

COCHRANE CONSUMERS & COMMUNICATION REVIEW GROUP

STUDY QUALITY GUIDE

GUIDE FOR REVIEW AUTHORS ON ASSESSING STUDY QUALITY

CONTENTS

<i>QUICK LINKS TO KEY SECTIONS</i>	2
<i>2.0 INTRODUCTION</i>	3
<i>2.1 WHAT IS 'STUDY QUALITY'?</i>	4
2.1.1 GENERAL POINTS TO REVIEW AUTHORS ON REPORTING OF QUALITY	7
2.1.2 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	8
<i>2.2 WHY ASSESS STUDY QUALITY?</i>	9
2.2.1 REASONS TO INCLUDE STUDY QUALITY ASSESSMENT IN SYSTEMATIC REVIEWS	9
2.2.2 STUDY DESIGN VERSUS STUDY QUALITY.....	10
2.2.2.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	12
2.2.4 QUALITY OF STUDIES VERSUS THE QUALITY OF REPORTING OF STUDIES.....	13
2.2.4.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	14
<i>2.3 BIAS: WHAT YOU SHOULD CONSIDER ABOUT INCLUDED STUDIES</i>	15
2.3.1 SELECTION BIAS	15
2.3.1.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	17
2.3.2 PERFORMANCE BIAS	18
2.3.2.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	19
2.3.3 ATTRITION BIAS	19
2.3.3.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	21
2.3.4 DETECTION (OR ASCERTAINMENT) BIAS	22
2.3.4.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	22
2.3.5 OTHER SOURCES OF BIAS	22
<i>2.4 INCLUDING ASSESSMENTS OF STUDY QUALITY IN SYSTEMATIC REVIEWS</i>	24
2.4.1 GENERAL POINTS ON THE ASSESSMENT AND INCORPORATION OF STUDY QUALITY	24
2.4.1.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	25
2.4.2 BASELINE IMBALANCES IN GROUPS: POINTS FOR REVIEW AUTHORS TO CONSIDER.....	26
<i>2.5 QUALITY ASSESSMENT: RANDOMISED CONTROLLED TRIALS</i>	28
2.5.1 QUALITY ITEMS TO BE ASSESSED: RCTS.....	28
<i>2.6 QUALITY ASSESSMENT: NON-RANDOMISED CONTROLLED STUDIES</i>	33
2.6.1 NON-RANDOMISED CONTROLLED STUDIES (INCLUDING QUASI-RANDOMISED STUDIES).....	33
2.6.2 CONTROLLED BEFORE-AND-AFTER STUDIES (CBAS)	37
2.6.3 INTERRUPTED TIME SERIES (ITS) STUDIES	38
<i>2.7 QUALITY ASSESSMENT: QUALITATIVE STUDIES</i>	40
<i>2.8 HOW TO REPORT STUDY QUALITY: TIPS FOR REVIEW AUTHORS</i>	42
2.8.1 GENERAL POINTS ON REPORTING QUALITY	42
2.8.1.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS	43
<i>APPENDIX A –ADDITIONAL SOURCES OF INFORMATION</i>	45
<i>REFERENCES</i>	50

This guide was prepared by Rebecca Ryan and edited by Dominique Broclain, Dell Horey, Sandy Oliver and Megan Pictor. This paper was approved by Sophie Hill, Coordinating Editor of the Consumers and Communication Review Group on 5 March 2007. **Last updated: 5 March 2007.**

How to cite this Guide:

Ryan R, Hill S, Broclain D, Horey D, Oliver S, Pictor M; Cochrane Consumers and Communication Review Group. *Study Quality Guide*. March 2007.

www.latrobe.edu.au/cochrane/resources.html (accessed *DATE*).

The authors have submitted this Guide to the Quality Advisory Group of the Cochrane Collaboration for peer review to ensure that it is up-to-date with developments in the critical appraisal of trials. People using this guide should check the website <http://www.latrobe.edu.au/cochrane/resources.html> regularly to ensure they have the latest version.

QUICK LINKS TO KEY SECTIONS

- Suggested wording for the Methods section of your protocol/review [page 25](#)
 - Criteria for assessing RCTs [page 29](#)
 - Criteria for assessing non-randomised (including quasi-randomised) controlled trials [page 34](#)
 - Criteria for assessing Controlled Before and After studies [page 37](#)
 - Criteria for assessing Interrupted Time Series studies [page 38](#)
-

2.0 INTRODUCTION

The primary purposes of this guide are:

- To provide authors and referees with general guidance on study quality issues;
- To provide authors and referees with information about the reporting of study quality in order to improve the consistency of reporting;
- To assist authors to make decisions about appropriate quality assessment criteria for studies included in their reviews

This guide is essential reading at both the title development and the protocol stages for people conducting Cochrane systematic reviews. It is also highly recommended at the review stage.

Structure

- **Section 1:** What is study quality?
- **Section 2:** Why assess study quality?
- **Section 3:** Biases and how to avoid them
- **Section 4:** How to include quality assessment in systematic reviews
- **Section 5:** Quality assessment: randomised controlled trials
- **Section 6:** Quality assessment: non-randomised controlled studies
- **Section 7:** Quality assessment: qualitative studies
- **Section 8:** How to report study quality: tips for authors.

This document should be read in conjunction with the [Cochrane Handbook for Systematic Reviews of Interventions](#) (Higgins 2006), which gives extensive guidance on the conduct of systematic reviews. Referencing to appropriate sections of the *Handbook* is given throughout. Readers are also referred to the [Cochrane Collaboration Open Learning Material for Reviewers](#), Version 1.1, November 2002, which provide additional information on methodological issues in conducting a systematic review.

Additional sources of information for review authors, with brief descriptions of their key points, are provided in Appendix A. Appendices to the accompanying **Study Design Guide** contain glossaries of useful terms for study design and the characterisation of study design. These are adapted primarily from the [Cochrane Handbook for Systematic Reviews of Interventions Glossary](#), with additional material from other sources.

Quality assessment of studies included in systematic reviews is currently an area of active research. This guide is a working document that will be updated periodically as the field develops. Several sections of this guide for authors are still under development: these are noted in the text and authors are referred to other resources that may be of assistance.

2.1 WHAT IS 'STUDY QUALITY'?

- See [HCochrane Handbook for Systematic Reviews of Interventions](#)^H, Section 6: Assessment of Study Quality
- See CRD Report Number 4: Phase V Study quality assessment (available at [Hwww.york.ac.uk/inst/crd/report4.htm](http://www.york.ac.uk/inst/crd/report4.htm))^H
- See [HCochrane Open Learning materials](#)^H, Module 1

Quality is itself a difficult concept to qualify or to quantify. It a complex idea that means different things to different people¹, and there are numerous tools in existence with which to assess study quality. However, there is little evidence available to help guide the selection of quality assessment tools to be used in systematic reviews². Further, the reporting of quality in published studies is often poor, which can increase the difficulty of assessing relevant information.

Despite these difficulties, most definitions of quality or validity³ used in systematic reviews involve some measure of the methodological strength of the relevant study, or how able it is, through its design and its conduct, to prevent systematic errors, or **bias**.

The following is taken directly from the [Cochrane Handbook for Systematic Reviews of Interventions](#), Sections 6 and 6.1.

6. Assessment of study quality

Quality assessment of individual studies that are summarised in systematic reviews is necessary to limit bias in conducting the systematic review, gain insight into potential comparisons, and guide interpretation of findings. Factors that warrant assessment are those related to applicability of findings, validity of individual studies, and certain design characteristics that affect interpretation of results. Applicability, which is also called external validity or generalisability by some, is related to the definition of the key components of well-formulated questions outlined in section 4 [of the Handbook]. Specifically, whether a review's findings are applicable to a particular population, intervention strategy or outcome is dependent upon the studies selected for review, and on how the people,

¹ For example, it can mean the internal and/or external validity of a study, the clinical relevance of the research question that is studied, the appropriateness of the data analysis and the presentation of the findings, or the ethical implications of the interventions under evaluation.

² There is limited empirical evidence on the relationships between specific aspects of study quality and the actual findings of studies. There is some evidence that studies which lack allocation concealment and proper blinding may overestimate the effects of interventions, but the effects of other elements of quality on outcomes has not yet been established empirically.

³ **Validity**, in general terms, is the extent to which the result (of a measurement or of an intervention/ study) is likely to be true and free from systematic errors.

interventions and outcomes of interest were defined by these studies and the authors (reviewers).

Interpretation of results is dependent upon the validity of the included studies and other characteristics. For example, a review may summarise twenty valid trials that evaluate the effects of antiischemic agents on symptoms of chest pain in adults with prior myocardial infarction. However, the trials may examine different preparations and doses of antiischemic agents and may have varying durations. These latter issues would affect interpretation though they may not be directly relevant to the internal validity of the trials. Examples of how to abstract data related to applicability and design factors likely to affect the interpretation are in section 7 [of the Handbook]. The remainder of this section will focus on assessing the validity of individual studies included in a systematic review. As most Cochrane reviews focus on randomised trials, it concentrates on how to appraise the validity from these studies.

6.1 Validity

In the context of a systematic review, the validity of a study is the extent to which its design and conduct are likely to prevent systematic errors, or bias (Moher 1995). An important issue that should not be confused with validity is precision. Precision is a measure of the likelihood of chance effects leading to random errors. It is reflected in the confidence interval around the estimate of effect from each study and the weight given to the results of each study when an overall estimate of effect or weighted average is derived. More precise results are given more weight.

Variation in validity can explain variation in the results of the studies included in a systematic review. More rigorous studies may be more likely to yield results that are closer to the 'truth'. Quantitative analysis of results from studies of variable validity can result in 'false positive' conclusions (erroneously concluding an intervention is effective) if the less rigorous studies are biased toward overestimating an intervention's effectiveness. They might also come to 'false negative' conclusions (erroneously concluding no effect) if the less rigorous studies are biased towards underestimating an intervention's effect (Detsky 1992).

It is important to systematically complete critical appraisal of all studies in a review even if there is no variability in either the validity or results of the included studies. For instance, the results may be consistent among studies but all the studies may be flawed. In this case, the review's conclusions would not be as strong as if a series of rigorous studies yielded consistent results about an intervention's effect.

Variations in study quality can explain differences in the findings of studies that are included in a systematic review. As a result, the quality of a study will reflect the strength of the evidence that can be drawn from it. In other words, it reflects the confidence that we can have that the results of a study reflect the 'truth'. By extrapolation, the quality of studies included in a systematic review will affect the confidence we can have in the results of the systematic review. See resources listed in **Appendix A** of this guide for further information on quality and validity.

Generally, assessment of study quality includes assessment of at least some elements of the **internal validity** of the study⁴. This is the degree to which the design and the conduct of the study avoid bias (Jadad 1998). Simply put, it is the degree to which we can have confidence that the results of the study reflect what is 'true.' Higher quality studies are more likely to produce results that are closer to the true result, as they are less prone to bias or distortions from the true value. Evaluating the internal validity of a study therefore involves assessing the measures that were used to try to prevent or minimise bias. These might include evaluation of elements such as randomisation, allocation concealment and blinding.

The guidance in this document focuses mainly on the assessment of internal validity and related issues for studies included in systematic reviews. Internal validity is the least context-dependent aspect of study quality, and some elements of internal validity are empirically related to study results. For example, trials with inadequate allocation concealment have been shown to consistently overestimate the effects of interventions. **Assessment of study quality should therefore *always* include an assessment of internal validity.**

Other aspects of study quality may also be considered and evaluated according to how relevant they are to the scope and to the particular question(s) addressed by the review. For instance, studies included in systematic reviews should also be evaluated in terms of **external validity**. This is the extent to which the results of the study can be applied, or **generalised**, to the population outside the study; in other words how the study's results apply to the real world.

That said, evaluating internal validity is necessary, but *not* sufficient, to comprehensively evaluate a study included in a systematic review. The guide that will follow on the analysis and interpretation of the results of studies included in a systematic review [still in development, 5 March 2007] deals in more detail with issues relating to the interpretation and external validity of studies. The current document will focus primarily on aspects of quality relating to internal validity. It will address the quality assessment of randomised controlled trials (RCTs) and of

⁴ **Internal and external validity:** *Internal* validity is the extent to which the study's design, conduct, analysis and presentation have minimised or avoided biased comparisons of the intervention under investigation. *External* validity is the precision and extent to which it is possible to generalise the results of the study to other settings (ie. to real life clinical situations) (Jadad 1998).

studies with non-randomised design, as both are relevant to the scope of the Cochrane Consumers and Communication Review Group and may be included in systematic reviews coordinated by the Group.

Much of the research on quality assessment to date has focussed on the evaluation of RCTs. As a result, this is where the most comprehensive advice is currently available. However, different aspects of quality apply to studies of different designs, and recently there has been increasing interest in methods of assessing and reporting quality for a range of different study types. This document attempts to provide guidance to authors on quality assessment of both randomised and non-randomised studies.

2.1.1 GENERAL POINTS TO REVIEW AUTHORS ON REPORTING OF QUALITY

An important point to be aware of when conducting a systematic review is that the methods you have employed throughout the review need to be **transparently reported**. You should aim to produce a review that is **reproducible** by anyone who reads it. That is, anyone should be able to conduct the review on the basis of what you have written and come to similar conclusions. This is only possible if you **explicitly** report the details of each decision you make throughout the review process.

All decisions that you make throughout the course of conducting and writing your review need to be logical and justified. They must also be clearly and consistently reported. While these are important issues throughout the entire review process, they may be especially important when you start to consider how you will evaluate the quality of included studies (which criteria you will evaluate studies against); how you will report study quality as a component of the systematic review; and how you will incorporate quality assessment into interpreting the results of the review.

For example, if your review includes only RCTs, how will you decide if a study is really an RCT? For studies that report that they 'randomly' assigned participants but do not provide any details about their randomisation method, will you decide to include or exclude them from your review? Will you await response to author contact to confirm that they used a truly randomised technique? What will you do if the authors are not contactable? Decisions like these that you make during the review process need to be explicitly reported, so that the logical processes you have followed throughout the review are clear to readers.

There are many complex decisions to make when conducting a systematic review. These decisions may be further complicated by variable quality of included studies, poor quality of reporting, complex interventions, complex outcomes or means of outcome assessment, or by

other factors. Such factors may make the task of systematically reviewing the evidence on a particular intervention an involved one. However, these factors are not in themselves problematic, as long as you are **transparent** and **systematic** in the way that you report these issues and the decisions that you reach as you progress through the review.

2.1.2 SUMMARY: KEY POINTS FOR REVIEW AUTHORS

- Assessment of study quality should include explicit and systematic evaluation of **internal** validity. This should be related to the review's findings, its limitations, and to the overall question being addressed by the review. Specific elements of study quality that should be considered are covered in later sections of this guide, *see Section 2.4 onwards*.
 - Aspects of the study's **external** validity should also be systematically considered and reported when examining the broader questions of the relevance and applicability of the study.
 - The aim is to produce a systematic review that could be replicated on the basis of what you have written; therefore all decisions made throughout the review process must be explicitly reported in a **clear, logical** and **transparent** way. This should include reporting of the various decisions you made when conducting the review. For example, how did you decide whether to include a study or not? What decisions did you have to make about how studies were reported?
-

2.2 WHY ASSESS STUDY QUALITY?

- See [HCochrane Handbook for Systematic Reviews of InterventionsH](#), Section 6
- See [HCochrane Open Learning materialsH](#), Module 9
- See *CRD Report Number 4: Phase V Study quality assessment* ([Hwww.york.ac.uk/inst/crd/report4.htmH](http://www.york.ac.uk/inst/crd/report4.htm))
- See Spiegelhalter et al (2000)
- See Moher et al (1999)

Studies, even well-designed ones, are never perfect in their design or execution, and there are many different ways in which they can be compromised or limited. These limitations (or various biases, see the following **Section 2.3**) can affect the results of a study. Biased studies are more likely to produce misleading results than those that are rigorously designed and conducted.

A systematic review is only as good as the studies upon which it is built. Including biased studies in a systematic review can therefore produce misleading results. Even if high quality methods are followed for the conduct of the review itself, if studies with serious biases are included and these are not adequately accounted for or acknowledged, poor quality evidence will arise from the review.

Assessing the quality of a study that might be included in a Cochrane review can be useful in a number of ways. Authors can set a minimum threshold of quality in order for a study to be included in the review. For example, review authors may decide to exclude all RCTs that do not have adequate randomisation. Note that such decisions should be made and reported in advance, or *a priori*. A second use for study quality assessment can be to initially include studies of poorer quality, but to successively remove the poorer quality studies from the analysis. This can allow review authors to determine how much, if at all, the result is affected. This approach is called a **sensitivity analysis** as it tests how 'sensitive' the results are to the quality of the included studies. These two different approaches can also be used together to examine the effects of study quality on study findings. Finally, quality assessment of the included studies can be used to help guide and structure the discussion and interpretation of the review's results.

2.2.1 REASONS TO INCLUDE STUDY QUALITY ASSESSMENT IN SYSTEMATIC REVIEWS

- To use a minimum quality level as a threshold for the selection of studies to be included in the review.

- To explore the possibility that differences in the quality of studies may help to explain differences in the results of the studies included in the review.
- To use the quality of studies as a basis for weighting the results obtained from them if combining the studies' results in a formal meta-analysis.
- To guide the interpretation of the results and to assist in determining the strength of the evidence that can be drawn from the review.
- To help to guide recommendations and directions for future research.

Adapted from *CRG guidelines Phase 5 Study quality assessment*, available at: www.york.ac.uk/inst/crd/report4.htm; see report for further information.

Variations in the validity of studies can explain differences in the results of the studies that are included in a systematic review. More rigorous, or higher quality studies, are more likely to yield results that are closer to the 'truth'. Analysis of results from studies of variable validity can result in false conclusions being drawn about the effects of an intervention⁵. Even if there is no variability in the results of the studies included in a systematic review, it is still important to **systematically evaluate**, or **critically appraise**, all of the included studies. In such an instance, while the results may all be consistent, (that is, they report similar effects of the intervention on specific outcomes), the studies from which they come may still have serious methodological flaws. This means that the conclusions drawn from the review would be much weaker: you could have less confidence that they accurately reflect the true effect of the intervention than if the studies had all been rigorously designed and conducted and had yielded consistent results.

Therefore it is vital to be able to recognise and systematically evaluate the likely contribution of different sources of bias in included studies. It is essential to assess **whether** and **how** these may have affected the results in order to determine how confident you can be in the study's results, and that they reflect reality rather than the influence of other factors.

2.2.2 STUDY DESIGN VERSUS STUDY QUALITY

While design and quality are related characteristics of any study, they actually represent different components of a study. **Design** refers to how the study has been set up, or its framework. **Quality**, in comparison, reflects how well the study was designed *and* how well it was executed. Hence, while the design of a study affects its quality (in part, how able it is to prevent bias), both quality and design determine the strength of the evidence that can be drawn from the results of any particular study.

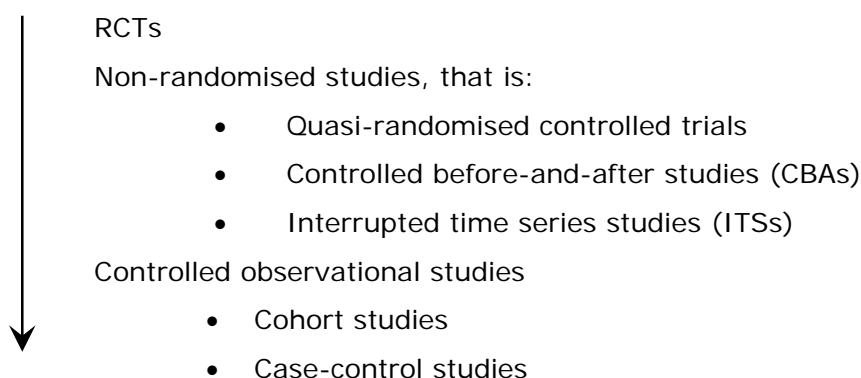
⁵ For example, if the study *overestimates* the effects of an intervention, the study may make a '**false positive**' conclusion (ie. wrongly conclude that the intervention works). If, in comparison, the study *underestimates* the effects of an intervention, it may make a '**false negative**' conclusion (ie. wrongly conclude that there is no effect of the intervention).

The **ranking** or **hierarchy** of different study designs depends on the question that is being asked (see the Group's **Study Design Guide** for more information, and see also *the Cochrane Handbook for Systematic Reviews of Interventions*, Section 4.2.4). When considering studies of effectiveness (that is, the effectiveness of therapy or some other type of intervention), the questions tend to focus on comparisons, or how an intervention compares with no intervention or with alternative intervention(s). To answer effectiveness questions, comparative studies that minimise bias will be the highest in the hierarchy, or the most suitable types of studies to investigate these questions.

It is for these reasons that RCTs are regarded so highly. RCTs are considered to be the 'gold standard' for addressing questions of effectiveness, as they are designed to minimise bias. For example, randomly assigning participants to treatment groups reduces the likelihood that the groups will differ on important baseline characteristics. This includes those characteristics of which the investigator may not be aware. By using chance it is likely that the two groups will be equivalent on important characteristics. Although it is possible to control for known confounders using other study designs,⁶ randomisation is the only way to control for confounders that are not known. This means that selection bias is likely to be minimised in RCTs; whereas the means of allocating participants in other study designs may not be as 'fair,' and groups of participants may differ on important characteristics at baseline.

Many different ways of classifying or ranking studies have been developed⁷. Although these schemes differ, broad similarities also exist. Below (Figure 1) is a general classification scheme or hierarchy of studies in terms of the suitability of their design to answer questions of effectiveness.

Figure 1: General hierarchy of study designs to answer questions of effectiveness



⁶ For example, by 'matching' participants in the different study groups on important characteristics.

⁷ See *CRD Phase V paper* Section 2.5.5 for more discussion. The exact schema adopted is not essential, but it is important to have some idea of the general level of evidence that different types of study may represent.

Many Cochrane systematic reviews specifically include RCTs only, as they are considered the strongest type of evidence to answer effectiveness questions. Reviews that fall within the scope of the Consumers and Communication Review Group, however, include many interventions for which an RCT may not exist. Sometimes interventions have not or cannot be subjected to an RCT, or even to a quasi-RCT, for practical or ethical reasons. Further, many kinds of research question may benefit from the inclusion of non-randomised studies, such as public health and health promotion interventions.

We have therefore decided to follow the advice developed by the Cochrane Effective Practice and Organisation of Care (EPOC) Review Group (see also www.epoc.uottawa.ca or email amayhe2@uottawa.ca). This advice allows review authors to decide whether to include a limited range of experimental study designs other than RCTs in their analysis and review.

These guidelines state that the following study designs are eligible for consideration for inclusion in systematic reviews of complex interventions:

RCTs (including cluster RCTs)

Non-randomised studies

- Quasi-randomised controlled trials
- Controlled before-and-after studies (CBAs)
- Interrupted time series (ITS)

It should be noted that any hierarchy of study designs should be treated with flexibility and used as a **guide** only. For example, RCTs should rank at the top of such a hierarchy only when they are well-conducted.⁸

The nature and type of the review question should determine the choice of studies to include in a review. In terms of providing answers to a review question, the suitability of different studies should be obvious if the question is formulated clearly and explicitly. It should be noted that if authors decide to include studies other than RCTs in their review, that their reasons for doing so must be **clearly** and **explicitly** stated.

2.2.2.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS

- Think about the type of question you want to answer with your systematic review. What types of study design would be most appropriate to answer this question?
- On the basis of the above, what studies will you include in your review? Will you include studies other than RCTs? If you decide that it is necessary to do so in order to answer your review question, you will need to explicitly **justify your decision** in the review.

⁸ If there are serious flaws or limitations in the design or conduct of an RCT, it should be considered as being of lower quality than one that is performed well.

This must be stated clearly at the protocol stage of your review, together with a rationale for your decision.

- As study designs can be fundamentally different in their structure and conduct, they also need to be differentially assessed for quality. That is, different elements of quality, or quality criteria, need to be considered for different study designs. **Specific quality items to be assessed for each different type of study design to be included in the review must be clearly and systematically described at the protocol stage.** These items must be described clearly and separately for each of the different study designs to be included in the review.

-
- For further information, see *CRD Report Number 4: Phase V Study quality assessment* (available at [Hwww.york.ac.uk/inst/crd/report4.htm](http://www.york.ac.uk/inst/crd/report4.htm))

2.2.4 QUALITY OF STUDIES VERSUS THE QUALITY OF REPORTING OF STUDIES

See *HCochrane Handbook for Systematic Reviews of InterventionsH*, Section 6.11

A major problem when assessing study quality is that the reporting of studies is often poor. Authors often neglect to provide adequate details of the methods they used to reduce bias. Even if the methods are reported, many authors simply report such measures as being 'done,' rather than providing detail that might allow an independent evaluation of the adequacy of the approach they used. This often makes the systematic assessment of study quality quite difficult. In cases where there is inadequate reporting, a review author may decide to assume that adequate measures were taken, or alternatively, that they were not taken⁹. Either decision may, however, be incorrect and may lead to an inaccurate representation of the study when included in the systematic review.

Given that many journals and other publications impose strict word limits on study reports, it is perhaps not surprising that authors often fail to elaborate on their methods. Standards for the reporting of trials have now been developed, and will help to standardise and improve the reporting of future studies¹⁰. However, those studies published previously remain likely to pose significant problems for review authors, when insufficient detail is reported to fully assess

⁹ That is, as a review author, you could decide to consistently rate those instances where there is insufficient detail available in a consistent manner as either done or not done. In either case you should report this decision in the text of your review.

¹⁰ For example, the Consolidated Standards of Reporting Trials (CONSORT) statement was developed to improve the reporting of RCTs (see [Hwww.consort-statement.org](http://www.consort-statement.org)); while the Transparent Reporting of Evaluations with Non-randomised Designs (TREND) statement provides guidelines for the standardisation of reporting for non-randomised studies (see [Hwww.TREND-statement.org](http://www.TREND-statement.org)). See [HAppendix AH](#) of this guide for further information.

study quality¹¹. This is often true in the case of Cochrane reviews, where the inclusion of studies is typically not restricted by the year of publication and it is common for a review to include studies dating back several decades.

Rather than assuming that something was not done if it was not reported, a more thorough approach should be taken. Review authors should contact the study authors in the first instance to request further information. If this is not possible or not successful¹², the reporting of quality should reflect this underlying uncertainty. For example, you may wish to score a particular quality item on a scale of 'done', 'not done' or 'unclear' (that is, that there is not enough information available to determine whether something was or was not adequately performed). Building in this capacity for reporting uncertainty may allow more accurate reporting on the status of the study than simply assuming that critical elements of the study were done or not done. This may, in turn, be more meaningful when it comes time to interpret the study's results and to incorporate them into the review's overall findings.

2.2.4.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS

- Variations in study quality (validity) can explain differences in the results of studies included in a systematic review. More rigorous or higher quality studies are more likely to closely estimate the truth. Study quality should therefore be assessed **systematically, transparently** and in an **explicit** manner. If in doubt, report more detail rather than less.
- Poor reporting of elements of a study that relate to quality is not the same as poor quality of the study itself.
- The Consumers and Communication Review Group recommends that if not enough detail is provided to assess quality criteria for included studies:
 - Review authors should contact the study authors to see if more information about the design and/or conduct of the study can be obtained (for sample letters contact the Review Group Coordinator).
 - Where further information is not available, a scheme for reporting on quality items that allows the reporting of uncertainty may be the most meaningful and accurate approach. For example, you could score a particular quality item on a scale of 'done', 'not done' or 'unclear' (not enough information to determine whether something was or was not adequately done). Building in this capacity allows accurate reporting on the status of the study. The design and use of such reporting schemes for different study designs are covered in detail from Section 2.4 onwards of this document.

¹¹ While the reporting of study quality is always improving, older studies may be more limited in their reporting. As a review author, you need to be consistent and clear about how you intend to deal with studies where a lack of information or detail is problematic in their evaluation.

¹² For example, for studies conducted a long time ago (10 or 20 years), even if study authors can be contacted they may no longer have access to the data or the specific details of the conduct of the study.

2.3 BIAS: WHAT YOU SHOULD CONSIDER ABOUT INCLUDED STUDIES

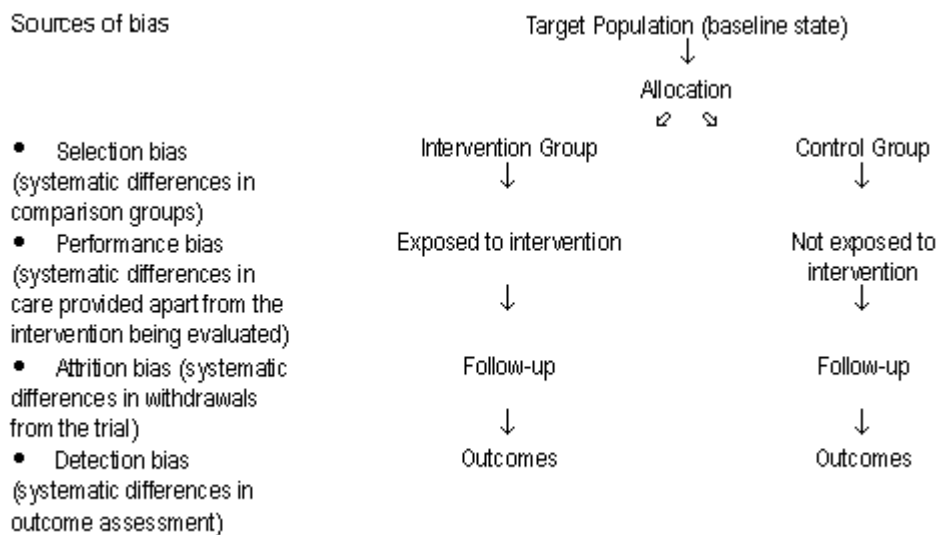
- See [HCochrane Handbook for Systematic Reviews of InterventionsH](#), section 6.2
- See also *Jadad (1998): Chapters 3 and 4*

Bias is any systematic error or factor, whether it is recognised or not, that distorts the results or inferences of a study. Bias can arise from many different sources. There are four major sources of bias that can arise in studies examining the effects of healthcare interventions:

- Selection bias
- Performance bias
- Attrition bias
- Detection bias.

Figure 2 (below) is taken directly from *The Cochrane Handbook for Systematic Reviews of Interventions*, section 6.2:

Figure 2: Sources of bias in healthcare interventions



2.3.1 SELECTION BIAS

- See [HCochrane Handbook for Systematic Reviews of InterventionsH](#), section

Selection bias occurs at the point of allocating participants to groups at the beginning of a trial, and results from systematic differences in the characteristics of the participants in each study

group. These differences can relate to prognostic factors, or can be due to the differential responsiveness of participants to the intervention.

Minimising selection bias:

- **Random assignment** of large numbers of participants to study groups can minimise this type of bias (see *Study Design Guide, section 1.3.2.3* for more information on randomisation). However, allocation must also be **adequately concealed**, so that neither the investigator nor the participant can influence which group a participant is entered into. Randomisation without adequate allocation concealment does *not* adequately protect against selection bias.
- **Allocation concealment** means that the process of allocating participants or actually placing them to the different groups to which they have been randomly assigned must be concealed from the person recruiting participants into the trial. If the allocation is not concealed, there is potential for systematic differences in characteristics between participants allocated to the different arms or treatment groups of the study. Evaluating allocation concealment involves a judgement about how the randomisation sequence was applied when participants were actually placed in the study groups, and whether it was possible for the researcher to subvert the randomisation process. For example, if random numbers are picked from an open envelope, there is potential to replace slips and to re-choose until the 'right' allocation is selected for a particular participant. Such practices subvert the randomisation process, so that it is no longer truly random.
- **Allocation concealment is different to blinding**¹³. Allocation concealment is achieved when the randomisation sequence is concealed *before* and up until the point at which people are allocated to groups. This means that no-one should know who has been assigned to the different groups before it actually occurs. In comparison, blinding (whether of participants, providers or outcome assessors) refers to measures that are taken *after* people have been assigned to groups. This means that no-one knows which participant belongs to the different groups throughout the course of the study. An important difference between the two is that allocation concealment can *always* be implemented, whereas blinding *cannot* always be successfully achieved.
- Allocation concealment has been shown to impact importantly on the results of trials. Several studies have demonstrated that when allocation concealment is inadequate or unclear, the treatment effect is overestimated.

¹³ **Blinding** prevents participants, providers and/or outcome assessors from knowing which participants are receiving the intervention(s) in the study.

- Allocation concealment can be achieved in different ways, but different approaches can have varying levels of success; that is, they are more or less likely to be interfered with or subverted.
-

2.3.1.1. SUMMARY: KEY POINTS FOR REVIEW AUTHORS

RANDOMISATION

'Truly' random methods of generating the randomisation sequence (ie. methods that produce a non-predictable assignment pattern):

- Computer-generated random numbers
- Random number tables
- Coin toss (for a trial with two groups) or die toss (where there are more than two groups)

Note: Trials employing a truly random sequence generation method are designated as RCTs.

Inadequate approaches (methods that produce a predictable assignment pattern):

- Alternation
- Case record numbers
- Birth dates
- Week days

Note: Trials employing such sequence generation methods are designated as quasi-RCTs.

ALLOCATION CONCEALMENT

Adequate allocation concealment approaches (sequence for allocating participants to groups is truly hidden from investigators):

- Central computer randomisation and allocation (for example, allocation by a central office that is unaware of the participants' characteristics);
- On-site computer from which assignment can only be determined after entering the patient's data;
- Pre-numbered or coded identical containers administered serially to participants;
- Serially (sequentially) numbered, opaque sealed envelopes.

There are also other approaches, similar to those above, that if administered by a different person to the one who generated the allocation scheme can also be considered adequate. Certain other approaches might also be new or innovative and not fit with those described above but can still provide adequate allocation concealment.

Inadequate approaches (sequence may be accessed or predicted by investigators before allocation to groups has occurred):

- Any 'open methods' (ie. transparent before allocation) eg lists on noticeboards, open lists, open envelopes;
- Non-opaque envelopes;
- Odd or even date or medical record number;
- Dates of birth or days of week;
- Alternation.

In cases where studies do not report any approach to concealing the allocation sequence, it should be assumed that the adequacy of approaches is **unclear**. Examples of this might include cases where studies state that a list or table or random numbers was used but do not provide further details; or stating that sealed envelopes were used .

Consider:

- Was there any way that those conducting the study could have known what the random allocation sequence was? Could allocation to groups be manipulated?¹⁴
- Were the study groups comparable at the start of the study, just after randomisation?

2.3.2 PERFORMANCE BIAS

- See [HCochrane Handbook for Systematic Reviews of InterventionsH](#), section 6.4

This bias arises from systematic differences in the way that care is provided, or from exposure to factors other than the intervention that is being studied.

- Performance bias occurs during the treatment and/or delivery of the intervention(s).
- It arises as the result of differences in the way that the intervention is delivered to the different study groups; that is, not only does the intervention differ between groups, the method of delivering it also differs. The impact of the intervention alone therefore cannot be assessed.

Minimising performance bias:

- Performance bias can be avoided when both the **participants** and the intervention **provider** are **blind** to the comparison group to which the participant belongs.
- Caregivers/providers, outcomes assessors, analysts and the participants themselves can all be eligible for blinding, leading to trials being described as single, double, triple, or even quadruple blind¹⁵.

¹⁴ If the answer to this questions is 'yes,' the method of allocation concealment can be considered inadequate.

¹⁵ There is very little consensus regarding what actually constitutes single/ double/ triple blind, and although these terms are commonly used there is little consistency in the way that they are used. See [HAppendix AH](#) of this guide fore more on blinding and interpretations of blinding.

2.3.2.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS

Consider:

- Were there any other factors that might have affected the study's results?
 - Of those involved in the study, who was blinded?
 - Participants?
 - Provider/ caregivers?
 - Outcome assessors?
 - Data analysts?
 - We suggest that you **avoid** the use of terms like 'double blind' etc. unless you clearly and explicitly define them in your review (eg. 'we will use the term double blind to mean...'). Instead we suggest that you specifically report on each component of blinding **separately** (eg. whether participants were blind and how adequate the approach to blinding was; whether providers were blinded adequately, and so on). This makes it very clear who was blinded throughout the study; allows comment on the adequacy of each possible component of blinding; and allows the potential bias introduced by inadequate blinding to be identified clearly both by you the review author and the readers of your review.
-

2.3.3 ATTRITION BIAS¹⁶

- See [HCochrane Handbook for Systematic Reviews of InterventionsH](#), section 6.5

This bias arises due to systematic differences between study groups in the withdrawals or exclusion of people entered into the study. That is, it is due to systematic differences between study groups in the numbers of participants who are 'lost' during the study (ie. it occurs *after* the allocation and inclusion of participants in the study). Attrition bias occurs over the duration of the trial.

For example, in a study examining the effects of a particular drug, more people receiving the active drug may leave a study due to side effects than those assigned to the control group. This could lead to an *overestimation* of the effectiveness of the intervention, as only those people who tolerated the intervention well remained in the study and were assessed. Attrition bias could also lead to an *underestimation* of the adverse effects or harms associated with the intervention, if those with the most severe side effects left the study and were not measured and included in analysis.

¹⁶ Attrition bias is also known as **exclusion bias**, as it may result from the post-randomisation exclusion of participants from the study. See [HAppendix AH](#) of this guide for more information on attrition bias.

In the case of RCTs, random assignment is not itself sufficient to make inferences infallible. The validity of inferences depends on the assumptions that random assignment produced comparable groups at baseline, and that systematic attrition over the study period did not occur. Studies often fail to report how many participants who were enrolled in the study were followed throughout it, which makes sample attrition difficult to assess. Even if not reported however, it is important to note the attrition of the samples as a possible source of bias; and to interpret the results of studies with large rates of attrition with some caution.¹⁷

Minimising attrition bias:

- Attrition bias can be minimised when the **proportion** and **characteristics** of the participants lost to follow-up (that is, those lost from the trial *after* their inclusion in the trial and assignment to groups) are described and are comparable for the different study groups. Note that this should include an account of **everyone** who was enrolled and randomised by the study. The reasons for participants dropping out of the study can provide valuable qualitative information when interpreting and discussing the results of the study. Studies should account for every participant initially enrolled in the study for **each group**. Ideally, analysis should always be based on the intention-to-treat (ITT).
- **Intention-to-treat (ITT) analysis** involves analysing participants in the groups to which they were allocated at the start of the study, even if they didn't receive the intervention, if they deviated from the protocol or they withdrew from the study. Note that ITT analysis is most often applied to RCTs and not to other types of study design, although a similar approach can be taken with non-randomised studies.

ITT analysis is a way of preserving the equivalence of the groups established at the start of the study (baseline) by the process of random allocation. If the study participants are randomly assigned to groups and there are adequate methods of allocation concealment, the groups should (theoretically) be similar at the start of the study and should be analysed on this basis. In order to prevent attrition bias, the groups also need to be similar at the end of the trial (Heritier *et al* 2003). The alternative approach, to analyse participants according to whether the participants did or did not receive the intervention, can markedly change the characteristics of the different study groups, and so make the process of random allocation ineffective.

Often studies report outcomes only for those participants that were followed up or completed the study; or they do not provide enough information to determine whether ITT

¹⁷ The way that losses of participants throughout the study are handled has much potential to introduce bias. However, reported attrition following allocation has not been found to be clearly related to bias. It is not yet clear exactly what the relationship between attrition and bias are, so review authors should be cautious when dealing with this aspect of study execution.

analysis was used. Even if ITT analysis is **not** used, it is important that the participants lost from the study (after they were enrolled in the study) are clearly accounted for. The way in which they were dealt with in the study's analysis must also be clearly and completely described. Many studies fail to report any details on these issues.

- See the [HCochrane Collaboration Open Learning MaterialsH](#), Module 14;
- See [HAppendix AH](#) of this guide, for additional sources of information (listed under **attrition bias** and **ITT analysis**).

2.3.3.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS

- Often the numbers of participants involved throughout studies (participant flow) are poorly reported. Review authors should contact study authors to request further information about participant flow if it is not clear from the study report.
- Note that even where participant flow is reported, participant numbers are often implied rather than explicitly stated or reported at each stage of the study. Authors of such studies should be contacted to provide further information to clarify participant flow. If information is not available, the potential bias introduced by participant attrition from the study should be noted in the review.

Consider:

- Was follow-up adequate?
 - How many participants completed the study for each group and for each outcome?
 - How many participants did not complete the study for each group and for each outcome?
 - Were the study groups comparable at the end of the study?
 - Were those followed up comparable to those who were lost from the study?
 - Was ITT analysis used?
 - If ITT analysis was not used, how were those who dropped out (or were excluded post randomisation) treated in the analyses?
-

2.3.4 DETECTION (OR ASCERTAINMENT) BIAS

- See [HCochrane Handbook for Systematic Reviews of InterventionsH](#), section 6.6

This bias arises due to differences in the way that outcomes are assessed. Detection bias occurs at the point of follow-up (that is, at the time of outcome assessment).

Minimising detection bias:

- Detection bias can be minimised when the outcome assessor is **blind** to participant groups. A lack of blinding can exaggerate the estimated effect of treatment. It is especially important that outcome assessors be blind to treatment allocation in cases where the outcomes are subjective (eg pain, satisfaction)¹⁸.

2.3.4.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS

Consider:

- Were outcome assessors blind to the treatment allocation?
-

2.3.5 OTHER SOURCES OF BIAS

- See *Egger et al (2003) and Song et al (2004) for more information*

There are many other sources of bias in studies of healthcare interventions. These arise whether the researcher is aware of them or not. Use of the term 'bias' does not imply that the researcher is actively seeking to distort the study's results.

For example, various types of bias can be introduced when conducting a systematic review if the search strategy is not comprehensive¹⁹. These different types of bias include the following:

- **Publication bias:** which occurs when trials in which the intervention is shown to be no more effective than the control are less likely to be published (that is, 'negative' studies are less likely to be published by many journals). Systematic reviews that do

¹⁸ Note that allocation concealment helps to prevent selection bias and protects the randomisation sequence *before* and *until* the interventions are given to the study participants. In contrast, blinding helps to prevent detection bias, by protecting the randomisation sequence *after* allocation.

¹⁹ **Publication and associated biases** can be formally evaluated using various statistical techniques, including **funnel plots**. See the [HCochrane Collaboration Open Learning MaterialsH](#), Module 15 for details; also *CRD Report Number 4: Phase VII, section 2.7.6* ([Hwww.york.ac.uk/inst/crd/report4.htm](http://www.york.ac.uk/inst/crd/report4.htm)) for further details.

not include unpublished studies may be more likely to overestimate the effects of the intervention.

- **Duplication bias:** which occurs when trials are published more than once and the results are counted as though they came from different studies. Different aspects of a trial are often published separately, and may include different parts of the complete data set. They can be mistakenly attributed as different studies, especially if the initial report does not cover the entire study period.
- **Location bias:** which occurs when new or exciting results are more likely to be published in prestigious and/or well-known journals or books, while results that are thought to be less exciting may be published only in second-tier or local journals that may not be indexed to major databases.
- **Language bias:** which occurs when there is failure to search for and retrieve studies in languages other than English.

While there is empirical evidence that certain types of bias can substantially alter study results, the effects of other biases on results are not yet clear. This complicates the assessment of study quality, as ideally we only want to evaluate those sources of bias that have been shown empirically to affect results. While the effects of some potential sources of bias on study results are not yet clear, it is logical to suspect that certain biases, particularly selection bias, performance bias, attrition and detection biases, may influence study results. Hence, it is important to systematically and transparently assess studies against criteria that evaluate the likely impact of each of these biases.

- For more information, see the [HCochrane Handbook for Systematic Reviews of Interventions](#) sections 6.2 to 6.9
- See also *CRD Report Number 4: Phase V and Phase VII* ([Hwww.york.ac.uk/inst/crd/report4.htmH](http://www.york.ac.uk/inst/crd/report4.htm))

2.4 INCLUDING ASSESSMENTS OF STUDY QUALITY IN SYSTEMATIC REVIEWS

- See the [HCochrane Handbook for Systematic Reviews of InterventionsH](#), Section 6.7
- See the [HCochrane Open Learning materialHs](#), Modules 7 and 9
- See *CRD Report Number 4: Phase V Study quality assessment*, available at

2.4.1 GENERAL POINTS ON THE ASSESSMENT AND INCORPORATION OF STUDY QUALITY

Many methods have been developed to assist people in assessing study quality, in particular for assessing the quality of randomised controlled trials (RCTs). Review authors can choose to assess and report study quality in many different ways. This may include assessment and reporting of quality for both RCTs and non-randomised studies, which are covered separately in following sections of this document. The guidance is based on that formulated by the Cochrane Collaboration, with additional information relevant to the scope of the Consumers and Communication Review Group.

The specifics of study quality and its incorporation into systematic reviews will often vary according to the type of interventions/questions being investigated by the review, but the following aims to give some general directions that can be followed by authors undertaking reviews through the Consumers and Communication Group. Specific aspects of quality assessment for studies of different designs are covered in discrete sections below. This is followed by some general advice to review authors on incorporating quality assessment into a systematic review.

Many scales have been developed for quantifying the quality of trials. These include both **quality scales**, where individual quality items or criteria are scored numerically and then combined to give an overall quality score; and **quality checklists**, where separate quality items or criteria are marked individually as done/ not done and do not have numerical values attached to them. Quality can also be assessed and reported via **individual quality items**. There is currently little agreement on the best method for scoring quality²⁰.

Although previous reviews may have reported quality by using an overall numerical score, this is no longer accepted as an optimal approach and the Consumers and

²⁰ There is even less consensus on the most appropriate quality scale to use. Many quality scales have been shown to have unreliable results; or to be limited in the information that they provide and how meaningful this information is to readers of reviews.

Communication Review Group will no longer accept reviews that report quality via a numerical quality score.

The approach adopted by authors is obviously dependent to some degree on the type of question being investigated. **At the Consumers and Communication Group, we require the use of a component-type approach to quality assessment; that is, one that reports specifically on individual quality items, rather than numerically scoring and summing these items to produce a quality score.** This approach ensures that reporting of individual aspects of quality is **clear** and **explicit** for each quality element. The approach is also suited to the incorporation of both randomised and non-randomised studies in systematic reviews, as this may allow greater ease of comparison between and across different study types.

2.4.1.1 SUMMARY: KEY POINTS FOR REVIEW AUTHORS

In the Consumers and Communication Review Group, we recommend an **individual quality item** approach for assessing study quality. This assessment should be reported explicitly in the review.

- We suggest that review authors use individual components of a quality checklist and score each as being done, not done or unclear (or other similar terms). This allows for accurate reporting of studies that includes the uncertainty arising from poor reporting of studies (see *Section 2.2.2* of this guide for more on this issue).
- **Review authors may include the following text (tailored as necessary) in the Methods section of their Protocol/Review:**

We will assess and report on the methodological quality of included studies in accordance with the guidelines of the Cochrane Consumers and Communication Review Group (Ryan 2007), which recommends the explicit reporting of the following individual quality elements for RCTs: randomisation; allocation concealment; blinding (participants, providers, outcomes assessors, data analysts); baseline comparability; follow-up; intention-to-treat analysis; validation of tools; and other sources of bias. If studies other than RCTs (that is, quasi-randomised controlled trials, CBA and ITS studies) are to be included in the review, we will additionally assess the quality of these studies systematically and according to the criteria outlined in the guidelines of the Cochrane Consumers and Communication Review Group. In all cases, two authors will independently assess the quality of included studies, with any disagreements resolved by discussion and consensus. We will contact study authors for additional information about the included studies, or for clarification of the study methods as required. We will incorporate the results

of the quality assessment into the review through systematic narrative description and commentary about each of the quality items, leading to an overall assessment the quality of included studies and a judgement about the internal validity of the review's results.

- The Review Group's generic data extraction template for review authors is available at: www.latrobe.edu.au/cochrane/resource.html. This template includes a section on study quality assessment, using specific items to evaluate the study's avoidance of bias.
- The template, in its generic form, is most suited to the quality assessment of RCTs; however, elements can be adapted for use in the assessment of non-randomised studies. The following section (*Section 2.5*) deals predominantly with RCT quality assessment; for assessment of study quality for quantitative studies other than RCTs and qualitative studies respectively, see *Section 2.6* and *Section 2.7*.
- We suggest that studies with diverse designs be addressed separately in the review. Study quality should be assessed differently for different study designs as different criteria are relevant. Likewise, we recommend that results be treated separately for different types of studies. For the sake of clarity, to readers of the review, the reporting of the results might follow a similar structure to that used in the quality assessment section of the review. This issue will be covered in more detail in the guide on results and analysis (*still in development at 5 March 2007*).
- We will **not** accept numerically scored quality checklists.

2.4.2 BASELINE IMBALANCES IN GROUPS: POINTS FOR REVIEW AUTHORS TO CONSIDER

- See *EPOC Methods Paper: Issues Related to Baseline Measures of Performance*, available at [Hwww.epoc.uottawa.ca/methods.htm](http://www.epoc.uottawa.ca/methods.htm) or email

Baseline measures of performance are measures taken prior to the introduction of the intervention. They can be important because they allow an estimation of the size of the problem that is being studied. For example, if performance or outcome scores are low at the start of the study (prior to the intervention), this might suggest that there is a lot of room for improvement in performance or the outcome. In contrast, if scores are high at baseline (before the intervention) this may suggest that there is little room for improvement and that a ceiling effect may impact on the results of the study (ie. the results can only improve up to a certain point, as they are not measured beyond that point). This can distort the results, as the measure of the outcome is not appropriate for assessing the whole spectrum of responses or outcomes possible.

Baseline comparability of groups is important in comparative studies, whether they are randomised or non-randomised. While the following sections deal with quality assessment issues associated with particular study designs, baseline imbalance has implications for all comparative studies, and should be explicitly assessed for **all** comparative studies.

It is also important for review authors to realise that although RCTs which randomly assign participants should theoretically produce equivalent groups, this may not always be the case, especially if the numbers of participants or clusters that are randomised is small. The likelihood of baseline imbalance is even greater when the allocation of participants to groups is not random. Baseline comparability of groups therefore always needs to be assessed as a potential source of bias in reviews of the effects of interventions, whether simple or complex.

For more information on baseline imbalance, worked examples and methods for adjusting for differences between groups at baseline, see *EPOC Methods Paper: Issues Related to Baseline Measures of Performance*, available at www.epoc.uottawa.ca/methods.htm or email amayhe2@uottawa.ca

2.5 QUALITY ASSESSMENT: RANDOMISED CONTROLLED TRIALS

- See *Consumers and Communication Review Group data extraction template*, available at: [Hwww.latrobe.edu.au/cochrane/resource.html](http://www.latrobe.edu.au/cochrane/resource.html)^H
- See *HCochrane Handbook for Systematic Reviews of Interventions*^H, Section 6.7

RCTs are considered to represent the 'gold standard' study design for investigating the effects of interventions. This is because their design helps to avoid or to minimise many different kinds of bias²¹ (see **Study Design Guide** for more details). However, it is important to remember that not all RCTs are well-designed or well-conducted. As a result, it is possible to have RCTs of variable quality. This in turn affects the strength of the evidence that can be drawn from their findings (see **Section 2.2.2: Study design versus study quality**).

2.5.1 QUALITY ITEMS TO BE ASSESSED: RCTS

The assessment of the quality of RCTs involves evaluating whether the study has avoided various types of bias. Because there are many sources of potential bias, this assessment needs to be carried out systematically and transparently. We recommend that you use the Review Group's [data extraction template](#) as a starting point, and consider the following as a basis for your quality assessment. You may also choose to build additional criteria into your evaluation, if you consider them to be important to the overall design and conduct of the study²².

Each of the following items should be assessed when considering the quality of a study included in a systematic review. Each of these items should be rated as done, not done, or unclear.²³ The rules used to make each of these decisions must be **explicitly stated** in your review. For example, if authors stated that participants were randomly assigned but gave no details of the methods used, you may wish to rate the study as 'unclear' on randomisation method: this decision should be clearly stated in the review.

Further, review authors are encouraged to provide a brief description of what the triallists reported in relation to each quality element, using verbatim quotes from the trial report where appropriate.

²¹ Note that bias is taken here to mean the various different sources of bias covered in earlier sections of this guide.

²² Note that the data extraction template is a generic one and this form is not directly applicable to all reviews coordinated through the Consumers and Communication Review Group. However, it can be adapted to suit particular reviews, and review authors should consider which additional data may be useful to extract from studies (in terms of the interpretation of their results, their generalisability beyond the study population, the context to which the results may be applied and so on. These issues are covered in detail in the forthcoming *Study Analysis Guide*).

²³ Or other similar wording can be used; for example, each item could be rated as yes, no or unclear.

(i) Randomisation:

- Was the method really random?
- What method was used to generate the random sequence?

Consider:

- Did each participant in the study have a truly equal chance of being assigned to either/ any group in the trial?
- Were the methods truly random? (See **Study Design Guide, Section 1.3.2.**)

Action:

- *Rate this point as done, not done or unclear (or similar wording). Describe or quote relevant text from the trial report, if appropriate.*
- **NOTE:** *If your review is to include only true randomised controlled trials (RCTs) then you must **EXCLUDE** any studies for which randomisation is rated as 'not done' or 'unclear'. Seek clarification from trial authors for any studies which you initially rate as 'unclear.'*

(ii) Allocation concealment²⁴:

- Was allocation of participants to groups adequately concealed?
- What method was used to conceal allocation?

Consider:

- Was group assignment (the sequence generated by randomisation) concealed from those involved in the trial until the point of allocation?
- Was it possible that the allocation sequence could have been tampered with or influenced?
- Were the methods adequate to prevent this? See **Section 2.3.1** of this guide, **Allocation concealment** (under *Selection bias*).

Action:

- *After considering the points above, rate allocation concealment as one of the following:*
 - *A Adequate*
 - *B Unclear*
 - *C Inadequate*
 - *D Not used*

Describe or quote relevant text from the trial report, if appropriate.

(iii) Blinding:

²⁴ Note that allocation concealment is often thought to be the most important aspect of study quality, as failure to adequately do so has been shown empirically to significantly distort the results of studies (overestimate the effects of the intervention). Blinding has also been shown to alter the results of studies. See **Appendix A** for further information on blinding and allocation concealment.

- Were the following people involved in the trial adequately blinded to group allocation?
 - Participants?
 - Providers?
 - Outcome assessors?
 - Data analysts?
- What methods were used and were they effective?²⁵

Consider:

- Who involved in the study was blinded? That is, who was unaware of group allocation after participants had been assigned to groups?
- How was blinding achieved and was it effective?
- We suggest that you report on each component of blinding that is relevant to studies included in your review.²⁶ See also **Section 2.3.2 Blinding** (under *Performance bias*).

Action:

- *Rate the blinding of each group (participants, providers, outcome assessors, data analysts) involved in the trial separately as done, not done or unclear (or other similar wording). Describe or quote relevant text from the trial report, if appropriate.*

(iv) Baseline comparability:

- Were the groups comparable at the start of the study on important characteristics?

Consider:

- Which characteristics are most important in relation to the question that your review is trying to answer?

In considering this you might want to think about the characteristics of participants that might be expected to influence the effect of the intervention if they are not accounted for (that is, if the groups were not balanced on these characteristics at the start of the study, how they might affect the outcomes of the study). These factors might include characteristics such as the age or sex of participants, the disease severity, previous episodes of medical care, literacy, or any other factors that you identify as relevant.

Action:

- *Specify the important baseline characteristics of your study sample.*

²⁵ Note that while it is always possible to adequately conceal allocation, it is not always possible to blind all people involved in the study. For example, it may not always be possible to blind the providers of the intervention as to who is receiving the intervention and who is not. However, it is still important to explicitly make note of this in your review. For example, you may state that it was not possible to blind providers, and hence, this was not formally assessed, but that this may be a source of bias in (some or all of the) included studies.

²⁶ Review authors may wish to omit the assessment of data analyst blinding from their assessment unless it is especially important to their review topic and the types of analyses included in studies. This decision should be explicitly reported in the review text.

- *Assess the equivalence of groups on these characteristics at baseline. It is especially important for you to note any significant differences between the study groups; these may indicate a source of bias in the study.*
- *Describe or quote relevant text from the trial report, if appropriate.*

(v) Follow-up:

- Was follow-up complete? What level of follow-up is acceptable (should be decided *a priori*)?²⁷
- How many participants completed the study for each group (and each outcome)?
- How many participants did not complete the study for each group (and each outcome)?

Consider²⁸:

- The point of randomisation is to produce groups that are equivalent at the start of the study (baseline). Ideally there should not be systematic differences between study groups in the withdrawals or exclusions of people enrolled in the study. As participants who are not followed up, do not complete the study or who are not included in the assessment of outcomes may be different from those completing the study and included in outcome assessments, this can be a source of bias.
- Were the study groups comparable at the end of the study?
- Were those followed up comparable to those who dropped out of the study?
- You may want to consider whether particular characteristics of participants made them more or less likely to withdraw or be lost to follow-up from the study.

Action:

- *Assess and specify how many participants completed the study; calculate the number that was lost over the course of the study for each group and for each outcome of interest.*
- *If possible, specify whether those who were lost from the study or dropped out were different to those who stayed in the study. Note any differences you detect as these may indicate a source of bias. Describe or quote relevant text from the trial report, if appropriate.*

(vi) Intention-to-treat analysis:

- Were participants analysed in the groups to which they were assigned? That is, were all participants analysed according to their allocated group, not by whether they actually received the intervention, or whether they remained in the study?

²⁷ You should also state how you determined whether follow-up was relatively complete (that is, what do you mean by 'complete?'). For example, you may decide *a priori* to exclude studies with loss to follow-up of >40%. This should be **clearly stated** in your review and a **rationale** provided.

²⁸ See **Appendix A** this guide for more on **participant flow** through trials and **attrition bias**.

Consider:

- Was ITT analysis used?²⁹
- If ITT analysis was not used, how were those who dropped out treated in the analyses?
- If there has been no ITT analysis, is there sufficient information to enable data to be reanalysed by ITT?

Action:

- *Specify the following:*
 - *The number of participants lost from each group for each outcome.*
 - *How losses were treated in the analysis of the study results. (Describe or quote relevant text from the trial report, if appropriate.)*
- *Rate the study as to whether ITT analysis was used (done, not done, unclear, or other similar wording)*
- *If sufficient information is provided, consider reanalysis of data based on ITT principles.*

(vii) Validation of tools:

- Were the tools used to measure the outcome validated? You may wish to consider whether the tools used are commonly used by the research community, and whether the tools are appropriate to the specific context of the study.

Consider:

- How was each of the outcomes assessed?
- Was each tool evaluated for its ability to measure what it was supposed to? How was this done?

Action:

- *Specify the following:*
 - *Each of the outcomes assessed by the study.*
 - *Whether each of the tools used were validated; and rate each as done, not done or unclear (or other similar wording). Describe or quote relevant text from the trial report, if appropriate.*

²⁹ See **Appendix A** this guide for more on ITT analysis.

2.6 QUALITY ASSESSMENT: NON-RANDOMISED CONTROLLED STUDIES

- See the [HCochrane Handbook for Systematic Reviews of InterventionsH](#), Section 6.8 and Appendix 6a
- See EPOC resources, available at [Hwww.epoc.uottawa.caH](http://www.epoc.uottawa.ca) or email [Hamayhe2@uottawa.caH](mailto:Hamayhe2@uottawa.ca)
- See Deeks et al (2003), available at [Hwww.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=1117](http://www.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=1117)

Studies of different design have different issues associated with their design and conduct. Criteria used to evaluate the quality of RCTs cannot be applied directly to studies of other designs. However, elements of the quality criteria used to assess RCTs can be adapted to systematically assess non-randomised studies. Reporting quality via a component approach, as suggested in the previous section, also helps to **standardise** the reporting of quality both within and across systematic reviews coordinated by the Consumers and Communication Review Group. Apart from anything else, it helps to make the process of assessing quality a transparent and explicit one.

2.6.1 NON-RANDOMISED CONTROLLED STUDIES (INCLUDING QUASI-RANDOMISED STUDIES)

Because controlled studies that are not randomised are similar in their design in many respects to RCTs, they can be assessed using a similar but slightly adapted set of criteria. Assessment of the randomisation method and suitability is obviously not relevant for a non-randomised study. However, as allocation to intervention and control groups is not determined by randomisation, review authors should pay particular attention to how the groups are chosen, the possibility of bias arising from the non-random allocation of participants, and the potential influence of baseline differences in groups upon the outcomes of the study.

The following are adapted from the guidelines for quality assessment of non-randomised controlled trials developed by EPOC (see www.epoc.uottawa.ca or email amayhe2@uottawa.ca) and from the Consumers and Communication Review Group data extraction template and guidelines (available at www.latrobe.edu.au/cochrane/resource.html).

(i) Allocation concealment:

- Was allocation of participants to groups adequately concealed?
- What method was used to conceal allocation?

Consider:

- Was group assignment concealed from those involved in the trial up until the point of allocation?
- Were the methods adequate to prevent this? See **Section 2.3.1** of this guide, **Allocation concealment** (under *Selection bias*).

Action:

- *After considering the points above, rate allocation concealment as one of the following:*
 - *A Adequate*
 - *B Unclear*
 - *C Inadequate*
 - *D Not used*

Describe or quote relevant text from the trial report, if appropriate.

(ii) Blinding:

- Were the following people involved in the trial adequately blinded to group allocation?
 - Participants?
 - Providers?
 - Outcome assessors?
 - Data analysts?
- What methods were used and were they effective?

Consider:

- Who involved in the study was blinded? That is, who was unaware of group allocation after the groups had been assigned?
- How was blinding achieved and was it effective?
- We suggest that you report separately on each component of blinding that is relevant to studies included in your review. See also **Section 2.3.2 Blinding** (under *Performance bias*).

Action:

- *Rate the blinding of each group (participants, providers, outcome assessors, data analysts) involved in the trial separately as done, not done or unclear (or other similar wording).*
- *Describe or quote relevant text from the trial report, if appropriate.*

(iii) Baseline comparability:

- Were the groups comparable at the start of the study on important characteristics?

Consider:

- Which characteristics are most important in relation to the question that your review is trying to answer? You should consider which characteristics of participants might influence the effect of the intervention if not accounted for (that is, if the groups were not balanced on these characteristics at the start of the study, that they might affect the outcomes of the study). Such characteristics might include the age or sex of participants, disease severity, previous episodes of medical care, literacy, or other factors that you identify as relevant.

Action:

- *Specify the important baseline characteristics of your study sample.*
- *Assess the equivalence of groups on these characteristics at baseline. It is especially important for you to note any significant differences between the study groups, as these may be a source of bias in the study. Describe or quote relevant text from the trial report, if appropriate.*

(iv) Follow-up:

- Was follow-up complete? What level of follow-up is acceptable (should be decided *a priori*)?
- How many participants completed the study for each group (and each outcome)?
- How many participants did not complete the study for each group (and each outcome)?

Consider:

- Groups should be comparable both at the start and at the end of a study. Participants who are not followed up, do not complete the study or who are not included in the assessment of outcomes may be different from those completing the study and included in assessments. This can be a source of bias.
- Were the study groups comparable at the end of the study?
- Were those followed up comparable to those who dropped out?
- You should consider whether there were particular characteristics of participants that made them more or less likely to withdraw or be lost to follow-up.

Action:

- *Assess and specify how many participants completed the study; calculate the number that was lost over the course of the study for each group and for each outcome of interest.*
- *If possible, specify whether those who were lost or dropped out from the study were different to those who stayed in the study. Note any differences that you*

detect; these may be a source of bias. Describe or quote relevant text from the trial report, if appropriate.

(v) Intention-to-treat (ITT) analysis:

- Were the groups analysed in the groups to which they were assigned? That is, were they analysed according to whether they were allocated to the intervention or control group, regardless of whether they actually received the intervention, whether they dropped out or did not complete the study?

Consider:

- Was ITT³⁰ analysis used?
- If ITT analysis was not used, how were those who dropped out treated in the analyses?
- If ITT analysis was not used, is there sufficient information for data to be reanalysed based on ITT?

Action:

- *Specify the following:*
 - *The number of participants lost from each group for each outcome.*
 - *How losses were treated in the analysis of the study results.*
- *Rate the study as to whether ITT analysis was used (done, not done, unclear; or other similar wording). Describe or quote relevant text from the trial report, if appropriate.*
- *If sufficient information is provided, consider reanalysis of the data based on ITT principles.*

(vi) Validation of tools:

- Were the tools used to measure the outcome validated? You may wish to consider whether the tools used are commonly used by the research community, and whether the tools are appropriate to the specific context of the study.

Consider:

- How was each of the outcomes assessed?
- Was each tool evaluated for its ability to measure what it was supposed to? How was this done?

Action:

- *Specify the following:*
 - *Each of the outcomes assessed by the study.*
 - *Whether each of the tools used were validated; and rate each as done, not done, or unclear (or other similar wording). Describe or quote relevant text from the trial report, if appropriate.*

³⁰ See [HAppendix AH](#) this guide for more on ITT analysis.

2.6.2 CONTROLLED BEFORE-AND-AFTER STUDIES (CBAS)

See EPOC resources, available at [Hwww.epoc.uottawa.caH](http://www.epoc.uottawa.ca) or by emailing [Hamayhe2@uottawa.caH](mailto:Hamayhe2@uottawa.ca); in particular:

- Data collection template ([Hwww.epoc.uottawa.ca/data%20abstraction%20form%20-%202002.docH](http://www.epoc.uottawa.ca/data%20abstraction%20form%20-%202002.doc))
- Data collection checklist ([Hwww.epoc.uottawa.ca/checklist2002.docH](http://www.epoc.uottawa.ca/checklist2002.doc))

Tailored advice for review authors with the Consumers and Communication Review Group is currently under development for controlled before-and-after studies. Authors are referred to the guidance provided by EPOC on CBA quality assessment until these resources become available in full.

The EPOC Review Group specifies that, to be included in a systematic review, the CBA study design must meet two key criteria. These are:

- The timing of the periods for study for the control and intervention groups should be comparable (that is, the pre- and post- intervention periods of measurement for the control and intervention groups should be the same).
- The intervention and control groups should be comparable on key characteristics.

CBA studies that do not meet these criteria should be excluded from a Cochrane review.

If CBA studies meet these criteria, they are eligible (at least on study design criteria) for inclusion in a systematic review and so need to be systematically assessed for quality.

The EPOC guidelines for quality assessment of CBA studies are included in a generic template for data extraction developed by the group. This template includes elements of quality assessment specific to the CBA study design.

Note that the scope of EPOC focuses more on healthcare structures and organisational aspects than does the scope of the Consumers and Communication Review Group. Review authors should bear this in mind, and will need to **adapt** the general guidance contained in the EPOC template to suit the needs of their specific review question. It may, for example, be appropriate in many cases for review authors to adopt the quality assessment sections of the data extraction template (with appropriate adaptation) developed by EPOC and to include these quality assessment criteria in the Consumers and Communication Group's data extraction template (that is, to replace the quality assessment criteria on the Consumers and

Communication Group's data extraction template, which are most appropriate for RCTs, and adapt and insert the EPOC quality assessment guidelines for controlled before-and-after studies in place of these).

Action:

- *Examine the data extraction template formulated by EPOC.*
- *Adapt this template appropriately to suit the question addressed by your review. It may be appropriate to adapt the quality assessment criteria section of the EPOC data extraction template only, and to substitute this in place of the quality assessment criteria contained in the generic Consumers and Communication Group data extraction template. Seek advice from the editorial base if needed.*
- *Included studies should be rated on each of the quality assessment criteria. Each criteria should be rated as done, not done, or unclear.*

2.6.3 INTERRUPTED TIME SERIES (ITS) STUDIES

Tailored advice for review authors with the Consumers and Communication Review Group is currently under development for ITS studies. Authors are referred to the guidance provided by EPOC on ITS quality assessment until these resources become available in full.

See EPOC resources, available at [Hwww.epoc.uottawa.ca](http://www.epoc.uottawa.ca) (or email Hamayhe2@uottawa.ca); in particular:

- *Data collection template*
([Hwww.epoc.uottawa.ca/data%20abstraction%20form%20-%202002.doc](http://www.epoc.uottawa.ca/data%20abstraction%20form%20-%202002.doc))
- *Data collection checklist* ([Hwww.epoc.uottawa.ca/checklist2002.doc](http://www.epoc.uottawa.ca/checklist2002.doc))
- *EPOC Methods Paper: Including Interrupted Times Series (ITS) designs in an EPOC review* ([Hwww.epoc.uottawa.ca/inttime.pdf](http://www.epoc.uottawa.ca/inttime.pdf))

The EPOC Review Group specifies that, to be included in a systematic review, studies of the ITS design must meet two key criteria. These are

- There must be a clearly defined point in time at which the intervention occurred, and this should be reported by the researchers.
- There should be collection of at least three data points before and three after the intervention was introduced.

ITS studies that do not meet these criteria should be excluded from a Cochrane review.

If CBA studies meet these criteria, they are eligible (at least on study design criteria) for inclusion in a systematic review and so need to be systematically assessed for quality.

The EPOC guidelines for quality assessment of ITS studies are included in a generic template for data extraction developed by the group. This template includes elements of quality assessment specific to the ITS study design.

Note that the scope of EPOC focuses more on healthcare structures and organisational aspects than does the scope of the Consumers and Communication Review Group. Review authors should bear this in mind, and will need to **adapt** the general guidance contained in this template to suit the needs of their specific review question. It may, for example, be appropriate in many cases for review authors to adopt the quality assessment sections of the data extraction template (with appropriate adaptation) developed by EPOC and to include these quality assessment criteria in the Consumers and Communication Group's data extraction template (that is, to replace the quality assessment criteria on the Consumers and Communication Group's data extraction template, which are most appropriate for RCTs, and adapt and insert the EPOC quality assessment guidelines for ITS studies in place of these).

Action:

- *Examine the data extraction template formulated by EPOC.*
- *Adapt this template appropriately to suit the question addressed by your review. It may be appropriate to adapt the quality assessment criteria for ITS section of the EPOC data extraction template only, and to substitute this in place of the quality assessment criteria contained in the generic Consumers and Communication Group data extraction template.*
- *Seek advice from the editorial base if needed.*
- *Included studies should be rated on each of the quality assessment criteria. Each criteria should be rated as done, not done, or unclear.*

2.7 QUALITY ASSESSMENT: QUALITATIVE STUDIES

- See CRD Report Number 4 Phase V, [Hwww.york.ac.uk/inst/crd/report4.htm](http://www.york.ac.uk/inst/crd/report4.htm)
- See also Spencer et al (2003). *Quality in Qualitative Evaluation: A framework for assessing research evidence*, [Hwww.strategy.gov.uk/downloads/su/qual/downloads/qqe_rep.pdf](http://www.strategy.gov.uk/downloads/su/qual/downloads/qqe_rep.pdf)
- See also Cochrane Health Promotion and Public Health Field. *Systematic Reviews of Health promotion and Public Health Interventions*, [Hwww.vichealth.vic.gov.au/cochrane](http://www.vichealth.vic.gov.au/cochrane)

Sometimes it is both appropriate and useful to include qualitative data as part of a systematic review. It is now becoming more common to include data from qualitative research alongside quantitative data in systematic reviews of effectiveness. Qualitative data can, for example, help to provide depth and explanation for observed outcomes; help to decide whether the intervention of interest is suitable for a particular target population; help to examine factors that might have influenced the results if the effect of the intervention is not in the expected direction; and so on. Such contributions can add meaning to the results of a systematic review and can help to address some of the questions which are likely to be important to the users of systematic reviews (for details, see **Study Design Guide**, *Quantitative and qualitative designs*).

Note that inclusion of qualitative data in the analysis section of a review must be clearly justified by the authors and is subject to editorial approval.

Just as it is desirable to adopt a structured approach to the quality assessment of quantitative research, clear and transparent approaches to the assessment of qualitative research also need to be adopted. Several frameworks for assessing the quality of qualitative research now exist³¹.

Like quantitative research, qualitative research needs to be transparently and systematically assessed. There is, however, little consensus as yet on appropriate hierarchies of evidence arising from qualitative research. There is also little agreement on whether the criteria against which to assess qualitative data should attempt to parallel those developed for quantitative data.

³¹ For more, see *Quality in Qualitative Evaluation: A framework for assessing research evidence*. Available at [Hwww.strategy.gov.uk/downloads/su/qual/downloads/qqe_rep.pdf](http://www.strategy.gov.uk/downloads/su/qual/downloads/qqe_rep.pdf), section 6; see also CRD Report Number 4 Phase V, Box 5.11.

As the conduct and design of qualitative research is also fundamentally different to quantitative research, in that qualitative research does not attempt to 'answer' pre-formulated research questions, there is also little consensus on when quality assessment of qualitative research should occur (that is, whether it should occur prior to or during the synthesis of data from included studies). Despite these differences, there is agreement that the aim of quality assessment of qualitative research should be to adopt a structured and transparent approach.

The Cochrane Qualitative Research Methods Group and other groups worldwide are in the process of developing standards for the assessment and reporting of qualitative data. As there is presently little consensus as to how to evaluate the quality of qualitative data, guidelines and recommendations for review authors will be developed and this document updated when they become available.

Tailored advice for review authors with the Consumers and Communication Review Group is currently under development for qualitative data. In the meantime, authors should contact the editorial group directly if they wish to incorporate qualitative data into their review.

2.8 HOW TO REPORT STUDY QUALITY: TIPS FOR REVIEW AUTHORS

2.8.1 GENERAL POINTS ON REPORTING QUALITY

- See also CRD Report Number 4 Phase V paper Section 2.5.4, available at [Hwww.york.ac.uk/inst/crd/report4.htm](http://www.york.ac.uk/inst/crd/report4.htm)

A structured and explicit approach to assessing and reporting quality is an important component of a systematic review. The frameworks outlined in previous sections help to guide authors in structuring their assessment of included studies. However, these guidelines provide only a structure or framework with which to help build a systematic report on the quality of included studies. While this information must be included in the review (for example, as a component of the 'Table of Included Studies' or as a separate table if the inclusion of more detail is necessary), this information is *not* all that is needed.

The text of the systematic review should attempt to integrate the quality assessment by drawing out key themes, emphasising important similarities and key differences between studies and relating them back to the overall aims of the systematic review (that is, the assessment of study quality should be systematically related back to the questions that the review is trying to answer). As a result, **we discourage review authors from simply reporting a quality 'grade.'** In part, this is because only individual quality items, such as allocation concealment, have been clearly related to measured treatment effects, and scoring systems have been shown to yield unreliable results (Juni, 1999). Commenting clearly on individual quality items can therefore help the reader to obtain a clear picture of the evidence that is included in the review.

We encourage review authors to comment on individual quality items for each of the included studies, and also to attempt to **integrate** the quality assessment of included studies into the wider aims of the review. This is important, as the point of performing quality assessment is to come to an overall decision about how believable the evidence you have collected and analysed is, and how applicable it might be to the real world. This type of synthesis therefore involves considering both the internal and external validity of the studies included in the review.

In terms of internal validity, as well as discussing the individual quality items considered for different studies, you might attempt to draw this together by considering some of the following, or similar, types of questions:

- How confident are you that the results of the included studies reflect what is true?

- Are there limitations in the design or execution of the studies that might render the results less believable? How severe are these limitations?
- Were there aspects of the studies that were particularly well designed and so increase your confidence in the results?

In terms of the external validity or generalisability of the studies, you might wish to consider and discuss some of the following types of questions when considering the body of evidence you have gathered and appraised as a whole:

- What works for whom? Which groups of the population do the results apply to? (eg. Was the study only of children? Of adults? Of people with a particular disease but no comorbidities?)
- Has the intervention been evaluated in a particular setting only? (eg. Was the intervention delivered in a hospital setting only? Are the results likely to be applicable to the community setting?)
- Do some versions of the intervention work better or best? (eg. Is the intensity or frequency of delivery of the intervention likely to affect results? Are some versions ineffective? Are some versions effective only in some people?)

2.8.1.1. SUMMARY: KEY POINTS FOR REVIEW AUTHORS

We encourage authors to attempt to integrate their quality assessment of individual studies into a narrative synthesis. This might include some or all of the following:

- **Narratively describe** and comment upon the quality issues associated with the included studies for each of the quality items. Some issues to consider when trying to formulate an overall description of the body of evidence gathered in the review might include the following:
 - Did most of the included trials report adequate detail on their methods of randomisation (or other allocation methods)?
 - Have you had to assume that they used adequate methods due to lack of information?
 - Did many studies adequately conceal allocation, and what methods did they use? How does this compare to those that did not adequately conceal allocation?
 - Was blinding done adequately (for participants, providers, outcome assessors and analysts)?; or was it not possible to blind all of those involved in the study?
 - Were outcomes assessed using validated tools? Were there differences between studies' findings depending on the tools used to measure outcomes? Were the tools validated for use in the population (or for the condition) studied?
 - Was follow-up of participants throughout the study period described; and was it adequate? Was analysis ITT or was there a large attrition of participants?

- Draw some general comments on the **overall quality** of studies included in the review. The purpose of doing so is to help to set the scene for the Results and Discussion sections of the review. As the results of the studies need to be interpreted in the light of the quality of the studies, this is a critical part of your review. Some of the things you may wish to consider might include:
 - Are there particular limitations in the design, execution or reporting of the included studies that render the results less believable?
 - How severe are these limitations?
 - Were there aspects of the included studies that were particularly well designed and so increase your confidence in the results?

It may be appropriate at this point, for example, to highlight any particular strengths and/or limitations of the included studies. The main focus should be to discuss and attempt to narratively synthesise the issues associated with the quality of the included studies.

- For reviews where studies other than RCTs have also been included, we suggest that studies of different designs be discussed as **separate sections** in terms of their quality and each of their respective strengths and weaknesses.
 - For example, it might be appropriate to discuss the quality of RCTs and non-randomised controlled studies separately: as their designs are different, the discussion will need to deal with different aspects of quality relating to design. Once this is done, it may then be appropriate to contrast the different types of studies, and to draw comparisons between the different categories.
 - It can be especially interesting, for example, to compare the findings of studies of different design and varying levels of quality to comment on the consistency of the findings of studies included in the review. For example, do non-randomised studies report the same direction and size of results as RCTs? Do high quality studies report the same direction and size of results as lower quality studies?
-

APPENDIX A –ADDITIONAL SOURCES OF INFORMATION

(SEE ALSO APPENDIX C IN THE STUDY DESIGN GUIDE)

ALLOCATION CONCEALMENT AND SUBVERTING RANDOMISATION

Schulz KF. Subverting randomisation in controlled trials. *JAMA* 1995; 274(18): 1456-8.

- This paper discusses the necessity of adequate allocation concealment in order to adequately randomise participants to groups in a RCT, giving examples, recommendations and further references on adequate approaches.

ATTRITION BIAS

Tierney JF. and Stewart LA. Theory and Methods: Investigating patient exclusion bias in meta-analysis. *International Journal of Epidemiology* 2005; 34 (1): 79-87.

Juni P. and Egger M. Commentary: Empirical evidence of attrition bias in clinical trials.

International Journal of Epidemiology 2005; 34 (1): 87-8.

- This study (and commentary) analyses the effects of participant exclusion from trials (post-randomisation), and directly analyses the effects of attrition bias upon trial effect estimates.
- It includes a useful glossary of terms.
- This study demonstrates that pooled analysis of trials that excluded patients post-randomisation may be prone to bias. The authors conclude that analysis (in trials, meta-analyses and systematic reviews) should be based on all participants randomised at the start of the trial.

BASELINE PERFORMANCE MEASURES

Cochrane Effective Practice and Organisation of Care (EPOC) Review Group. *EPOC Methods Paper: Issues Related to Baseline Measures of Performance* (www.epoc.uottawa.ca or email amayhe2@uottawa.ca)

- This paper gives further detail on issues associated with baseline performance measures in primary studies. It includes a rationale for measuring baseline values, the problem of imbalances at baseline, and a worked example of baseline imbalance.
- It also provides guidance on ways of adjusting analysis and data extraction to include baseline values; and how to incorporate information about baseline values into a review.

BLINDING

Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, Bhandari M and Guyatt GH. Physician interpretations and textbook definitions of blinding terminology in randomised controlled trials. *JAMA* 2001; 285(15): 2000-5.

- As physicians commonly use study blinding as a quality assessment criteria for studies they read, this study examined the definitions and interpretations of blinding used by physicians and by textbooks.

- The study reports little consensus about the use of blinding and associated terms (single, double, etc); and discusses the findings in terms of implications for the quality of trials.

INCONSISTENCY: HETEROGENEITY AND HOMOGENEITY OF EVIDENCE

Higgins JPT, Thompson SG, Deeks JJ and Altman DG. Measuring inconsistency in meta-analysis. *BMJ* 2003; 327: 557-60.

- This article discusses the concept of consistency of results in meta-analysis, the desirability of consistent results across studies; and the need for tests of heterogeneity in order to assess how generalisable the results of a meta-analysis are.
- Specifically, this paper discusses the merits of the I^2 test for heterogeneity among the results of studies (which is included in Cochrane systematic reviews with meta-analysis); and compares this test of consistency with other measures.

INTENTION-TO-TREAT (ITT) ANALYSIS

Heritier SR, Gebiski VJ and Keeck AC. Inclusion of patients in clinical trials analysis: the intention-to-treat principle. *MJA*, 2003; 179: 438-40.

- This article provides a definition and examples of ITT analysis. It examines some of the advantages and the disadvantages of ITT analysis. This also includes examples where ITT analysis is not a highly desirable option for analysis.

Also refer to the studies listed under *Attrition bias*.

NON-RANDOMISED STUDIES AND QUALITY

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al.* (2003) Evaluating non-randomised intervention studies. *Health Technol Assess* 7(27).

www.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=1117

- This paper presents an empirical comparison of randomised and non-randomised studies. It concludes that results from these different study types are sometimes, but not always, different; and discusses cases where non-randomised studies may produce seriously misleading results.
- This study also evaluates quality assessment tools for non-randomised studies and identifies key components that should be included in any quality assessment.

Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. (2000) Bayesian methods in health technology assessment: a review. *Health Technol Assess* 4(34).

www.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=930

- This paper assesses the association between study (methodological) quality and estimates of effect sizes. In particular, it examines the differences between effect sizes generated by RCTs in comparison with quasi-randomised and observational studies. It discusses differences in effect estimates across various different non-randomised study designs.
- This study also recommends that further studies to directly compare the results of RCTs and quasi-randomised and observational studies be performed; and that further instruments be

developed to assess study quality and the accuracy of effect estimates in relation to study quality.

PARTICIPANT FLOW AND FOLLOW-UP

Cakir B, Gebiski VJ and Keech AC. Flow of participants in randomised studies. *MJA* 2003; 178: 347-9.

- Outlines the components of participant flow through a trial that should be included in a trial report and should be assessed when including such as study in a systematic review.
- This article draws clear distinctions between participant loss at different stages of a trial, and emphasises the implications of this for the quality of the study and the degree of confidence that should be placed in the results of the study.

PUBLICATION BIAS

Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000; 4(10) www.hta.nhsweb.nhs.uk/execsumm/summ410.htm

- This paper systematically examines and evaluates studies on methodological issues associated with publication bias. It attempts to identify empirical evidence about the existence and consequences of publication bias and other forms of dissemination-based biases.

PUBLIC HEALTH INTERVENTIONS

Weightman A, Ellis S, Cullum A, Sander L and Turley R. Grading evidence and recommendations for public health interventions: developing and piloting a framework, *Health Development Agency* 2005, www.hda.nhs.uk

- This paper is an extensive evaluation of different types of research design and their appropriateness for answering questions about the efficacy of public health interventions.
- This paper presents the results of the literature and development of the evidence grading framework, including a detailed description of the criteria used to assess studies. It also presents the results of the piloting of the provisional framework developed for grading evidence about public health interventions.

For further discussion of issues associated with public health interventions, also see the Cochrane Health Promotion and Public Health Field. *Systematic Reviews of Health Promotion and Public Health Interventions*. www.vichealth.vic.gov.au/cochrane.

QUALITY ASSESSMENT

Jüni P, Altman DG and Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42-6.

- This paper discusses the various different elements of quality that are necessary to assess for studies included in systematic reviews.

- It provides a detailed description of different sources of bias in studies of healthcare interventions; as well as a description of the ways in which study quality can be incorporated into meta-analysis and systematic reviews.

Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assessment* 1999; 3(12) www.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=943

- This paper examines and discusses the various different elements of quality that are necessary to assess for studies included in systematic reviews.
- It discusses in detail the different aspects of study quality; which are likely to be most important (ie. to impact the most upon results), and makes recommendations for assessing study quality as part of systematic review conduct.

Schultz KF, Chalmers I, Hayes RJ and Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273(5): 408-12.

- This review assessed the relationships between different aspects of trial quality and trial results.
- It reports empirical evidence that several elements of study quality (allocation concealment and blinding), if not adequately performed, tend to overestimate the estimates of effect in trials.

Undertaking systematic reviews of research on effectiveness. *CRD's guidance for those carrying out or commissioning systematic reviews*. CRD Report Number 4 (2nd Edition). NHS Centre for Reviews and Dissemination, University of York. March 2001 ;

(www.york.ac.uk/inst/crd/report4.htm); see Phase V Study quality assessment in particular.

- These resources give an overview of study quality issues, bias and why it is necessary to assess the quality of studies included in systematic review. It also provides ways of approaching quality assessment of studies, with discussion in terms of answering different types of research questions.

RANDOMISED CONTROLLED TRIALS

Jadad A. (1998) *'Randomised controlled trials: A user's guide.'* BMJ Books, London.

- This book provides a detailed overview of RCTs, from basic definitions of RCTS, methods of randomisation and different types of RCTs to the assessment of RCT quality, sources of bias and reporting of individual trials.

SEARCHING

Egger M. (2003) *The importance of comprehensive literature searches and quality assessment in systematic reviews: empirical study.* *HTA Methodology* 7(26).

www.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=1099

- This paper examines the characteristics of trials that are difficult to locate by searching. It also examines the effects of including only easily accessible studies, compared with more difficult to locate studies (that require comprehensive search techniques), on reported treatment effects and on pooled treatment effects (from meta-analysis).
- It concludes that trials that are difficult to locate tend to be smaller and of poorer methodological quality than those that are easily accessible; and that the inclusion or exclusion of trials of low methodological quality has a substantial impact on the results and the conclusions reached by systematic reviews and meta-analyses.

See also Jadad 1998.

CONSORT STATEMENT

Moher D., Schulz K.F. and Altman D.G. The CONSORT statement: revised recommendations for improving the quality of report of parallel-group randomised trials. *The Lancet* 2001; 357 (9263): 1191-8.

- This paper outlines the CONSORT checklist, which has been developed to improve the reporting of RCTs, so that readers will be better able to determine why the study was undertaken, and how it was conducted and analysed.

TREND STATEMENT

Des Jarlais D.C., Lyles C., Crepaz N. and the TREND Group. Improving the reporting quality of nonrandomised evaluations of behavioural and public health interventions: The TREND statement. *American Journal of Public Health* 2004; 94(3): 361-6.

- This paper notes that non randomised trials and other studies should be included in systematic reviews, as excluding all data from studies other than RCTs will bias the evidence base towards those interventions which are 'simpler' to evaluate (ie. they can be evaluated using an RCT). This forms the basis for the development of the TREND statement, a consensus statement for the reporting of nonrandomised intervention studies.
- The authors describe the TREND checklist and its development (as companion to the CONSORT statement for RCTs), which provides guidelines for the transparent reporting of studies, including aspects of intervention and comparisons, the research design, as well as methods of adjusting for possible sources of bias.

REFERENCES

- Cochrane Collaboration *Open Learning Material for Reviewers*, Version 1.1, November 2002. Eds, P Alderson and S Green, www.cochrane-net.org/openlearning/index.htm
- Cochrane Health Promotion and Public Health Field. *Systematic Reviews of Health Promotion and Public Health Interventions*. www.vichealth.vic.gov.au/cochrane.
- Cochrane Effective Practice and Organisation of Care (EPOC) Review Group. *The data collection checklist*. www.epoc.uottawa.ca/checklist2002.doc or email amayhe2@uottawa.ca
- Cochrane Effective Practice and Organisation of Care (EPOC) Review Group. *The data collection template*. www.epoc.uottawa.ca/data%20abstraction%20form%20-%202002.doc or email amayhe2@uottawa.ca
- Cochrane Effective Practice and Organisation of Care (EPOC) Review Group. *EPOC Methods Paper: Including Interrupted Times Series (ITS) designs in an EPOC review*. www.epoc.uottawa.ca or email amayhe2@uottawa.ca
- Cochrane Effective Practice and Organisation of Care (EPOC) Review Group. *EPOC Methods Paper: Issues Related to Baseline Measures of Performance*. www.epoc.uottawa.ca or email amayhe2@uottawa.ca
- Undertaking systematic reviews of research on effectiveness. *CRD's guidance for those carrying out or commissioning systematic reviews*. CRD Report Number 4 (2nd Edition). NHS Centre for Reviews and Dissemination, University of York. March 2001. www.york.ac.uk/inst/crd/report4.htm
- Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al.* (2003) Evaluating non-randomised intervention studies. *Health Technol Assess* 7(27). www.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=1117
- Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. (1999) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 7(1). www.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=1099
- Health Care Practice Research Development Unit (HCPREDU) *Evaluation tool for qualitative studies*. www.fhsc.salford.ac.uk/hcprdu/tools/qualitative.htm
- Heritier SR, Gebiski VJ and Keeck AC (2003). Inclusion of patients in clinical trials analysis: the intention-to-treat principle. *MJA*, 179: 438-40.
- Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]. In: *The Cochrane Library*, Issue 4, 2006. Chichester, UK: John Wiley & Sons, Ltd.
- Jadad A. (1998) 'Randomised controlled trials: A user's guide.' BMJ Books, London.
- JuniP, Witschi A, Bloch R, Egger M. (1999) The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*; 282: 1054-60.
- Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, *et al.* (1999) Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assessment* 3(12).

- Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000; 4(10) www.hta.nhsweb.nhs.uk/execsumm/summ410.htm
- Spencer L, Ritchie J, Lewis J, Dillon L. *Quality in Qualitative Evaluation: A framework for assessing research evidence*. Government Chief Social researcher's Office. Crown Copyright, 2003. www.strategy.gov.uk/downloads/su/qual/downloads/qqe_rep.pdf
- Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. (2000) Bayesian methods in health technology assessment: a review. *Health Technol Assess* 4(34). www.hta.nhsweb.nhs.uk/ProjectData/3_project_record_published.asp?PjtId=930
- Torgerson DJ. and Sibbald B. What is a patient preference trial? *BMJ* 1998; 316: 360.