may introduce bias. However, if an obvious reason for the outlying result is apparent, the study might be removed with more confidence. Since usually at least one characteristic can be found for any trial in any meta-analysis which makes it different from the others, this criterion is unreliable because it is all too easy to fulfill. It is advisable to perform analyses both with and without outlying trials as part of a sensitivity analysis (see section 4 below and The Cochrane Handbook Section 8.10 "Sensitivity analysis"). Whenever possible, potential sources of clinical diversity that might lead to such situations should be specified in the protocol.

3. SUBGROUP ANALYSES

a) Pre-specify, in the protocol, planned subgroup analyses, keep them simple and justify on mechanistic or trial variability grounds.

b) Ensure that subgroups are mutually exclusive

c) Label as such all a posteriori sub-group analyses.

d) When subgroup differences are detected, interpret them in light of whether they were proposed a priori, are supported by plausible causal mechanisms, are important (qualitatively vs quantitatively), and are consistent across studies.

e) We do not propose statistical adjustment for multiple significance testing at this juncture. These procedures are controversial with opinions ranging from "they should never be done" to "always do them". Some might argue that a priori stratification does not need it while a posteriori does. Your written commentary should indicate appropriate need for caution when interpreting the results of all sub-group analyses.

4. SENSITIVITY ANALYSES

a) Test the robustness of the results relative to features of the primary studies and to key assumptions and decisions in your review.

b) Test for bias due to the retrospective nature of systematic review (e.g. with/without trials which meet specified inclusion criteria, methodologic standards, published or unpublished).

c) Consider assessing the fragility of results by determining the effect of small shifts in the number of events between intervention and control groups; i.e. how many additional events would it take to change the statistical or clinical significance of the results in either direction.

d) Consider using cumulative meta-analysis to explore the relationship between effect size and study quality or other relevant features.

5. CROSS-OVER TRIALS

At the March 2005 meeting of CNRG editors, the following practice was adopted for the meta-analysis of cross-over and cluster trials.

The inverse variance method (IVM in REVMAN 4.2 or later) provides an opportunity to take the effect estimate (categorical or continuous data) directly from the individual cross-over trial and enter it into meta-analysis. The following method assumes that the individual RCT has been correctly analyzed as a cross-over trial (for example, does the analysis account for all periods and have treatment by period interactions - sometimes called carry over effects - been considered). Professional statistical advice may be needed to determine whether this assumption is justified. The IVM is described in the Reviewer's Handbook (RH 8.6.2). It is recommended that the entire section on data extraction (RH 8.5) be read before conducting any analyses.

For ratio measures (RR, OR or Hazard Ratio) the data are entered as logarithms with the standard error (se) of the log odds ratio. It is unusual for the se of RR, OR and Hazard Ratio to be provided in original reports. This can be calculated using the confidence interval (RH 8.5.6.2) or the p-value (HR 8.5.6.1).

For absolute measures of effect (RD) the se may not be provided in the original report. This can be estimated using the confidence intervals or from the p-value (RH 8.5.6.1)

For continuous data, the actual treatment effect is simply the observed - expected value. If the effect estimate is a change score and no measure of variance is provided, this can be calculated (RH 8.5.2.9, 10). If the effect estimate is a group mean, such as a final score, then a measure of variance can be estimated from t-tests or confidence intervals (RH 8.5.2.4) or p-values (RH 8.5.2.5).

Because the IVM program displays the summary data as the estimate from each trial, the actual data used in analysis (numerators and denominators, or means and SD's) for each treatment group should be provided as an Additional Table.

The protocol should reflect anticipated use of the IVM. It unlikely, however, that the detailed methodology will be anticipated until the individual trial reports have been examined. The procedures used in analysis should be fully documented in the Statistical Methods section of the review.

The inverse variance method allows combined meta-analysis of cross-over and parallel trial designs. It is also possible to meta-analyze all parallel trials using this methodology but this is not recommended by CNRG.

References
For Randomized Cross-Over Trials:

Senn S. Cross-over trials in clinical research. Wiley, 2nd Edit. 2002

Elbourne DR, Altman DA, Higgins JPT, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: Methodological issues. International Journal of Epidemiology 2002;31:140-149.

6. CLUSTER OR GROUP TRIALS

The meta-analysis of cluster trials proposed for CNRG is analogous to that for cross-over trials (above). The inverse variance method (IVM) is used by analysing the effect estimate from each individual cluster trial. Standard errors are calculated as described for cross-over trials. The IVM assumes that the individual cluster trial has been correctly analyzed (for example, the unit of analysis is the cluster not individuals and the analysis takes into account the correlation between clusters). Professional statistical advice may be needed to make this assumption.

In theory, the IV methodology permits combined meta-analysis of cluster and non-cluster trials. However, interpretation of such a combined analysis is not straightforward as cluster trials makes inferences about a group (eg. clinic, hospital or neighborhood) rather than individuals. Therefore, combined meta-analysis of cluster and non-cluster trials is not recommended.

References
For Cluster Randomized Trials:

Murray DM. Design and analysis of group randomized trials. Oxford University press, 1998

Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. Statistics in Medicine 2002;21:2971-2980.

7. NEED TO CHANGE REPORTED UNITS OF MEASUREMENT

In order to conduct a meta-analysis, reviewers may find it necessary to change reported units of measurement. For example, if some trials report growth per day and others per week, daily growth could be multiplied by 7. When this is done, it is important to adjust the Standard Deviation by the same multiplier as used for the mean value.

8. LOOKING FOR PUBLICATION BIAS

We recommend making use of the funnel plot option in RevMan 4.2. If your analysis contains sufficient trials to make visual inspection of the plot meaningful (no standard for this, but 5 trials may be minimum), the presence of asymmetry in the inverted funnel suggests a systematic difference between large and small trials in their estimates of treatment effect, as may occur for example because of publication bias, and may merit comment in the discussion section. In its most common form, asymmetry is seen at the lower (wide) end of the funnel where the smaller trials are plotted, and on the right-hand side of the line representing the typical effect size which is where the trials with the least favorable results would be shown.

9. PAIRED STUDY DESIGNS

Paired study designs may be found in both parallel and cross-over trials.

a) Parallel Trials

Reviewers may identify a type of trial where subjects were entered as a pair. After one patient was randomized to active therapy or placebo - the next patient entering the trial is given the alternative therapy. Although there is a deterministic element to this method of randomization, it is conceptually a blocked design of two, and no different than a (more commonly used) larger block of, say, six (where the sixth patient is also given a treatment determined by the treatments of the prior 5 subjects in the block). The CNRG considers the paired design, where the first of the pair is allocated therapy by randomization, to be a properly randomized trial.

An apparently related but different design is where the investigator treated a pair of subjects, one received the experimental intervention and the other the control, but without any indication that the choice of treatment in the first subject in the pair was chosen in a random manner. This design should be considered a quasi-randomized trial.

b) Cross-Over Trials

The same principle applies to cross-over trials. If the allocation of the first pair to treatment is random, it is considered a randomized controlled trial.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Adapted from:

Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. J Clin Epidemiol 1995;48:167-171.

**Reporting of reviews**

Discussion:
In the discussion of the review the reviewer will consider the validity of the results. State any important methodologic limitations, both of the primary studies and of this systematic review. If present, consider whether they are sufficient to pose a serious threat to validity. Consider the potential clinical importance of the results. Relevant here is the size of treatment effect in reducing adverse outcomes compared to side effects caused and, if relevant, increased economic costs. If relevant, consider additional issues such as consistency of treatment effect, dose-response relationship, limits of applicability of results.

Conclusions:
Consider the strength of inference regarding the clinical implications of the results, which varies directly with the methodological quality of the primary trials on which the review was based, degree of consistency of results among the trials, and comprehensiveness of search for all relevant trials.

For results which have the potential of influencing clinical practice, state their potential clinical importance in terms of
- the beneficial effects vs any unwanted side effects or increased economic costs
- limits of applicability, i.e. in whom the treatment should be recommended (e.g. baseline risk above which benefits are likely to outweigh harms)

For results which have the potential of influencing research, consider which questions have been well answered (further trials not warranted), which questions remain important because they have not been answered clearly (further trials warranted), and which questions remain important in only certain populations (further trials in selected populations warranted). Consider hypotheses generated by data-driven (a posteriori) analyses which now require testing in future trials. Consider new questions that arise from the reviewed research (e.g. new interventions, modification of dose, combination of therapies).