



The Cochrane
Renal Group

Additional information on preparing a protocol or a review for the Cochrane Renal Group

Last updated: 6 March 2001

- **This information is to be read with the protocol or review checklists as an aid to reviewers and referees.**
- **It is not a substitute for the Cochrane Handbook but highlights our own way of doing things**
- **The checklists and the additional information are being continually updated so let us know any comments you might have.**
- **If you would like a copy of any reference please ask and we will send it.**

Objective

- A well formulated objective should be specified in terms of the interventions, the population with the clinical problem and the outcomes.
- In the renal group we have broad inclusion criteria but we separate trials at the analysis stage.
- The objective should be sufficiently broad to allow for variations in the interventions (different drugs, different doses of the same drug, different durations of the same drug) and populations (different age groups, different sexes) unless there are good reasons not to do so. Defining how broad or how narrow a review should be can be difficult. A narrow review is easier for the reviewer but is less useful for the reader. Generally there should be only one clinical problem but in the renal group we try to be as broad as we can for populations, interventions and outcomes. This is mainly because we want to provide as much assistance as we can to the decision making of health care providers and consumers. Limiting a review to one intervention when there are many possible means that different places have to be searched to obtain all of the information, and means that trial data are duplicated in many reviews. In renal medicine, there are usually only a small number of trials on any one intervention and so a “lumping” approach is usually feasible. Having a range of populations and interventions means that variations in treatment effects can also be explored which makes application of the review results to clinical practice easier.
- All clinically relevant outcomes both beneficial and harmful should be included.

Types of studies

- In the renal group (as in most groups) we only include randomised and quasi-randomised trials, because they provide unconfounded comparisons of the effects of intervention.
- A randomised controlled trial uses a truly random sequence generation such as computer-generated or using a random number table.
- In quasi- (or pseudo) randomised trials, the allocation process is known to all so that the physician or subject could influence the treatment group to which the subject is

allocated. Examples of quasi-randomised trials are where patients are allocated to groups according to odd and even medical record numbers or month of birth.

Interventions/comparisons

- In general a systematic review is of more value to a user if all interventions for the clinical problem under study are included. Thus all possible interventions available for the clinical problem should be included unless this is not feasible.
- Just because all interventions are included in a review does not mean that all interventions will be pooled to achieve a (potentially) meaningless summary effect measure.

Types of outcomes

- The outcomes to be considered should be those that are of clinical relevance and of relevance to the patient. Surrogate outcomes, which measure a change in a laboratory measurement such as protein excretion, are less valuable than a clinical outcome such as requiring dialysis for end stage renal failure even though the surrogate is believed to be a marker for the clinical outcome. Clinical outcomes such as survival tend to be those considered most important by investigators. They may however not capture all the outcomes considered important by the patient. For example interventions that improve patient survival may result in adverse reactions that the patient may consider so detrimental to his/her quality of life that he/she may prefer not to have the intervention. One way to construct this list of benefits and harms is to consider yourself as the patient and think of all of the outcomes that would matter to you. You can then add the biochemical markers if you like. Remember that timing of outcomes also needs to be considered. Long term outcomes are generally going to be more important for patients even though they are more difficult for triallists to obtain.
- The outcomes included should include all effects of the intervention - both the benefits and harms.

Search strategy

- We are in the process of developing a clean and comprehensive register of all nephrology trials which will mean that once you provide us with your title, we can provide you with a list of all relevant trials. Until then our trials search coordinator (ruthm4@chw.edu.au) will work with you as you develop your search strategy.
- The aim of the search strategy is to locate all relevant trials. The strategy used by the Cochrane Renal Group for searching MEDLINE and EMBASE is shown in Appendix 1. However neither database provides a complete list of published trials and, of course, they do not include trials that were never published.
- The biases that result from an incomplete literature search are publication bias, duplication bias, location bias and language bias.
- Publication bias: A trial in which the experimental intervention proves to be no more effective than the control intervention is less likely to be published. Failure to locate such studies and include them in a systematic review could result in the summary estimate of effect overestimating the true effect of an intervention due to publication

bias. Hand-searching conference proceedings and reference lists of textbooks and previous trials and contacting experts in the field may help to identify unpublished studies. A graphical test that can be used to detect publication bias if large numbers of studies are available is described later in the “Methods of the Review” section.

- **Duplication bias:** Many trials are published more than once under the same or different authors and may include different amounts of data especially if initial reports do not cover the entire period of follow-up. Where there are multiple reports of the same trial, with different data sets included in each publication, you may need to extract data from several different publications to obtain all the outcomes. It is important to specify in the protocol how duplicate publication will be dealt with.
- **Location bias:** New or exciting results with large treatment effects are more likely to be published in major journals. Results deemed to be less interesting may be published in second line or local journals that may not be indexed in the major databases.
- **Language bias:** Failure to search for and include trials published in a language other than English can result in language bias.

Quality assessment

- Often we assume that all trials are equally valid (accurate), but this is not the case. Some trials are better than others. Assessment of the quality of a trial is necessary because quality can affect (bias) the results of a trial. Because we want our systematic reviews to limit bias we need to analyse trial quality. Low quality trials tend to overestimate the benefits of interventions¹. Important quality domains are allocation concealment, blinding (of investigators, subjects and those assessing outcomes), intention-to-treat analysis and completeness of follow-up.
- Adequate allocation concealment is the most important factor in reducing bias in RCTs. Trials with inadequate allocation concealment have been shown to overestimate the effectiveness of intervention by 30-40% while trials that are not double-blinded overestimate effectiveness by 17%². Open outcome assessment has been shown to overestimate effectiveness by 35%³. Intention-to-treat analysis and loss to follow-up seem to be less important in contributing to bias.

Quality items

The quality items to be assessed are shown in the appendix 2. Trial quality is a difficult but important area, so we will spend some time on it.

- **Allocation concealment:** It is important to determine if the random allocation of a subject to an intervention group in a trial could be influenced by the investigator (or subject) resulting in biased selection of subjects to one group or the other. This is prevented if there is no way of knowing what the next subject is going to get, and is called “adequate allocation concealment”. If the investigator is aware or could be aware of the intervention group to which the subject could be assigned, it is possible for that knowledge to influence which intervention the subject receives. This is obviously the situation in quasi-randomised trials. It may be less obvious in RCTs but may occur if allocation is simply using sealed but non-opaque envelopes. If the investigator preferred one intervention over another, he/she might decide not to enter his patient to the trial at that time but wait until an envelope containing the preferred treatment group was opened. Alternatively the investigator could change

the order of envelopes to achieve his/her preferred sequence. The randomisation method needs to be described in detail (e.g. central computerised randomisation system) for a reviewer to determine whether allocation was concealed or not. In many trials, the method of randomisation is not described and allocation concealment has to be considered “not stated” unless more information can be obtained from the investigators.

- *Blinding*: Ideally trials should be quadruple blind – the treatment assignment should be unknown to the investigators, the subjects, those evaluating outcomes and those analysing the data. In practice most trials only consider blinding of investigators, participants and outcome assessors so those are considered here. Double-blinded trials mean neither the investigator or the patient are aware of treatment assignment but they do not refer to blinded outcome assessment, which is probably the most important. Blinding prevents performance or cointervention bias (differences in care provided apart from the intervention being evaluated) and detection bias (differences in outcome assessment). Blinding becomes particularly important with subjective outcomes such as pain. It may not be possible to implement completely. For example it is not possible to blind the investigator and subject if a surgical intervention (ureteric reimplantation for ureteric reflux) is compared with a medical intervention (antibiotic prophylaxis). In that situation, it is essential that those evaluating the outcomes (new cortical renal scars on intravenous pyelogram) are blinded to the intervention. In some trials, the investigators and subjects may be blinded but the investigators may be able to determine the intervention group from the adverse effects experienced by the patients.
- *Intention-to-treat analysis*: Adequate allocation concealment will generally ensure that the intervention groups are similar at entry to a trial (if the trial is large enough). Publications of most trials will include a table demonstrating the similarity of treatment groups at the beginning of the trial. It is important that the groups be similar at the completion of the evaluation to avoid attrition bias. This requires that the results should be analysed with the patients in the intervention groups to which they were allocated even if they did not receive the intervention, if they deviated from protocol or if they withdrew from the study but follow-up continued. Frequently a publication report will only include data on those patients who actually received the interventions and completed the study.
- *Completeness of follow-up*: Losses of participants to follow-up will also lead to differences between treatment groups at the end of the study. In assessing a study, it is important to account for all the participants in the study.

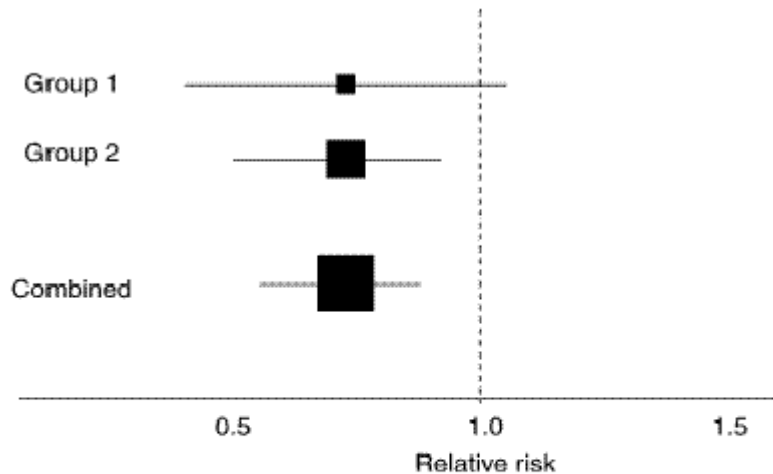
Reporting on quality items

- In the Cochrane Renal Group we assess each quality item separately rather than combine quality items in a scoring system. This is because only individual items of quality have been found to be associated with observed treatment effects, and scoring systems give unpredictable and unreliable results.³

Methods of the Review

- To reduce errors, two reviewers should carry out each step of the review. Simple differences should be resolved in discussion between the two reviewers but more substantial differences and unresolved differences should be resolved in discussion with a third reviewer.

- Each reviewer should extract data on to a standard data collection form. A specimen data extraction form is shown in Appendix 3.
- If a large number of studies are available for review, it is useful to investigate for publication bias using a funnel plot⁴. In a funnel plot, individual study weights are plotted on the Y axis against the study relative risks on the X axis. Smaller studies have less precision than large studies so are more likely to have a result further from the “true” result than a large study. If no publication bias is present, the graph should have an inverted funnel shape with the largest study at the apex of the funnel and smaller studies distributed equally on either side of the large study. If publication bias is present, the bottom right of the funnel (assuming that relative risks less than one indicate positive results) will be missing reflecting that small studies with a negative result are not included. Bias is also suggested if the summary effect estimate is situated to the left of the relative risk for the largest study.
- Reviewers have the opportunity to explore whether there is a constant intervention effect or whether it varies across studies based upon plausible factors like quality (see above), population (age, gender), clinical problem (severity of disease), intervention (dose, duration), and outcome (particularly timing). Reviewers need to decide which of these factors they would like to explain based upon biological plausibility, before they look at the trial results. Sub-group analysis is used to see whether these potential factors actually do influence treatment effect, by comparing the treatment effect of studies with the factor of interest with those studies without the factor of interest. This is formally tested as for effect modification, but you will get a reasonable idea by seeing whether the 95% confidence intervals of the two groups of studies cross.



↳ Hypothetical study showing combined and subgroup analysis: subgroups 1, 2 and the combined effect are all equivalent, but only group 2 and the combined groups are statistically significant

The general approach to subgroup analysis should be to assume similarity unless a difference can be demonstrated. Thus individual subgroups should not be tested for significance of their main effects, but should be tested to see whether the subgroups differ significantly. In the figure, we have a hypothetical study which is clearly significant (the 95%CI do not cross 1). Now if subdivided into 2 groups, group 1 is no longer significant even though the intervention effect is obviously not different from that in group 2.

- Above we emphasised the importance of specifying the subgroups to be analysed before looking at trial results. It is always tempting to look for subgroups in which the trial intervention is effective especially if the overall result suggests that the trial intervention is no better than the control intervention. Subgroup analysis has two potential problems. The first problem occurs when multiple subgroup analyses are performed. In this situation it is likely that a difference from the overall result will be found in one or more comparisons even if none exists (type 1 error). In the ISIS 2 study (Lancet 1988;2:349-360), overall streptokinase alone, aspirin alone and the combination of aspirin and streptokinase reduced mortality in acute myocardial infarction. However subdivision of patients according to their astrological sign appeared to indicate that for patients born under Gemini or Libra, there was a slightly adverse effect of aspirin on mortality while in patients born under the other astrological signs, there was a striking benefit equivalent to that seen overall. The second problem occurs if the study is not adequately powered to detect differences in subgroups. Though a randomised controlled trial may be powered to find a difference between the treatment and control groups, it is usually not powered to find differences between smaller subgroups so that a real difference in a subgroup may not be detected (type 2 error). If after looking at trial results by subgroups, which were not prespecified, you do find a comparison suggesting a treatment effect, you should record this but point out that the result would need investigation in future randomised controlled trials.

Statistical analysis

- By convention the outcome we choose is a bad one, which we aim to reduce with an intervention. This will mean, that if beneficial, an intervention will reduce the number of adverse events and the point estimate will be to the left of the line of no effect.
- Relative risk and random effects models:
You don't have to be a statistician to do a systematic review, as RevMan does much of the work for you. You simply need to decide which option to choose. Our default position is to use the relative risk measure for individual study and summary results, with the random effects model used for the summary effect measure. The random effects model incorporates possible between study differences as well as within study differences and so is more conservative. The fixed effects model assumes that no between study differences exist. As expected both models give very similar results unless there is significant between study differences (heterogeneity). We have chosen relative risk (rather than odds ratios or risk difference) because we believe it provides the best trade-off in ease of understanding, consistency across studies, and mathematical properties^{7,8}.
- Heterogeneity

Revman automatically tests for significant between study differences (heterogeneity). Heterogeneity will also be evident by looking at the forest plot (the plot of the individual trial results). It is important to distinguish between qualitative and quantitative differences. If all study results show benefit but there is heterogeneity in terms of how much benefit then the clinical implication is different from the situation where some studies show benefit and others show harm. Heterogeneity should be welcomed and explored using subgroup analysis of the factors already stipulated in the protocol. If heterogeneity cannot be explained then it is probably better to disregard the pooled result and discuss the results of the study qualitatively. Cochran's Q is the test statistic for heterogeneity. In Revman this is provided with its degrees of freedom. The corresponding p-value can be found in any statistical table. As a rough guide, there is no significant heterogeneity if the degrees of freedom equal or exceed the Q statistic.

Results and discussion

Limitations of the study

These should be discussed in relation to trial quality, publication bias, precision of results and uncertainty of harms. Trial quality and publication bias have already been dealt with.

- Precision of results: The precision of results will depend on the sample size of the study and on the confidence intervals. This becomes particularly important when there appears to be no evidence that the trial intervention is effective and one has to decide whether "there is no evidence of effect" or whether "there is evidence of no effect". For example in a comparison of cyclosporin and cyclophosphamide in preventing relapse in steroid responsive nephrotic syndrome, it is tempting to state that both drugs were equally effective in preventing relapse (evidence of no effect) since the summary relative risk is 1.07 and the 95% confidence intervals (0.48, 2.35) cross 1. However by taking the upper and lower bounds of the 95% confidence intervals of the summary relative risk, it appears that cyclosporin treatment could reduce the risk of relapse by 50% or could more than double the risk. Thus the large confidence intervals indicate that significant clinically important differences in treatment effects have not been excluded so the study has found no evidence of effect.
- Uncertainty of harms: The primary aim of most randomised controlled trials is to determine the efficacy of an intervention. As a result most trials are not powered to determine all adverse effects. In addition many studies devote little of their publication to reporting harms and may only report the most serious adverse effects such as death⁵. It is important to recognise that the number of serious adverse effects reported may be an underestimate so that the results of harms may not be directly applicable to larger groups of children treated under non trial conditions.

Applicability of results

- If the results indicate that the trial intervention is more effective than the control intervention, it is valuable to try to apply these results to patients with different baseline risks of the outcome⁶. This is demonstrated in the table below. Based on the results of a systematic review of non-corticosteroid therapies in children with steroid responsive nephrotic syndrome, the summary relative risk for relapse following cyclophosphamide treatment in comparison with prednisone alone is 0.4.

Therefore if the risk of relapse with prednisone is 100%, only 40 of 100 children will relapse after cyclophosphamide and 60 will benefit. For the 60 who benefit, one will suffer a serious infection and 4 cystitis – thus the benefits can be felt to justify the harms. However if the risk of relapse with prednisone is only 10%, then only 6 will benefit but the risk of harms is unchanged so clearly the benefits do not justify the harms.

Table demonstrating applicability of results to children at different risks of relapse of steroid responsive nephrotic syndrome

Patient subgroups	Risk of relapse if not given CPA	Risk of relapse if given CPA	No. of patients with no relapse for every 100 treated	No. with life threatening infection for every 100 treated	No. with cystitis for every 100 treated
Low risk	10%	4%	6	1	4
Medium risk	50%	20%	30	1	4
High risk	100%	40%	60	1	4

References

1. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; **352**: 609-613.
2. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**: 408-412.
3. Jüni P, Witschi A, Bloch, R, Egger M. The hazards of scoring the quality of clinical trials for Meta-analysis. *JAMA* 1999; **282**: 1054-1060.
4. Egger M, Davey-Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ* 1997; **315**: 629-634.
5. Ioannidis JPA, Lau J. Completeness of safety reporting in randomised controlled trials: An evaluation of 7 medical areas. *JAMA* 2001; **285**: 437-443.
6. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995; **311**: 1356-1359.
7. Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! *Evidence-Based Medicine* 1996; **1**: 164.
8. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods of combining event rates from clinical trials. *Statistics in Medicine* 1989; **8(2)**: 141-151.

Appendix 1

Medline Search Strategy for RCTs

- 1 RANDOMIZED CONTROLLED TRIAL.pt.
- 2 CONTROLLED CLINICAL TRIAL.pt.
- 3 RANDOMIZED CONTROLLED TRIALS.sb.
- 4 RANDOM ALLOCATION.sb.
- 5 DOUBLE BLIND METHOD.sb.
- 6 SINGLE BLIND METHOD.sb.
- 7 1 or 2 or 3 or 4 or 5 or 6
- 8 (ANIMAL not HUMAN).sb.
- 9 7 not 8
- 10 CLINICAL TRIAL.pt.
- 11 exp CLINICAL TRIALS/
(clin\$ adj25 trial\$).ti,ab.
- 13 ((singl\$ or doubl\$ or trebl\$ or tripl\$) adj25 (blind\$ or mask\$)).ti,ab.
- 14 PLACEBOS.sb.
- 15 placebo\$.ti,ab.
- 16 random\$.ti,ab.
- 17 RESEARCH DESIGN.sb.
- 18 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
- 19 18 not 8
- 20 19 not 9
- 21 9 or 20

EMBASE Search Strategy for RCTs

- 1 exp controlled study/ or controlled study.ti,ab,hw,tn,mf.
- 2 exp statistical analysis/ or clinical study.ti,ab,hw,tn,mf.
- 3 exp major clinical study/ or major clinical study.ti,ab,hw,tn,mf.
- 4 exp randomized controlled trial/ or randomised controlled study.ti,ab,hw,tn,mf.
- 5 random\$.ti,ab,hw,tn,mf.
- 6 exp double blind procedure/ or double blind procedure.ti,ab,hw,tn,mf.
- 7 exp single blind procedure/ or single blind procedure.ti,ab,hw,tn,mf.
- 8 exp multicenter study/ or multicenter study.ti,ab,hw,tn,mf.
- 9 exp placebo/ or placebo.ti,ab,hw,tn,mf.
- 10 or/1-9
- 11 (human not animal).sb,de,hw.
- 12 10 and 11

Appendix 2

Quality Checklist

1. Allocation Concealment

- Adequate (randomisation method described that would not allow investigator/participant to know or influence intervention group before eligible participant entered in the study)
- Unclear (randomisation stated but no information on method used is available)
- Inadequate (method of randomisation used such as alternate medical record numbers or unsealed envelopes; any information in the study that indicated that investigators or participants could influence intervention group)

2. Blinding

- Investigator (yes, not stated)
- Participant (yes, not stated)
- Outcome assessor (yes, not stated)

3. Intention to treat analysis

- Stated (or implied if no drop outs due to protocol violations or participant crossover)
- Unstated

4. Completeness of follow-up:

- Number of participants with outcome data/total number of participants randomised, expressed as a percentage