Cochrane Reviewers' Handbook 4.2.0

Updated March 2003

Contents

ABOUT THE HANDBOOK

Editors

Mike Clarke and Andrew Oxman

How to cite this version of the Handbook

Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software. Updated quarterly.

or

Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. http://www.cochrane.dk/cochrane/handbook/handbook.htm (accessed 30 April 2003).

When referring to a specific section or subsection refer to it by the title and section number, NOT page numbers. For example:

Clarke M, Oxman AD, editors. Formulating the problem. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]; Section 4. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software. Updated quarterly.

or

Clarke M, Oxman AD, editors. Formulating the problem. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]; Section 4. http://www.cochrane.dk/cochrane/handbook/handbook.htm (accessed 30 April 2003).

Contact addresses

Mike Clarke
UK Cochrane Centre
Summertown Pavilion
Middle Way
Oxford OX2 7LG
United Kingdom
Phone: +44 1865 516300
Fax: +44 1865 516311
Email: mclarke@cochrane.co.uk

Dr Andy Oxman
Department for Health Services Research
Norwegian Directorate for Health and Social Welfare
P.O. Box 8054 Dep

Oslo
N-0031
NORWAY
Phone: +47 24 16 32 90
Fax: +47 24 16 30 13
Email: oxman@online.no

Internet addresses

Germany
http://www.cochrane.de
UK
http://www.update-software.com/ccweb/

USA
http://www.cochrane.org

Sources of support

Present sources of support

National Health Service Research & Development Programme, England
National Institute of Public Health, Norway
Copenhagen Hospital Corporation, Denmark
Health Research Board, Ireland

Previous sources of support

Health Services Research and Development Service and the University of Texas Health Science Center, San Antonio, USA
US Veterans Health Administration, USA
Oxford Regional Health Authority, UK
Nuffield Provincial Hospitals Trust, UK
LW Frohlich Fund, USA.
Norwegian Ministry of Health and Social Affairs, Norway
Norwegian Research Council, Norway
Glaxo Wellcome, Norway
Pfizer, Norway

Acknowledgements

Many people have contributed to this and preceding editions of the Cochrane Reviewers' Handbook, both directly and indirectly (through workshops, helpful suggestions and inspiration). Grateful acknowledgement is offered to everyone who has made its preparation worthwhile, by enthusiastically receiving the Cochrane Collaboration and committing his or her own time to making it a reality.

The Handbook is co-edited by Mike Clarke and Andy Oxman, with advice from the Handbook Advisory Group and technical support provided by Jacob Riis. In addition to the editors, the current membership of this Group is Phil Alderson, Lisa Askie, Chris Cates, Jon Deeks, Matthias Egger, Frances Fairman, Alex Jadad, Philippa Middleton, Jim Neilson, Ole Olsen, as well as the convenors of the Cochrane Methods Groups. We would

also like to thank the following people for their continuing contributions to this edition of the Handbook: Doug Altman, Hilda Bastian, Lisa Bero, Davina Ghersi, Sally Green, Jeph Herrin, Andrew Herxheimer, Julian Higgins, Monica Kjeldstrøm, Carol Lefebvre, Eric Mannheimer, Rasmus Moustgaard, Melissa Ober, Drummond Rennie, Jacob Riis, Lesley Stewart, Luke Vale, Veronica Yank.

The first edition of the Handbook (1994) was developed by Andy Oxman, Iain Chalmers, Mike Clarke, Murray Enkin, Ken Schulz, Mark Starr, Kay Dickersin, Andrew Herxheimer and Chris Silagy with administrative support from Sally Hunt.

The second and third editions of the Handbook (1995 and 1997) were edited by Andy Oxman and Cynthia Mulrow.

Section 5 of the Cochrane Reviewers' Handbook was prepared first, in 1994, by Kay Dickersin and Carol Lefebvre as a document entitled *Establishing and Maintaining Registers of RCTs*). Many others contributed to it and Kay Dickersin was the editor. This document was updated by Kay Dickersin and Kristen Larson in 1995. In 1997, a major revision was undertaken by Mike Clarke for inclusion in version 3 of the Cochrane Reviewer's Handbook. The version included in version 4.2.0 was largely rewritten in 2002 by Eric Manheimer; with input from Kay Dickersin and members of the Handbook Advisory Group.

WHAT'S NEW?

Corrections and changes in version 4.2.0 (March 2003) of the Handbook

Major corrections and changes:

Section 5 and Appendices 5: these have been revised and updated.
Section 8: this has been amended slightly to mention the addition of the generic inverse variance method to RevMan 4.2 and the ability to include additional figures.

Minor corrections:

Acknowledgements: the help of additional people in the preparation of this version of the Handbook has been acknowledged.
Additional figures: relevant sections have been amended to note the ability to include additional figures in Cochrane reviews (using RevMan 4.2).
Appendix 2a, section 2a.1: the categories for sources of support have been changed to 'internal' and 'external'.
Appendix 2a, section 2a.3: the way to describe a search of the 'Cochrane Central Register of Controlled Trials (CENTRAL)' in the abstract for a Cochrane review has been clarified.
Appendix 2c: the permission to publish form has been removed while it is being revised (this has also led to a change in section 2.2.3).
Some typographical mistakes have been corrected.

Corrections and changes in version 4.1.6 (January 2003) of the Handbook

Major corrections and changes:

Section 1: this has been updated and information has been added on The Cochrane Collaboration Open Learning Material for Cochrane Reviewers.
Section 3.2: the policy on the withdrawal of protocols has been updated.
Section 10.10: the policy on the withdrawal of reviews has been added.
Appendix 1: this has been added to provide information on The Cochrane Collaboration Open Learning Material for Cochrane Reviewers.

Minor corrections:

Some typographical corrections have been made.
The name for the 'Cochrane Controlled Trials Register (CENTRAL/CCTR)' has been changed to the 'Cochrane Central Register of Controlled Trials (CENTRAL)'.
Sources of support: this has been updated to reflect the support from the National Health Service Research & Development Programme, UK and the Health Research Board, Ireland.

Acknowledgements: the help of additional people in the preparation of this version of the Handbook has been acknowledged.
Section 4.5: an additional example (children versus adults) has been added of why separate reviews might be done.
Section 9.7: this has been amended to clarify the distinction between 'no evidence of an effect' and 'no evidence of effect'.
Appendix 2a, 2a.4 Text, Results: this has been amended to clarify the distinction between 'no evidence of an effect' and 'evidence of no effect'.
Appendix 2a, 2a.5 Conflict of interest: suggested wording if there no known conflicts of interest has been changed to 'None known'.

Corrections and changes in version 4.1.5 (April 2002) of the Handbook

Internet addresses: the list of Internet addresses has been reduced to the three official Cochrane Collaboration sites that are mirrors of each other (i.e. www.cochrane.de, www.cochrane.org and www.update-software.com/ccweb).

Appendix 2a, Section 2a.2: it has been clarified that help with synopses should be sought directly from the Cochrane Consumer Network, rather than the Australasian Cochrane Centre.

Corrections and changes in version 4.1.4 (October 2001) of the Handbook

Major corrections:

Section 2.3: the suggested wording to use when versions of Cochrane reviews are published in paper journals has been revised.

Minor corrections:

Section 9.7: advice has been added on the balanced interpretation of analyses when the confidence interval for the effect estimate overlaps the null value.
Section 10.11: the address of the Comments and Criticisms web page has been updated.

Corrections and changes in version 4.1.3 (June 2001) of the Handbook

Minor corrections:

Section 9.7: this has been expanded to include more discussion of the interpretation of results that are not statistically significant.
Appendix 5a: a contact address has been added for the International Register of Clinical Trials Registers.

Corrections and changes in version 4.1.2 (March 2001) of the Handbook

Major corrections:

Appendix 6: this has been replaced with an updated version.

Minor corrections:

Section 1.0: the new name (Cochrane Methodology Register) has replaced "Cochrane Review Methodology Database".
Glossary: Three terms have been added: inter-rater reliability, intra-rater reliability and N of 1 randomised trial.

Corrections and changes in version 4.1.1 (December 2000) of the Handbook

Major corrections:

Section 10.10: the revised Cochrane Collaboration policy that reviews should be updated at least every 2 years (instead of every year) has been added. This policy was agreed by the Steering Group in October 2000.

Minor corrections:

Section 10.10: The Collaboration policy that protocols that have not been converted into full reviews within two years should generally be withdrawn from the CDSR (stated in section 3.2) has been restated here.
Section 10.11: the mention that software is being developed to help Criticism Editors to coordinate the reviewers' responses to comments and criticisms has been deleted.
Appendix 5a: The list of registers has been replaced by the URLs for online registers of registers

Corrections and changes in version 4.1.0 (June 2000) of the Handbook

Major corrections and changes

Chapter 2: additional guidance has been added on the publication of Cochrane Reviews in journals.
Chapter 5: this has been updated.
Chapter 6: this has been updated.
Chapter 11: a new section (11.6) has been added on the conversion of reviews that used individual patient data into Cochrane Reviews

Appendix 2a, synopses: The guidance on preparing synopses has been changed to reflect the new policy that responsibility for the approval of the synopsis to be included in a Cochrane review rests solely with the relevant review group.
Appendix 2a: a section has been added to show the elements of Cochrane protocols and reviews that should be published.

Minor corrections

Acknowledgements: the help of additional people in the preparation of this version of the Handbook has been acknowledged.
Appendix 2a, Text: The importance of keeping searches up-to-date has been added to the guidance on the content of the Search strategy section of the text of a Cochrane Review.
Appendix 2a, references: the title of the Flanagin 1998 reference has been corrected.

Corrections and changes in Version 4.0.0 (July 1999) of the Handbook

The Handbook has been thoroughly revised to take account of the changes in RevMan. We have also taken the opportunity to update several other sections of the Handbook.

Corrections and changes in Version 3.0.2 (September 1997) of the Handbook (Cochrane Library 1997, Issue 4)

Major corrections and changes

1. In appendix 2c, 'Conditions of publication', it has now been specified that a new 'Conditions of publication' form should be filled out with each substantive revision of a review.
2. In order to keep version numbers of the Handbook consistent with version numbers of RevMan, the Handbook will now make use of three digits:
- the first digit indicates a new release of RevMan and the Handbook,
- the second digit indicates an interim release of RevMan and the Handbook,
- the third digit indicates changes to the Handbook only.

Minor corrections and changes

1.Section 5.5 on handsearching has been updated to take account of the development of the control register on studies that might be relevant for inclusion in Cochrane Reviews (CENTRAL).
2. The glossary has four additions; *CENTRAL*, *trend*, *Trials Register Development Group* and *peer review*. The terms *Handsearching* and *Cochrane Controlled Trials Register (CCTR)* have been updated.
3. Synapse Publishing Inc. have put a version of the Handbook on the WWW at the following address: hhtp://www.medlib.com/cochranehandbook/.
4. Corrections have been made to the references in appendix 2a.6.
5. The list of handbook versions and related resources has been updated.
6. About the Handbook and What's New have been updated.

Corrections and changes in Version 3.0.1 (December 1996) of the Handbook (Cochrane Library 1997, Issue 3)

Major corrections and changes

1. Appendix 11a, 'Practical methodology of meta-analyses (overviews) using updated individual patient data', was added to the Handbook.
2. Appendix 5a, 'Registers of clinical trials', was updated.

Minor corrections and changes

1. All references to publications included in the Cochrane Library were updated ('How to cite the Handbook'; references: section 1; references: section 3; references: section 4; references: section 6; references: section 8; Appendix 5b; Appendix 5c.).

Corrections and changes in Version 3.0.0 (October 1996) of the Handbook (Cochrane Library, Issue 1)

1. Editorial responsibility

Responsibility for maintaining material formerly contained in Sections I to V of The Cochrane Collaboration Handbook was devolved as described below. The Handbook now consists solely of what was formerly Section VI: Preparing and Maintaining Systematic Reviews (Oxman, 1995). Cynthia Mulrow, director of the San Antonio Cochrane Center, joined Andy Oxman as co-editor. The entire Handbook was revised in response to suggestions we have received regarding the previous edition of the Handbook and the Training Manual prepared by the San Antonio Cochrane Center.

Editorial responsibilities for written material prepared on behalf of the Cochrane Collaboration has been evolving and it became clear in 1995 that new arrangements were required to deal with new circumstances. At its meeting 27 February 1996 in San Francisco the Steering Group established an Editorial Board to oversee the preparation of written material prepared on behalf of the Collaboration. This is one of five groups responsible for core functions that report directly to the Steering Group. The other groups responsible for core functions are the Software Development Group, the Trials Registers Development Group, a group responsible for forthcoming Colloquia, and the editorial team for the Handbook.

Further changes in editorial responsibility were proposed by Iain Chalmers and Andy Oxman to accommodate several developments, including:

- potential duplication of effort, and confusion regarding the roles of the Editorial Board and the Handbook editorial team
- the availability of CDSR and the development of modules in CDSR for Cochrane Centres, Fields, MGs and the Consumer Network as well as for CRGs

- the establishment of an elected Steering Group with representatives for each type of entity and the formation of groups responsible for core functions, which are directly responsible to the Steering Group

The proposed changes were circulated to all registered groups and approved by the Steering Group at its meeting 19 August 1996. The new arrangements are as follows:

Material about	**Responsible group**	**Current co-ordinator**
The Collaboration	Editorial Board	Jos Kleijnen
Core Functions:		
Handbook	Handbook Advisory Group	Andy Oxman & Cynthia Mulrow
Software	Software Development Group	Monica Fischer
Trials registers	Trials Registers Development Group	Kay Dickersin & Jean-Pierre Boissel
Registered groups:		
CRGs	CRG reps on Steering Group	CRG reps to decide
Cochrane Centres	Centre directors on Steering Group	Peter Gøtzsche
Fields	Field rep on Steering Group	Field rep to decide
Consumer Network	Consumer reps on Steering Group	Consumer reps to decide
MGs	MG rep on Steering Group	Andy Oxman

2. Abstracts

Abstracts are no longer optional and the subheadings used in abstracts have been changed to:

- Objectives
- Search strategy
- Selection criteria
- Data collection & analysis
- Main results

(see section 2a.2 in appendix 2a .)

3. Descriptions of methods used by Collaborative Review Groups

All reviews should state specifically when the register of trials maintained by the CRG responsible for the review was last searched for relevant studies. Descriptions of the methods used to develop and maintain CRG registers of trials are included in CRG modules published in the *Cochrane Database of Systematic Reviews* (*CDSR*). Other standardised methods used by a CRG should also be described in the group's module. Reviewers should state explicitly that they have used these methods and when they have used methods that differ from the standard methods used by a group.

4. Reviews of non-experimental evidence

Some CRGs, Fields and Methods Groups (MGs) have begun to explore ways of incorporating non-experimental evidence in reviews when this is appropriate. These

developments are reflected in changing the terminology from 'trials' to 'studies' and adding 'Types of studies' as a new subheading under 'Selection criteria'.

5. Links between the Handbook and related resources

The Handbook is being linked to several related resources (see 'About the Handbook'). These include: the *Cochrane Review Methodology Database*, the San Antonio Cochrane Center's Training Manual, Review Manager, a glossary, a frequently asked questions (FAQ) list, a library of examples, a library of slides, a register of empirical methodological studies, systematic reviews of those studies, and modules prepared by MGs for inclusion in *CDSR*.

6. Conflict of interest

A conflict of interest statement will be included in all Cochrane Reviews beginning with the second issue of the Cochrane Library in 1997 (see section 2a.2 and section 2a.4 in appendix 2a).

Proposed changes that have not yet been implemented

Conclusions

Because the results of a review can be interpreted differently depending on one's perspective and circumstances, the Steering Group decided in 1994 to separate the conclusions of a review from the rest of the review, and to attach conclusions or commentaries from Fields and other entities representing different perspectives (along with the reviewers' conclusions) to reviews in *CDSR*. However, practical arrangements for preparing and maintaining these have not yet been developed. Reviewers' conclusions, including implications for practice and implications for research, are currently maintained as part of the text of each review.

1. INTRODUCTION

Healthcare providers, consumers, researchers, and policy makers are inundated with unmanageable amounts of information. We need systematic reviews to efficiently integrate valid information and provide a basis for rational decision making (Mulrow 1994). Systematic reviews establish where the effects of healthcare are consistent and research results can be applied across populations, settings, and differences in treatment (e.g. dose); and where effects may vary significantly. The use of explicit, systematic methods in reviews limits bias (systematic errors) and reduces chance effects, thus providing more reliable results upon which to draw conclusions and make decisions (Antman 1992, Oxman 1993b). Meta-analysis, the use of statistical methods to summarise the results of independent studies, can provide more precise estimates of the effects of healthcare than those derived from the individual studies included in a review (Oxman 1993a, Sacks 1987, L'Abbe 1987, Thacker 1988).

Wider recognition of the key role of reviews in synthesising and disseminating the results of research has prompted people to consider the validity of reviews. In the 1970s and early 1980s, psychologists and social scientists drew attention to the systematic steps needed to minimise bias and random errors in reviews of research (Light 1971, Glass 1976, Rosenthal 1978, Jackson 1980, Cooper 1982). It was not until the late 1980s that people drew attention to the poor scientific quality of healthcare review articles (Mulrow 1987, Yusuf 1987, Oxman 1988). However, recognition of the need for systematic reviews of healthcare has grown rapidly and continues to grow, as reflected by the number of articles about review methods and empirical studies of the methods used in reviews (CMR 2003), the number of systematic reviews published in healthcare journals (NHS CRD 2003), and the rapid growth of The Cochrane Collaboration (CDSR 2003).

This Handbook builds on the work of a large number of people, including those represented in the Cochrane Methodology Register (CMR 2003), input from Cochrane Methods Groups, practical experience and feedback from Collaborative Review Groups (which have taken on the daunting task of systematically reviewing the effects of healthcare within their areas of interest), and Cochrane Centres (which provide training for reviewers).Whenever possible recommendations made here are based on empirical evidence and advice from Cochrane Methods Groups.

Our aim is to help reviewers make good decisions about the methods they use, rather than dictate arbitrary standards. However, where the Cochrane Collaboration has laid down policy, which must be followed by Cochrane reviewers, this is made clear. The guidelines provided here are intended to help reviewers to be systematic and explicit (not mechanistic!) about the questions they pose and how they derive answers to those questions. These guidelines are not a substitute for good judgement.

The Cochrane Collaboration and the Cochrane Reviewers' Handbook focus particularly on systematic reviews of randomised controlled trials (RCTs) because they are likely to provide more reliable information than other sources of evidence on the differential effects of alternative forms of healthcare (Kunz 2003). Systematic reviews of other types of evidence can also help those wanting to make better decisions about healthcare, particularly forms of care where RCTs have not been done and may not be possible or appropriate. The basic principles of reviewing research are the same, whatever type of

evidence is being reviewed. Although we focus mainly on systematic reviews of RCTs, we address issues specific to reviewing other types of evidence when this is relevant. Fuller guidance on such reviews is being developed.

Cochrane Reviews have a standard format that we describe in the next section (section 2). Those preparing a review should begin by developing a protocol (Section 3). The seven succeeding sections are organised according to the steps of preparing and maintaining a systematic review:

- Formulating the problem
- Locating and selecting studies
- Quality assessment of studies
- Collecting data
- Analysing and presenting results
- Interpreting results
- Improving and updating reviews

In the last section we take up specific issues about using individual patient data in reviews.

In 2002, the *Cochrane Collaboration Open Learning Material for Cochrane Reviewers* was prepared to accompany the *Cochrane Reviewers' Handbook* in helping people who are working on a Cochrane Review. It does not replace the *Handbook*, instead it provides a framework to progressing through the *Handbook*, supplementing it with examples and activities along the way (http://www.cochrane-net.org/openlearning/). More information on this material is available in Appendix 1.

1.1 References

Antman 1992. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 1992; 268:240-8.

CDSR 2003. Cochrane Database of Systematic Reviews. In: The Cochrane Library, Issue 1, 2003. Oxford: Update Software. Updated quarterly.

CMR 2003. Cochrane Methodology Register. In: The Cochrane Library, Issue 1, 2003. Oxford: Update Software. Updated quarterly.

Cooper 1982. Cooper HM. Scientific guidelines for conducting integrative research reviews. Rev Educ Res 1982; 52:291-302.

Glass 1976. Glass GV. Primary, secondary, and meta-analysis of research. Educ Res 1976; 5:3-8.

Jackson 1980. Jackson GB. Methods for integrative reviews. Rev Educ Res 1980; 50:438-60.

Kunz 2003. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials (Cochrane Methodology Review). In: The Cochrane Library, Issue 1, 2003. Oxford: Update Software.

L'Abbe 1987. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987; 107:224-33.

Light 1971. Light RJ, Smith PV. Accumulating evidence: procedures for resolving contradictions among different research studies. Harv Educ Rev 1971; 41:429-71.

Mulrow 1987. Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987; 106:485-8.

Mulrow 1994. Mulrow CD. Rationale for systematic reviews. BMJ 1994; 309:597-9.

NHS CRD 2003. NHS Centre for Reviews and Dissemination. Database of Abstracts of Reviews of Effectiveness. In: The Cochrane Library, Issue 1, 2003. Oxford: Update Software. Updated quarterly.

Oxman 1988. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988; 138:697-703.

Oxman 1993a. Oxman AD. Meta-statistics: Help or hindrance? ACP J Club 1993; 118:A-13.

Oxman 1993b. Oxman AD, Guyatt GH. The science of reviewing research. Ann NY Acad Sci 1993; 703:125-33.

Rosenthal 1978. Rosenthal R. Combining results of independent studies. Psychol Bull 1978; 85:185-93.

Sacks 1987. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450-5.

Thacker 1988. Thacker SB. Meta-analysis: a quantitative approach to research integration. JAMA 1988; 259:1685-9.

Yusuf 1987. Yusuf S, Simon R, Ellenberg S (eds). Proceedings of 'Methodologic issues in overviews of randomized clinical trials'. Stat Med 1987; 6:217-409.

2. FORMAT OF A COCHRANE REVIEW

The format of a Cochrane Review has several objectives. It helps readers to find the results of research quickly and to assess the validity, applicability and implications of those results. It guides reviewers to report their work explicitly and concisely, and minimises the effort required to do this. The format is also suited to electronic publication and updating, and it generates reports that are informative and readable when viewed on a computer monitor or printed.

Mike Clarke, Murray Enkin, Chris Silagy, and Mark Starr developed the original format of a Cochrane Review, with input from many others. The format is flexible enough to fit different types of reviews, including those making a single comparison, those making multiple comparisons and those prepared using individual patient data. Modifications of the format of Cochrane Reviews may be desired for a variety of reasons. However, because of the huge effort it can take to change the structure of reviews in *The Cochrane Database of Systematic Reviews* (*CDSR*), the format must be well defined and fixed. Some minor changes have been made from the format described in the first (1994) edition of the Handbook. These changes have been made based on the experience of Collaborative Review Groups, feedback from users of Cochrane Reviews and suggestions brought forward through the Review Manager (RevMan) Advisory Group, which has developed specifications for the software that is used to prepare Cochrane Reviews. The RevMan software is designed to help reviewers in constructing reviews in the appropriate format and to prepare files required to transfer reviews electronically.

Each review consists of:

- a cover sheet - giving the title, citation details and contact addresses
- a synopsis – prepared in collaboration with the Australasian Cochrane Centre
- an abstract - using a structured format
- the text of the review - consisting of an introduction (background and objective), materials (selection criteria and search strategy) and methods, results (description of studies, methodological quality, and results), discussion and reviewers' conclusions.
- tables and figures - showing characteristics of the included studies, specification of the interventions that were compared, the results of the included studies, a log of the studies that were excluded, and additional figures relevant to the review.
- references

Standard headings and tables guide reviewers when preparing their report and make it easier for readers to identify information that is of particular interest to them. The headings are listed below. The content that should follow each heading and the elements that will be published are described in appendix 2a (Guide to the format of a Cochrane Review).

2.1 Outline of a Cochrane Review

Cover sheet:

Title
Reviewers
Sources of support

What's New

Text of review:

- Synopsis
- Abstract
  - Background
  - Objectives
  - Search strategy
  - Selection criteria
  - Data collection & analysis
  - Main results
  - Reviewers' conclusions
- Background
- Objectives
- Criteria for selecting studies for this review
  - Types of studies
  - Types of participants
  - Types of interventions
  - Types of outcome measures
- Search strategy for identification of studies
- Methods of the review
- Description of studies
- Methodological quality
- Results
- Discussion
- Reviewers' conclusions
  - Implications for practice
  - Implications for research
- Acknowledgements
- Conflicts of interest

References:

- References to studies
  - Included studies
  - Excluded studies
  - Studies awaiting assessment
  - Ongoing studies
- Other references
  - Additional references
  - Other published versions of this review

Tables and figures:

- Characteristics of included studies
- Characteristics of excluded studies
- Characteristics of ongoing studies
- Comparisons, data and graphs
- Additional tables
- Additional figures

2.2 Conflict of interest

2.2.1 General principle

Cochrane Reviews should be free of any real or perceived bias introduced by the receipt of any benefit in cash or kind, any hospitality, or any subsidy derived from any source that may have or be perceived to have an interest in the outcome of the review. It is a matter of Cochrane Collaboration policy that direct funding from a single source with a vested interest in the results of the review is not acceptable.

2.2.2 Recommendations

The following recommendations are taken from the Collaboration's Code of Conduct for Avoiding Potential Financial Conflicts of Interest (see appendix 2b):

- Receipt of benefits from any source of sponsored research must be acknowledged and conflicts of interest must be disclosed in *CDSR* and other publications that emanate from the Collaboration.
- If a proposal raises a question of serious conflict of interest, this should be forwarded to the local Cochrane Centre for review (and the Steering Group notified accordingly).
- It is not mandatory to send funding proposals to the local Cochrane Centre or Steering Group prior to accepting them. However, this would be desirable in the cases of restricted donations, or any donation that appears to conflict with the general principle noted above.

2.2.3 Conflict of interest statements in reviews

Under the heading 'Conflict of interest' reviewers should report any conflict of interest capable of influencing their judgements, including personal, political, academic and other possible conflicts, as well as financial conflicts. It is impossible to abolish conflict of interest, since the only person who does not have some vested interest in a subject is somebody who knows nothing about it (Smith 1994). Financial conflicts of interest cause the most concern, can and should be avoided, but must be reported if there are any. Any secondary interest (such as personal conflicts) that might unduly influence judgements made in a review (concerning, for example, the inclusion or exclusion of studies, assessments of the validity of included studies or the interpretation of results) should be reported.

Disclosing a conflict of interest does not necessarily reduce the worth of a review and it does not imply dishonesty. However, conflicts of interest can influence judgements in subtle ways. Reviewers should let the editors of their Collaborative Review Group know of potential conflicts even when they are confident that their judgements were not or will not be influenced. Editors may decide that disclosure is not warranted or they may decide that readers should know about such a conflict of interest so that they can make up their own minds about how important it is. Decisions about whether or not to publish such information should be made jointly by reviewers and editors.

To help ensure the integrity and perceived integrity of Cochrane Reviews, all reviewers must sign the relevant statements in the form giving the Cochrane Collaboration permission to publish their review in addition to disclosing conflicts of interest (see appendix 2c).

2.3 Publication of Cochrane Reviews in print journals and books

Reviewers may wish to seek co-publication of Cochrane Reviews in peer-reviewed medical journals, particularly in those journals that have expressed enthusiasm for co-publication of Cochrane Reviews. For the Cochrane Collaboration, there is one essential condition of co-publication: Cochrane Reviews must remain free for dissemination in any and all media, without restriction from any of them. To ensure this, Cochrane reviewers grant the Collaboration world-wide licences for these activities, and do not sign over exclusive copyright to any journal or other publisher. A journal is free to request a non-exclusive copyright that permits it to publish and re-publish a review, but this cannot restrict the publication of the review by the Cochrane Collaboration in whatever form the Collaboration feels appropriate.

Reviewers are strongly discouraged from publishing Cochrane Reviews in journals before they are ready for publication in CDSR. This applies particularly to Centre directors and editors of Review Groups. However, journals will sometimes insist that the publication of the review in CDSR should not precede publication in print. When this is the case, reviewers should submit a review for publication in the journal after agreement from their CRG editor and before publication in CDSR. Publication in print should not be subject to lengthy production times, and reviewers should not unduly delay publication of a Cochrane Review either because of delays from a journal or in order to resubmit their review to another journal.

Journals can also request revision of a review for editorial or content reasons. External peer review provided by journals may enhance the value of the review and should be welcomed.

Journals generally may require shorter reviews than those published in CDSR. Selective shortening of reviews may be appropriate, but there should not be any substantive differences between the review as published in the journal and CDSR. If a review is published in a journal, it should be noted that a fuller and maintained version of the review is available in CDSR. Typically, this should be done by including a statement such as the following in the introduction: 'A more detailed review will be published and updated in the Cochrane Database of Systematic Reviews.Reference' The reference should be to the protocol for the review published in CDSR. A similar statement should be included in the introduction if a review is published in CDSR prior to publishing a version of the review in a journal. After a version of a Cochrane Review has been published in a journal, a reference to the journal publication must be added under the heading 'Other published versions of this review'. Reviewers are also encouraged to add the following statement to versions of Cochrane Reviews that are published in journals: "This paper is based on a Cochrane review published in The Cochrane Library YYYY, Issue X (see www.CochraneLibrary.net for information). Cochrane reviews are regularly updated as new evidence emerges and in response to comments and criticisms, and The Cochrane Library should be consulted for the most recent version of the review." This

statement should refer to the latest version of The Cochrane Library in which the Cochrane review appears.

Reviewers whose primary affiliation is a Cochrane entity should include the following sentence when publishing an article that is not about the Cochrane Collaboration or does not reflect official policy: "The views expressed in this article represent those of the authors and are not necessarily the views or the official policy of the Cochrane Collaboration". In addition, the following modification of the disclaimer published in The Cochrane Library should be added to Cochrane Reviews published in journals. "The results of a Cochrane Review can be interpreted differently, depending on people's perspectives and circumstances. Please consider the conclusions presented carefully. They are the opinions of review authors, and are not necessarily shared by the Cochrane Collaboration."

The passage below can be provided to journal editors upon submission of a review for publication, and the letter of submission should be copied to the CRG editors for information. This policy and procedure may be new to some journal editors and may require direct discussion with the journal editor. The CRG editors should be informed of any problems encountered in this process.

The following passage is suggested for inclusion in letters of submission to journal editors:

This systematic review has been prepared under the aegis of the Cochrane Collaboration, an international organisation that aims to help people make well-informed decisions about healthcare by preparing, maintaining and promoting the accessibility of systematic reviews of the effects of healthcare interventions. The Collaboration's publication policy permits journals to publish reviews, with priority if required, but permits the Cochrane Collaboration also to publish and disseminate such reviews. Cochrane Reviews cannot be subject to the exclusive copyright requested by some journals.

2.4 Publication of previously published reviews as Cochrane Reviews

Most reviews that have been previously published (referred to as 'previously published reviews' here) require substantial additional work before they can be published as a Cochrane Review in *CDSR*. In light of this additional work and substantial differences from the previously published review, the Cochrane Review can be considered a new publication. The previously published version of the review must be referenced in the Cochrane Review under the heading 'Other published versions of this review'. However, it is generally not necessary to seek permission from the publisher of the previously published review.

Occasionally a Cochrane Review will be similar enough to a previously published review that the only change is in the formatting of the review. In these cases reviewers should obtain permission from the publisher of the previously published review prior to publishing the review in *CDSR*. If reviewers are in doubt about whether they should request permission, they are encouraged to do so. This is unlikely to present a problem, provided it is done well in advance of the planned submission to *CDSR*. If it is known in advance that there is interest in publishing a version of a Cochrane Review in a journal,

reviewers should not assign exclusive copyright to the journal (see section 2.3). The Cochrane Collaboration does not require exclusive copyright. It is therefore not a problem to publish a version of a Cochrane Review in a journal after it has been published in *CDSR*, provided it is not called a Cochrane Review and that it is acknowledged that it is based on a Cochrane Review (see section 2.3).

The conversion of individual patient data reviews into Cochrane Review is discussed in section 11.6.

2.5 References

Smith 1994. Smith R. Conflict of interest and the BMJ. BMJ 1994; 308:4-5.

3. DEVELOPING A PROTOCOL

3.1 Rationale for protocols

Preparing a review is a complex process that comprises many judgements, as well as decisions about the process and the resources needed (see appendix 3a). As in any scientific endeavour, the methods to be used should be established beforehand. However, reviews are by their nature, retrospective, since the studies included are usually identified after they have been completed and reported. Therefore, it is important to make the process as rigorous and well defined as possible (Light 1984b) while maintaining a practical perspective. The reviewer's knowledge of the results of the study may influence:

- the definition of a systematic review question
- the criteria for study selection
- the comparisons for analyses
- the outcomes to be reported in the review

Just as protocols for randomised trials must sometimes be changed to adapt to unanticipated circumstances (such as problems with participant recruitment, data collection or unexpected event rates), changes in a review protocol are sometimes necessary. While every effort should be made to adhere to a predetermined protocol, it should be recognised that this is not always possible or appropriate. Changes in the protocol should not be made on the basis of how they effect the results of the review. *Post hoc* decisions (such as excluding selected studies) that are made when the impact on the results of the review is known are highly susceptible to bias and should be avoided. As a rule, changes in the protocol should be documented and reported, and 'sensitivity analyses' (see section 8.10) of the impact of such decisions on the results of the review should be made when possible.

The protocol for a Cochrane Review should consist of the following sections:

- Cover sheet
- Background
- Objectives
- Criteria for considering studies for this review
- Search strategy for identification of studies
- Acknowledgements
- Conflict of interest
- Methods of the review
- Additional references
- Additional tables

The content of these sections for a full review is described in the appendix to section 2 (Guide to the format of a Cochrane Review). Guidelines for specific methodological decisions regarding problem formulation, the identification, selection and assessment of studies, data collection and analysis are given in section 4, section 5, section 6, section 7 and section 8.

3.2 The background for a review

Well-formulated questions usually do not appear out of thin air. They occur in the context of an already formed body of knowledge. This context should be addressed in the background section of the review. It may include information about the biology, epidemiology, public health importance, clinical relevance, or current practice regarding the topic that will be addressed by the review. It should refer to any previously published systematic reviews that address the same question or to existing controversy. This background helps set the rationale for the review, and should explain why the questions being asked are important. The background and objectives along with the proposed search strategy and plans for collecting and analysing data form the basis of the protocol of a Cochrane Review. Editors of Collaborative Review Groups appraise and give feedback on these protocols before actual reviews are conducted. The protocol will also be published in the *Cochrane Database of Systematic Reviews* and may be subject to comments and criticisms from users of this.

Systematic reviews can, in general, be motivated by a number of factors. For example, they can be conducted in an effort to resolve conflicting evidence, to answer questions where the answer is uncertain or to explain variations in practice. While Cochrane Reviews might be motivated by any of these and other factors, their primary aim should be to summarise and help people to understand the evidence. Reviewers must be careful not to impose their own values and preferences on others when answering the questions they pose. They should help people make practical decisions about healthcare. This has important implications for deciding whether or not to undertake a Cochrane Review, how to formulate the problem that a review will address, how to develop the protocol and how to present the results of the review.

- Questions should address the choices (practical options) people face when deciding about healthcare.
- Reviews should address outcomes that are meaningful to people making decisions about healthcare.
- Reviewers should describe how they will address adverse effects as well as benefits
- The methods used in a review should be selected to optimise the likelihood that the results will provide the best current evidence upon which to base decisions.
- It is important to let people know when there is no reliable evidence, or no evidence about particular outcomes that are likely to be important to decision makers.
- It is not helpful to include evidence for which there is a high risk of bias in a review, even if there is no better evidence.
- Similarly, it is not helpful to focus on trivial outcomes simply because those are what researchers have chosen to measure.
- So far as is possible, it is important to take an international perspective. The evidence collected should not be restricted by nationality or language without good reason.

When the protocol is converted into a full review, the fact that this review was preceded by a published protocol should be noted. It is Collaboration policy that protocols that have not been converted into full reviews within two years should generally be withdrawn from the CDSR. If a protocol is withdrawn for any reason other than it being superseded by a review, a withdrawal notice should be published in CDSR for one issue. Thereafter, information on the withdrawal of the protocol should be noted in the CRG's module.

3.3 References

Light 1984b. Light RJ, Pillemer DB. Organizing a reviewing strategy. In: Summing Up: The Science of Reviewing Research. Cambridge, Massachusetts: Harvard University Press, 1984; 13-31.

4. FORMULATING THE PROBLEM

4.1 Rationale for well-formulated questions

Poorly focused questions lead to unclear decisions about what research to include and how to summarise it.

As with any research, the first and most important decision in preparing a review is to determine its focus (Light 1984b). This is best done by asking clearly framed questions. Such questions are essential for determining the structure of a review (Jackson 1980, Cooper 1984, Hedges 1994). Specifically, they will guide much of the review process including strategies for locating and selecting studies or data, for critically appraising their relevance and validity, and for analysing variation among their results.

In addition to guiding the review process, a review's questions and objectives are used by readers in their initial assessments of relevance. The readers use the stated questions and objectives to judge whether the review is likely to be interesting and directly relevant to the issues they face.

4.2 Key components of a question

There are several key components to a well-formulated question (Richardson 1995, Counsell 1997) and these should be set in the Criteria for selecting studies section of the review. A clearly defined question should specify the types of people (participants), types of interventions or exposures, and the types of outcomes that are of interest. In addition, the types of studies that are relevant to answering the question should be specified. In general the more precise one is in defining components, the more focused the review. Equal precision in addressing each component is not necessary. For example, one might wish to concentrate on various treatments for a particular stage of breast cancer, or alternately to focus on a particular drug for any stage of breast cancer. In the former example the stage and severity of the disease would be defined very precisely within the Types of participants. Whereas, in the latter example, the treatment formulation would be defined very precisely within the Types of intervention.

An overview of the key components follows with examples of useful issues to consider for each component. Reviewers need to ensure that they understand the terminology used to describe these components in different places and settings.

4.2.1 What types of people (participants)?

It is often helpful to consider the types of people that are of interest in two steps. First, define the diseases or conditions that are of interest. Explicit criteria sufficient for establishing the presence of the disease or condition should be developed. Second, identify the population and setting of interest. This involves deciding whether one is interested in a special population group determined on the basis of factors such as age, sex, race, educational status, or the presence of a particular condition such as angina or shortness of breath. One might also be interested in a particular setting on the basis of

factors such as whether people are living in the community; are hospitalised, in nursing homes or chronic care institutions; or are outpatients.

Any restrictions with respect to specific population characteristics or settings should be based on sound evidence. For example, focusing a review of the effectiveness of mammographic screening on women between 40 and 50 years old can be justified on the basis of biological plausibility, previously published systematic reviews and existing controversy. On the other hand, focusing a review on a particular subgroup of people on the basis of their age, sex or astrological birth-sign simply because of personal interests when there is no underlying biologic or sociological justification for doing so should be avoided. When there is uncertainty about whether there are important differences in effects among various subgroups of people, it is probably best to include all of the relevant subgroups and then test for important and plausible differences in effect in the analysis (see section 4.5 below and section 8).

4.2.2 What types of comparisons (interventions)?

The next key component of a well-formulated question is to specify the interventions that are of interest. It is also important to define the interventions against which these will be compared, such as the types of control groups that are acceptable for the review. Give thought to whether persons in a control group might receive interventions other than a placebo, and whether those interventions overlap in any way with the active intervention being tested. This issue is discussed further in the section on assessing the quality of studies (section 6).

4.2.3 What types of outcomes?

The third key component of a well-formulated question is the delineation of particular outcomes that are of interest. While all important outcomes should be included in Cochrane reviews, trivial outcomes should not be included. Reviewers need to avoid overwhelming readers with data that is of little or no importance. At the same time that they must be careful not to leave out important data. If explicit criteria are necessary for establishing the presence of those outcomes these should be specified. Likewise if combinations of outcomes will be considered these need to be specified. For example, if a study only has data on nonfatal and fatal strokes combined, will this be included if the question specifically relates to stroke death?

In general, Cochrane Reviews should include all reported outcomes that are likely to be meaningful to people making a decision about the healthcare problem the review addresses. Beyond this, it may be important to specify outcomes that are important to decision makers, even when it is unlikely that data will be found. For example, quality of life is an important outcome, perhaps the most important outcome, for people considering whether or not to use chemotherapy for advanced cancer, even if the available studies only report survival data. In addition, reviewers should indicate how they will try to include data on adverse effects in their review. In regard to this, rather than including an exhaustive list of adverse outcomes it may be more informative to summarise 'severe' (e.g. severe enough to require withdrawal of treatment) and minor adverse outcomes and include appropriate description of these.

It is sometimes possible to acquire unpublished data from investigators in order to disentangle combined outcomes, as well as for other purposes (see section 7). Before excluding a study that seems to meet criteria for relevance, but has not reported results in a way that is adequate for the review, it is worth considering trying to obtain the necessary information from the investigators.

4.2.4 What types of study designs?

Certain study designs are superior to others when answering particular questions. Randomised controlled trials (RCTs) are considered by many the *sine qua non* when addressing questions regarding therapeutic efficacy, whereas other study designs are appropriate for addressing other types of questions. For example, questions relating to aetiology or risk factors may be addressed by case-control and cohort studies. Reviewers should consider up-front what study designs are likely to provide reliable data with which to answer their questions.

Other aspects relevant to study design that are worth initial consideration are whether to review studies that: have a placebo comparison group, evaluate outcomes in an unbiased manner, or have a certain length of follow-up. The more restrictive reviewers are in matching questions to particular aspects of design, the less likely they are to find data specific to the restricted question. However, reviewing studies that are unlikely to provide reliable data with which to answer the question is a poor use of time and can result in misleading conclusions. If, for example, one is interested in whether a therapy improves survival in patients with a chronic condition, it might be inappropriate to look at studies of very short duration, except to make explicit the fact that they cannot address the question of interest.

Because Cochrane Reviews address questions about the effects of healthcare, they focus primarily on RCTs. There are two reasons why one should be cautious about including non-randomised studies in a review of the effects of healthcare, both relating to bias. First, although it is possible to control for confounders that are known and measured using other study designs, randomisation is the only way to control for confounders that are not known or not measured. For clinical interventions, deciding who receives an intervention and who does not is influenced by many factors, including prognostic factors. Empirical evidence suggests that, on average, non-randomised studies tend to overestimate the effects of healthcare (Sacks 1982, Chalmers 1983, Schulz 1995). However, a systematic methodology review has shown that the extent and even the direction of bias in non-randomised studies is often impossible to predict (Kunz 1998).

Second, although it is often difficult to locate RCTs (Dickersin 1994) and reviews that fail to include unpublished trials may be biased towards overestimating the effectiveness of an intervention (Dickersin 1993). The efforts of the Cochrane Collaboration to identify RCTs have not been matched for the identification of other types of studies. Consequently, including studies other than controlled trials in a review may require additional efforts to identify studies and to keep the review up-to-date, and might increase the risk that the result of the review will be influenced by publication bias.

Despite the above concerns, it may sometimes be appropriate to conduct a systematic review of non-randomised studies of the effects of healthcare. For example, occasionally the course of a disease is so uniform or the effects of an intervention are so dramatic that

it is unnecessary and unethical to conduct RCTs. Under such circumstances it would be senseless to restrict a review to RCTs. While attention to the risk of bias should guide decisions about what types of study designs to include in a review, individual reviewers and Collaborative Review Groups must decide what types of studies are best suited to specific questions.

4.3 Using the key components of a question to locate and select studies

Once one has a well-formulated question, one should determine which key components to focus on in initial searching strategies. For Cochrane Reviews searching for studies is greatly facilitated by the availability of specialised registers compiled by CRGs. However, the extent to which these registers are developed varies and it may be necessary for reviewers to conduct supplemental searches.

Searches that demand the simultaneous presence of several components or very specific formulations of certain components are likely to be too specific and miss important information. For example, if one searches for studies addressing long-term effects of insulin therapy on renal function in type II diabetics by demanding that they be indexed as 'type II diabetes', 'insulin', 'renal function' and 'long-term', relevant studies are likely to be missed. On the other hand if 'insulin' or 'type II diabetes' is used alone as a search term, hundreds of irrelevant reports are likely to be identified.

In general, useful key components to use when searching include the condition or disease of interest and the intervention or exposure being evaluated. Although one may be specifically interested in a particular setting, studies are often not indexed by the type of setting in electronic databases. Also, multiple outcomes may be evaluated in studies, some of which may be relevant to the review, but not part of the indexing of the article. This issue is discussed further in the next section on locating and selecting studies (section 5).

Whatever search strategies are used, it will be necessary to go through a number of reports and decide which ones are relevant and which ones are not relevant. Formulating a question in terms of the types of participants, interventions, outcomes and study designs of interest will lead naturally to specifying the criteria that will be used to select studies. However, some additional effort is often needed to clarify the selection criteria and develop decision rules that are sensible and reproducible. If, for example, you are reviewing studies of therapies for constipation, you must decide if you will review studies addressing acute and/or chronic constipation as well as acceptable criteria for acute and chronic. Are you interested in the entire spectrum of severity of constipation or only in severe constipation and how will you define 'severe'? Do you want to review studies that define constipation on the basis of a certain frequency of bowel movements per week or limit yourself to studies that define constipation on the basis of symptoms such as straining and hard stools? Will you only review studies that have determined the underlying pathophysiologic mechanism of constipation or limit your review to certain specific pathophysiologic disorders? Will you consider studies that merely state that participants were 'constipated'.

4.4 Using the key components of a question to guide data collection

Details relevant to key components of questions are what reviewers will be collecting from individual studies. Thus well-formulated questions are directly linked to the data collection process because they guide: determination of final criteria that will be used to select appropriate studies for review, and what data should be abstracted from studies meeting those selection criteria. Components of questions may also be directly related to how one chooses to present and analyse data. These issues are discussed further in section 6, section 7 and section 8.

4.5 Broad versus narrow questions

The questions that a review addresses may be broad or narrow in scope. For example, a review might address a broad question regarding whether antiplatelet agents in general are effective in preventing thrombotic events in humans. Alternatively, a review might address whether a particular antiplatelet agent, such as aspirin, is effective in decreasing the risks of a particular thrombotic event, stroke, in elderly persons with a previous history of stroke. As another example, separate reviews might be done to investigate the effectiveness of antibiotics to treat respiratory tract infections in young children and adults.

Determining the scope of a review question is a decision dependent upon multiple factors including perspectives regarding a question's relevance and potential impact; supporting theoretical, biologic and epidemiological information; the potential generalisability and validity of answers to the questions; and available resources.

There are several advantages and disadvantages to initially asking broad or narrow questions. Narrowly focused reviews may not be generalisable to multiple settings, populations and formulations of an intervention. They can also result in spurious or biased conclusions in the same way that subgroup analyses sometimes do (see section 8.7). For example, a review of the effectiveness of aspirin for preventing strokes in women could lead to a false conclusion that aspirin was not effective in women when in truth there were not enough data to detect any significant difference in effect between men and women. A narrow focus is at high risk of resulting in biased conclusions when the reviewer is familiar with the literature in an area and narrows the inclusion criteria in such a way that one or more studies with results that are in conflict with the reviewer's beliefs are excluded. There is also a danger that the known results of a series of studies of a class of interventions might influence the choice of a specific intervention from this class for a narrow review.

The validity of very broadly defined reviews may be criticised for mixing apples and oranges, particularly when there is good biologic or sociological evidence to suggest that various formulations of an intervention behave very differently or that various definitions of the condition of interest are associated with markedly different effects of the intervention. It is fine to mix apples and oranges, if your question is about fruit, but not if your question is about vitamin C and you know that apples and oranges are different with respect to vitamin C.

Searches for data relevant to broad questions may be more time-consuming and more expensive than searches relevant to narrowly defined questions. As broad questions may be addressed by large sets of heterogeneous studies, the synthesis and interpretation of data may be particularly challenging. Broadly focused reviews can also become unwieldy to present, maintain and understand.

One option that has been found useful is to build a broadly focused review on the basis of a series of more narrowly focused reviews. For example, healthcare providers and pregnant women who want to quit smoking are likely to want to know which smoking cessation strategy to use - a broad question. A review that helps them to answer this question could be built upon a series of more focused reviews that ask what the effectiveness of a specific strategy, such as behaviour modification, is. Whether it makes most sense to start with narrower questions and build up to a broader question, or to start with a broad question and then divide it into a number of smaller questions depends on the nature of the problem (e.g. how complex it is, how well understood it is, how much research is available) and the particular circumstances of the reviewers and their CRG (e.g. how well developed their specialised register is, the availability of resources, time and interest).

4.6 Changing questions

While questions should be posed in the protocol before initiating the full review, these questions should not become a straightjacket that prevents exploration of unexpected issues (NHS CRD 1996). Reviews are analyses of existing data that are constrained by previously chosen study populations, settings, intervention formulations, outcome measures and study designs. It is generally not possible to formulate an answerable question for a review without knowing some of the studies relevant to the question, and it may become clear that the questions a review addresses need to be modified in light of evidence accumulated in the process of conducting the review.

Although a certain fluidity and refinement of questions is to be expected in reviews as one gains a fuller understanding of the problem, it is important to guard against bias in modifying questions. *Post-hoc* questions are more susceptible to bias than those asked *a priori*, and data-driven questions can generate false conclusions based on spurious results. Any changes to the protocol that result from revising the question for the review should be documented. When refining questions it is useful to ask the following questions:

- What is the motivation for the refinement?
- Was it made after you had seen and been influenced by results from a particular study or was it simply that you had not initially considered alternate but acceptable ways of defining the participants, interventions or outcomes of interest?
- Are your search strategies appropriate for the refined question (especially any that have already been undertaken)?
- Is your data collection tailored to the refined question?

4.7 References

Chalmers 1983. Chalmers TC, Celano P, Sacks HS, Smith H, Jr. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983; 309:1358-61.

Cooper 1984. Cooper HM. The problem formulation stage. In: Cooper HM, editor. Integrating Research. A Guide for Literature Reviews. Newbury Park: Sage Publications, 1984; 19-37.

Counsell 1997. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med 1997; 127: 380-7.

Dickersin 1993. Dickersin K, Min YI. NIH clinical trials and publication bias. Online J Curr Clin Trials [serial online] 1993; Doc No 50.

Dickersin 1994. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994; 309:1286-91.

Hedges 1994. Hedges LV. Statistical considerations. In: Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York: Russell Sage Foundation, 1994; 30-3.

Jackson 1980. Jackson GB. Methods for integrative reviews. Rev Educ Res 1980; 50:438-60.

Kunz 1998. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised trials. BMJ 1998; 317: 1185-90.

Light 1984. Light RJ, Pillemer DB. Organizing a reviewing strategy. In: Summing Up: The Science of Reviewing Research. Cambridge, Massachusetts: Harvard University Press, 1984; 13-31.

NHS CRD 1996. NHS Centre for Reviews and Dissemination. Undertaking Systematic Reviews of Research on Effectiveness (CRD Report 4). York: The University of York, 1996.

Richardson 1995. Richardson WS, Wilson MS, Nishikawa J, Hayward RSA. The Well-built Clinical Question: A Key to Evidence Based Decisions. ACP J Club 1995; A12-3.

Sacks 1982. Sacks HS, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. Am J Med 1982; 72:233-40.

Schulz 1995. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408-12.

5. LOCATING AND SELECTING STUDIES FOR REVIEWS

Systematic reviews of the effects of health care interventions generally focus on reports from randomized controlled trials (RCTs), when such data are available, because of the general acceptance that this study design will lead to the most reliable estimates of effects. A comprehensive search for relevant RCTs, which seeks to minimize bias, is one of the essential steps in doing a systematic review, and one of the factors that distinguishes a systematic review from a traditional review.

A 'quick and dirty' search of, for example MEDLINE, is generally not considered adequate. Studies have shown that only 30 - 80% of all known published RCTs were identifiable using MEDLINE (depending on the area or specific question) (Dickersin 1994). Even if relevant records are in MEDLINE it can be difficult to retrieve them easily. A comprehensive search is important not only for ensuring that as many studies as possible are identified but also to minimize selection bias for those that are found. Relying exclusively on a MEDLINE search may retrieve a set of reports unrepresentative of all reports that would have been identified through a comprehensive search of several sources. For example, the majority of the journals indexed in MEDLINE are published in English. If studies showing an intervention to be effective are more likely to be published in English, then any summary of only the English language reports retrieved through a MEDLINE search may result in an overestimate of effectiveness due to a language bias (Gregoire 1995; Moher 1996; Egger 1997; Juni 2002). In addition, the results of many studies are never published, and most of these probably remain unknown. If studies showing an intervention to be effective are more likely to be published, then any summary of only the published reports may result in an overestimate of effectiveness due to a publication bias (Simes 1986; Dickersin 1987; Simes 1987; Begg 1988; Hetherington 1989; Easterbrook 1991; Dickersin 1993; Song 2000).

This section contains information about locating and selecting studies for systematic reviews. The first section describes some of the sources and approaches that can be used. The second section provides guidance on developing and documenting search strategies and organizing the records retrieved.

5.1 Searching for studies

5.1.1 Electronic databases

A search for relevant studies generally begins with health-related electronic bibliographic databases. Searches of electronic databases are generally the easiest and least time-consuming way to identify an initial set of relevant reports. Some electronic bibliographic databases, such as MEDLINE and EMBASE, include abstracts for the majority of recent records. Often a searcher can determine an article's relevance to a review based on the abstract, and can thereby avoid retrieving the full journal article, if the reported study is clearly not eligible for inclusion. Another advantage of these databases is that they can be searched electronically, for either words in the title and abstract, or using standardized subject related indexing terms that have been assigned to the record. For example, the MEDLINE indexing term RANDOMIZED-CONTROLLED-TRIAL (Publication Type)

was introduced in 1991 and allows a user to search for articles describing individual randomized trials.

Hundreds of electronic bibliographic databases exist. Some databases, such as MEDLINE/PubMed and EMBASE, cover all areas of health care and index journals published from around the world. Other databases, such as the Australasian Medical Index, the Chinese Biomedical Literature Database, the Latin American Caribbean Health Sciences Literature (LILACS), and the Japan Information Centre of Science and Technology File on Science, Technology and Medicine (JICST-E) index journals published in specific regions of the world. Others, such as the Cumulative Index of Nursing and Allied Health (CINAHL) and AIDSLINE, focus on specific areas of health. The Cochrane Collaboration has been developing an electronic database of reports of controlled trials ("CENTRAL") that is now the best single source of information about records that relate to studies, which might be eligible for inclusion in Cochrane Reviews (Dickersin 2002). Details of other databases that might contain eligible records are available in the Gale Directory of Online, Portable and Internet databases (http://www.dialog.com). The three electronic bibliographic databases generally considered as the richest sources of trials - MEDLINE, EMBASE, and CENTRAL - are described in more detail below.

5.1.1.1 MEDLINE and EMBASE

Index Medicus (published by the US National Library of Medicine (NLM)) and Excerpta Medica (published by Elsevier) are indexes of healthcare journals that are available in electronic form as MEDLINE and EMBASE, respectively. MEDLINE indexes about 4600 journals from the United States and 70 other countries, and in February 2002 contained over 11 million records from 1966 forward. (Some pre-1966 records have been added recently.) PubMed is a free, online MEDLINE database that also includes up-to-date citations not yet indexed (http://www.ncbi.nlm.nih.gov). EMBASE, which is often considered the European counterpart to MEDLINE, indexes nearly 4000 journals from over 70 countries and, in May 2002, contained approximately 9 million citations.

The overlap in journals covered by MEDLINE and EMBASE has been estimated to be approximately 34% (Smith 1992). The actual degree of reference overlap depends on the topic, with reported overlap values in particular areas ranging from 10% to 75% (Kleijnen 1992; Odaka 1992; Smith 1992; Rovers 1993; Ramos-Remus 1994). Studies comparing searches of the two databases have generally concluded that a comprehensive search requires that both databases be searched. Although MEDLINE and EMBASE searches tend not to identify the same sets of references, they have been found to return similar numbers of relevant references.

MEDLINE and EMBASE can be searched using standardized subject terms assigned by indexers employed by the publishing organization. Standardized subject terms (as part of a "controlled vocabulary") are useful because they provide a way of retrieving articles that may use different words to describe the same concept and because they provide information beyond what is simply contained in the words of the title and abstract. Using the appropriate standardized subject terms, a simple search strategy can quickly identify articles pertinent to the topic of interest. This approach works well if the goal is to identify a few good articles on a topic or to identify one particular article. However, when searching for studies for a systematic review the precision with which subject terms are

applied to references should be viewed with healthy skepticism. Authors may not describe their methods or objectives well, indexers are not always expert in the subject area of the article that they are indexing, and indexers make mistakes, like all people. In addition, the available indexing terms might not correspond to the terms the searcher wishes to use. The controlled vocabulary search terms for MEDLINE and EMBASE are not identical. Search strategies need to be customized for each database. One way to begin to identify controlled vocabulary terms for a particular database is to retrieve articles from that database, which meet the inclusion criteria for the review and to note common text words and the terms the indexers had applied to the articles, which could then be used for a full search.

Assuming that search results from each database are of approximately equal value, the choice of which to search first may often be a matter of cost, with MEDLINE typically being the less costly option. As noted earlier, PubMed provides free online access to MEDLINE. Other NLM databases, including AIDSLINE, and HealthSTAR are being phased out and their unique journal citations are migrating to PubMed. PubMed also provides links to full-text versions of articles on other publishers' web sites. A particularly useful feature of PubMed is that a list of 'Related articles' can be obtained for each relevant record identified. The NLM is developing a new database, called the Gateway, which allows users to search PubMed and multiple other NLM retrieval systems simultaneously. The current Gateway (http://gateway.nlm.nih.gov/gw/Cmd) searches PubMED, OLDMEDLINE, LOCATORplus, MEDLINEplus, DIRLINE, AIDS Meetings, Health Services Research Meetings, Space Life Sciences Meetings, and HSRProj.

5.1.1.2 The Cochrane Central Register of Controlled Trials (CENTRAL)

The Cochrane Central Register of Controlled Trials (CENTRAL) serves as the most comprehensive source of records related to controlled trials. As of January 2003, CENTRAL contained just over 350,000 citations to reports of trials and other studies potentially relevant to Cochrane Reviews. CENTRAL includes citations to reports of controlled trials that might not indexed in MEDLINE, EMBASE or other bibliographic databases; citations published in many languages; and citations that are available only in conference proceedings or other sources that are difficult to access (Dickersin 2002). Guidance on searching CENTRAL has been prepared as part of the CENTRAL Management Plan (http://www.cochrane.us/manage.htm). Many of the records in CENTRAL have been identified through systematic searches of MEDLINE and EMBASE, as described in the paragraph below.

The US Cochrane Center (as the former New England Cochrane Center, Providence Office) and the UK Cochrane Centre have searched MEDLINE for publication years 1966-2000 using phases 1 and 2 of the Cochrane highly sensitive search strategy (Appendix 5b) (Dickersin 1994). Each year, the US Cochrane Center updates this searching of MEDLINE. Hundreds of thousands of records have been retrieved and reviewed to date. If, on the basis of their title and abstract, the retrieved citations were judged to meet the Cochrane definitions for reports of randomized controlled trials (RCTs) and controlled clinical trials (CCTs), they have been assigned the Publication Type RANDOMIZED CONTROLLED TRIAL or CONTROLLED CLINCIAL TRIAL in MEDLINE and also included in CENTRAL (with the permission of the NLM) (see Appendix 5a.1 for NLM and Appendix 5a.2 for Cochrane definitions of RCT and CCT).

Similarly, in an ongoing project, the UK Cochrane Centre is retrieving records from EMBASE, checking their titles and abstracts and submitting these for inclusion in CENTRAL when appropriate (with the permission of Elsevier). A search of EMBASE using five free text terms (ie, random*, crossover*, cross-over*, factorial*, and placebo*), and covering the years 1974-1999, was run in 1999 to identify reports of trials. The results of this search are published in each quarterly release of CENTRAL. Additional searching of EMBASE began in December 2000, and this stage of the project includes searching using additional free text terms and EMBASE (EMTREE) thesaurus terms (Dickersin 2002).

Other general healthcare databases published in Australia, China, and Brazil are undergoing similar systematic searches to identify reports of trials for CENTRAL. The Australasian Cochrane Centre is coordinating the search of the National Library of Australia's Australasian Medical Index; the Chinese Cochrane Centre is coordinating the search of the Chinese Biomedical Literature Database; and the Brazilian Cochrane Centre is coordinating the search of the Pan American Health Organization's database LILACS (Latin American Caribbean Health Sciences Literature).

Each Collaborative Review Group (CRG) is responsible for the development of a subject specific specialized register of trials, which serves to ensure that individual reviewers within the CRG have easy and reliable access to the maximum possible number of studies relevant to their review topic. Typically, the editorial team will assume at least some, if not all, responsibility for examining new studies and forwarding them to appropriate reviewers. CRGs use all the methods described in this chapter to identify trials for their specialized registers, with the exception of generalized searches of MEDLINE and EMBASE, which, as described above, are performed by the US Cochrane Center and the United Kingdom Cochrane Centre. Many CRGs also have systems to ensure that reports identified by reviewers for their review(s) are contributed to the CRG's specialized register. The registers should, in turn, be submitted for inclusion in CENTRAL. Thus, records included in the specialized register of one CRG become accessible to all other CRGs through CENTRAL.

More detailed information about the development and contents of CENTRAL is included in a recent article (Dickersin 2002) and *The Cochrane Library* help file for CENTRAL.

5.1.1.3 SciSearch

SciSearch is an electronic database that lists published "source" articles from 4500 major scientific and technical journals and the articles that cite them. SciSearch can be used to identify studies for a review by identifying in the database a known relevant source article, and checking each of the articles citing the source article, to see if it is also relevant to the review. It is a way of searching forward in time from the publication of an important article. SciSearch also includes reference lists for records it indexes.

5.1.2 Handsearching

Handsearching involves a manual page-by-page examination of the entire contents of a journal issue to identify all eligible reports of trials, whether they appear in articles, abstracts, news columns, editorials, letters or other text. Handsearching health care journals is a necessary adjunct to searching electronic databases for at least two reasons:

1) not all trial reports are included on electronic bibliographic databases, and 2) even when they are included, they may not be indexed with terms that allow them to be easily identified as trials. Each journal year should be handsearched thoroughly and competently by a well-trained handsearcher for all reports of trials so that once a journal year has been handsearched, it will not need to be searched again. A recent study has found that a combination of handsearching and electronic searching is necessary for full identification of relevant reports published in journals that are indexed in MEDLINE, especially for articles published before 1991 when the NLM system for indexing trial reports was not as well developed as it is today and for those articles that are in parts of journals (such as supplements and correspondence) which are not indexed in MEDLINE (Hopewell 2002).

To facilitate the identification of all published trials the Cochrane Collaboration has organized extensive handsearching efforts. Overall coordination of the Collaboration's handsearch of the world's medical literature is managed by the US Cochrane Center, which oversees prospective registration of all potential handsearching on the Master List of Journals being Searched (http://www.cochrane.us/cochranemainpage.asp). Almost 2200 journals have been, or are being, searched within the Collaboration, and are included in the Master List. "Stand-alone" conference proceedings being searched are also included. The Master List enables search progress to be recorded and monitored for each title and also serves to prevent the duplication of effort that might otherwise arise if journals or conference proceedings in overlapping specialties were to be searched by more than one group or individual.

Cochrane entities and reviewers can prioritize handseaching based on where they expect to identify the most trial reports. This prioritization can be informed by searching CENTRAL, MEDLINE, and EMBASE in a topic area and identifying which journals appear to be associated with the most retrieved citations. Preliminary evidence suggests that most of the journals with a high yield of trial reports are indexed in MEDLIINE (Dickersin 2002), but this may reflect the fact that Cochrane contributors have concentrated early efforts on searching these journals.

Conference proceedings are important to handsearch because individual conference. These abstracts are not included on MEDLINE and are not usually included in other databases. Abstracts and other grey literature have been shown to be sources of approximately 10% of the studies referenced in Cochrane Reviews (Mallett 2002). Over one-half of trials reported in conference abstract never reach full publication, and those that are eventually published in full have been shown to be systematically different than those that are never published in full (Scherer 2003). In addition, grey literature in general has been found to be more likely than health care journals to contain 'negative' reports (McAuley 2000). Thus, failure to identify trials reported in conference proceedings might affect the results or threaten the validity of a systematic review.

Reviewers who wish to handsearch journals or conference proceedings to identify reports of studies for their review should first consult with the editorial based of their CRG. The CRG's Trials Search Coordinator/Review Group Coordinator can determine whether the journal or conference proceedings has already been searched, and, if it has not, the Coordinator can register the search on the Master List and provide training in handsearching. Training material is available on the US Cochrane Center web site (http://www.cochrane.us/hsmain.htm) All correspondence regarding the initiation of a journal search, progress of a journal search, status of a search etc needs to be between

staff at the US Cochrane Center and the Trials Search Coordinator/Review Group Coordinator.

5.1.3 Checking reference lists

Reviewers should check the reference lists of articles obtained (including those from previously published systematic reviews) to identify relevant reports. The process of following up references from one article to another is generally an efficient means of identifying studies for possible inclusion in a review. Because investigators may selectively cite studies with positive results (Gotzsche 1987; Ravnskov 1992), reference lists should never be used as a sole approach to identifying reports for a review, but rather as an adjunct to other approaches.

5.1.4 Checking other reviews

Some of the most convenient and obvious sources of references to potentially relevant studies are existing reviews. Copies of previously published reviews on the topic of interest should be obtained and checked for references to the original studies. As well as *the Cochrane Database of Systematic Reviews, The Cochrane Library* includes the Database of Abstracts of Reviews of Effects (DARE) a database produced by the NHS Centre for Reviews and Dissemination in York, UK, that provides information on previously published reviews of the effects of healthcare. MEDLINE, EMBASE and other bibliographic databases can also be used to identify review articles. In MEDLINE, the most appropriate review articles would be indexed under the Publication Type terms META-ANALYSIS and REVIEW, ACADEMIC. Search strategies have been developed to enhance identification of these types of publication (Boynton 1998).

5.1.5 Print versions of electronic databases

While MEDLINE and EMBASE include citations from 1966 and 1974 to the present, respectively, Index Medicus and Excerpta Medica, the print versions of these databases, include citations from 1879 and 1948, respectively. Searching the earlier printed subject indexes may be worthwhile, especially if there is reason to believe that there were early studies of the intervention being reviewed.

Science Citation Index is the print version of SciSearch (see Section 5.1.1.3) and is used for the same general purpose, i.e. for listings of where a published article was subsequently cited. Science Citation Index is more comprehensive than SciSearch, which began in 1974.

5.1.6 Identifying unpublished studies

Some completed studies are never published. If it could be assumed that unpublished studies of a given intervention were comparable to published studies on the same intervention, the failure to identify unpublished results would not be an important threat to the validity of a systematic review. However, an association between significant results and publication has been documented across a number of studies (Dickersin 1998). Finding out about unpublished studies, and including them in a systematic review, when

eligible, may be important to minimizing bias. Unfortunately, there is no easy way to obtain information about studies that have been completed but never published.

Colleagues can be an important source of information about unpublished studies, and informal channels of communication can sometimes be the only means of identifying unpublished data. Formal letters of request for information can also be used to identify completed but unpublished studies. One way of doing this is to send a comprehensive list of relevant articles along with the inclusion criteria for the review to the first author of reports for included studies, asking if they know of any additional studies (published or unpublished) that might be relevant. It may also be desirable to send the same letter to other experts and pharmaceutical companies or others with an interest in the area. However, it should be borne in mind that asking researchers for information about completed but never published studies has not typically been fruitful (Hetherington 1989; Horton 1997).

Identifying ongoing studies may also be important so that when a review is later updated, these can be assessed for possible inclusion. Unfortunately no single, central register of ongoing randomized trials currently exists and instead there are hundreds of distinct, predominantly online registers that vary widely in content, quality, and accessibility. These may have limited use as a means of identifying studies relevant to systematic reviews. Various efforts have been made by independent groups to begin to provide central access to ongoing trials, mostly through web sites that provide links to hundreds of registers of ongoing clinical trials. Two such examples are TrialsCentral™ (www.trialscentral.org) and Current Controlled Trials (www.controlled-trials.com). Current Controlled Trials also has a searchable database of information about thousands of ongoing and completed trials, including those registered on ClinicalTrials.gov.

5.1.7 Evidence on adverse effects

The first sources to investigate for information on adverse effects are reports from trials or other studies included in the systematic review. Excluded reports might also provide some useful information.

There are a number of sources of information on adverse effects of drugs, including Current Problems produced by the UK Medicines Control Agency (http://www.open.gov.uk/mca), MedWatch produced by the US Food and Drug Administration, and the Australian Adverse Drug Reactions Bulletin (http://www.health.gov.au/). Other regulatory authorities and the drug manufacturer may also be able to provide some information. Information on adverse effects might also be sought from other types of studies than those considered appropriate for the systematic review (e.g. cohort and case-control studies, uncontrolled trials, case series and case reports). However, all such studies and reports are subject to bias to a greater extent than randomized trials, and findings must be interpreted with caution.

5.2 Developing and documenting a search strategy for studies and organizing search results

5.2.1 Developing a search strategy

The ultimate goal in developing a specialized register for a CRG is that it can serve as an all-inclusive source of reports relevant to the CRG's scope and topic area, such that a relatively simple search using some key words related to the intervention could be run against the specialized register to identify all relevant studies. Most CRG specialized registers have not yet reached this point of comprehensiveness. Nevertheless, for many CRGs, the specialized register is still the best available source of studies for a given review. Different CRGs have different systems of ensuring reviewers have access to reports included in their specialized registers. Many Trials Search Coordinators/Review Group Coordinators search their CRG's specialized register for reviewers on request. Specialized registers can also be searched through CENTRAL, which contains a recent version of the registers for most CRGs.

It is always necessary to strike a balance between comprehensiveness and precision when developing a search strategy. Increasing the comprehensiveness of a search entails reducing its precision and retrieving more non-relevant articles. Developing a search strategy is an iterative process in which the terms that are used are modified, based on what has already been retrieved. There are diminishing returns for search efforts; after a certain stage, each additional unit of time invested in searching returns fewer references that are relevant to the review. Consequently there comes a point where the rewards of further searching may not be worth the effort required to identify the additional references. The decision as to how much to invest in the search process depends on the question a review addresses, the extent to which the CRG's specialised register is developed, and the resources that are available.

It is a good idea to search other electronic bibliographic databases regardless of whether CENTRAL or a CRG's specialized register is searched. If reviewers wish to conduct their own additional searches, information specialists with expertise in electronic searching should be sought to design and run the search strategy. The assistance of an information specialist should help to avoid many errors, and ensure that database-specific search term syntax will be appropriate and that advanced searching techniques (e.g. 'exploding' controlled vocabulary terms) can be employed where available. If information specialists are involved in developing the search strategy, they should be made aware of the greater importance of high recall (i.e. sensitivity) as compared to precision in searching for studies for systematic reviews. Ideally, reviewers should be present when the search is done. There are often costs associated with searching each database and with each record that is downloaded. Therefore, judgments about what to download often need to be made while the search is being done. The exact search performed and material retrieved for each search should be recorded in the Search Strategies for Identification of Studies section of the Cochrane review.

An electronic search strategy should generally have three sets of terms: 1) terms to search for the health condition of interest; 2) terms to search for the intervention(s) evaluated; and 3) terms to search for the types of study design to be included (typically randomized trials). The exception to this is CENTRAL, which aims to contain only reports with study

designs possibly relevant for inclusion in Cochrane Reviews, so searches of CENTRAL should be based on health condition and intervention only. A good approach to developing an electronic search strategy is to begin with multiple terms that describe the health condition of interest and join these together with the Boolean 'OR' operator. This means you will retrieve articles containing at least one of these search terms. You can do likewise for a second set of terms related to the intervention(s) and for a third set of terms related to the appropriate study design. These three sets of terms can then be joined together with the 'AND' operator. This final step of joining the three sets with the 'AND' operator limits the retrieved set to articles of the appropriate study design that address both the health condition of interest and the intervention(s) to be evaluated. A note of caution about this approach is warranted however: if an article does not contain at least one term from each of the three sets, it will not be identified. For example, if an index term has not been added to the record for the intervention or the intervention is not mentioned in the title and abstract, the article would be missed. A possible remedy is to omit one of the three sets of terms and decide which records to check on the basis of the number retrieved and the time available to check them.

No language restrictions should be included in the search strategy. Date restrictions should be applied only if it is known for certain that relevant studies could only have been reported during a specific time period.

A Trials Search Coordinators or information specialist can often be helpful in suggesting terms for the health condition and intervention. In general, both controlled vocabulary terms and text words (i.e. those found in the title or abstract) should be used. You should assume that earlier articles are harder to identify. For example, abstracts are not included in MEDLINE for most articles published before 1976 and, so, text word searches will only apply to titles. In addition, few MEDLINE indexing terms relating to study design were available before the 1990s. In designing a search strategy, it may be helpful to look at published papers on the same topic and check the controlled vocabulary terms and text words. Although a research question may address particular populations, settings or outcomes, these concepts are often not well indexed with controlled vocabulary terms and generally do not lend themselves well to searching.

The Cochrane highly sensitive search strategy for MEDLINE (Dickersin 1994; Robinson 2002) was developed specifically with the needs of Cochrane reviews in mind. The earliest version of this search strategy was developed in 1994 and subsequent versions have been developed, each with a different syntax, specific to the version of MEDLINE being searched (e.g. Silver Platter MEDLINE, OVID MEDLINE, PubMed) (Appendix 5b).

As noted in Section 5.1.1.2, the first two phases of the strategy have already been applied to search MEDLINE for all years from 1966 to 2000. Records resulting from the search were downloaded, printed out, and classified as definite or possible randomized or quasi-randomized trials, or not using the information in the title and abstract. If no abstract was available, the decision was based on the title alone. Because identification relied solely on the titles and, where available, the abstracts, some relevant articles may not have been identified. Therefore, it may still be worthwhile for reviewers to search MEDLINE using the Cochrane highly sensitive search strategy and to obtain and check the full reports of possibly relevant citations.

None of the terms from phase 3 of the Cochrane highly sensitive search strategy were used for generalized searching for controlled trial reports on MEDLINE noted above because of a pilot assessment which showed an unfavorable ratio of effort and expense to results (Clarke 1999).

CRGs typically use phases 1-3 of the Cochrane highly sensitive search strategy plus subject matter terms (using the Boolean "AND") for searching MEDLINE. In developing a search strategy for other electronic bibliographic databases, the terms used to identify trials would generally be similar or the same as terms from the Cochrane highly sensitive search strategy. If an information specialist is assisting with developing a search strategy, she should be made aware of the Cochrane highly sensitive search strategy and how it is used.

5.2.2 Documenting a search strategy

5.2.2.1 Electronic databases

The search strategy for electronic databases should be described in sufficient detail in a review that the process could be replicated. The following information should be included for each electronic bibliographic database each time it is searched, including CENTRAL and specialized registers:

- Title of database searched (e.g. MEDLINE)
- Name of the host (e.g. Silver Platter version 2.0)
- Date search was run (month, day, year)
- Years covered by the search
- Complete search strategy used, including all search terms (preferably cut and pasted rather than retyped)
- One or two sentence summary of the search strategy indicating which lines of the search strategy were used to identify records related to the health condition and intervention, and which lines were used to identify studies of the appropriate design
- The absence of any language restrictions

A description of a search strategy for electronic databases is included as Appendix 5c.

5.2.2.2 Journal Handsearching

Any journal years searched specifically for the review should be listed in the Search Strategies for Identification of Studies section of the review, by journal title, in alphabetical order. Ideally the full titles should be used for the journals. The months and years searched should be stated.

Example: British Journal of Surgery January 1948 - December 1998

5.2.2.3 Conference Proceedings

Details of the conference proceedings searched for the review should be provided as follows:

Proceedings with a title in addition to the conference name:

Child abuse and neglect: a medical community response. 1st AMA National Conference on Child Abuse and Neglect; 1984 Mar 30-31; Chicago.

Proceedings without a separate title:

Symposium on Nasal Polyp; 1984 Oct 5-6; Tokyo.

Proceedings in a language other than English:

Patologia de cancer de higado. Primera Reunion Germano-Espanola de Anatomia Patologica [Pathology of liver cancer. 1st German-Spanish Meeting on Pathological Anatomy]; 1988 Sep 23-25; Granada, Spain.

Proceedings also published as part of a journal:

Symposium on Vaccination against Hepatitis B; 1990 Sep 9; Goteburg, Sweden. (Scandinavian Journal of Infectious Diseases.1991 Supplement; 38).

Note whether the printed proceedings were handsearched or an electronic database was searched.

5.2.2.4 Efforts to identify unpublished studies

Provide a brief summary including databases searched (e.g. SIGLE, National Research Register, HSRProj), giving database details as described in 5.2.2.1. Include also efforts to contact investigators for information about unpublished studies.

5.2.2.5 Other sources

Provide a brief summary of other sources searched (e.g. bibliographies, reference lists and web sites) specifically for the review, giving details of date searched, search terms used, and web sites if relevant.

The search strategies used to develop the specialized register of a CRG are described in their module and should not be reported in the text of Cochrane reviews, but it is helpful to include details of the strategy used to search the specialized register.

5.2.3 Selecting studies

It is generally for reviewers to decide which study design(s) to include in their review. Most Cochrane reviews include only randomized or quasi-randomized trials (Appendix 5a). Some reviews are more restrictive, and include only randomized trials, while others are less restrictive, and include other study designs as well, particularly when few randomized trials addressing the topic of the review are identified. For example, many of the reviews from the Cochrane Effective Practice and Organization of Care (EPOC) Collaborative Review Group include before-and-after studies and interrupted time series in addition to randomized and quasi-randomized trials.

The process by which studies will be selected for inclusion in a review should be described in the review protocol. The selection of studies for consideration for inclusion in a review is a process that involves several stages. The first stage of checking the results of an electronic search involves assessing titles and abstracts to determine whether each article might meet predetermined eligibility criteria. Reviewers must decide if more than one of them will assess the records retrieved by electronic databases. There is evidence

that using at least two reviewers has an important effect on reducing the possibility that relevant reports will be discarded (Edwards 2002). If, given the information available, it can be determined that an article definitely does not meet inclusion criteria, it can be rejected. If the title or abstract leave room for doubt that the article cannot definitely be rejected, the full text of the article should be obtained. Reading the full text may lead the reviewers to exclude the study because it does not meet inclusion criteria. If the article is not rejected, information from it may then be formally extracted as described in Section 7. At all but the last stage of the selection process it is important to err on the side of over-inclusion because once a study has been excluded from the selection process it is unlikely to be reconsidered. Articles about which there is some doubt which are included at one stage can be excluded at a latter stage when more information becomes available.

All reports of studies that are identified as potentially eligible must be assessed to see whether they meet the inclusion criteria for the review. Reviewers must decide:

- whether more than one reviewer will assess the relevance of each report
- whether the decisions concerning relevance will be made by content area experts, non-experts, or both
- whether the people assessing the relevance of studies will know the names of the authors, institutions, journal of publication and results when they apply the inclusion criteria
- how disagreements will be handled if more than one reviewer applies the criteria to each article

Decisions about which studies to include in a review often involve judgment. To help ensure that these judgments are reproducible, it is desirable for more than one reviewer to apply the inclusion criteria to all the potentially relevant reports that are retrieved. However, the approach used varies from review to review. Whatever the case, the number of people assessing the relevance of each report should be stated in the Methods section of the review (if it is not stated in a description of the methods used by all of the reviewers in a particular CRG).

Experts in a particular area frequently have pre-formed opinions that can bias their assessments of both the relevance and validity of articles (Cooper 1989; Oxman 1993). Thus, while it is important that at least one reviewer is knowledgeable in the area under review, it may be an advantage to have a second reviewer who is not an expert in the area.

Some reviewers may decide that assessments of relevance should be made by people who are blind or masked to the journal from which the article comes, the authors, the institution, and the magnitude and direction of the results by editing copies of the articles (Berlin 1997a; Berlin 1997b). However, this takes much time, and may not be warranted given the resources required and the uncertain benefit in terms of protecting against bias (Berlin 1997b).

Disagreements about whether a study should be included can generally be resolved by discussion. Often the cause of disagreement is a simple oversight on the part of one of the reviewers. When the disagreement is due to a difference in interpretation, the issue should be resolved by consensus. Occasionally, it will not be possible to resolve disagreements about whether to include a study without additional information. In these cases, reviewers may choose to categorize the study in their review as one that is awaiting assessment until the additional information is obtained.

For most reviews it will be worthwhile to pilot test the inclusion criteria on a sample of articles (say ten to twelve papers, including ones that are thought to be definitely eligible, definitely not eligible and questionable). The pilot test can be used to refine and clarify the inclusion criteria, train the people who will be applying them and ensure that the criteria can be applied consistently by more than one person.

One approach to determining which studies to identify in the review as 'excluded' is to list any studies about which it is plausible to expect that a reader would question why the study was not included. This covers all studies that apparently meet the selection criteria but have had to be excluded and also any that do not meet all of the criteria but are well known, in the same general area as the review and likely to be thought relevant by some readers. By listing such studies as excluded and giving the reason for exclusion, the reviewer can show that consideration has been given to these studies.

5.2.4 Keeping track of identified studies

Specially designed reference management systems such as ProCite, Reference Manager, and EndNote are useful and relatively easy to use to keep track of reports of studies. ProCite is the most widely used package and the one for which support to editorial bases is most widely available. It is also the preferred database for submitting controlled trials and specialized registers to CENTRAL. ProCite eases the work of identifying duplicate references. In addition, it facilitates storage of information about the methods and process of a search. For example, separate unused fields in ProCite can be used to store 1) when and from whom an article was ordered, and the date of article receipt; 2) reasons for article exclusion; and 3) name of electronic bibliographic database source from which an article was identified.

General database packages such as Access and FoxPro include powerful query capabilities and lend themselves well to customisation, but require some programming and database design skills to set up. An Access-based software (called 'MeerKat') has been developed by the UK Cochrane Centre, in association with Update Software, to address the specific needs of CRGs in managing their specialised registers (http://www.update-software.com/meerkat/). MeerKat allows for a specialized register to be organized around studies, instead of the publications or reports generated from these studies. Each study may have several associated reports. For example, a single randomized trial may have reports that relate to plans for the trial, baseline characteristics of the trial participants, initial results from the trial, and final results from the trial. In MeerKat, each of these reports can be associated with the corresponding study. MeerKat has also been designed specifically to facilitate the work of the Review Group Coordinator/Trials Search Coordinator. For example, MeerKat can produce tables to indicate which records have been assigned to a particular reviewer or topic, and which records have been submitted to CENTRAL. MeerKat also allows complex database searches, including wildcard searches, Boolean searches, and searches of only specific fields. If adopted, MeerKat may ease the task of managing references within a CRG.

5.3 Summary

Conducting a comprehensive, objective, and reproducible search for studies can be the most time-consuming and challenging task in preparing a systematic review. Yet it is also

one of the most important. Identifying all relevant studies, and documenting the search for studies with sufficient detail so that it can be reproduced is, after all, largely what distinguishes a systematic review from a traditional narrative review. Although currently it is necessary to search multiple sources to identify relevant published studies, it is envisioned that CENTRAL will eventually become a comprehensive source for published studies, thus reducing the searching burden for reviewers. Identifying ongoing studies, however, will continue to remain a challenge until a comprehensive, searchable, ongoing trials register is produced to track, organize, and disseminate reports for ongoing studies, as CENTRAL doing for reports of studies that have been published (Lefebvre 2001).

5.4 References

Begg 1988. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. Journal of the Royal Statistical Society A 1988;151:445-63.

Begg 1989. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. Journal of the National Cancer Institute 1989;81:107-15.

Berlin 1997a. Berlin JA, Miles CG, Crigliano MD. Does blinding of readers affect the results of meta-analyses? Results of a randomized trial. Online Journal of Current Clinical Trials [serial online] 1997.

Berlin 1997b. Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. Lancet 1997;350:185-6.

Boynton 1998. Boynton J, Glanville J, McDaid D, Lefebvre C. Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design. Journal of Information Science 1998;24:137-57.

Clarke 1999. Clarke M. Feasibility study of the search terms in phase 3 of the Highly Sensitive Search Strategy for the MEDLINE recode project. The Cochrane Collaboration Methods Groups Newsletter 1999;3:20-1. (http://www.cochrane.de/newslett/1999mg.pdf, accessed 28 February 2003)

Cooper 1989. Cooper HM, Ribble RG. Influences on the outcome of literature searches for integrative research reviews. Knowledge 1989;10:179-201.

Dickersin 1987. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. Controlled Clinical Trials 1987;8:343-53.

Dickersin 1992. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. JAMA 1992;263:374-8.

Dickersin 1993. Dickersin K, Min YI. NIH clinical trials and publication bias. Online Journal of Current Clinical Trials [serial online] 1993; Doc No 50.

Dickersin 1994. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994;309:1286-91.

Dickersin 2002. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S, and the CENTRAL Development Group. Development of the Cochrane Collaboration's CENTRAL Register of Controlled Clinical Trials. Evaluation and the Health Professions 2002;25:38-64.

Easterbrook 1991. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991;337:867-72.

Edwards 2002. Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. Statistics in Medicine 2002; 21:1635-40.

Egger 1997. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997;350:326-9.

Gotzsche 1987. Gotzsche PC. Reference bias in reports of drug trials. BMJ 1987;295:654-6.

Gregoire 1995. Gregoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a tower of Babel bias? Journal of Clinical Epidemiology 1995;48:159-63.

Hetherington 1989. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. Pediatrics 1989;84:374-80.

Hopewell 2002. Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M. A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. Statistics in Medicine 2002;21:1625-34.

Horton 1997. Horton R. Medical editors trial amnesty. Lancet 1997;350:756.

Juni 2002. Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. International Journal of Epidemiology 2002;31:115-23.

Kleijnen 1992. Kleijnen J, Knipschild P. The comprehensiveness of MEDLINE and EMBASE computer searches. Searches for controlled trials of homeopathy, ascorbic acid for common cold and ginkgo biloba for cerebral insufficiency and intermittent claudication. Pharmaceutisch Weekblad, Scientific Edition 1992;14:316-20.

Lefebvre 2001. Lefebvre C, Clarke M. Identifying randomized trials. In: Egger M, Davey Smith G, Altman D, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing Group, 2001:69-86.

Mallett 2002. Mallett S, Hopewell S, Clarke M. Grey literature in systematic reviews: The first 1000 Cochrane systematic reviews. 4th Symposium on Systematic Reviews: Pushing the Boundaries, Oxford, UK, July 2-4, 2002.

McAuley 2000. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 2000;356:1228-31.

Moher 1996. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996;347:363-6.

Odaka 1992. Odaka T, Nakayama A, Akazawa K, Sakamoto M, Kinukawa N, Kamakura T, et al. The effect of a multiple literature database search -- a numerical evaluation in the domain of Japanese life science. Journal of Medical Systems 1992;16:77-81.

Oxman 1993. Oxman AD, Guyatt GH. The science of reviewing research. Annals of the New York Academy of Sciences 1993;703:125-33.

Petrosino 2000. Petrosino A, Boruch RF, Rounding C, McDonald S, Chalmers I. The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to facilitate the preparation and maintenance of systematic reviews of social and educational interventions. Evaluation and Research in Education 2000;14:206-19.

Ramos-Remus 1994. Ramos-Remus C, Suarez-Almazor M, Dorgan M, Gomez-Vargas A, Russell AS. Performance of online biomedical databases in rheumatology. Journal of Rheumatology 1994;21:1912-21.

Ravnskov 1992. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. BMJ 1992;305:15-9.

Robinson 2002. Robinson KA. Dickersin K. Development of a highly sensitive search strategy for the retrieval of controlled trials using PubMed. International Journal of Epidemiology 2002;31:150-3.

Rovers 1993. Rovers JP, Janosik JE, Souney PF. Crossover comparison of drug information online database vendors: Dialog and MEDLARS. Annals of Pharmacotherapy 1993;27:634-9.

Scherer 2003. Scherer RW, Langenberg P. Full publication of results initially presented in abstracts (Cochrane Methodology Review). In: The Cochrane Library, Issue 1, 2003. Oxford: Update Software.

Simes 1986. Simes RJ. Publication bias: the case for an international registry of clinical trials. Journal of Clinical Oncology 1986;4:1529-41.

Simes 1987. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. Statistics in Medicine 1987;6:11-29.

Smith 1992. Smith BJ, Darzins PJ, Quinn M, Heller RF. Modern methods of searching the medical literature. Medical Journal of Australia 1992;157:603-11.

Song 2000. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. Health Technology Assessment 2000;4(10).

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. 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, interventions and outcomes of interest were defined by these studies and the 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. 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

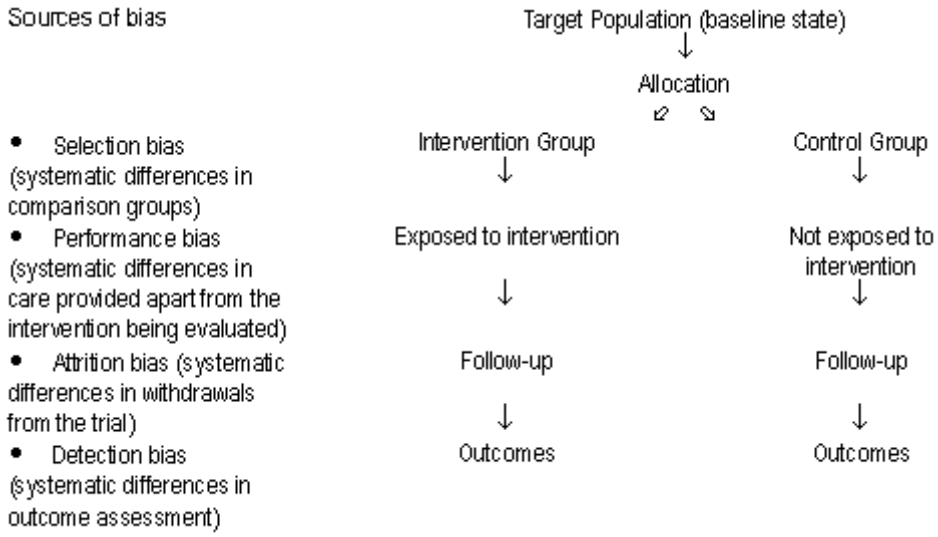
In the context of a systematic review, the validity of a study is the extent to which its design and conduct are likely to prevent systematic errors, or bias (Moher 1995). An important issue that should not be confused with validity is precision. Precision is a measure of the likelihood of chance effects 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.

6.2 Sources of bias in trials of healthcare interventions

There are four sources of systematic bias in trials of the effects of healthcare: selection bias, performance bias, attrition bias and detection bias (see figure below). Unfortunately, we do not have strong empirical evidence of a relationship between trial outcomes and specific criteria or sets of criteria used to assess the risk of these biases (Moher 1995, Moher 1996b). There is, however, a logical basis for suspecting such relationships and good reason to consider these four potential biases when assessing studies for a review (Feinstein 1985).



6.3 Selection bias

One of the most important factors that may lead to bias and distort treatment comparisons is that which can result from the way that comparison groups are assembled (Kunz 1998). Using an appropriate method for preventing foreknowledge of treatment assignment is crucially important in trial design. When assessing a potential participant's eligibility for a trial, those who are recruiting participants and the participants themselves should remain unaware of the next assignment in the sequence until after the decision about eligibility has been made. Then, after assignment has been revealed, they should not be able to alter the assignment or the decision about eligibility. The ideal is for the process to be impervious to any influence by the individuals making the allocation. This will be most securely achieved if an assignment schedule generated using true randomisation is administered by someone who is not responsible for recruiting subjects, such as someone based in a central trial office or pharmacy. If such central randomisation cannot be organised, then other precautions are required to prevent manipulation of the allocation prosess by those involved in recruitment.

The process of concealing assignment until treatment has been allocated has sometimes been referred to as 'randomisation blinding' (Chalmers 1983). This term does not clearly distinguish concealed allocation from blinding of patients, providers, outcome evaluators and analysts and is unsatisfactory for three reasons. First, the reason for concealing the assignment schedule is to eliminate selection bias. In contrast, blinding (used after the allocation of the intervention) reduces performance and detection biases. Second, from a

practical standpoint, concealing allocation up to the point of assignment is always possible, regardless of the study question, but blinding after allocation may be impossible, as in trials comparing surgical with medical treatment. Third, control of selection bias is relevant to the trial as a whole, and thus to all outcomes being compared. In contrast, control of detection bias is often outcome-specific and may be accomplished successfully for some outcomes in a study but not others. Thus, blinding up to allocation and blinding after allocation are addressing different sources of bias, are inherently different in their practicability and may apply to different components of a study. To clearly distinguish these different forms and purposes of 'blinding', we will refer to the process of concealing assignments as allocation concealment and reserve blinding for measures taken to reduce bias after the intervention has been assigned.

Empirical research has shown that lack of adequate allocation concealment is associated with bias (Chalmers 1983, Schulz 1995, Moher 1998a). Indeed, concealment has been found to be more important in preventing bias than other components of allocation, such as the generation of the allocation sequence (e.g., computer, random number table, alternation). Thus, studies can be judged on the method of allocation concealment. Information should be presented that provides some assurance that allocations were not known until, at least, the point of allocation. The method for assigning participants to interventions should be robust against patient and clinician bias and its description should be clear. The following are some approaches that can be used to ensure adequate concealment schemes.

- centralised (e.g. allocation by a central office unaware of subject characteristics) or pharmacy-controlled randomisation
- pre-numbered or coded identical containers which are administered serially to participants
- on-site computer system combined with allocations kept in a locked unreadable computer file that can be accessed only after the characteristics of an enrolled participant have been entered
- sequentially numbered, sealed, opaque envelopes

Other approaches may include approaches similar to ones listed above, along with reassurance that the person who generated the allocation scheme did not administer it. Some schemes may be innovative and not fit any of the approaches above, but still provide adequate concealment.

Approaches to allocation concealment that should be considered clearly inadequate include: alternation; the use of case record numbers, dates of birth or day of the week, and any procedure that is entirely transparent before allocation, such as an open list of random numbers. When studies do not report any concealment approach, adequacy should be considered unclear. Examples include merely stating that a list or table was used, only specifying that sealed envelopes were used and reporting an apparently adequate concealment scheme in combination with other information that leads the reviewer to be suspicious. When reviewers enter studies into RevMan they are required to indicate whether allocation concealment was adequate (A), unclear (B), inadequate (C), or that allocation concealment was not used (D) as a criterion to assess validity.

6.4 Performance bias

Performance bias refers to systematic differences in the care provided to the participants in the comparison groups other than the intervention under investigation. To protect against unintended differences in care and placebo effects, those providing and receiving care can be 'blinded' so that they do not know the group to which the recipients of care have been allocated. Some research suggests that such blinding is important in protecting against bias (Karlowski 1975, Colditz 1989, Schulz 1995). Studies have shown that contamination (provision of the intervention to the control group) and cointervention (provision of unintended additional care to either comparison group) can affect study results (CCSG 1978, Sackett 1979b). Furthermore, there is evidence that participants who are aware of their assignment status report more symptoms, leading to biased results (Karlowski 1975). For these reasons, reviewers may want to consider the use of ' blinding' as a criterion for validity. This can be done with the following questions: Were the recipients of care unaware of their assigned intervention? Were those providing care unaware of the assigned intervention?

A third question addressing blinding and detection bias is often added: Were persons responsible for assessing outcomes unaware of the assigned intervention? This addresses detection bias, as noted below.

Reviewers working on topics where blinding is likely to be important may want to develop specific criteria for judging the appropriateness of the method that was used for blinding. In some areas it may be desirable to use the same criterion across reviews, in which case a Collaborative Review Group (CRG) might want to agree to a standard approach for assessing blinding (Chalmers 1989, Schulz 1995, Jadad 1996, Moher 1996b).

6.5 Attrition bias

Attrition bias refers to systematic differences between the comparison groups in the loss of participants from the study. It has been called exclusion bias. It is called attrition bias here to prevent confusion with pre-allocation exclusion and inclusion criteria for enrolling participants. Because of inadequacies in reporting how losses of participants (e.g. withdrawals, dropouts, protocol deviations) are handled, reviewers should be cautious about implicit accounts of follow-up. The approach to handling losses has great potential for biasing the results and reporting inadequacies cloud this problem. What is reported, or more frequently implied, in study reports on attrition after allocation has not been found to be consistently related to bias (Schulz 1995). Thus reviewers should be cautious about using reported follow-up as a validity criterion, particularly when it is implied rather than explicitly reported. This is a general recommendation, however, and may not apply to certain topic areas that have higher quality reporting or where it is possible to obtain missing information from investigators.

6.6 Detection bias

Detection bias refers to systematic differences between the comparison groups in outcome assessment. Trials that blind the people who will assess outcomes to the intervention allocation should logically be less likely to be biased than trials that do not. Blinding is likely to be particularly important in research with subjective outcome measures such as

pain (Karlowski 1975, Colditz 1989, Schulz 1995). However, at least two empirical studies have failed to demonstrate a relationship between blinding of outcome assessment and study results. This may be due to inadequacies in the reporting of studies (Reitman 1988).

Bias due to the selective reporting of results is different from bias in outcome assessment. This source of bias may be important in areas where multiple outcome measures are used, such as evaluations of treatments for rheumatoid arthritis (Gotzsche 1989). Therefore, reviewers may want to consider specification of predefined primary outcomes and analyses by the investigators as indicators of validity. Alternatively, selective reporting of particular outcomes could be taken to suggest the need for better reporting and efforts by reviewers to obtain missing data.

6.7 Approaches to summarising the validity of studies

6.7.1 Simple approaches

There are several ways to rate validity. One is to rate individual criteria as 'met', 'unmet', or
'unclear' and to use individual criteria, such as adequacy of allocation concealment, in sensitivity analyses (see section 8.10). However, if several explicit criteria are used to assess validity, it is desirable to summarise these so as to derive an overall assessment of how valid the results of each study are. A simple approach to doing this is to use three categories such as the following:

Risk of bias	Interpretation	Relationship to individual criteria
A. Low risk of bias	Plausible bias unlikely to seriously alter the results	All of the criteria met
B. Moderate risk of bias	Plausible bias that raises some doubt about the results	One or more criteria partly met
C. High risk of bias	Plausible bias that seriously weakens confidence in the results	One or more criteria not met

The relationships suggested above will most likely be appropriate if only a few assessment criteria are used and if all the criteria address only substantive, important threats to the validity of study results. In general and when possible, reviewers should obtain further information from the authors of a report when it is unclear whether a criterion was met.

6.7.2 'Quality' scales and checklists

David Moher and his colleagues identified 25 scales and 9 checklists that have been used to assess the validity and 'quality' of randomised controlled trials (Moher 1995, Moher 1996b). These scales and checklists include anywhere from 3 to 57 items and take from 10 to 45 minutes to complete. Almost all of the items in the instruments are based on suggested or 'generally accepted' criteria that are mentioned in clinical trial textbooks. Many of the instruments are liable to confuse the quality of reporting with the validity of

the design and conduct of a trial. Moreover, scoring is based on whether something was reported (such as how participants were allocated) rather than whether it was done appropriately in the study. Many also contain items that are not directly related to validity, such as whether a power calculation was done (an item that relates more to the precision of the results) or whether the inclusion and exclusion criteria were clearly described (an item that relates more to applicability than validity).

Because there is no 'gold standard' for the 'true' validity of a trial, the possibility of validating any proposed scoring system is limited. While it is possible to apply basic principles of measurement to the development of a scale for assessing the validity of randomised trials, the relationship between such a score and the degree to which a study is free from bias is not obvious. None of the currently available scales for measuring the validity or 'quality' of trials can be recommended without reservation. If reviewers or CRGs choose to use such a scale, it must be with caution.

Most of the available scales for assessing the validity of randomised controlled trials derive a summary score by adding the scores (with or without differential weights) for each item. While this approach offers appealing simplicity, it is not supported by empirical evidence (Emerson 1990, Schulz 1995). Notably, scales with multiple items and complex scoring systems take more time to complete than simple approaches. They have not been shown to provide more reliable assessments of validity (Jüni 1999). They may carry a greater risk of confusing the quality of reporting with the validity of the study. They are more likely to include criteria that do not directly measure internal validity, and they are less likely to be transparent to users of the review. For these reasons, it is preferable to use simple approaches for assessing validity that can be fully reported (i.e. how each trial scored on each criterion).

6.8 Bias in non-experimental studies

The Non-randomised Studies Methods Group are preparing guidance on the use of non-randomised studies in Cochrane Reviews (appendix 6). In the meantime, this section describes some issues that should be considered in assessing the validity of non-randomised studies. The logical reason for focusing on randomised controlled trials in Cochrane Reviews is that randomisation is the only means of allocation that controls for unknown and unmeasured confounders as well as those that are known and measured. Differences between comparison groups in prognosis, responsiveness to treatment or exposure to other factors that affect outcomes can distort the apparent magnitude of effects of the intervention of interest. It is possible to control or adjust for confounders that are known and measured in observational studies, such as case-control and cohort studies. However, it is not possible to adjust for those factors that are not known to be confounders or that were not measured. Unfortunately it can rarely, if ever, be assumed that all important factors relevant to prognosis and responsiveness to treatment are known, and for those that are known difficulties can arise in measuring and accounting for them in analyses. Empirical evidence supports these logical concerns (Kunz 1995). Selection bias can distort effects in either direction, causing them to appear either larger or smaller than they are. It is generally not possible to predict the magnitude, and often not even the direction of this bias in specific studies. However, on average, selection bias tends to make treatment effects appear larger than they are and the size of these distortions can be as large or larger than the size of the effects that are being measured (Kunz 1995).

Despite these concerns, there is sometimes good reason to rely on observational studies for information about the effects of healthcare interventions, and to include such studies in Cochrane reviews. For example, well designed observational studies have provided useful data regarding the effects of interventions such as mandatory use of helmets by motorcyclists, screening for cervical cancer, dissemination of clinical practice guidelines to change professional practice and rare adverse effects of medication.

Various criteria have been suggested to critically appraise the validity of observational studies (Horwitz 1979, Feinstein 1982, Levine 1994, Bero 1999). In general, the same four sources of bias noted above can be applied to other types of comparative studies, as illustrated below:

Source of bias	Cohort studies	Case-control studies
Selection bias	Control for confounders	Matching
Performance bias	Measurement of exposure	Measurement of exposure
Attrition bias	Completeness of follow-up	Completeness of follow-up
Detection bias	Blinding	Case definition

Concerns about attrition bias are similar in randomised trials, cohort studies and case-control studies and relate to the extent that all participants in a study are appropriately accounted for in its results. Concerns about detection bias are also similar for cohort studies, and are related to the case definition that is used in case-control studies (since people are entered into such studies based on knowledge of the outcome of interest). The major difference between randomised trials and observational studies has to do with selection bias and the need to identify and account for potential confounders in observational studies. To do this reviewers must make judgements about what confounders are important and the extent to which these were appropriately measured and controlled for. Assessing 'performance bias' is also more difficult in observational studies since it is necessary to measure exposure to the intervention of interest and ensure that there were not differences in the exposure of the comparison groups to other factors that could affect outcomes. In addition to considerations of blinding, which are similar to those in randomised trials, it is important to consider whether exposure was measured in a similar and unbiased way in the groups being compared. So, for example, in addition to concerns about bias due to confounders in cohort and case control studies of the effects of post-menopausal hormone replacement therapy, investigators and reviewers must ensure that use of hormones was measured in an unbiased way.

In summary, a great deal of judgement is necessary in assessing the validity of observational studies. Judgement is also needed when the validity of randomised trials is assessed, but the nature of observational studies makes them even more difficult to critically appraise. This requires a thorough understanding of both the problem that is the focus of the review and methodological considerations. Caution is needed.

6.9 Application of quality assessment criteria

Several basic decisions must be made regarding the assessment of studies, similar to those made regarding the process of selecting studies (section 5.7). A prime consideration is the number of reviewers. Should there be one or more than one? How many are necessary and how many are too many? Will reviewers review the same articles to maximise

reliability or mutually exclusive sets of reports to minimise workload? A concomitant consideration is the backgrounds of the different reviewers and whether previous training and experience in study design or critical appraisal will be required.

Conducting systematic reviews with multiple reviewers is a two-sided coin. On the one hand it may limit bias, minimise errors and improve reliability of findings, but having more than one creates the potential for disagreement among reviewers. When multiple reviewers will be involved, there should be an explicit procedure or decision rule identified a priori for identifying and resolving disagreement. As a general rule, we recommend that at least two reviewers assess information that involves subjective interpretation and information that is critical to the interpretation of results (e.g., outcome data). Section 7 describes methods for reaching and monitoring consensus when more than one reviewer is used.

Regardless of the number of reviewers, it is important to first test any assessment criteria that are planned using a pilot sample of articles to ensure that the appraisal criteria can be applied consistently. Three to six papers that span a range of low to high risk bias might provide a suitable sample for this.

Should reviewers be especially trained in research methods, the content area of a review or both? Although experts in content areas may have pre-formed opinions that can bias their assessments (Oxman 1993b), they may nonetheless give more consistent assessments of the validity of studies than persons without content expertise (Jadad 1996). They may also have valuable insights that are different than those that someone with methodological expertise alone would have. It would seem intuitively desirable to use both content experts and non-experts and to ensure that both have an adequate understanding of the relevant methodological issues.

Reviewers must also decide whether those assessing study validity will be blinded to the names of the authors, institutions, journal and results of a study when they assess its methods. Some empirical evidence suggests that blind assessment of reports might produce lower and more consistent scores than open assessments (Jadad 1996). Other empirical evidence suggests little benefits from blind assessments (Berlin 1997). However, blinded assessments are very time consuming. Reviewers must weigh their potential benefits against the costs involved when deciding whether or not to blind the reviewers. Further research is underway comparing blind and open assessments of study validity and these results may help guide this decision.

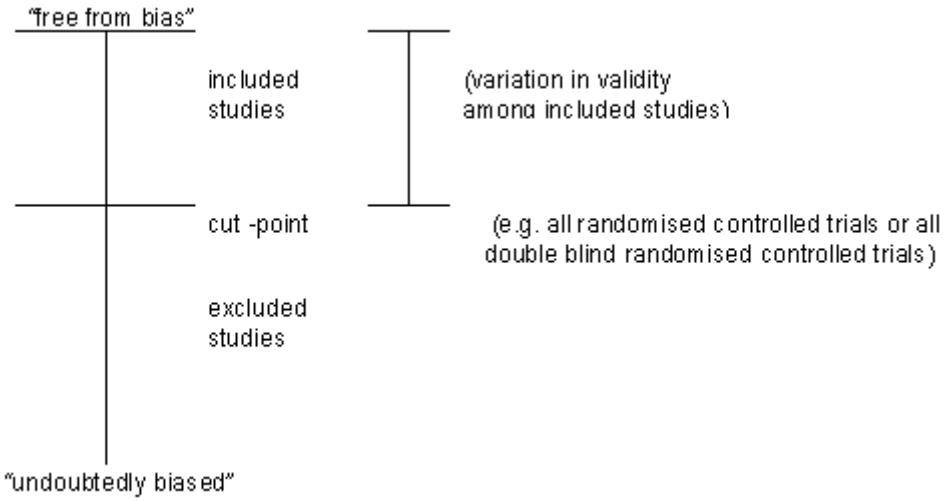
6.10 Incorporating assessments of study validity in reviews

There are several ways in which validity assessments can be used in a review:

- as a threshold for inclusion of studies
- as a possible explanation for differences in results between studies
- in sensitivity analyses
- as weights in statistical analysis of the study results

Failure to meet one or more validity criteria may indicate such a high risk of bias in some reviews that it constitutes grounds for exclusion of those studies. For example, for highly subjective outcomes such as pain, reviewers may decide to include only studies that

prevent 'performance bias' by blinding participants. The decision about where to set the cut point for inclusion can be conceptualised as existing on a continuum between 'free from bias' and 'undoubtedly biased' as illustrated below:



If reviewers raise the methodological cut-point for including studies, there will be less variation in validity among the included reports. Assessments of validity would then categorise studies by the risk of bias within the range above the inclusion cut-point. With a sufficiently high cut-point, any variation in validity among included studies may be too small to be important.

There are several methods to examine whether validity may explain differences among study results (Detsky 1992). Visual plots of the results arranged in order of their validity can be used. A second approach is to analyse subgroups of studies above a methodological cut-point, which should, preferably, be specified a priori, in the protocol of the review. This approach can be used whether or not the study results are heterogeneous, by doing a sensitivity analysis to determine if the overall results are the same when only studies with little risk of bias are included in the analysis. A third approach is to combine the results of each study sequentially in order of their assessed validity ('cumulative meta-analysis'), examining the impact on the overall results as trials of decreasing validity are included (see section 8.11.6).

A fourth approach is to use statistical methods to weight studies according to their assessed validity or to use 'meta-regression' to explore the relationship between validity and the magnitude of effect across studies (see section 8.8.1). Statistical methods for combining the results of studies generally weight the influence of each study by the inverse of the variance for the estimated measure of effect. In other words, studies with more precise results (narrower confidence intervals) are given more weight. It is also possible to weight studies according to validity so that more valid studies have more influence on the summary result. The main objection to this approach is that there is no empirical basis for determining how much weight to assign to different validity criteria or for quantitatively relating differences on 'quality' scales to differences in the risk of bias between studies.

It is possible using RevMan 4.0 to order studies according to either adequacy of concealment of allocation or 'user defined' assessments of validity. Subgroup analyses

based on assessments of validity can be done, although a test of statistical significance of differences between subgroups of studies has not been implemented. RevMan does not include an option for weighting studies by methodological validity and meta-regression is not possible using RevMan 4.0.

6.11 Limitations of quality assessment

There are two major difficulties with assessing the validity of studies. The first is inadequate reporting of trials (SORT 1994, Schulz 1994, WGRR 1994, Begg 1996). It is possible to assume if something was not reported it was not done. However, this is not necessarily correct. Reviewers should attempt to obtain additional data from investigators as necessary, but this may be difficult. The application of standards for reporting trials (SORT 1994, WGRR 1994, Begg 1996) should facilitate the assessment of study validity in the future.

The second limitation, which in part is a consequence of the first, is limited empirical evidence of a relationship between parameters thought to measure validity and actual study outcomes. As noted above, there is empirical evidence suggesting that, on average, both inadequate concealment of allocation and lack of double blinding result in over-estimates of the effects of treatment. Clearly much more research needs to be done to establish which criteria for assessing validity are indeed important determinants of study results and when. Improved reporting of methods will facilitate such research. Meanwhile, reviewers should avoid the use of 'quality scores' and undue reliance on detailed quality assessments. It is not supported by empirical evidence, it can be time-consuming and it is potentially misleading.

6.12 References

Begg 1996. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637-9.

Berlin 1997. Berlin JA, on behalf of the University of Pennsylvania meta-analysis blinding study group. Does blinding of readers affect the results of meta-analyses? Lancet 1997; 350: 185-6.

Bero 1999. Bero L, Grilli R, Grimshaw J, Mowatt G, Oxman A, Zwarenstein M (editors). Effective Practice and Organisation of Care Module. In: The Cochrane Library, Issue 2, 1999. Oxford: Update Software.

CCSG 1978. The Canadian Cooperative Study Group. The Canadian trial of aspirin and sulfinpyrazone in threatened stroke. N Engl J Med 1978; 299:53-9.

Chalmers 1983. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983; 309:1358-61.

Chalmers 1989. Chalmers I, Hetherington J, Elbourne,D, Keirse MJNC, Enkin M. Materials and methods used in synthesizing evidence to evaluate the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Keirse MJNC, editors. Effective Care in Pregnancy and Childbirth. Oxford: Oxford University Press, 1989; 39-65.

Colditz 1989. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: medical. Stat Med 1989; 8:441-54.

Detsky 1992. Detsky AS, Naylor CD, O'Rourke K, McGreer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. J Clin Epidemiol 1992; 45:255-65.

Emerson 1990. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. Controlled Clin Trials 1990; 11:339-52.

Feinstein 1982. Feinstein AR, Horwitz RI. Double standards, scientific methods, and epidemiological research. N Engl J Med 1982; 307:1611-7.

Feinstein 1985. Feinstein AR. Clinical Epidemiology: The Architecture of Clinical Research. Philadelphia: Saunders, 1985: 39-52.

Gotzsche 1989. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. Controlled Clin Trials 1989;10:3-56.

Horwitz 1979. Horwitz RI, Feinstein AR. Methodological standards and contradictory results in case-control research. Am J Med. 1979; 66:556-64.

Jadad 1996. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Controlled Clin Trials 1996; 17:1-12.

Jüni 1999. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282: 1054-60.

Karlowski 1975. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold: a prophylactic and therapeutic trial. JAMA 1975; 231: 1038-42.

Kleijnen 1997. Kleijnen J, Gotzsche P, Kunz RA, Oxman AD, Chalmers I. So what's so special about randomization? In: Chalmers, I, Maynard, A, editors. Non-Random Reflections on Health Services Research. London: BMJ, 1997; 93-106.

Kunz 1995. Kunz RA, Oxman AD. Empirical evidence of selection bias in studies of the effects of health care: a systematic review. Presented at the Cochrane Colloquium, Oslo, 5-8 October, 1995.

Kunz 1998. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 1998; 317:1185-90.

Levine 1994. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, IV: how to use an article about harm. JAMA 1994; 271:1615-9.

Moher 1995. Moher D, Jadad A, Nichol G, Penman M, Tugwell T, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Controlled Clin Trials 1995; 16:62-73.

Moher 1996b. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. Int J Tech Assess in Health Care 1996; 12:195-208.

Moher 1998a. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.

Oxman 1993b. Oxman AD, Guyatt GH. The science of reviewing research. Ann NY Acad Sci. 1993;703:125-33.

Reitman 1988. Reitman D, Chalmers TC, Nagalingam R, Sacks H. Can efficacy of blinding be documented by meta-analysis? Presented to the Society for Clinical Trials, San Diego, 23-26 May, 1988.

Sackett 1979b. Sackett DL. Bias in analytic research. J Chronic Dis 1979; 32:51-63.

Schulz 1994. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA 1994;272:125-8

Schulz 1995. Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. JAMA 1995; 273:408-12.

SORT 1994. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. JAMA 1994; 272:1926-31.

WGRR 1994. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Ann Intern Med 1994; 121:894-5.

7. COLLECTING DATA

7.1 Rationale for data collection forms

The data collection form is a bridge between what has been reported by primary investigators (e.g journal articles, project reports, personal communications) and what is ultimately reported by a reviewer. The data collection form serves at least three important functions. First, the data collection form is directly linked to the formulated review question and planned assessment of included studies and, therefore, provides a visual representation of these. Second, the data collection form is the historical record of the multitude of decisions (and changes to decisions) that occur throughout the review process. Third, the data collection form is the data repository from which the analysis will emerge.

Given the important functions of data collection forms, ample time and thought should be invested in their designs. Because each review is different, data collection forms will vary across reviews. However, there are similarities regarding types of information that are important, and forms can be adapted from one review to the next.

7.2 Electronic versus paper data collection forms

Should reviewers design paper data collection forms or automate the review process with electronic data collection forms? Paper forms can be easier to design because electronic forms require computer programming knowledge. On the other hand, large amounts of data from reviews involving large numbers of studies are more easily stored and retrieved with electronic than paper forms. Electronic forms eliminate the need for data entry separate from data abstraction. They also can be used to calculate simple variables or conversions (e.g. pounds to kilograms) for data that is presented in various formats in different studies. Both electronic and paper forms can be designed to provide an historical record of decisions and refinements that occur throughout the review process.

Many reviewers use a double-abstraction process whereby two independent assessments of each study can be compared and reconciled if necessary. When using a paper data collection form, the comparison process is simple: one form is used to mark and correct errors and disagreements. Comparing double-abstractions using electronic forms is fast but requires the writing and testing of programs within the structure of the database being used. Identifying and addressing errors and disagreements among reviewers may be more difficult with electronic than paper forms. This is because fields or areas of data collection forms that allow open-ended responses are not easily compared electronically. Amendments or changes to original forms may be more difficult with electronic than paper forms because of programming issues. A final potential drawback to electronic data collection forms is whether they will be compatible with Review Manager (RevMan) which is used to generate and store the final review. Although there are ways to transfer data from electronic data collection forms to RevMan, this might not be straightforward and should, ideally, be planned in advance.

If an electronic form is used, consider the following guides. First, do not program the electronic form until you have designed, piloted, and refined a paper copy of the form.

Such pilot tetsting ideally involves more than one reviewer and several articles. Second, when designing the data collection form, consider the needs of the data entry person, structure the form in a logical manner and make coding of responses as consistent and straightforward as possible. Third, when choosing an electronic database or spreadsheet, check whether it can create an electronic file that will be transferable to RevMan. Fourth, don't forget to develop quality control mechanisms for assessing and correcting data entry errors.

7.3 Data management and software

A variety of software and data management programs may be helpful in the systematic review process. Spreadsheet software such as QuatroPro, Excel and Lotus or database programs such as FoxPro or DataEase can be used for electronic data collection forms. Software such as DBMSCOPY may be useful for converting such database files into files compatible with data analysis, if analyses not available in RevMan are planned (see section 8.8).

7.4 Key components of a data collection form

There is no single correct way to design a data collection form. The following suggestions are
based on experience. When adapting or designing a data collection form reviewers first should consider how much information they want to collect. Overly detailed collection can result in forms that are longer than original study reports, tedious and boring to complete, and wasteful of reviewer time. On the other hand, if forms are not sufficiently detailed and omit key data, reviewers may have to re-abstract studies using supplemental data collection forms. Having to review a study a second time can be frustrating and time-consuming.

7.4.1 Information about study references and reviewers

Because data collection forms are adaptable across reviews and some reviewers participate in multiple reviews, a clear title of the review is needed and the name of the reviewer who is abstracting data should be recorded. It is useful to leave space after the title so reviewers can write notes specific to the study being abstracted. This avoids placing notes, questions or reminders on the last page of the form where they are least likely to be noticed. Important notes may be entered into RevMan in the 'notes' column of the Characteristics of Included Studies table, or in the text of the review. Every Cochrane Review is assigned a unique identifier. This should be included next to the title on the data collection form. Forms occasionally have to be revised. Coding the form with a revision date or version number reduces the chances of erroneously using an outdated form by mistake.

Each included study must be given a study identifier that is used in RevMan. Reviewers may need to collect data from multiple reports of the same trial. It is a good idea to record the source of key information, including where it was found in a report or if information was obtained from unpublished sources and personal communications. Any unpublished information that is used should be written and coded in the same way as published information.

7.4.2 Verification of study eligibility

Although the search and selection process should have weeded out most ineligible studies, it is good to verify study eligibility at the time of data abstraction or collection. Verification information should occur early because the remainder of the form pertains to studies which meet inclusion criteria and the extraction of data from studies that will be excluded is a waste of resources.

Cochrane Reviews include an excluded studies table for studies that appear to meet the inclusion criteria and which others might believe to be relevant, but upon closer inspection were excluded. The verification information on the data collection form can be a mechanism for coding reasons why such studies were excluded. For example, a reviewer may only include truly randomised trials in a review. A verification query on the data collection form might be: Randomised? Yes, No, Unclear. If the study used alternate allocation, the answer to the query is no, and this information would be entered in RevMan as the reason for exclusion.

7.5 Study characteristics

When assessing each study, it is necessary to code specific study characteristics. These can be categorized into groups that match information that will be entered into RevMan: methods, participants, interventions, and outcomes. Information under participants might include details relevant to the study setting and diagnostic criteria for the condition of interest. The development of this part of the data collection form deserves careful thought and pilot testing. Data that is collected should be directly linked to the review question(s) and planned analysis strategies. It should be collected in a format conducive to logical entry into RevMan.

7.5.1 Methods

Different research methods can influence study outcomes by introducing bias and artefacts in study results. For example, whether allocation was adequately concealed is important, as discussed in section 6. When entering information about particular studies in RevMan, it will be necessary to code allocation concealment as adequate (A), unclear (B), inadequate (C) or not used (D). Data collection forms should reflect these assessments. Other methods features that may be relevant include study duration; type of trial such as parallel or cross-over design; patient, provider and outcome assessor blinding; amount of drop-outs and cross-overs; cointerventions and other potential confounders. The methods part of the data collection form should include any validity criteria that are used.

7.5.2 Participants

Characteristics of participants may vary substantially across studies and some Collaborative Review Groups (CRGs) have developed standards regarding which characteristics should be collected. Typically, items that should be collected are those that could affect study results or help users assess applicability. For example, if the reviewer has reason to suspect important treatment effect differences between various ethnic populations, this information should be collected. If treatment effects are thought constant over ethnic groups, and if such information would not be useful to help apply

results, it should not be collected. Items that are often useful for assessing applicability include age and sex. Occasionally, other sociodemographic items such as education level are important as well as items addressing the presence of important comorbid conditions.

If the settings of studies are likely to influence treatment effects or applicability, they should be assessed. Typical settings that are involved in healthcare intervention studies are: acute care hospitals, emergency facilities, offices or clinics, extended care facilities such as boarding and nursing homes, and communities. Sometimes studies are conducted in different geographical regions that have important differences in cultural characteristics that could affect delivery of an intervention and its outcomes. Sometimes temporal settings indicate important technology differences. If such items are important for the interpretation of the review, they should be assessed.

Diagnostic criteria that were used to define the condition of interest can be a particularly important source of clinical heterogeneity and should be described. For example, in a review of drug therapy for congestive heart failure, it is important to know how the definition and severity of heart failure was determined in each study (e.g. systolic or diastolic dysfunction, severe systolic dysfunction with ejection fractions of < 20%, etc.). Similarly, in a review of antihypertensive therapy, it is important to describe baseline levels of blood pressure of participants.

7.5.3 Interventions

The intervention and how it was delivered should be described. For trials of pharmaceutical agents, routes of delivery (e.g., oral, intravenous), doses, and timing (e.g. within 24 hours of diagnosis) may be assessed. Treatment length also may be recorded here, particularly if it was different than study follow-up length and was not recorded under methods. For complex interventions such as those that evaluate psychotherapy, behavioural and educational approaches or healthcare delivery strategies, it is important to collect information that will help to disentangle the underlying relationships. This includes information about who delivered the intervention, its contents, format, timing, etc.

For trials that do not utilise placebos and those that evaluate complicated interventions, it is also important to collect information regarding what was given to the control group. This will help guide later decisions about whether it is reasonable to combine data across studies; since marked heterogeneity in what is received by control groups may be a reason for not combining studies, or for doing sensitivity analyses.

7.5.4 Outcome measures and results

What may appear to be obvious and simple may in fact be one of the more difficult sections of the data collection form to design. Reports of studies often include more than one outcome (mortality, morbidity, quality of life, etc.), may report the same outcome using different measures, may include outcomes for subgroups and may report outcomes measured at different points in time. The reviewer needs to integrate what type of outcome information is needed to answer the review's question(s) with what is likely to be in the reports of studies. To avoid hidden mistakes outcomes should be collected in the format they were reported and transformed in a subsequent step. For cross-over trials and

trials with outcome assessments at various periods of follow-up, decisions will need to be made about which outcomes to assess (see section 8.11.5 and section 8.9.1 respectively).

Reviewers should consider formatting the forms to match RevMan data tables. For example, if the reviewer plans to use continuous data, the following information is required for each comparison group: the number of participants, the mean and the standard deviation. However, these data fields may be insufficient because there is great variation in what researchers report and fail to report. In this example, investigators may have reported a confidence interval for the mean difference and not reported any standard deviations, or they may simply have reported the value of a test statistic (t test, F test, chi-square test, etc.) or a p-value. Data collection forms should incorporate flexibility for addressing this type of variability in outcome assessments. For more detail, regarding what outcome information is necessary for specific types of analyses, see section 8

7.6 Coding format and instructions for coders

Accurate coding is extremely important. The coding should not be so complicated that the abstractor is easily confused or likely to make poor decisions. Reviewers need instructions and decision rules on the data collection form. There are varying preferences regarding where instructions should be included. One approach is to insert the instruction adjacent or near to the data field that is to be coded. In some cases, instructions can be lengthy and may have to be placed on a separate page. Regardless of the approach used (most likely it will be a mixture), it is crucial for reviewers to practice using the form and receive, or give, training if the form was designed by someone other than the person using it.

7.7 Pilot testing and form revisions

All forms should be pilot tested using a representative sample of the studies to be reviewed. This test is likely to identify data that are not needed or are missing. Abstractors may provide feedback that certain coding instructions are confusing or incomplete (e.g. all of the types of responses might not be described). When multiple reviewers are participating on a project, there may need to be a consensus among them before the form is modified to avoid any misunderstandings or later disagreements. Depending on the complexity of the review and the experiences from piloting, additional pilot tests may be necessary.

Problems with the data collection form will occasionally surface after pilot testing has been completed and the form has been revised. In fact, it is rare for a data collection form to not require any modifications after it has been piloted. When changes have to be made to the form or coding instructions, be sure to correct the forms of those studies that have already been reviewed. In some situations, it may only be necessary to clarify coding instructions without modifying the actual data collection form.

7.8 Reliability of data collection

Reliability refers here to the degree to which different people review a study in the same way. For example, did each reviewer agree on the presence of comorbidity among subjects in a specific trial? Did reviewers agree on the outcome data in each comparison

group? When more than one person is reviewing data, there will inevitably be disagreements. Multiple reviewers need to develop a plan for comparing information in their data collection forms and for reaching consensus when there are disagreements. Consensus can be achieved by discussion among reviewers or by using an additional independent arbitrator. It is also important to plan how the 'consensus' agreement will be recorded. There are at least three possibilities: 1) use one reviewer's form and record changes after consensus in red ink; 2) use a separate printed form; or 3) enter only the consensus data onto an electronic form. Keeping the 'consensus' information separate is essential for assessing the reliability of coding.

It may not be important to formally examine reliability for all of the collected data; for example, a reviewer may elect to limit the evaluation of reliability to the coding of outcomes and for validity assessments. There is no fixed standard for the degree of reliability that is adequate or how to assess reliability. However, it is important to examine reliability throughout the data collection process. For example, if after reaching consensus on the first few studies, the reviewers note a frequent disagreement for specific data, then coding instructions may need modification. Reviewers may display 'coder drift' (a change over time in how information is coded), indicating a possible need for re-training or re-coding. This can only be identified when reliability is examined throughout the project.

7.9 Blinded data extraction

UNDER CONSTRUCTION – A section is being prepared on the issue of whether data extraction should be done blinded; for example to the authors and journal and to the results when assessing quality. Although there is some evidence that blinded assessments of the quality of trials may be more reliable and different from assessments that are not blinded (Jadad 1996, Moher 1998b), blinding is difficult to achieve, time consuming and may not substantially alter the results of a review (Berlin 1997a, Berlin 1997b).

7.10 Collection of data from investigators

Reviewers will often find that they are unable to extract all of the information they are interested in from published reports, both with regard to the details of the study and its numerical results. In such circumstances, the reviewers need to determine how to collect the missing information. They might wish to contact the original investigators and should, for example, consider whether they will contact them with an open-ended request, send them their standard data collection form, request individual patient data (see section 11) or do something else.

7.11 References

Jadad 1996. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Controlled Clin Trials 1996; 17:1-12.

Moher 1998b. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses?. Lancet 1998; 352:609–13.

Berlin 1997a. Berlin JA, Miles CG, Crigliano MD. Does blinding of readers affect the results of meta-analyses? Results of a randomized trial. Online J Curr Clin Trials 1997.

Berlin 1997b. Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. Lancet 1997; 350: 185-6.

8. ANALYSING AND PRESENTING RESULTS

8.1 Do not start here!

Please read sections 3 - 7 before reading this section. It is sometimes tempting to jump directly into the analysis when undertaking a review. This is hazardous if appropriate attention has not been given to formulating the question (section 3 and section 4), identifying, selecting and critically appraising studies (section 5 and section 6) and collecting data (section 7).

8.2 Planning the analysis

The reason for conducting reviews systematically is to ensure that the results are valid. The role of statistical analysis in reviews may be less clear. Sometimes the use of statistics (meta-analysis) may seem and, in fact, be more of a hindrance than a help to those who are not familiar with the techniques. There is also a danger in focusing too much attention on the statistical bottom line or the diamond used at the bottom of a graph to sum up the individual study results in a meta-analysis. Sometimes reviewers and users of reviews, in their rush to get to this, lose sight of the importance of reflection and judgement in the steps preceding analysis, in the analysis itself and in drawing conclusions. In this section we try to clarify what the analysis of a review entails, the role of statistics in this process and the need for judgement by the reviewer and user.

8.2.1 Why perform a meta-analysis in a review?

To prepare a review, one must collect data from individual studies and statistical methods can be used to analyse and summarise these data. If used appropriately, these methods provide a powerful tool for deriving meaningful conclusions from the data and help prevent errors in interpretation.

A common error when statistical methods are not used in a review is to compare the number of 'positive' studies with the number of 'negative' studies. This 'vote counting' should be avoided. It is unreliable, since whether a study is counted as 'positive' or 'negative' may depend on how the results are interpreted by the reviewers and it gives no consideration to the relative weight of reliable evidence contributed by each study. There is also a tendency to overlook small but clinically important effects when counting votes, particularly when counting studies with statistically non-significant results as 'negative' (Cooper 1980, Antman 1992). It is also possible for a reviewer to introduce bias by inappropriately stressing the results of one study over another when using non-quantitative methods.

Of course, the use of statistical methods does not guarantee that the results of a review are valid, any more than it does for a primary study. Moreover, like any tool, statistical methods can be misused.

8.2.3 What does a meta-analysis entail?

While the use of statistical methods in reviews can be extremely helpful, the most essential element of an analysis is a thoughtful approach, to both its qualitative and quantitative elements. This entails consideration of the following questions:

- What comparisons should be made?
- What study results should be used in each comparison?
- Are the results of studies similar within each comparison?
- What is the best summary of effect for each comparison?
- How reliable are those summaries?

The first step in addressing these questions is to decide which comparisons to make. The next step is to prepare tabular summaries of the characteristics and results of the studies that are included in each comparison. It is then possible in a systematic way to derive estimates of effect across studies, to conclude how much confidence should be placed in those estimates and to investigate differences among studies.

8.2.4 Which comparisons should be made?

The first and most important step in planning the analysis is to specify the comparisons that will be made. These should relate clearly and directly to the questions or hypotheses that are posed when the objective of the review is formulated (see section 4). It should be possible to specify the main comparisons that will be made in the protocol of a review. However, it will often be necessary to modify comparisons and add new ones in light of the data that are collected. For example, it may not be possible to know before data are collected which outcome measures can sensibly be analysed together, and important variations in the intervention may only be discovered after data are collected.

Decisions about which study results are similar enough that they should be grouped together require an understanding of the problem that the review addresses and judgement by the reviewer and the user. The formulation of the questions that a review addresses is discussed in section 4. Essentially the same considerations apply to deciding what comparisons to make, what outcomes to combine and what key characteristics (of study design, participants, interventions and outcomes) to consider when investigating heterogeneity (variation in effects). These considerations must be addressed when setting up the Comparisons in RevMan and in deciding what information to put in the table of Characteristics of Included Studies.

8.2.5 What study results are needed for each comparison?

Having specified the comparisons that will be made, the next step in the analysis is to prepare tabular summaries of the data. For dichotomous data (i.e. outcomes that are either present or not, such as death), the data that need to be summarised are the number of people who experienced the event or outcome in each comparison group and the total number in each group. Consideration should be given to the order in which the results will be presented (e.g. by year, alphabetically by study ID, or based on the results). The order can be modified both in RevMan and in the *Cochrane Database of Systematic Reviews* (*CDSR*), but the default display for each outcome in each review is determined by the reviewer.

Because concealment of allocation is a validity criterion that applies to all outcomes in a study, and there is good evidence to support its use, this is assigned when studies are entered in RevMan and is one characteristic by which studies can be sorted in the analyses. It is also possible for reviewers to order studies based on other assessments of validity or other, user-defined characteristics.

The data that need to be summarised for continuous data (such as blood pressure) are the number of people in each group, the mean value for the outcome in each group and the standard deviation for each mean. Unfortunately, standard deviations are often not reported. Although it may be possible to calculate standard deviations from other measures of variation that are reported (e.g. confidence intervals), this is not always possible. Options for handling this and other types of missing data are discussed below in section 8.11.4.

Sometimes results for a continuous outcome will be reported as changes in some studies (i.e. the mean difference between pre- and post-intervention measurements) and post-intervention results in other studies. There is no single best solution for summarising results when this is the case. In general, reviewers should first determine which way of reporting results is most likely to be intuitively understandable and helpful to users of their review. So far as possible, results should be transformed or data should be sought from investigators so that all of the study results can be presented in the same way. Various options are available for imputing results, but they all rely on assumptions and should be used cautiously, if at all, and only if the assumptions are likely to hold true. Sensitivity analyses should be done to test the impact of changing the assumptions (see section 8.10.)

Occasionally reviewers encounter a situation where data for the same outcome are presented in some studies as dichotomous data and in other studies as continuous data. For example, scores on depression scales can be reported as means or as the percentage of patients who were depressed at some point after an intervention (i.e. with a score above a specified cut-point). This type of information is often easier to understand and more helpful when it is dichotomised. However, deciding on a cut-point may be arbitrary and information is lost when continuous data are transformed to dichotomous data. There are several options for handling combinations of dichotomous and continuous data. Generally, it is useful to summarise results from all the relevant, valid studies in a similar way, but this is not always possible. It may be possible to collect missing data from investigators so that this can be done. If not, it may be useful to summarise the data in three ways: by placing the continuous data in a Continuous Data Table, Dichotomous data in a dichotomous Data Table and all of the data in an Other data Data Table. Another option is to present all of the data as standardised mean differences. Alternatively, the log-odds-ratio can be used as a measure of treatment difference (Whitehead 1999).

A number of other situations arise when it might not be appropriate to summarise data in either a dichotomous or continuous Data Table. One example is when studies address the same question, but use different outcome measures. For example, if one is interested in synthesising research into the effects of audit and feedback on professional practice, it can be helpful to summarise all valid studies that compared the effects of audit and feedback versus no intervention on objective measures of performance, even though the types of professionals, settings and outcomes vary substantially from study to study. Although it

might be possible to transform results in this case to standardised mean differences, these might be difficult to interpret and misleading. It might also be possible to standardise results in some other way (e.g. percent change). Even then, people making decisions are likely to want a summary of the results in 'natural units' that are intuitive.

If individual patient data have been collected from the studies in the review, this can be analysed outside of RevMan to produce summary statistics for each study. These summary statistics, along with the number of events and participants in each group, are input to an individual patient data Data Table in RevMan. The analysis program then uses the summary statistics that were input, rather than the crude numbers.

There is no standard set of data that should be included in a Data Table for 'Other' types of data; i.e. data that do not neatly fit into Dichotomous, Continuous or Individual Patient Data Data Tables. In general, five types of information are likely to be useful in summarising the results and facilitating their analysis:

- key variables that characterise the participants
- key variables that characterise the interventions
- key variables that characterise the outcome measures
- the results in natural units that are easily understood
- standardised results, if these can be derived in a way that will facilitate comparisons across studies without being misleading

The first three types of information are 'independent variables' that can help to explain the results. A fourth type of independent variable that might also be included is the internal validity of each study. The choice of specific variables depends on an understanding of the problem that the review addresses. They should, however, correspond to any *a priori* hypotheses that were made in the protocol regarding important differences between studies.

If identified studies contribute no data for a particular outcome, the reviewer should consider whether or not this lack of data might be biased. For example, some outcomes will not have been collected in certain studies because the relevant means of measurement were not available at the time of the study and, so, the reason for their unavailability cannot have been influenced by the study's results. However, if data on the relevant outcomes were collected in the study but are not available for the review, this might be due to bias. The reviewer must decide whether studies with no data should be included with the associated comparison in the review. The inclusion of such studies helps to inform the reader of how much evidence is missing for the outcome of interest. This is particularly important if the relevant data were collected in a study but are not available.

8.3 Summarising effects across studies

An aim of most Cochrane Reviews is to provide a reliable estimate of the effects of an intervention, based on a weighted average of the results of all the available relevant studies. Typically, the weight given to each study is the inverse of its variance, i.e. more precise estimates (from larger studies with more events) are given more weight (Laird 1990). It is also possible to give studies more or less weight based on other factors such as their methodological quality, but this is rarely done (Detsky 1992).

If it makes sense to combine the results of a group of studies and the observed differences between the results of the studies are not statistically significant or practically important, it is relatively straightforward to combine the results. Each study is summarised using a measure of effect (such as an odds ratio, a relative risk or a mean difference) that represents the within study comparison of the intervention and control groups. In this way participants in each study are only compared to other participants in the same study.

Occasionally, a meta-analysis may be seen in which an 'effect size' has been calculated for each intervention group in each randomised trial (based on the difference in outcomes before and after the intervention) and groups from different studies are then compared with each other. In this way, each intervention group becomes the equivalent of a before-and-after study and the benefits of the original randomisation in each trial are lost. The results of any such meta-analysis need to be interpreted with a great deal of caution and it should not be regarded as a meta-analysis of randomised trials. This approach to synthesising data is not supported in RevMan and should be avoided.

8.4 Statistical methods available in Review Manager and CDSR

Meta-analyses use a variety of techniques. Unfortunately, there is not one, single 'correct' technique. The choice of technique depends on the nature of the data being analysed. Fortunately, as with other uses of statistics by reviewers, a conceptual understanding of the principles is more important than detailed knowledge of the specific techniques. The statistical methods available in RevMan are listed below:

Type of data	**Summary statistic**	**Model**	**Method**
Dichotomous	odds ratio (O-E)	fixed effect	Peto Mantel-Haenszel
		random effects	DerSimonian and Laird
	relative risk	fixed effect	Mantel-Haenszel
		random effects	DerSimonian and Laird
	risk difference	fixed effect	Mantel-Haenszel
		random effects	DerSimonian and Laird
Continuous	Weighted mean difference	fixed effect	inverse variance
		random effects	DerSimonian and Laird
	Standardised mean difference	fixed effect	inverse variance
		random effect	DerSimonian and Laird
Individual patient data	odds ratio (O-E)	fixed effect	Peto

NOTE: RevMan 4.2 also includes a new statistical method: generic inverse variance. More information on this method will be added to a future revision of the Handbook.

8.4.1 Dichotomous data

For dichotomous (or binary) data there are pros and cons for each of the summary statistics available: odds ratio, relative risk (or relative effect), and risk difference (see appendix 8a). Although the odds ratio has been commonly used in meta-analysis, there are concerns about the potential for it to be interpreted incorrectly (Sinclair 1994). The odds ratio has statistical advantages relating to its sampling distribution and its suitability for modelling (Fleiss 1993), but these advantages are not always important in meta-analyses. Moreover, there is no solid basis for assuming that the odds ratio is more constant across studies than other summary statistics (Sinclair 1994). The odds ratio and the relative risk are similar if the outcome is relatively rare.

The statistical method that is used and the presentation of results can be considered as separate issues. For example, odds ratios and relative risks are relative measures and can be used to combine studies, but the risk difference is an absolute measure and is better used when applying the results to a particular healthcare situation (Egger 1997a). The number needed to treat (NNT) (Laupacis 1988), is often a useful way to re-express results. It can be obtained easily from the risk difference and, with more difficulty, from the relative risk or odds ratio. Meta-analyses should never be done using the NNTs from individual studies.

Absolute measures can be more informative than relative measures because they reflect the baseline risk as well as the change in risk with the intervention. When absolute measures, such as the NNT, are reported, it is useful to report them for a range of baseline risks (say, the range observed in the included studies) (Smeeth 1999). At the same time, obtaining confidence intervals for absolute measures are problematic (Altman 1998), in part because they also vary with baseline risk. Also, NNTs are specific to a follow-up period, and should only be calculated when trials of similar follow-up periods are combined (Smeeth 1999). Information on this should be included in the review. An NNT calculator is not available in RevMan 4.0 but methods for calculating NNTs can be found in Appendix 8b.

It is often desirable to describe results both as a relative measure (odds ratio, relative risk or relative risk reduction) and an absolute measure (NNT or risk difference). In general, relative risk should be used whenever possible.

8.4.2 Continuous data

For continuous data, the weighted mean difference should be used whenever outcomes are measured in a standard way across studies. This has the advantage of summarising results in natural units that are easily understood. Occasionally, it may be desirable to summarise results across studies with outcomes that are conceptually the same but measured in different ways (e.g. different measurement scales in schizophrenia). Under these circumstances standardised mean differences can be used. Because this approach makes it possible to combine the results of highly dissimilar studies, reviewers must be particularly cautious in deciding whether such an analysis makes sense and in interpreting the results.

Reviewers should be aware of the possibility of skewed data and the potential problems it presents. To detect skew in measurements that are always positive (e.g., blood pressure),

divide the mean by the standard deviation. If the result is less than 1.64 there is some positive skew.

8.4.3 Individual patient data

For individual patient data, the summary statistics for each study are calculated by a re-analysis of raw data. This analysis is not possible within RevMan and should be done in a separate analysis program with assistance from a statistician. The summary statistics can then be input to RevMan and combined using the Peto Odds Ratio.

8.4.4 Fixed effect versus random effects analysis

There are different techniques for summarising the results of studies, and, unfortunately, there is disagreement on which is best, or whether the results should be combined at all in some situations, such as when there is substantial heterogeneity between the results of the studies (Schulz 1993). (Heterogeneity is discussed in section 8.6.) The primary disagreement that arises in meta-analyses is whether to incorporate between-study variation (heterogeneity) in estimating summaries of effect size. If there is little between-study variation (i.e. the test for heterogeneity results in a large P value, say greater than 0.10), this choice will usually make little difference in the results. If, however, there is significant between-study variation, an analysis that ignores this variation (a 'fixed effect' approach) will give a narrower confidence interval than one that does not (a 'random effects' approach). For each approach, there are situations where the result is counter-intuitive, and so it may be appropriate that results are analysed using both methods. Beyond this, there is some common ground on at least four points.

First and foremost, reviewers should alert readers of their review when there is substantial heterogeneity, and they should encourage cautious interpretation of the aggregated results. They might even decide that the heterogeneity is such that it would be preferable not to combine the results of the individual studies. Reviewers should attempt to explore the reasons for the heterogeneity and try to explain it, but any interpretations must be cautious because these analyses are usually *post hoc*. Ideally, factors that might lead to differences between the results of individual studies should be specified *a priori* in the protocol for the review. These might include differences in the formulations of the interventions, the participants, the means of assessing the outcomes or the methodological quality of the studies. Differing methodological quality could produce bias and accompanying heterogeneity. If there are important differences in the quality of studies, reviewers should consider the effect of excluding poorer quality studies from the review. Subgroup analyses, which can be used to explore differences between studies, are discussed in section 8.7 (Yusuf 1991, Oxman 1992).

Secondly, it is generally agreed that the fixed effect approach is valid as a test of significance of the overall null hypothesis (i.e. 'no effect in all studies'). A statistically significant result indicates that there is an effect in at least one of the studies.

Thirdly, whether heterogeneity is present or not, the overall result from a fixed effect approach is an average measure of treatment effect in the studies in the analysis.

Fourthly, the random effects approach relies on assumptions that the studies are a random sample from a hypothetical population of studies and that the heterogeneity between

studies can be represented by a single variance. Moreover, a pragmatic concern is that the random effects approach gives more weight to the results of smaller studies than the fixed effect approach, and such studies are often of poorer quality and may be more susceptible to publication bias than larger studies.

In practice, results can be analysed using both a fixed effect and a random effects model. If this is done, it is best to decide in advance how the results will be reported to avoid bias in deciding which model to emphasise. The random effects model will tend to give a more conservative estimate (i.e. with wider confidence intervals), but the results from the two models should basically agree where there is no heterogeneity. If the test of heterogeneity is statistically significant and the differences in results are of practical importance, then more emphasis should be placed on the random effects model. However, reviewers should be cautious about over-interpreting small differences in the confidence intervals from different analyses.

In addition to the alternative models (fixed versus random effects), there are different methods available for analyses using each which a reviewer might wish to explore with assistance from a statistician. Generally, not much difference exists between the results from the various fixed effect approaches (Thompson 1991). Less is known about the various random effects approaches.

8.4.5 Summary

While it may be desirable to maintain consistency across reviews with regard to the use of statistical methods and presentation of results, it is difficult to justify restricting analyses to any particular method. However, to help protect against bias and misuse of statistical methods:

- Each CRG should have a policy regarding which statistical method(s) to use.
- If a reviewer deviates from the CRG's policy, the reason for this should be reported in their review.

8.5 Displaying results

The analytic methods that are used in a review should be described in the Methods section and the results should be described in the Results section. Reviewers should ensure that the Results section concentrates on the objectives and the comparisons specified in the protocol of the review. *Post hoc* analyses should be identified as such.

When Cochrane Reviews are published, they contain the data entered into the Data Tables by reviewers. Tables and graphs that are displayed in RevMan and *CDSR* are generated using a special program, MetaView. This allows both reviewers and those using *CDSR* to select alternative statistical methods and to view the results in ways other than those specified in the Methods section of the review. To avoid confusion, the default for each outcome in each review is specified by the reviewers. This ensures that the results displayed are consistent with what is described in the text, unless someone using *CDSR* actively chooses to override this default.

The convention in *CDSR* is that if the odds ratios or relative risks for unfavourable dichotomous outcomes are less than one (and risk differences are less than zero), this

indicates that treatment is better than control. This is represented in the graphs by estimates to the left of the vertical line that indicates no difference. Both reviewers and users of the reviews can be confused when this convention is not maintained and there is a greater danger that errors will be made. However, there are circumstances where it may not be appropriate to adhere to this convention of showing 'bad' outcomes. It sometimes makes more sense to present the results for 'good' outcomes; for example, livebirth after treatment for infertility instead of 'no livebirth', or smoking cessation after an intervention to help people to quit smoking instead of 'failure to quit smoking'. For some outcomes deciding whether an outcome is 'good' or 'bad' may reflect personal values. In circumstances such as these, reviewers may elect not to adhere to the convention of reporting results for 'bad' outcomes. In addition, in some analyses two different interventions might be compared, so that deciding which is 'treatment' and which is 'control' is arbitrary. To accommodate these issues, RevMan 4.0 allows reviewers to specify the labels used for 'treatment' and 'control' groups in each outcome. These labels are then used in *CDSR*.

RevMan 4.2 includes the ability to add additional figures to Cochrane reviews.

8.6 Investigating heterogeneity

It is important to ask whether the results of studies are similar within each comparison or, put another way: Are the differences among the results of the studies greater than could be expected by chance? One way of doing this is to look at a graphical plot of the results. If the confidence intervals for the results of each study (typically presented by horizontal lines) do not overlap, it suggests that the differences are likely to be statistically significant (Walker 1988). Tests of homogeneity are formal statistical analyses for examining whether the observed variation in study results is compatible with the variation expected by chance alone. The more significant the results of the test (the smaller the p-value), the more likely it is that the observed differences were not due to chance alone.

When there is statistically significant heterogeneity (a low probability, perhaps less than 1 in 10 or 1 in 20), this suggests that the observed differences in the results of individual studies might be caused by factors other than chance. When looking for possible explanations for the differences, however, it is important to be cautious. Studies usually differ in many ways, and it may be incorrect to attribute a difference in results to any single factor.

RevMan will automatically test the homogeneity of the results of the individual studies being combined for each comparison of dichotomous, continuous or individual patient data. For all reviews, though, including those that have used 'Other data', the underlying biology, psychology and sociology should be kept in mind when considering any differences between studies.

8.7 Subgroup analyses

Irrespective of the extent of heterogeneity in the results, it is often of interest to examine a particular category of participants in a review (e.g. women, a particular age group, or those with a specific disease subtype). These examinations, or subgroup analyses, are exceedingly common, but they are also often misleading (Yusuf 1991, Oxman 1992) Conclusions based on subgroup analyses can do harm both when they lead to a particular

category of people being denied effective treatment (a 'false-negative' conclusion), and when they lead to ineffective or even harmful treatment being given to a particular category of people (a 'false-positive' conclusion). Subgroup analyses can also generate misleading recommendations about directions for future research that, if followed, would waste scarce resources.

Because of these risks and the frequency of their occurrence, reviewers need to be cautious about doing subgroup analyses and about interpreting the results of the ones that they feel compelled to do. The following guidelines can be used to help decide whether to do a subgroup analysis, whether an apparent difference between the effect of the intervention in these subgroups is real and, consequently, when to include a conclusion based on a subgroup analysis in the review (Oxman 1992). They can also be used to guide decisions about whether apparent differences between the effect of different formulations of the intervention (such as the specific drug or dose that was used) are real.

- Is there indirect evidence in support of a difference?

Another way of asking this question is: Is it plausible to expect there will be a difference, based on what is known about the intervention and the problem? If not, the subgroup analysis probably should not be done.

- Did the hypothesis about the difference precede rather than follow the analysis?
- Is the subgroup analysis one of a small number of hypotheses tested?

Reviewers should, so far as possible, attempt to state any hypotheses they have about potentially important differences in subgroup response in the protocol for the review and the number of planned subgroup analyses should be kept to a minimum. The greater the number of hypotheses tested, the greater the number of differences one will find by chance alone. Decisions about which analyses to do and which to report are much more likely to be data driven with *post hoc* analyses, and thereby more likely to be spurious. On the other hand, when a hypothesis has been clearly and unequivocally suggested by a different data set, the subgroup analysis rests on stronger ground.

- Is the difference suggested by comparisons within rather than between studies?
- Is the difference consistent across studies?

Reviewers should be extremely cautious about examining and drawing conclusions about differences in subgroup response based on differences in results between studies. There may be many other factors, aside from one that is the basis of the inference being made, which could explain the difference in results between studies. For instance, differences in the specific drugs, doses or treatment schedules used; the patients (whose risk of adverse outcomes may vary); the amount of cointervention; or the means of assessing the outcome, could all explain differences between the results of individual studies. Stated simply, inferences based on between-study comparisons are based on comparisons between non-comparable groups. Even if the review only includes randomised trials, it must be remembered that patients were not randomised to one study or another. Differences in subgroup response observed within a single study provide a stronger basis for drawing conclusions, and if a difference is observed across a number of high quality studies it strengthens a conclusion that the difference is real.

- Is the magnitude of the difference practically important?

If not, it makes no difference whether recommendations are based on the subgroup analysis or the overall analysis.

- Is the difference statistically significant?

RevMan 4.0 does not provide an appropriate method for testing this. Statistical techniques for conducting subgroup analysis include the Breslow-Day technique and regression approaches. With the Breslow-Day technique and similar approaches, it is possible to use a test for heterogeneity to estimate the probability that an observed interaction might have arisen by chance. However, as noted above, subgroup analyses based on differences between studies are discouraged.

Commonly, a number of comparisons for different subgroups are done without any formal testing for interactions. This practice, when combined with the reporting of just those subgroup analyses within which sizeable treatment differences are found, can lead to an overestimate of both the significance and the size of the difference. It should be avoided.

Regression models, such as logistic regression (Berlin 1994) can be used for analysis of interactions if the interactions are modelled by product terms. This allows the significance of an interaction to be tested while controlling for other factors. However, if there are many subgroup factors, the number of product terms necessary for an adequate modelling of the interactions may be higher than the number of observations, and an analysis of the interactions is impossible. An additional problem is deciding which of many possible interaction terms to enter into the model, and the potential for bias in their selection.

Reviewers without a statistical background who wish to formally test the significance of an apparently important difference in a subgroup analysis should consider asking a statistician for help. It is not important for reviewers to understand the details of the statistical approaches that can be used, but it is important to understand the concepts of statistical significance and power in subgroup analysis. Statistical analysis is a useful tool for assessing whether an observed difference in subgroup response might have occurred by chance alone, but it is not a substitute for good judgement, and the statistical methods must be appropriate.

8.8 Statistical methods not available in Review Manager

For some reviews it may be useful to undertake an analysis that is not supported by RevMan. When such an analysis is done it should be described and referenced in the Methods section of the review and the results should be reported in the Results section. Reviewers who wish to use tables for this can do so with the Additional tables facility in RevMan 4.0. There is currently no mechanism to publish additional graphs in *CDSR* but these could be made available in some other way, such as on an Internet site .

8.8.1 Meta-regression

Meta-regression can be used to assess statistically whether specific factors (covariates) influence the magnitude of effect across studies (Berlin 1994). Such analyses should best be used as hypothesis generating tools. The number of factors that can be included in an analysis is limited and there is a potential for bias when selecting which of many possible factors to include.

The covariates of interest may or may not be common for all patients in each study. For study characteristics that are not common for all patients (e.g. the mean age of the patients in each study), it is important to be cautious in their interpretation because of the risk of an ecological fallacy (Morgenstern 1982, Greenland 1994). The covariates that are investigated can also be either general (e.g. the year of the study) or problem-specific (e.g. the particular disease subtype). A variety of statistical methods including weighted least squares, logistic and hierarchical models can be used for meta-regression (Morris 1992, Berlin 1994, McIntosh 1995).

8.8.2 Bayesian meta-analysis

Bayesian approaches to meta-analysis, such as the Confidence Profile Method (Eddy 1990, Eddy 1992), are based on the principle that each observation or set of observations should be viewed in conjunction with a prior probability describing existing knowledge about the phenomenon of interest. The new observations alter this probability to generate a posterior probability. In Bayesian terms, traditional statistical methods used in meta-analysis assume that the prior probability distribution is uniform, with all outcomes being equally probable. Bayesian approaches also allow the incorporation of indirect evidence and opinions in the generation of the prior distributions. These approaches are controversial because they depend on opinions, and these will often vary considerably. To implement credible Bayesian analyses, additional development in eliciting prior probability distributions and conducting robust analyses is needed (Louis 1993).

8.8.3 'Exact' methods

Analysis of 2 X 2 tables with zeros (no events) in some cells presents a problem. One way of handling this is to add a fraction, such as ½ to each cell in the table. This is done in RevMan 4.0. A better approach might be to use exact calculations. Computational methods are available for computing an exact confidence interval for the common odds ratio (Mosteller 1992) and these are available in commercial software packages (e.g. Egret (Egret 1996)).

8.9 Other types of data

8.9.1 Time to event (survival) analysis

For reviews where survival time or time to an event is particularly important, individual patient data is highly desirable (see section 11). When this is not practical, it may be possible to extract and analyse data from published survival curves or to calculate a weighted average of median survival duration across studies (Parmar 1998) (see, for example Grilli 1993). Alternatively, analyses for different follow-up periods can be done using the different time-points as part of the outcome, thereby approximating a survival analysis (see, for example Pharoah 1999).

The analysis should be based on survival data for individual participants whenever possible. Not only is this approach more powerful (Buyse 1987), it also provides insight into the course of the disease and treatment effect over time (Clarke 1998). When these

survival data are not obtainable, it is still desirable to examine the distributions of survival times in those studies for which individual data are available.

8.9.2 Ordinal data

UNDER CONSTRUCTION – A new section is being prepared on methods for combining studies with ordinal data (ordered categories with no natural (numerical) distance between possible categories). One solution that is simple but not totally satisfactory is to treat ordinal scales as though they were continuous. A more complicated, but more robust solution is to use proportional odds models (Whitehead 1994).

8.10 Sensitivity analyses

Because there are different approaches to conducting a systematic review, reviewers should ask: How sensitive are the results of the analysis to changes in the way it was done? This provides reviewers with an approach to testing how robust the results of the review are relative to key decisions and assumptions that were made in the process of conducting the review. Each reviewer must identify how the key decisions and assumptions might conceivably have affected the results for a particular review. Generally, the types of decisions and assumptions that might be examined in sensitivity analyses include:

- changing the inclusion criteria for the types of study (e.g. using different methodological cut-points), participants, interventions or outcome measures
- including or excluding studies where there is some ambiguity as to whether they meet the inclusion criteria
- reanalysing the data using a reasonable range of results for studies where there may be some uncertainty about the results (e.g. because of inconsistencies in how the results are reported that cannot be resolved by contacting the investigators, or because of differences in how outcomes are defined or measured)
- reanalysing the data imputing a reasonable range of values for missing data
- reanalysing the data using different statistical approaches (e.g. using a random effects model instead of a fixed effect model, or *vice versa*)

The same cautions discussed for subgroup analyses apply to sensitivity analyses. In particular, since many sensitivity analyses involve between study subgroup comparisons, these findings need to be interpreted very carefully.

If the sensitivity analyses that are done do not materially change the results, it strengthens the confidence that can be placed in these results. If the results do change in a way that might lead to different conclusions, this indicates a need for greater caution in interpreting the results and drawing conclusions. Such differences might also enable reviewers to clarify the source of existing controversies about the effectiveness of an intervention, or lead them to hypothesise potentially important factors that might be related to the effectiveness of the intervention and warrant further investigation.

8.11 Special topics

8.11.1 Publication bias and funnel plots

As discussed in section 5, a particularly important component of a review is the identification of relevant studies. Publication bias has long been recognised as a problem in this regard since it means that the likelihood of finding studies is related to the results of those studies (Begg 1988, Begg 1989, Easterbrook 1991, Dickersin 1992b). One way to investigate whether a review is subject to publication bias is to prepare a 'funnel plot' and examine this for signs of asymmetry. RevMan 4.0 includes a facility to produce such a graph. However, if there is asymmetry, reasons other than publication bias should also be considered.

Funnel plots were first used in educational research and psychology (Light 1984a). They are simple scatter plots of the treatment effects estimated from individual studies (on the x axis) against some measure of each study's sample size (y axis). The name 'funnel plot' arises from the fact that precision in the estimation of the true treatment effect increases as the sample size of the component studies increases. Effect estimates from small studies will therefore scatter more widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias the plot should resemble a symmetrical inverted funnel (see panel A of the figure).

Relative measures of treatment effect (such as relative risks and odds ratios) are plotted on a logarithmic scale. This ensures that effects of the same magnitude but opposite directions (for example relative risks of 0.5 and 2) are equidistant from 1.0 (Galbraith 1988). Treatment effects have generally been plotted against sample sizes. However, the statistical power of a trial is determined both by its total sample size and the number of participants experiencing the event of interest. For example, a study with 100,000 patients and 10 events is less likely to show a statistically significant treatment effect than a study with 1000 patients and 100 events. The standard error (SE) or the variance of the effect estimate, rather than total sample size, have therefore been increasingly used for the y axis in funnel plots. RevMan 4.0 uses 1/SE, plotted against the effect size calculated by the statistical method chosen by the reviewer for the particular outcome.

If there is bias, for example because smaller studies without statistically significant effects (shown as open circles in the figure) remain unpublished, this will lead to an asymmetrical appearance of the funnel plot with a gap in a bottom corner of the graph (panel B). In this situation the effect calculated in a meta-analysis will overestimate the treatment effect (Villar 1997, Egger 1997b). The more pronounced the asymmetry, the more likely it is that the amount of bias will be substantial.

Publication bias has long been associated with funnel plot asymmetry (Light 1984a). However the funnel plot should be seen as a generic means of examining whether the smaller studies in a meta-analysis tend to show larger treatment effects and this may be due to reasons other than publication bias (Egger 1997a, Egger 1998b). Some of these are shown in the table:

Possible sources of asymmetry in funnel plots

1. Selection biases
Publication bias
Location biases
Language bias
Citation bias
Multiple publication bias
2. Poor methodological quality of smaller studies
Poor methodological design
Inadequate analysis
Fraud
3. True heterogeneity
Size of effect differs according to study size (for example, due to differences in the intensity of interventions or differences in underlying risk between studies of different sizes)
4. Artefactual
5. Chance

Even if a study has been published, the probability of finding it is also influenced by its results. For example, language bias (the preferential publication of studies without significant findings in languages other than English), makes it less likely that such 'negative' studies will be found (Grégoire 1995, Egger 1997c). Citation bias leads to 'negative' studies being referred to less often and they are therefore more likely to be missed when searching for relevant trials (Gotzsche 1987, Gotzcshe 1989, Ravnskov 1992). Conversely, results of 'positive' trials are sometimes reported more than once, increasing the probability that they will be located (multiple publication bias) (Gotzsche 1989, Huston 1996, Tramèr 1997).

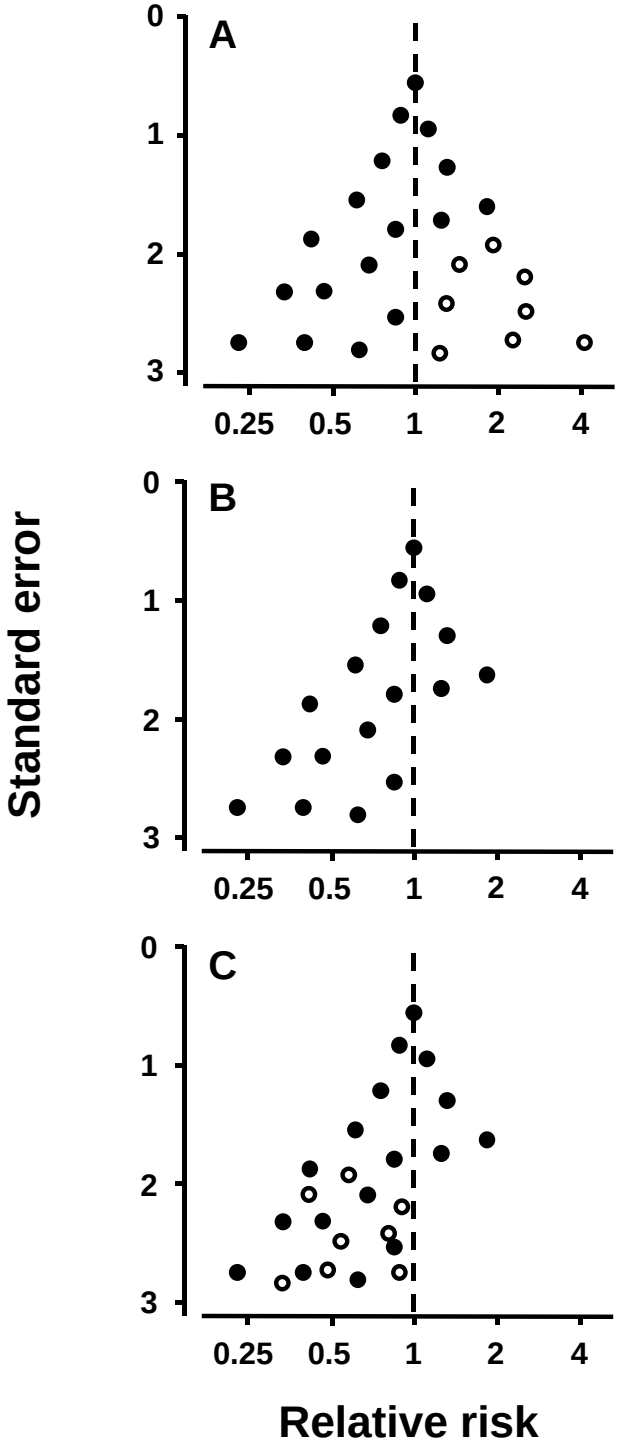
Another source of asymmetry arises from differences in methodological quality. Smaller studies are, on average, conducted and analysed with less methodological rigour than larger studies. Trials of lower quality also tend to show larger treatment effects (Schulz 1995, Moher 1998a). Trials which, if conducted and analysed properly, would have been 'negative' may thus become 'positive' (panel C).

True heterogeneity in treatment effects may also lead to funnel plot asymmetry. For example, substantial benefit may be seen only in patients at high risk for the outcome which is affected by the intervention and these high risk patients are usually more likely to be included in early, small studies (Davey Smith 1994, Glasziou 1995). In addition, small trials are generally conducted before larger trials are established and in the intervening years standard treatment may have improved. Furthermore, some interventions may have been implemented less thoroughly in larger trials and, therefore, have resulted in smaller estimates of the treatment effect (Stuck 1998). It has also been argued that funnel plot asymmetry may be artefactual (Irwig 1998), but simulation studies have shown that this will occur infrequently, if the overall treatment effect is very large and the outcome of interest is rare (Sterne, unpublished). Finally, it is, of course, possible that an asymmetrical funnel plot arises merely by the play of chance.

Symmetry or asymmetry is generally defined informally, through visual examination, but the visual interpretation of funnel plots may vary between observers (Villar 1997). More formal statistical methods to examine associations between the study effects and size have been proposed (Begg 1994, Egger 1997b). At present there is debate regarding the statistical properties, potentials and limitations of these tests (Naylor 1997, Irwig 1998, Seagrott 1998, Egger 1998b). No such tests are available in RevMan 4.0. Methodological work examining these issues is currently underway, but it is clear that both visual examination and statistical analysis of funnel plots have limited power to detect bias if the number of studies is small.

Reviewers should look at the relevant funnel plot whenever they do a meta-analysis. If asymmetry is present, likely reasons should be explored. The power of this method is, however, at its most limited in those situations when bias is most likely to distort the results of the meta-analyses: when it comprises only a few small studies. Finally, it should be remembered that although funnel plots may alert reviewers to a problem which needs considering, they do not provide a solution to this problem. The only satisfactory way to address reporting bias and the inadequate quality of individual trials is through prospective registration of trials (Simes 1986, Dickersin 1988, Anonymous 1991, Dickersin 1992a) and improvements in the quality of the conduct, analysis and reporting of studies, meta-analyses and systematic reviews (Begg 1996, Moher 1995b).

Legend to figure: Hypothetical funnel plots. Panel A: symmetrical plot in the absence of bias; Panel B: asymmetrical plot in the presence of reporting bias, Panel C: asymmetrical plot in the presence of bias due to low methodological quality of smaller studies.



8.11.2 Unit of analysis errors (cluster randomised trials)

In some studies people are allocated to the different interventions in clusters (e.g. by practice, hospital or community) to avoid contamination or for convenience. Often when this is done the unit of allocation is different from the unit of analysis with people allocated by groups but analysed as though they had been allocated individually. This is sometimes called a 'unit of analysis error' (Whiting-O'Keefe 1984). Effectively, using individuals as the unit of analysis when groups of people are allocated increases the power of the studies by increasing the degrees of freedom. This can result in false positive conclusions that the intervention had an effect when in truth it did not. In the context of a review, it can lead to studies having narrower confidence intervals and receiving more weight than is appropriate. This situation can also arise if people are allocated to interventions that are then applied to parts of them (e.g. eyes, teeth or bunions) but the separate parts are analysed rather than the people.

Reviewers should collect information about the unit of randomisation and the unit of analysis when this is relevant. To avoid a unit of analysis error, analyses should be conducted at the same level as allocation but this might considerably, and unnecessarily, reduce the power of the study. If possible, an estimate of the relative variability within and between clusters (the 'intra-class correlation coefficient') should be obtained and it may be possible to use this to adjust the results (Cornfield 1978, Donner 1981).

8.11.3 Multiple comparisons and the play of chance

For reviews and studies with several subgroup analyses and multiple outcomes, if enough analyses are done some are likely to produce 'statistically significant' results by chance alone. For example, using the conventional test of statistical significance of $P = 0.05$, one in 20 comparisons is likely to be 'statistically significant' even when there is truly no difference between the interventions being compared. The more analyses that are done, the more likely it is that some will be found to be statistically significant by chance alone.

Because of this problem, the strength of evidence from a review or a study depends on how focused its questions were at the outset. Unfortunately, when the results of a study are presented, it is not always possible to know how many comparisons were made. It is likely that interesting findings will have been selected for presentation or publication from a larger number of uninteresting ones, leading to spurious conclusions.

Reviewers should, if possible, distinguish between data that were not collected in a study and data that were collected but not reported. This is particularly important for studies reporting multiple outcomes, because of the potential for biased reporting (Gotzsche 1989, Gotzsche 1990).

There is no simple or totally satisfactory solution to the problem of multiple comparisons, but the following advice can be offered:

- Keep subgroup analyses to a minimum.
- State *a priori* in the protocol for the review, which analyses and outcomes are of particular interest (the fewer the better).
- Although it is recommended that all outcomes from included studies that are likely to be important to people should be reported in Cochrane Reviews, the strength of

evidence is less if there are multiple comparisons, particularly for outcomes that were not pre-specified.
- Interpret any findings that were not hypothesised in advance cautiously, even when they are 'statistically significant'.

8.11.4 Missing data

Three types of data may be missing from a review: unidentified studies, information for estimating the magnitude of effects, and information about characteristics that may be related to the magnitude of effects. The first type of missing data is discussed in section 8.11.1 in relation to publication bias. Methods for handling missing information in the studies that are identified include simply excluding studies for which data are missing, and imputations or complex strategies for assuming values for the missing data (Follmann 1992, Pigott 1993). No simple solutions exist for the problem of missing data. In general:

- Whenever possible, the investigators should be contacted to ensure that no data are missing for their study.
- The assumptions of whatever method is used to cope with missing data should be made explicit.
- Analyses should be done to test how sensitive the results are to reasonable changes in the assumptions that are made (see section 8.10).

8.11.5 Cross-over trials

The analysis of cross-over trials is problematic, particularly if both cross-over and parallel group trials are included together. One approach is to analyse the cross-over trials in the same fashion as parallel group trials. That is, results from the treatment period are used as if they came from one group of patients and results from the control period are used as if they came from an independent group of patients. However, this approach ignores the fact that the same patients will be in both arms of the study and they are clearly not independent of each other. An alternative, more appropriate approach is to use techniques specifically developed for paired designs and for combining paired designs with unpaired designs (Duffy 1989). If both approaches yield similar results, the results of the first approach (which can be done using RevMan 4.0) are on much stronger ground. For an example of this approach see (Jaeschke 1990).

8.11.6 Cumulative meta-analysis

In cumulative meta-analysis studies are added one at a time in a specified order (e.g. according to their date of publication or methodological quality) (Lau 1992). This approach can be used as a sensitivity analysis to determine whether results have been robust over time or across studies of differing quality. Cumulative meta-analysis can be done in RevMan 4.0 by de-selecting all studies from the analyses and then selecting them in the specified order, running an analysis as necessary. However, it is not possible to reproduce these analyses easily when the review is published in *CDSR*. If a cumulative meta-analysis is to be done, this should be stated in the protocol for the review.

8.11.7 Dose response analysis

The same principles described for subgroup analyses can be applied to dose response and duration response analyses. Meta-regression methods can be used to estimate dose-response parameters (Greenland 1992, Berlin 1993). Conclusions about differences in effect due to differences in dose are on strongest ground if participants are randomised to one dose or another within a study and a consistent relationship is found across similar studies. While reviewers should consider dose effects, particularly as a possible explanation for heterogeneity, they should be cautious about drawing conclusions about dose or duration effects based on between-study differences.

8.12 Where to go for help

The first point of contact for reviewers in need of statistical support should be the coordinator of their CRG. Each CRG should establish policies regarding statistical analyses and, ideally, there should be a statistician working with each CRG. It may, however, be more convenient for some reviewers to consult with a local statistician. If a reviewer needs statistical help, or other help with their analyses, which cannot be met directly through the editorial team of their CRG or local support, a Cochrane Centre can be asked for help. As a rule, requests for help should not be taken directly to the Statistical Methods Group (SMG). The SMG is responsible for providing policy advice and training, but does not have the resources necessary to support individual reviewers directly. Statisticians working with CRGs are members of the SMG and will consult with others in that Methods Group as needed. Similarly, Cochrane Centres will forward enquiries to the SMG when this is appropriate. However, to make the best use of what is a relatively scarce resource, it is important that the SMG is not flooded with what, for the statisticians at least, are routine enquiries.

Additional help can be found in the RevMan user guide, references cited in this section of the Handbook, other references in the *Cochrane Methodology Register* and in the frequently asked question (FAQ) list for RevMan. If you discover that there are important gaps in the help that can be found in the Handbook or the RevMan user guide, or specific problems that are not addressed, please let us (Mike Clarke or Andy Oxman) know. If you have solutions to problems that others might encounter, please share them with us. The Handbook will continue to evolve based on the needs expressed by those who use it, their experience and new methodological developments and this is especially true of this section on Analysing and Presenting Results.

8.12 References

Altman 1992. Altman DG, Gardner MJ. Confidence intervals for research findings. Br J Obstet Gynaecol 1992; 99:90-1.

Altman 1998. Altman DG. Confidence intervals for the number needed to treat. BMJ 1998; 317:1309-12

Anonymous 1991. Anonymous. Making clinical trialists register. Lancet 1991; 338:244-5.

Antman 1992. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 1992; 268:240-8.

Begg 1988. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. J Roy Stat Soc A 1988; 151:419-63.

Begg 1989. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. J Natl Cancer Inst 1989; 81:107-15.

Begg 1994. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50:1088-99.

Begg 1996. Begg CB, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637-639.

Berlin 1989. Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. J Am Stat Assoc 1989; 84:381-92.

Berlin 1993. Berlin JA, Longnecker MP, Greenland S. Meta analysis of epidemiologic dose-response data. Epidemiology 1993; 4:218-28.

Berlin 1994. Berlin JA, Antman EM. Advantages and limitations of meta-analytic regressions of clinical trials data Advantages and limitations of metaanalytic regressions of clinical trials data. Online J Curr Clin Trials [serial online] 1994; Doc No 134.

Buyse 1987. Buyse ME, Ryan LM. Issues of efficiency in combining proportions of deaths from several clinical trials. Stat Med 1987; 6:565-76.

Clarke 1998. Clarke M, Godwin J. Systematic reviews using individual patient data: a map for the minefields? Ann Oncol 1998;9:827-33.

Cooper 1980. Cooper HM, Rosenthal R. Statistical versus traditional procedures for summarizing research findings. Psychol Bull 1980; 87:442-9

Cornfield 1978. Cornfield J. Randomization by group: a formal analysis. Am J Epidemiol 1978; 108:100-2.

Davey Smith 1994. Davey Smith G, Egger M. Who benefits from medical interventions? Treating low risk patients can be a high risk strategy. Br Med J 1994;308:72-4.

Detsky 1992. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta analysis. J Clin Epidemiol 1992; 45:255-65.

Dickersin 1988. Dickersin K. Report from the panel on the Case for Registers of Clinical Trials at the Eighth Annual Meeting of the Society for Clinical Trials. Controlled Clin Trials 1988; 9:76-81.

Dickersin 1992a. Dickersin K. Keeping posted. Why register clinical trials? - revisited. Controlled Clin Trials 1992; 13:170-7.

Dickersin 1992b. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. JAMA 1992; 263:374-8.

Donner 1981. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. Am J Epidemiol 1981; 114:906-14.

Duffy 1989. Duffy SW, Rohan TE, Altman DG. A method for combining matched and unmatched binary data. Application to randomized, controlled trials of photocoagulation in the treatment of diabetic retinopathy. Am J Epidemiol 1989; 130:371-8.

Easterbrook 1991. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991; 337:867-72.

Eddy 1990. Eddy DM, Hasselblad V, Schachter R. An introduction to a Bayesian method for meta-analysis: the confidence profile method. Med Decis Making 1990; 10:15-23.

Eddy 1992. Eddy DM, Hasselblad V, Shachter R. Meta-analysis by the confidence profile method. The statistical synthesis of evidence. Boston: Academic Press, 1992.

Egger 1997a. Egger M, Davey Smith G, Phillips AN. Meta-analysis: principles and procedures. Br Med J 1997;315:1533-7.

Egger 1997b. Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. Br Med J 1997;315:629-634.

Egger 1997c. Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997;350:326-329.

Egger 1998. Egger M, Davey Smith G. Meta-analysis: Bias in location and selection of studies. Br Med J 1998;316:61-6.

Egger 1998b. Egger M, Davey Smith G, Minder C. Authors' reply. Br Med J 1998;316:471.

Egret 1996. Egret statistical software. Seattle: Statistics in Epidemiology Research Corporation, 1996.

Fleiss 1993. Fleiss JL. Measures of effect size for categorical data. In: Cooper H, Hedges LV, editors. The Handbook Of Research Synthesis. New York: Russell Sage Foundation, 1993; 245-81.

Follmann 1992. Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. J Clin Epidemiol 1992; 45:769-73.

Galbraith 1988. Galbraith R. A note on graphical presentation of estimated odds ratios from several clinical trials. Stat Med 1988;7:889-94.

Glasziou 1995. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. BMJ 1995;311:1356-9.

Gotzsche 1987. Gotzsche PC. Reference bias in reports of drug trials. Br Med J 1987;295:654-6.

Gotzsche 1989. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis [published erratum appears in Controlled Clin. Trials 1989;50:356]. Controlled Clin Trials 1989; 10:31-56.

Gotzsche 1990. Gotzsche PC. Sensitivity of effect variables in rheumatoid arthritis: a meta-analysis of 130 placebo controlled NSAID trials [published erratum appears in J Clin Epidemiol 1991;44:613]. J Clin Epidemiol 1990; 43:1313-8.

Greenland 1992. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. Am J Epidemiol 1992; 135:1301-9.

Greenland 1994. Greenland S, Robins J. Invited commentary: ecologic studies - biases, misconceptions, and counterexamples. Am J Epidemiol 1994; 139:747-50.

Grégoire 1995. Grégoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? J Clin Epidemiol 1995;48:159-63.

Grilli 1993. Grilli R, Oxman AD, Julian JA. Chemotherapy for advanced non-small-cell lung cancer: How much benefit is enough? J Clin Oncology 1993; 11:1866-72.

Hedges 1992. Hedges LV. Modelling publication selection effects in meta-analysis. Stat Sci 1992; 7:246-55.

Huston 1996. Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. Lancet 1996;347:1024-6.

Irwig 1998. Irwig L, Macaskill P, Berry G. Graphical test is itself biased [letter]. Br Med J 1998;316:470.

Iyengar 1988. Iyengar S, Greenhouse JB. Selection models and the file drawer problem. Stat Sci 1988; 3:109-35.

Jaeschke 1990. Jaeschke R, Oxman AD, Guyatt GH. To what extent do patients with chronic heart failure and sinus rhythm benefit from digoxin therapy? A systematic overview (meta-analysis). Am J Med 1990; 88:279-86.

Laird 1990. Laird NM, Mosteller F. Some statistical methods for combining experimental results. Int J Tech Assess in Healthcare 1990; 6:5-30.

Lau 1992. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med 1992; 327:248-54.

Laupacis 1988. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988; 318:1728-33.

Light 1984. Light RJ, Pillemer DB. Summing up. The science of reviewing research. Cambridge, Massachusetts, and London, England: Harvard University Press, 1984.

Louis 1993. Louis T, Zelterman D. Bayesian approaches to research synthesis. In: Cooper H, Hedges LV, editors. The Handbook Of Research Synthesis. New York: Russell Sage Foundation, 1993; 411-22.

McIntosh 1995. McIntosh M. The population risk as an explanatory variable in research synthesis of clinical trials. Harvard Medical School 1995; 1-35.

Moher 1995. Moher D, Olkin I. Meta-analysis of randomised controlled trials. A concern for standards. JAMA 1995;274:1962-3.

Moher 1998. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.

Morgenstern 1982. Morgenstern H. Uses of ecological analysis in epidemiologic research. Am J Public Health 1982; 72:1336-44.

Morris 1992. Morris CN, Normand SL. Hierarchical models for combining information and for meta-analyses. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. Bayesian Statistics, 4th edition. New York: Oxford University Press, 1992.

Mosteller 1992. Mosteller F, Chalmers TC. Some progress and problems in meta-analysis of clinical trials. Stat Sci 1992; 7:227-36.

Naylor 1995. Naylor CD. The case for failed meta-analyses. J Evaluation Clin Practice 1995; 1:127-30.

Naylor 1997. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. Br Med J 1997;315 :617-9.

Oxman 1992. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992; 116:78-84.

Parmar 1998. Parmar MK, Torri V, Stewart LA. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17:2815-34.

Pharoah 1999. Pharoah FM, Mari JJ, Streiner D. Family intervention for schizophrenia (Cochrane Review). In: The Cochrane Library, Issue 2, 1999. Oxford: Update Software.

Pigott 1993. Pigott TD. Methods for handling missing data in research synthesis. In: Cooper H, Hedges LV, editors. The Handbook Of Research Synthesis. New York: Russell Sage Foundation, 1993; 163-75.

Ravnskov 1992. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. Br Med J 1992;305:15-9.

Schulz 1993. Schulz KF, Altman DG. Statistical methods for data synthesis. Cochrane Workshop Report. Oxford: UK Cochrane Centre, 1993.

Schulz 1995. Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.

Seagrott 1998. Seagrott V, Stratton I. Test had 10% false positive rate [letter]. Br Med J 1998;316:470.

Simes 1986. Simes RJ. Publication bias: the case for an international registry of clinical trials. J Clin Oncol 1986; 4:1529-41.

Sinclair 1994. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. J Clin Epidemiol 1994; 47:881-9.

Smeeth 1999. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses – sometimes informative, usually misleading. Brit Med J 1999;318:1548-51.

Sterne (submitted). Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: Simulation and prevalence studies (submitted).

Stuck 1988. Stuck AE, Rubenstein LZ, Wieland D. Bias in meta-analysis detected by a simple, graphical test. Asymmetry detected in funnel plot was probably due to true heterogeneity [letter]. Br Med J 1998;316:469-71.

Thompson 1991. Thompson SG, Pocock SJ. Can meta-analyses be trusted? Lancet 1991; 338:1127-30.

Tramèr 1997. Tramèr MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. Br Med J 1997;315:635-40.

Villar 1997. Villar J, Piaggio G, Carroli G, Donner A. Factors affecting the comparability of meta-analyses and largest trials results in perinatology. J Clin Epidemiol 1997;50:997-1002.

Walker 1988. Walker AM, Martin-Moreno JM, Artalejo FR. Odd man out: a graphical approach to meta-analysis. Am J Public Health 1988; 78:961-6.

Whitehead 1994. Whitehead A, Jones NMB. A meta-analysis of clinical trials involving different classifications of response into ordered categories. Stat Med 1994; 13:2503-15.

Whitehead 1999. Whitehead A, Bailey AJ, Elbourne D. Combining summaries of binary outcomes with those of continuous outcomes in a meta-analysis. J Biopharm Stat 1999;9:1-16.

Whiting-O'Keefe 1984. Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in medical care experiments. Med Care 1984; 22:1101-14.

Yusuf 1991. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; 266:93-8.

9. INTERPRETING RESULTS

Although it can be argued that the results of a systematic review should stand on their own, many people faced with a decision look to the Discussion and Reviewers' Conclusions for help interpreting the results. Indeed, many people prefer to go directly to the conclusions before looking at the rest of the review.

Discussion and conclusions about the following issues can help people to make decisions:

- The strength of the evidence
- The applicability of the results
- Other information, such as considerations of costs and current practice, that might be relevant to someone making a decision
- Clarification of any important trade-offs between the expected benefits, harms and costs of the intervention

Because Cochrane Reviews have an international audience, the discussion and reviewers' conclusions should, so far as possible, assume a broad international perspective, rather than addressing specific national or local circumstances. Reviewers should be particularly careful to bear in mind that different people might make different decisions based on the same evidence. The primary purpose of the review should be to present information, rather than to offer advice. The discussion and conclusions should be to help people to understand the implications of the evidence in relationship to practical decisions. Recommendations that depend on assumptions about resources and values should be avoided.

9.1 Strength of evidence

A good starting point for the discussion section of a review is to address any *important* methodological limitations of the included trials and the methods used in the review that might affect practical decisions about healthcare or future research. This should not be a detailed discussion of study or review methods. Information provided in the section of the review on methodological quality need not be repeated here.

It is often helpful to discuss how the included studies fit into the context of other evidence that is not included in the review. For example, for reviews of drug therapy it may be relevant to refer to dosage studies or non-randomised studies of the risk of rare adverse events. It should be stated clearly whether the other evidence that is referenced was systematically reviewed when other types of evidence are cited.

One type of evidence that can be helpful in considering the likelihood of a cause-effect relationship between an intervention and an important outcome is indirect evidence of a relationship. This includes evidence relating to intermediate outcomes (such as physiological or biochemical measures that are markers for risk of the outcome of interest), evidence from studies of different populations (including animal studies) and evidence from analogous relations (i.e. similar interventions).

Because conclusions regarding the strength of inferences about the effectiveness of an intervention are essentially causal inferences, reviewers might want to consider guidelines

for assessing the strength of a causal inference, such as those put forward by Hill (Hill 1971). In the context of a systematic review of clinical trials, these considerations might include:

- How good is the quality of the included trials?
- How large and significant are the observed effects?
- How consistent are the effects across trials?
- Is there a clear dose-response relationship?
- Is there indirect evidence that supports the inference?
- Have other plausible competing explanations of the observed effects (eg. bias or co-intervention) been ruled out?

More or less explicit approaches to grading the strength of evidence underlying a conclusion are available (CTFPHE 1979, Cook 1992, Gyorkos 1994 Guyatt 1995, US PSTF 1996), although there is no single approach that is universally accepted as being appropriate for the wide range of reviews included in the *Cochrane Database of Systematic Reviews*. A Collaborative Review Group (CRG) may elect to use a standard approach to grading the strength of evidence across its reviews. Over time, it may be possible for the Cochrane Collaboration as a whole to develop a more consistent and explicit approach to drawing conclusions about the overall strength of evidence for the main conclusions of a review. However, it is currently up to individual reviewers, in consultation with others in their CRG, to select an approach to summarising the strength of evidence that is appropriate for the question being reviewed.

9.2 Applicability

'A leap of faith is always required when applying any study findings to the population at large' or to a specific person. *'In making that jump, one must always strike a balance between making justifiable broad generalizations and being too conservative in one's conclusions.'* (Friedman 1985)

Users of Cochrane Reviews must decide, either implicitly or explicitly, how applicable the evidence is to their particular circumstances. To do this, they must first decide whether the review provides valid information about potential benefits and harms that are important to them. To the extent that this is the case, they then need to decide whether the participants and settings in the included studies are reasonably similar to their own situation. In addition, it will often be important for them to consider the characteristics of the interventions or additional care provided in the included studies in reaching conclusions about the applicability of the evidence.

Decisions about applicability depend on knowledge of the particular circumstances in which decisions about healthcare are being made. In addressing the applicability of the results of a review, reviewers should be cautious not to assume that their own circumstances, or the circumstances reflected in the included studies are necessarily the same as those of others. Reviewers can, however, help people to make decisions about applicability by drawing attention to the spectrum of circumstances to which the evidence is likely to be applicable, circumstances where the evidence is not likely to be applicable, and predictable variation in effects across different circumstances.

Generally, rather than rigidly applying the inclusion and exclusion criteria of studies to specific circumstances, it is better to ask whether there are compelling reasons why the evidence should not be applied under certain circumstances (Guyatt 1994, Dans 1996). Reviewers can sometimes help people making specific decisions by identifying important variation where divergence might limit the applicability of results, including:

- biologic and cultural variation
- variation in compliance
- variation in baseline risk

In addressing these issues, reviewers cannot be expected to be aware of, or address the myriad differences in circumstances around the world. They can, however, address differences of known importance to many people and, importantly, they should avoid assuming that other people's circumstances are the same as their own in discussing the results and drawing conclusions.

9.2.1 Biologic and cultural variation

Issues of biologic variation that might be considered include divergence in pathophysiology (e.g. biologic differences between women and men that are likely to affect responsiveness to a treatment) and divergence in a causative agent (e.g. for infectious diseases such as malaria). For some healthcare problems, such as psychiatric problems, cultural differences can sometimes limit the applicability of results.

9.2.2 Variation in compliance

Variation in the compliance of the recipients and providers of care can limit the applicability of results. Predictable differences in compliance can be due to divergence in economic conditions or attitudes that make some forms of care not accessible or not feasible in some settings, such as in developing countries (Dans 1996).

9.2.3 Variation in baseline risk

The net benefit of any intervention depends on the risk of adverse outcomes without intervention, as well as on the effectiveness of the intervention. Therefore, variation in baseline risk is almost always an important consideration in determining the applicability of results. However, it is important to distinguish between two issues. First, whether the relative benefits and harms are applicable. For example, there might be reasons to doubt whether results obtained in high-risk patients are applicable to low-risk patients, or whether they are applicable to patients with co-morbid conditions. If there is not a compelling reason to assume that the relative benefits and harms are applicable, it is possible to estimate the expected effect of an intervention (e.g. the number needed to treat) by applying the estimated relative effect of an intervention to a specific baseline risk. The second issue related to baseline risk that warrants consideration is the extent of variation that can be expected in the impact of the intervention. For example, it can be useful to consider the number needed to treat for the range of baseline risk observed in the control groups of the studies included in the review.

9.2.4 Variation in the results of the included studies

In addition to identifying limitations of the applicability of the results of their review, reviewers should discuss and draw conclusions about important variation in results within the circumstances to which the results are applicable. Is there predictable variation in the relative effects of the intervention, and are there identifiable factors that may cause the response or effects to vary? These might include:

- patient features, such as age, sex, biochemical markers
- intervention features, such as the timing or intensity of the intervention
- disease features, such as hormone receptor status

These features should be examined even if there is not statistically significant heterogeneity. This should be done by testing whether there is an interaction with treatment, and not by subgroup analysis. As discussed in section 8.7, differences between subgroups, particularly those that correspond to differences between studies, need to be interpreted cautiously. Some chance variation between subgroups is inevitable, so unless there is strong evidence of an interaction then it should be assumed there is none.

9.3 Other relevant information

It can be helpful for reviewers to discuss the results of a review in the context of other relevant information, such as epidemiological data about the magnitude and distribution of a problem, information about current clinical practice from administrative databases or practice surveys, and information about costs. However, this is often beyond the scope of Cochrane Reviews and can be done better on a national or regional basis; for example, by people developing clinical practice guidelines or undertaking a technology assessment. It must be kept in mind that evidence about the effects of healthcare is essential for well informed decisions, but it is not sufficient. Cochrane Reviews cannot and should not be expected to provide all of the information that is needed for people making decisions. On the other hand, reviewers can help people by clarifying other information, that might vary widely, which is likely to be important in making a decision.

9.4 Adverse effects

The discussion and conclusions of a review should note the strength of the evidence on adverse effects including the estimates of their seriousness and frequency in different circumstances. In particular, the causal relationship of an adverse effect to a particular intervention should be critically assessed, bearing in mind that under-ascertainment and under-reporting of adverse and unexpected effects are common. Reviewers may wish to comment on how adverse effects should be further investigated in their Implications for Research section.

9.5 Trade-offs

In addition to considering the strength of evidence underlying any conclusions that are drawn, reviewers should be as explicit as possible about any judgements about preferences (the values attached to different outcomes) that they make. Healthcare interventions generally entail costs and risks of harm, as well as expectations of benefit. Drawing conclusions about the practical usefulness of an intervention entails making

trade-offs, either implicitly or explicitly, between the estimated benefits and the estimated costs and harms (Eddy 1990b). It is beyond the scope of most Cochrane Reviews to incorporate formal economic analyses (although they might well be used for such analyses) (Mugford 1989, Mugford 1991) and this is discussed in appendix 9. However, reviewers should consider all of the potentially important outcomes of an intervention when drawing conclusions, including ones for which there may be no reliable data from the included trials. They should also be cautious about any assumptions they make about the relative value of the benefits, harms and costs of an intervention.

9.6 Implications

The above cautions about drawing conclusions not withstanding, CRGs (and users of Cochrane Reviews) may find it useful to categorise interventions into one of six mutually exclusive categories. This has been done by the Pregnancy and Childbirth Group (Enkin 1994), based on an earlier effort to classify interventions into four categories that drew a great deal of attention and praise. The first three categories of interventions, listed below, are ones for which there is sufficient evidence to reach relatively firm conclusions for practice. The last three are categories for which further research may be required before firm conclusions for practice can be drawn.

1. Forms of care for which there is sufficient evidence to provide clear guidelines for practice

    A) Forms of care that improve outcome

    B) Forms of care that should be abandoned in light of the available evidence

    C) Forms of care that involve important trade-offs between known benefits and known adverse effects

2. Forms of care for which the evidence is insufficient to provide clear guidelines for practice, but which should influence priorities for research

    A) Forms of care that appear promising, but require further evaluation

    B) Forms of care that have not been shown to have the effects expected from them, but which may require further attention

    C) Forms of care with reasonable evidence that they are not effective for the purpose for which they have been used

9.7 Common errors in reaching conclusions

A common mistake when there is inconclusive evidence is to confuse 'no evidence of an effect' with 'evidence of no effect'. When there is inconclusive evidence, it is wrong to claim that it shows that an intervention has ‘no effect’ or is ‘no different’ from the control intervention. It is safer to report the data, with a confidence interval, as being compatible with either a reduction or an increase in the outcome. When there is a ‘positive’ but statistically non-significant trend reviewers commonly describe this as ‘promising’,

whereas a 'negative' effect of the same magnitude is not commonly described as a 'warning sign'. Reviewers should be careful not to do this. Another mistake is to frame the conclusion in wishful terms. For example, reviewers might write 'the included studies were too small to detect a reduction in mortality' when the included studies showed a statistically non-significant increase in mortality. One way of avoiding errors such as these is to consider the results blinded; i.e. consider how the results would be presented and framed in the conclusions if you reversed the direction of the results. If the confidence interval for the estimate of the difference in the effects of the interventions overlaps the null value, the analysis is compatible with both a true beneficial effect and a true harmful effect. If one of the possibilities is mentioned in the conclusion, the other possibility should be mentioned as well.

Another common mistake is to reach conclusions that go beyond the evidence that is reviewed. Often this is done implicitly, without referring to the additional information or judgements that are used in reaching conclusions about the implications of a review for practice. Even when conclusions about the implications of a review for practice are supported by additional information and explicit judgements, the additional information that is considered is rarely systematically reviewed and implications for practice are often dependent on specific circumstances and values that must be taken into consideration (see section 9.5). Reviewers should always be cautious about reaching conclusions about implications for practice and they should avoid making recommendations.

In reaching conclusions about implications for research, platitudes like "more research is needed" should also be avoided. Reviewers should state exactly what research is needed and why. Opinions on how the review might be improved with additional data or resources can also be noted.

9.8 References

CTFPHE 1979. Canadian Task Force on the Periodic Health Examination. The periodic health examination. Can Med Assoc J 1979; 121:1193-254.

Cook 1992. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Antithrombotic Therapy Consensus Conference. Chest 1992; 102:305S-311S.

Dans 1996. Dans AL, Dans LF. Users' guides for the applicability of the results of clinical trials, perspective of clinicians from developing countries. Manila: UP College of Medicine, 1996 (unpublished manuscript).

Eddy 1990b. Eddy DM: Anatomy of a decision. JAMA 1990; 263:4413.

Enkin 1994. Enkin M, Keirse MJNC, Renfrew MJ, Neilson JP. A Guide to Effective Care in Pregnancy and Childbirth. 2nd edition. Oxford: Oxford University Press, 1994.

Friedman 1985. Friedman LM, Furberg CD, DeMets DL. Fundamentals of Clinical Trials. 2nd edition. Littleton, MA: John Wright PSG Inc, 1985.

Guyatt 1994. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, B: what were the results and will they help me in caring for my patients. JAMA 1994; 271:59-63.

Guyatt 1995. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, for the Evidence-Based Medicine Working Group. Users' guide to the medical literature IX: A method for grading health care recommendations. JAMA 1995; 274:1800-4.

Gyorkos 1994. Gyorkos TW, Tannenbaum TN, Abrahamowicz M, Oxman AD, Scott EA, Millson ME, et al. An approach to the development of practice guidelines for community health interventions. Can J Public Health 1994; 85(Suppl 1):S8-S13.

Hill 1971. Hill AB. Principles of Medical Statistics. 9th edition. London: Lancet, 1971: 312-20.

Mugford 1989. Mugford M, Kingston J, Chalmers I. Reducing the incidence of infection after caesarean section: implications of prophylaxis with antibiotics for hospital resources. BMJ 1989; 299:1003-6.

Mugford 1991. Mugford M, Piercy J, Chalmers I. Cost implications of different approaches to the prevention of respiratory distress syndrome. Arch Dis Child 1991; 66:757-64.

US PSTF 1996. US Preventive Services Task Force. Guide to Clinical Preventive Services. 2nd edition. Baltimore: Williams & Wilkins, 1996; xxxix-lv.

10. IMPROVING AND UPDATING REVIEWS

If Cochrane Reviews are to be useful to those who want to take more informed decisions in healthcare and research, then they must be trustworthy, and transparently so. As made clear throughout the Handbook, the Collaboration uses explicit methods to produce reviews and this feature alone will make them more useful to users than the vast majority of reviews that are currently available. Textbooks and review articles with 'Materials and Methods' sections remain rare.

Above a certain guaranteed minimum standard, the reviews contributed to *The Cochrane Database of Systematic Reviews* (*CDSR*) will vary in the level of methodological quality that it has been possible for the reviewers to achieve. The 'gold standard' will continue to be represented by systematic reviews, conducted by the responsible investigators, that are based on individual patient data for all patients entered into all of the trials meeting the entry criteria for the review (see section 11). Such reviews require not only substantial resources (including time), they also depend on the success of negotiations among the investigators. These factors should not be underestimated. Furthermore, because 'the best can be the enemy of the good', it will be important to do empirical research to learn more than is currently known about which methodological standards are essential, and which desirable, in attempts to avoid bias.

Mechanisms for maintaining and raising the standards of Cochrane Reviews include:

- Attracting dedicated participants and avoiding conflicts of interest
- Consumer involvement
- Ensuring access to studies
- Improving access to unpublished data
- Establishing and developing standards and guidelines
- Using rigorous review methods
- Software and informatics support
- Training
- Ongoing and open peer review
- Keeping reviews up-to-date

10.1 Attracting dedicated participants and avoiding conflicts of interest

The quality of the Cochrane Collaboration, viewed from any of a variety of perspectives, will reflect the characteristics of the individuals contributing to it. The community of people who have experience preparing systematic reviews of the effects of healthcare remains small, and among those who have this experience, only some will wish to commit themselves to taking on the substantial commitment that is expected of anyone who joins the Collaboration. In other words, the disincentives that confront those who are wondering whether they should become involved are a useful screening test in their own right and, in general, the people doing Cochrane Reviews have selected themselves.

To help ensure the integrity and perceived integrity of Cochrane Reviews the Collaboration has adopted a Code of Conduct for Avoiding Potential Financial Conflict of

Interest. Reviewers must avoid financial conflicts of interest and disclose conflicts of interest that cannot be avoided (see section 2.2). All reviewers must sign statements of responsibility and conflict of interest. The editorial team of each Collaborative Review Group (CRG) must also disclose any potential conflict of interest that they might have.

10.2 Consumer involvement

Healthcare consumers must be involved in developing Cochrane Reviews to help ensure that these:

- Are targeted at problems that are important to people
- Take account of outcomes that are important to those affected
- Are accessible to people making decisions
- Adequately reflect variability in the values and conditions of people, and the circumstances of healthcare in different countries

Relatively little is known about the effectiveness of various means of involving consumers in the review process or, more generally, in the spectrum of healthcare research. The Collaboration is dedicated to consumer involvement in principle. This is based on our values, good logic, and evidence that the views and perspectives of consumers often differ greatly from those of providers and researchers (Bastian 1998). Researchers and funders have generally failed to ensure that healthcare research adequately meets the needs of those ultimately affected. Because of conflicting values and interests, it is unlikely that this situation will improve substantially without appropriate mechanisms for involving consumers in decisions about research. However, to ensure the effectiveness of consumer involvement, creativity and a critical approach must be used to develop and evaluate the mechanisms that are used. This is being done in a variety of ways by CRGs, through the activities of the Cochrane Consumer Network and by other entities within the Collaboration. The Consumers and Communication CRG will be reviewing evidence on the effects of consumer participation in systematic reviews, as well as in research more generally. This practical experience and formal evaluations will provide a basis for guidelines on how to ensure that consumer involvement effectively contributes to ensuring the quality and accessibility of Cochrane Reviews.

Consumers are participating in the development of protocols and reviews in the following ways:

- Helping to determine topics and issues for reviews
- As co-reviewers
- As part of a consumer consultation during protocol and review development (including by questionnaire, direct dialogue or interview, in focus groups, and email discussion groups or teleconferences)
- As referees during the editorial process

Whenever consumers (or others) are consulted during the development of a protocol or review, their contribution should be acknowledged in the acknowledgement section of the protocol or review. Formal inclusion in the list of reviewers for citation may also be appropriate, as it is for other contributors.

Literature by consumers, or surveys and studies exploring consumers' views, can also be discussed within the review to ensure that issues of importance to consumers are addressed.

Many of these issues will also apply to other users of Cochrane reviews. The Collaboration is similarly committed to user involvement in principle, and encourages reviewers to seek and incorporate the views of users other than consumers in the development of protocols and reviews.

10.3 Ensuring access to studies

Because of the disarray of the medical literature, considerable efforts are required to locate the research that addresses the questions posed by a review (see section 5). The Collaboration is helping to ensure that relevant, valid studies are located by reviewers and included in their reviews by:

- Hand-searching the world's healthcare literature to identify trials
- Facilitating and supporting the development and maintenance of specialised registers by CRGs
- Providing training and support to those undertaking searching activities
- Developing the Cochrane Central Register of Controlled Trials (CENTRAL) to facilitate the transfer of trials between CRGs and other Cochrane entities, and to facilitate access to studies from other sources contributed to this register
- Working with the US National Library of Medicine to improve the coding of trials in MEDLINE and to develop an ancillary database of reports of trials not included in MEDLINE
- Developing and evaluating strategies to improve the coding and classification of trials

This work involves a large number of people engaged in a variety of activities through CRGs, Cochrane Centres, Fields and Methods Groups. The CENTRAL/CCTR Advisory Group, the New England Cochrane Center, Providence Office and the Information Retrieval Methods Group have key responsibilities for co-ordinating and guiding these activities.

10.4 Improving access to unpublished data

Improved access to unpublished data is needed to overcome problems with missing information in published reports and to protect against publication bias. In addition to the efforts undertaken by each CRG to help ensure access to relevant unpublished data within their scope, the Collaboration as a whole is working to develop strategic alliances with the pharmaceutical industry and others, and is actively promoting ethical standards that clarify the unacceptability of withholding unpublished data.

10.5 Establishing and developing standards and guidelines

The Handbook is the Cochrane Collaboration's most tangible manifestation of the development of standards and guidelines. As experience accumulates among people trying to apply these, and in the light of relevant research findings, the Handbook has been and will continue to be modified. In addition to building on the experience of CRGs,

multiple sources contribute to the development of the Handbook, including: empirical methodological studies, other methodological articles in *The Cochrane Review Methodology Database*, and advice from relevant Methods Groups.

Beyond what is found in the Handbook, each CRG must make decisions about standards and guidelines specific to the nature of the healthcare problems within their scope. Standard methods used by a CRG are published in the Group's module in *The Cochrane Library*.

10.6 Using rigorous review methods

It is neither feasible nor desirable to dictate the decisions that a reviewer should take. These will vary from review to review depending on the topic, the nature of the available evidence and the resources available to the reviewer. However, in general, the validity of Cochrane Reviews is ensured by:

- Searching as thoroughly as possible for studies meeting the inclusion criteria of a review, relying as much as possible on the Collaboration's efforts to ensure the thoroughness and efficiency with which randomised trials are identified
- Use of explicit criteria for selecting studies for inclusion in a review and for assessing the quality of these studies
- Application of these critieria by more than one reviewer where appropriate and feasible, to ensure the reproducibility of the judgements that are made
- Ongoing efforts to collect missing information that might contribute importantly to a review, to the extent possible depending on the availablity of resources and data
- Collection of individual patient data from investigators where appropriate and feasible, to the extent possible depending on the availability of resources and data
- Use of appropriate statistical techniques, where appropriate, to synthesize results
- Use of sensitivity analyses to test the robustness of the results of a review relative to any judgements or assumptions
- Cautious use of subgroup analyses and avoidance of over-interpretation of any subgroup analyses that are undertaken
- Carefully drawn conclusions, including implications for practice and research, based on cautious interpretation of results - taking into account the limitations of the review and variability in the values and conditions of people whose decisions might be influenced by the review
- Full reporting of the materials and methods used in undertaking the review

Just as it is possible to update Cochrane Reviews in the light of new evidence, it is possible to improve upon the methods. Moreover, because the methods are explicitly reported in Cochrane Reviews, users can judge for themselves how these might affect the validity of the results of a review.

10.7 Software and informatics support

The Cochrane Collaboration's Review Manager software (RevMan) is designed to assist reviewers in constructing reviews in the structured format described in section 2. This software has been and will continue to be developed to incorporate standards and guidelines for Cochrane Reviews, improved analytic methods, 'online' help and error

checking mechanisms, as these evolve. One way in which the software contributes to ensuring the validity of Cochrane Reviews is by facilitating registration and publication of protocols for planned reviews, as described in section 3.

The development of RevMan and the Collaboration's other software, Module Manager (ModMan) and the Parent Database (the suite of software programs used to prepare and compile *CDSR*) is directed by the Software Development Group with guidance from the relevant advisory groups. Specialised software to facilitate the management of their specialised register by CRGs (MeerKat) is currently being developed. Advice, support and training regarding the use of computers, the Internet and methods to meet the information needs of the Collaboration is provided by the Informatics Methods Group. The Canadian, Australasian and Nordic Cochrane Centres have each assumed specific responsibilities with regards to meeting these needs. Information about the activities of each of these entities is provided in their respective modules in *The Cochrane Library*.

10.8 Training

It is important to ensure that those contributing to the work of the Collaboration have the knowledge and skills that they need to do a good job. Training may be needed by reviewers, editors, Criticism edirots, CRG coordinators and Trial Search Coordinators, hand-searchers, trainers and users of Cochrane Reviews. We focus here on the training needs of reviewers and editors to help them to prepare and maintain high quality reviews.

10.8.1 Training for reviewers

While some reviewers who join a CRG have training and experience in conducting a systematic review, many do not. Cochrane Centres are responsible for developing training materials and organising training workshops for members of CRGs. Each CRG is responsible for ensuring that the members of the group have adequate training and methodological support. Training materials and opportunities for training will continue to be developed and will evolve to reflect the needs of the Collaboration and its standards and guidelines.

10.8.2 Training for editors

CRG editors need skills related to the area covered by their group, skill and experience in assessing studies of the effects of healthcare, an ability to edit scientific material for publication, and an ability to facilitate the inclusion of the results of comments and criticisms into reviews. A CRG editor may have to:

- Identify and define topics to be reviewed
- Identify potential reviewers who are capable of preparing a systematic review, and explain the implications of doing a Cochrane Review to them
- Support and help reviewers to overcome difficulties encountered in preparing their reviews
- Check and ensure that reviews make scientific sense and address relevant issues, using referees when appropriate
- Check and ensure that reviews are in the standard format used by the Cochrane Collaboration

- Ensure that existing reviews are updated by helping reviewers to do this.

Some of these tasks are common to many types of medical editorial work, others are specific to the work of the Cochrane Collaboration as a whole, and a few to particular areas or CRGs. The training of editors is therefore made up of general editorial training, training within the Cochrane Collaboration, and training on the job within CRGs.

General training is probably most efficiently provided in courses, workshops and seminars organised for science editors, for example by the European Association of Science Editors (EASE) or the North American Council of Biology Editors (CBE). The UK Cochrane Centre is a member of EASE; it may be useful for other Cochrane Centres to join EASE or CBE.

Training within the Cochrane Collaboration takes the form of workshops and seminars organised by Cochrane Centres and those responsible for the annual Cochrane Colloquia, with participation of experienced editors from established CRGs and beginners. Some of these workshops will be a general sharing of editorial experience, others may focus on particular specialised aspects of editing - such as the correct use of meta-analysis, how to plan the work of a CRG and how to help and support the reviewers.

Within each CRG, the coordinating editor or other senior editors should take responsibility for the systematic development of the editorial skills of the other editors. When a new editor joins a group, the senior editor(s) should assess her or his experience and skills. The new editor should begin by editing one or more reviews together with one of the established editors, who can act as supervisor and tutor. They can together make an outline plan for the new editor's training, revising it as the training proceeds. It may be useful for a CRG's editorial office to keep a formal or informal record of the work experience and training of each editor.

Every CRG must have a Criticism Editor who is responsible for screening and summarizing incoming criticisms and forwarding these summaries on to the Reviewers and the San Francisco Cochrane Center. Criticism Editors are also responsible for making sure that reviewers respond to their criticisms in a timely manner.

10.9 Peer review and the Criticism Management System

It is important to have efficient arrangements for criticising the reviews prepared by contributors to the Cochrane Collaboration, and for amending reviews in the light of valid criticisms. Developing these arrangements is facilitated if the potential of electronic publication is exploited imaginatively. Opportunities for criticising reviews before they are published in print are restricted by the number and competence of the referees selected by editors. After a review has been printed in a paper journal or book, opportunities for published criticism are usually limited to the few letters that editors can accept for publication, or to book reviews, that are often unhelpfully brief and non-specific. It is also frustrating that there is no straightforward way in which the authors of printed reviews can amend their reports after taking account of valid criticisms.

The Cochrane Collaboration has created a Criticism Management System through which successive versions of each review can be updated to reflect not only the emergence of

new data, but also valid criticisms. Successive versions of a particular review, together with any intervening criticisms, will be archived electronically.

10.9.1 Refereeing

Each CRG is required to publish a statement describing its pre-publication peer-reviewing policy in the 'Editorial process' section of their module in *The Cochrane Library*.

The main issues to consider when the title for a review is being considered for registration are whether there is any overlap or potential duplication of effort with another reviewer either within or outside the CRG; objectives are clearly phrased and include all of the components of a well-formulated question; and the review is likely to be feasible. This refereeing stage can often be accomplished quickly by a CRG's editorial team.

Refereeing protocols can be more time-consuming than the refereeing of the full review. This is done to ensure that background information is rational and clearly presented, and that appropriate methods are planned for identifying, collecting and synthesising data. Peer review at this stage is particularly important to prevent methodological errors that may not be easily remediable at later stages of the review. The refereeing of the full review will include a second critique of the review's methods as well as a critique of the actual results, presentation of results, discussion and conclusion.

Prior to publication, all reviews must be refereed by at least two people external to the editors of the CRG. The CRG editors should appoint a referee (or contact) editor(s) for each review. If they inform the San Francisco Cochrane Center (sfcc@sirius.com) of their choice, the Center can train and support this person. It is recommended that these referees have 1) methodological expertise, 2) content area expertise, and/or 3) are a potential consumer of the review. The two referees should be selected on the basis of having differing viewpoints. Referees should include people without direct financial or personal conflicts of interest concerning the topic addressed. The referees should be asked to submit courteous and constructive comments on the Review that identify its weaknesses or fatal flaws, as well as ways of improving it. They should also be requested to return these comments to the Referee Editor within, at most, a month.

Explicit standardised methods and checklists aimed at ensuring comprehensiveness and limiting bias should be encouraged among peer reviewers. Specific areas to address at each stage of peer review vary. Differences among referees' critiques should be elucidated and reconciled whenever possible. This could be done by arbitration by the CRG editors or the use of an additional independent referee. The referee editor should monitor the timeliness of returned comments, grade the quality of the comments, and, if necessary, appoint backup referees. They forward the comments from the referees, together with their own comments (if any), to the authors of the review or to the CRG Co-ordinator for distribution to the authors of the review and, if appropriate, the other editors. The referee editor, in concert with the editorial team, approves the final version of the review before it is published in *The Cochrane Library*.

The referee editor should keep records of all materials received and sent out during the refereeing process. An electronic refereeing system for keeping electronic records of these exchanges is being developed. Copies of the records will be requested and studied

periodically by the San Francisco Cochrane Center in order to improve the refereeing process.

10.9.2 Checklist for peer reviewers

Preparing a review involves judgements at each step in the review process. Both systematic and random errors can occur. Several checklists are available for peer reviewers to use as guides for detecting important errors in the review process. Some points to keep in mind are shown below. These have been extracted from multiple citations (Jackson 1980, Cooper 1982, Light 1984a, L'Abbe 1987, Mulrow 1987, Sacks 1987, Oxman 1988, Oxman 1994a, Oxman 1994b, Cook 1995)

Problem Formulation
- Are review questions well formulated with specified key components?
- Are any changes to the protocol well documented and justified?

Study Identification
- Is there a thorough search for relevant data using appropriate sources?
- Are the search strategies appropriate to the question posed?

Study Selection
- Are appropriate inclusion and exclusion criteria used to select studies?
- Are selection criteria applied in a manner that limits bias?

Assessment of Studies
- Is the validity of individual studies addressed in a reliable manner?
- Are important parameters (e.g., setting, study population, study design) that could affect study results systematically addressed?

Data Collection
- Is there a minimal amount of missing information regarding outcomes and other variables considered key to interpretation of results?

Data Synthesis
- Are reasonable decisions made concerning whether and how to combine data?
- Are important factors, such as study designs, considered in the synthesis?
- Are results sensitive to changes in the way the analysis was done?
- Is the precision of results reported?

Discussion
- Are limitations of studies and the review process stated?
- Are review findings integrated within the context of relevant indirect evidence?

Reviewer's Conclusions
- Are conclusions supported by the content of the review?
- Are plausible competing explanations of observed effects addressed?
- Is any interpretation of inconclusive evidence (i.e. no evidence of effect) and/or of evidence that a particular strategy did not work (i.e. evidence of no effect) appropriate?

- Are important considerations for decision-makers identified, including values and contextual factors that might influence decisions?

10.10 Updating reviews

When registering a review with the Cochrane Collaboration, reviewers agree to keep it up-to-date. This entails repeating, at periodic intervals, the steps involved in the original review. Some of the steps will require minimal effort (e.g. reviewing the research question to make sure it is still relevant) while others may require a substantial investment of time and effort.

The most logistically demanding aspect of keeping a review up-to-date is the identification of new studies. For CRGs that are sufficiently organised and funded, the periodic identification of relevant new studies is an ongoing function of the editorial team (usually the CRG's Coordinator or Trial Search Coordinator). In other instances, reviewers and editors must work out collaborative mechanisms to periodically identify new studies. At a minimum, strategies to identify new studies should include periodically checking the CRG's specialised register, CENTRAL and MEDLINE. The Cochrane Collaboration has a Criticism Management System which continues to develop and allows users of Cochrane Reviews to provide comments and criticisms of reviews, and this is discussed further in the next section. It is likely to provide an additional source of studies to be considered for the review.

Original data collection forms should be used to abstract new research evidence. If new research evidence addresses important variables that were not included in the original collection form, these may be modified. For example, if reviewers had originally only abstracted morbidity and mortality outcomes in trials addressing treatment of advanced cancer, and recent studies routinely report quality of life outcomes, the collection form could be amended. In such instances, reviewers may need to recheck whether any of their earlier identified studies had such information that was overlooked.

Occasionally, reviewers may decide to include a new analysis strategy in their updated review; for example, using statistical methods not previously available in RevMan. In general, new analysis strategies will represent substantive changes that merit editorial critique through the CRG's established editorial process.

How often reviews need updating will vary depending on the production of valid new research evidence. Reviewers should work with their editorial team to establish guides addressing when new research evidence is substantive enough to warrant a major update or amendment. The dates of such amendments must be recorded in the What's New section of the review. It is Collaboration policy that reviews should either be updated within two years or should have a commentary added to explain why this is done less frequently. It is also Collaboration policy that protocols that have not been converted into full reviews within two years should generally be withdrawn from the CDSR. Even if no substantive new evidence is found on annual review and no major amendment is indicated, this information should still be used to update the review by adding the date of the latest search for evidence to the review.

If a review needs to be suspended or withdrawn, this should be noted in the Published Notes section of the review. The review containing this suspension/withdrawl notice

should be submitted for publication in each issue of the CDSR, until the content of review is judged to be satisfactory by the reviewers and their CRG. If a review is merged with another review, a notice should be included in its Published Notes section to explain that it has been withdrawn for this reason.

10.11 Responding to criticisms

The electronic format of the *CDSR* offers a unique opportunity to respond to, and incorporate criticism from the users of Cochrane reviews. This will greatly increase the quality of the reviews, and also allow users to be brought into the reviewing process.

The reader should use the 'Comments/Criticisms' button to make constructive and courteous comments. These are automatically sent to the Criticism Editor of the relevant CRG. In an effort to prevent redundancy, the user should read criticisms that have already been received before sending in their own criticism. They can do this by visiting the 'Current Comments and Criticisms' Internet page: http://www.update-software.com/comcrit.htm .

When they receive the feedback, the Criticism Editor should summarize it and send a copy to the reviewers and to the San Francisco Cochrane Center so that it can be posted on the 'Current Comments and Criticisms' web page. The reviewers are responsible for responding to all criticisms in a timely fashion. They should provide a written response by using the criticism section of RevMan and update their review if appropriate. Software is being developed to help Criticism Editors to coordinate the reviewers' responses to comments and criticisms.

10.12 References

Bastian 1998. Bastian H. Speaking up for ourselves: the evolution of consumer advocacy in health care. International Journal of Technology Assessment in Health Care 1998;14:3-23.

Cook 1995. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol 1995; 48:167-71.

Cooper 1982. Cooper HM. Scientific guidelines for conducting integrative research reviews. Rev Educ Res 1982; 52:291-302.

Jackson 1980. Jackson GB. Methods for integrative reviews. Rev Educ Res 1980; 50:438-60.

L'Abbe 1987. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987; 107:224-33.

Light 1984. Light RJ, Pillemer DB. Summing Up: The Science of Reviewing Research. Cambridge: Harvard University Press, 1984.

Mulrow 1987. Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987; 106:485-8.

Oxman 1988. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988; 138:697-703.

Oxman 1994a. Oxman AD, Cook DJ, Guyatt GH. Users' guide to the medical literature, VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994; 272:1367-71.

Oxman 1994b. Oxman AD. Checklists for review articles. BMJ 1994; 309:648-51.

Sacks 1987. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450-5.

11. REVIEWS USING INDIVIDUAL PATIENT DATA

11.1 Rationale

If a systematic review is to contain a meta-analysis in which the results of separate studies will be brought together in a statistical synthesis, then the data for this could be collected in a variety of ways. These include extraction from published reports, collection of aggregate data from the responsible investigators or collection of individual patient data (IPD) from the investigators. The latter has been used in large-scale collaborative overviews in which data from all randomised trials in a particular disease area are brought together (EBCTCG 1992) and also in more restricted reviews in which data from a relatively small number of trials assessing a specific healthcare intervention are collected and combined (Jeng 1995). Systematic reviews based on IPD have been described as the yardstick against which all reviews should be measured (Chalmers 1993). Although they can require more time, resources and expertise than other forms of review, the process brings with it a number of advantages. Reviewers should consider the importance of these advantages to their particular systematic review when deciding whether to embark on such a project. Examples of IPD reviews are available in the *Cochrane Database of Systematic Reviews,* including some that were originally published in paper journals.

11.2 Methods Group on Individual Patient Data Reviews

To try to help with this decision and with the logistics of such projects, a Cochrane Collaboration Methods Group has been established to provide guidance to those wishing to conduct an IPD meta-analysis. This group (co-convened by Lesley Stewart and Mike Clarke) was formed following a UK Cochrane Centre sponsored workshop in April 1994 at which representatives of research groups involved in such projects were brought together for the first time. This allowed for discussion of areas such as protocol use and development, methods of data-checking, and resource requirements. And a detailed report from the workshop was published in Statistics in Medicine in October 1995 (Stewart 1995]. This report is included in full in this Handbook (Appendix 11a) with permission of the publisher.

11.3 What an IPD meta-analysis is and is not

As with any systematic review the fundamental principle for one which uses IPD is that as much as possible of the relevant, valid evidence is included. This means that the process of trial identification must be as thorough as possible and that the attempts to collect data must be equally thorough. The ultimate aim should be that all randomised participants, and no non-randomised participants, from all relevant studies are included and that they are analysed using the intention-to-treat principle. In this way, systematic biases and chance effects will be minimised. To this end, the data collection should be kept simple and straightforward, with the minimum amount of data being collected for the required analyses. It should be as easy as possible for the investigators to supply their data since this should increase the likelihood that data will be received for all relevant studies. In addition, investigators should know that any data supplied for the review will be held in confidence and will not be used for any other purpose without their permission, and that

the reports of the review will be published in the names of the collaborating investigators rather than the central co-ordinators.

The predominant difference between an IPD meta-analysis and meta-analysis based on aggregate data (whether extracted from published reports or supplied direct by investigators) is that the combined study results come from a central re-analysis of the raw data from each study. The necessary data items are sought and, after central processing, any inconsistencies or problems are discussed and hopefully resolved by communication with the responsible investigators. The finalised data for each study are then analysed separately to obtain summary statistics, which are combined to give an overall estimate of the effect of treatment. In this way, participants are only directly compared with others in the same study and the entire dataset is not pooled as though it came from a single, homogeneous study.

11.4 How can an IPD meta-analysis help?

If a systematic review relies solely on data from published studies, it is open to a number of problems. The most obvious of these is that unpublished studies will not be included, but the published data may be inadequate for other reasons also. For example, there may be insufficient information on the types of patient or outcome of interest in the review, the data are 'frozen-in-time' when important findings may come from longer follow-up or more detailed study, and the intention-to-treat principle may not have been followed (and, occasionally, this might not be clear from the published report). Collection of either aggregate or individual patient data from investigators will resolve some of these problems: unpublished trials can be included, updated data on specific types of participant and outcome can be requested, and whether the data are based on the randomised allocations can be clarified (if the studies are randomised trials).

Collecting IPD rather than aggregate data brings additional advantages. These include the ability to undertake survival and other time-to-event analyses; to undertake analyses using commonly defined subgroups to test and generate hypotheses; to ensure the quality of the randomisation and follow-up data used in the meta-analysis through detailed data checking and iterative correction of errors by communication with the investigators; and to update follow-up information through patient record systems (such as mortality registers) where available. In addition, it might be easier for an investigator to send individual patient, rather than aggregate, data particularly if they do not have sufficient data-management or statistical support to prepare the necessary tables. It will also be easier for a small amount of extra information to be supplied. For example, if further follow-up becomes available on some participants, the investigator can simply send these details instead of preparing new tables.

Furthermore, as IPD meta-analyses involve the collaboration of the investigators, they can have other benefits, some of which may also be found if the investigators are contacted for aggregate data. These include more complete identification and understanding of the studies; better compliance with providing missing data; more balanced interpretation of the results of the review; wider endorsement and dissemination of these results; a broader consensus on the implications for future practice and research; and possible collaboration in such research.

11.5 Where is the evidence?

One of the aims of the Methods Group on IPD based meta-analysis was to establish, and encourage the tackling of, a research agenda to investigate this approach to systematic reviews. Limited empirical evidence already exists for some of the advantages of IPD reviews over other types of review. Typically, these have involved comparison of the results from an IPD meta-analysis with those from a meta-analysis based on published material. They have shown the importance of the former in helping control publication bias, in ensuring the use of the intention-to-treat principle in the analysis, and in obtaining a fuller picture of the effects of different treatments over time (Stewart 1993, Pignon 1993, Jeng 1995, Clarke 1997).

11.6 Converting reviews that used individual patient data into Cochrane reviews

The conversion into Cochrane reviews of relevant, pre-existing reviews that have used individual patient data should be encouraged, unless a Cochrane review of higher quality can be prepared in some other way. However, these conversions can present particular challenges to reviewers and Collaborative Review Groups (CRGs).

IPD meta-analyses have generally been carried out by large**,** collaborative groups of trialists. Sometimes more than 100 people will be involved, including the trialists who provided their source data for re-analysis, an organisational secretariat and, in some cases, an advisory committee. However, the size of these groups, the social politics involved and prior paper publication can make it difficult to comply with certain Cochrane procedural, style and format recommendations. In particular:

- For pre-existing IPD reviews, a protocol can usually not be provided retrospectively and CRGs should not require one before accepting the review. However, IPD reviewers should try to submit protocols for ongoing projects at an early stage.

- The text of the IPD review will usually have been through many drafts and circulated to all members of the collaborative group for comment. Agreement on wording within such large groups is not always easy to achieve and so it may be difficult to change the text of a review for inclusion in the Cochrane Database of Systematic Reviews. The editors and peer reviewers of CRGs should be sympathetic to this constraint.

- The Study Identifiers or labels will usually have been chosen in collaboration with the trialists and it is unlikely to be possible to change these to reflect particular conventions.

- For a pre-published IPD review, the secretariat will already have obtained sufficient declarations of contribution and consent to authorship from each member of the collaborative group to satisfy publication of the review in a journal. It would be resource-intensive to further obtain Cochrane authorship contribution forms for each "author" and the authorship declarations submitted to the journal should be accepted by the CRG as an alternative.

IPD meta-analyses should be peer reviewed by the CRG's normal peer review process. However, the difficulties of making changes (discussed above) should be made clear to

the peer reviewers. They should also bear in mind that pre-existing IPD reviews will probably have been through an extensive peer review process prior to submission to the CRG. Manuscripts will have been scrutinised by the trialists' group, secretariat and advisory committee for the review, as well as by the peer review process of the journal in which the IPD review was published. As with all reviews, the final decision on whether an IPD review is acceptable for publication as a Cochrane review in the Cochrane Database of Systematic Reviews must rest with the editorial group of the CRG.

CRGs who would like advice relating to IPD reviews, for example in regard to their peer review, should contact the IPD meta-analysis Methods Group for help.

11.7 Prospective meta-analysis

Prospective meta-analysis are a special form of IPD meta-analysis. In these projects, a group of investigators agree, in advance of knowing the results of their studies, to pool their data in the future. A Cochrane Collaboration Methods Group has been established to address this issue and will provide training and support in the conduct of these projects (appendix 11b).

11.8 Further information

Many of the topics discussed here are expanded on in Stewart 1995 (Appendix 11a). That report also contains examples of how IPD meta-analyses have been conducted previously, which may be useful to reviewers planning one now. If Cochrane reviewers would like further information they should contact the Methods Group. In addition, a slide show that is used by the Methods Group in training workshops is available from the Collaboration's Internet site (http://www.cochrane.org/cochrane/training.htm).

11.9 References

Chalmers 1993. Chalmers I. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. Ann N Y Acad Sci 1993; 703: 156-65.

Clarke 1997. Clarke M, Stewart L. Individual patient data or published data meta- analysis: a systematic review [abstract]. *Proceedings of the Fifth Cochrane Collaboration Colloquium* 1997; 94, abstract 019.04.).

EBCTCG 1992. Early Breast Cancer Trialists' Collaborative Group. Systematic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Lancet 1992; 339:1-15, 71-85.

Jeng 1995. Jeng GT, Scott JR, Burmeister LF. A comparison of meta-analytic results using literature vs individual patient data: paternal cell immunization for recurrent miscarriage. JAMA 1995; 274: 830-6.

Pignon 1993. Pignon JP, Arriagada R. Meta-analysis. Lancet 1993; 341:964-5.

Stewart 1993. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is ther a difference? Lancet 1993; 341:418-22.

Stewart 1995. Stewart L, Clarke M, for the Cochrane Collaboration Working Group on meta-analyses using indivudal patient data. Practical methodology of meta-analyses (overviews) using updated individual patient data. Stat Med 1995; 14:2057-79.

APPENDICES

APPENDIX 1. THE COCHRANE COLLABORATION OPEN LEARNING MATERIAL FOR REVIEWERS

The *Cochrane Collaboration Open Learning Material for Cochrane Reviewers* is designed to accompany the *Cochrane Reviewers' Handbook* in helping people who are working on a Cochrane Review. It does not replace the *Handbook*, instead it provides a framework to progressing through the *Handbook*, supplementing it with examples and activities along the way. The first version of the *Open Learning Material for Reviewers* (Version 1.1) was made available on the Internet in November 2002. It can be accessed at http://www.cochrane-net.org/openlearning/.

Many training events and tools have been developed and published to help reviewers acquire the skills they need, however not all are accessible to all reviewers. This material is designed to help train reviewers in the methods and processes of performing a Cochrane review. Along with the *Handbook*, this material will stand alone, offering an alternative to face-to-face training, especially for those reviewers living and working away from easy access to the training offered by Cochrane Centres and Cochrane Collaborative Review Groups. For those able to access this face-to-face training, this material will serve as a useful resource to remind them of what they learned.

This material takes a step-by-step approach to Cochrane Reviews, exploring each step individually, signposting appropriate links and references and providing examples and activities to help you make sense of the information. The material is organised in modules, each module relating to a consecutive section of your review. It is a good idea to complete each module as you start working on the corresponding part of your review. There are also some additional modules relating to issues of reviewing that do not occur in all Cochrane Reviews.

Editors:
Phil Alderson, UK Cochrane Centre (palderson@cochrane.co.uk)
Sally Green, Australasian Cochrane Centre (sally.green@med.monash.edu.au)

Acknowledgments:
Jon Deeks (Centre for Statistics in Medicine, Oxford, UK), Julian Higgins (MRC Biostatistics Unit, Cambridge, UK), Jacob Riis (Nordic Cochrane Centre, Copenhagen, Denmark) and Chris Silagy (Australasian Cochrane Centre, Melbourne, Australia) contributed to the *Open Learning Material for Cochrane Reviewers*. The National Health Service Research and Development Programme, UK and the Department of Health and Ageing, Australia provided support for its preparation. We are grateful to Mike Clarke of the UK Cochrane Centre and Philippa Middleton of the Australasian Cochrane Centre for their comments on this material, and to Nicola Thornton and Sarahjayne Sierra of the UK Cochrane Centre for their help with copy editing.

Principles underlying use of this material

Use of the *Cochrane Collaboration Open Learning Material for Cochrane Reviewers* should adhere to the following set of principles.

- This material, developed by the Cochrane Collaboration for training reviewers, should be freely available to those reviewers with a registered Cochrane Review title.
- Profits generated from training non-Cochrane reviewers with this material should benefit the Cochrane Collaboration.
- Organisations utilising this material within their courses should acknowledge its source.
- Any suggestions for the improvement and updating of this material should be sent to the editors so that these suggestions can be considered in future revisions of the material.

APPENDIX 2a. GUIDE TO THE FORMAT OF A COCHRANE REVIEW

2a.1 Cover sheet

The cover sheet includes the following information:

Title: The title should succinctly state the focus of the review. It should make clear the intervention(s) reviewed and the problem at which the intervention is directed. Someone scanning the title should be able to decide quickly whether the review addresses a question of interest. At its most basic, a title should take the structure 'Intervention for condition'. Other structures are included in the Style Guidelines for Cochrane Reviews. Mention of specific outcomes should only rarely be retained within the title, usually as a subtitle separated by a colon from the main title.

Version: One version of each review must be marked as the primary version and this is the one that should be submitted for publication in the *Cochrane Database of Systematic Reviews* (*CDSR*).

Status: This specifies what stage the review is at: title, protocol or full review. Titles are only used internally, within CRGs, and are not included in *CDSR*.

Date edited: This date is entered automatically any time the review is amended.

Date of last substantive update: The reviewer(s) and/or editors of a Collaborative Review Group (CRG) should decide whether an amendment is substantive or not. Substantive amendments are ones which are sufficient to recommend that previous readers of the review should look at the updated version. For example, important changes in the conclusions of the review or the list of studies that are included or excluded, may qualify as substantive amendments.

Date next stage expected: This must be filled in for protocols to inform *CDSR* users when they can expect the completed review to be available. It can also be filled in for full reviews to inform *CDSR* users when an updated review is likely to be available.

Contact reviewer: This should be the contact details for the person to whom correspondence about the review should be addressed.

Co-reviewers: This should contain the contact details for any co-reviewers on the review.

Contributions: The names and contribution of all individuals who have contributed to a Cochrane Review should be described in this section. This might include the contributions of the editorial team of the CRG. One contributor should be identified as the guarantor of the review. All contributors should discuss and agree on their respective descriptions of contribution before the review is submitted for publication on *CDSR*. When the review is updated, this section should be checked and revised as necessary to ensure that it is accurate and up-to-date.

The following potential contributions have been adapted from (Yank 1999). This a suggested scheme and the section should describe what people did not simply try to identify which of these categories someone's contribution falls within. Ideally, the contibutors should describe their contribution in their own words:

Conceiving the review
Designing the review
Coordinating the review
Data collection for the review
- Developing search strategy
- Undertaking searches
- Screening search results
- Organising retrieval of papers
- Screening retrieved papers against inclusion criteria
- Appraising quality of papers
- Abstracting data from papers
- Writing to authors of papers for additional information
- Providing additional data about papers
- Obtaining and screening data on unpublished studies

Data management for the review
- Entering data into RevMan

Analysis of data
Interpretation of data
- Providing a methodological perspective
- Providing a clinical perspective
- Providing a policy perspective
- Providing a consumer perspective

Writing the review
Providing general advice on the review
Securing funding for the review
Performing previous work that was the foundation of current study

List of reviewers for citation: This can be considered the 'byline' for Cochrane reviews. Authorship of all scientific papers (including Cochrane protocols and reviews) establishes accountability, responsibility and credit (Rennie 1997, Flanagin 1998, Rennie 1998). When deciding on who should go in the byline for Cochrane reviews, it is important to distinguish individuals who have made a substantial contribution to the review (and who should be listed) and those who have made other contributions which should be noted in the Contributions section. This should be based on substantial contributions to the following three steps, based on (ICMJE 1997):

- conception and design of study, or analysis and interpretation of data
- drafting the review or revising it critically for important intellectual content
- final approval of the version to be published.

The list of reviewers for citations can be the name of an individual, several individuals or a collaborative group (e.g. Early Breast Cancer Trialists' Collaborative Group). Ideally, the order of authors should relate to their relative contributions to the review. The person who contributed most should be listed first.

Sources of support to the review: Reviewers should give details of grants that supported the review and other forms of support, such as support from their university or institution in the form of a salary. Sources of support are divided into 'internal' (provided by the institutions at which the review was produced) and 'external' (provided by other institutions or funding agencies).

What's new: This should describe the major changes to the review since it was last published in the *CDSR*. For example, you should describe briefly how much new information has been added to the review (e.g. number of studies, participants or extra analyses) and any important changes to the results of the review. The associated dates provide the reader of the review with information on when the review was last updated and why.

Issue protocol first published: The issue of *The Cochrane Library* where the protocol was first published (e.g. Issue 2, 1998).

Issue review first published: The issue of *The Cochrane Library* where the full review was first published (e.g. Issue 1, 1999).

Date of last substantive update: The reviewer(s) and/or editors of a Collaborative Review Group (CRG) should decide whether an amendment is substantive or not. Substantive amendments are ones which are sufficient to recommend that previous readers of the review should look at the updated version. For example, important changes in the conclusions of the review or the list of studies that are included or excluded, may qualify as substantive amendments.

Date of last minor update: The last date on which the review was updated, but this update is not sufficient to recommend that previous readers of the review should look at the new version.

Date review re-formatted: The last date on which structural changes were made to the review (e.g. the addition of a new fixed heading).

Date new studies sought but none found: The last date on which a search was done for new studies but none were found.

Date new studies found but not yet included or excluded: The last date on which a search was done for new studies and some were found and added to the list of studies awaiting assessment or ongoing studies.

Date new studies found and included or excluded: The last date on which studies were added to the list of included or excluded studies.

Date reviewers' conclusions section amended: The last date on which the Reviewers' Conclusions section was amended in such a way that it is recommended that previous readers of the review should look at the new version.

Date comment / criticism added: The last date on which a comment or criticism was added to the review.

Date response to comment / criticism added: The last date on which a reply to a comment or criticism was added to the review.

Unpublished CRG notes: These notes will not be published in the *CDSR* but can be used for sending messages to co- reviewers or the CRG's editorial team.

Published notes: These notes will be published in the *CDSR*.

Amended sections: These boxes can be checked to make it easier for co-reviewers or the CRG's editorial team to locate changes in the review. This information is not published in the *CDSR*.

2a.2 Synopses

The synopsis is a brief summary of the results of the review in plain language for consumers and non-specialist readers. The synopsis does not replace the abstract but is an additional product. It will be published as part of the Cochrane Review in the Cochrane Database of Systematic Reviews. The synopsis should enhance the accessibility of the review, disseminate its findings to a wide community internationally and act as an aid to browsing in The Cochrane Library. The Consumer Network will consider each synopsis and those that are suitable, will additionally be made available on the Cochrane Collaboration's Internet site, and available on paper from the Consumer Network and some other Cochrane entities. Synopses will be translated into several non-English languages. Stand-alone publications of Cochrane Review synopses will include reference to the full review and will be replaced when the Cochrane Review (including synopsis) is updated.

Reviewers may either draft the synopsis themselves (ideally with consumer input on content and readability), or send a draft or final copy of their review to the Cochrane Consumer Network. To use this service, reviews can be sent to the Cochrane Consumer Network at any time. The Network will provide a draft synopsis as soon as possible after receipt of a review, usually within two weeks. CRGs may wish to develop their own policy about this process. Draft synopses prepared by reviewers or CRGs can also be sent

to the Cochrane Consumer Network to ensure consistency in content, style and length. Drafts prepared or edited by the Cochrane Consumer Network will be sent to the contact reviewer and their CRG editorial base for comment and revision as necessary. Editorial approval of the synopsis to be published with a Cochrane Review will be the responsibility of the CRG. Assistance with drafting the synopsis, whether from consumers or the Cochrane Consumer Network, may be acknowledged in the 'Acknowledgements' section of the review. When a review update requires amendment of an existing synopsis, the new version can be sent to the Cochrane Consumer Network for editing. This will also enable the Network to notify those maintaining translations of synopses that an update is required.

The synopsis has two parts: a short, single sentence 'headline' of up to 25 words (in lower case apart from the first letter of the first word); followed by a single paragraph summary of the context and findings of the review (50 to 100 words). These should be separated by a blank line. In order to keep the 'headline' short, some abbreviations and technical terms may be inevitable. However, these should be expressed in plain language in the body of the synopsis (with either the technical or simple description in parentheses).

Sentences should be short and use the most easily understandable language possible. The text should briefly cover: the context and alternatives relating to the problem and interventions; the potential benefits and risks of the interventions; and the main findings of the review. The synopsis should simply present the evidence, and not advise any particular action. However, where major dangers or benefits are identified, these can be noted. Brief statements about the strength of the evidence or limitations in terms of generalisability should be included when these are critical. The synopsis must be consistent with the review and written in the third person (e.g. 'the reviewers' instead of 'we'). The number of studies and full statistical results with confidence intervals should be avoided, although a narrative description of the results can be included if this might be helpful to readers.

2a.3 Abstract

All full reviews must include an abstract of not more than 400 words. It should be kept as brief as possible without sacrificing important content. Abstracts are made freely accessible on the Internet and will often be read as stand-alone documents. They should, therefore, summarise the key methods and content of the review and not contain any material that is not in the review. The content must be consistent with the text, data and conclusions of the review and not include references to any information outside the abstract. A hypothetical example is included at the end of this section.

Abstracts should be made as readable as possible without compromising scientific integrity. They should primarily be targeted to healthcare decision makers (clinicians, consumers and policy makers) rather than just researchers. Terminology should be reasonably comprehensible to a general rather than a specialist medical audience. Abbreviations should be avoided, except where they are widely understood (e.g. HIV). Where essential, other abbreviations should be spelt out (with the abbreviations in brackets) on first use. Names of drugs and interventions which can be understood internationally should be used wherever possible. One way to measure the readability of an abstract is to use the Flesch Reading Ease Score that is part of the Tools component in Word and is easy to access and apply. There may be better tests, but Flesch is convenient

and can identify the most difficult pieces to read. The higher the Flesch score, the better the readability. For instance, a readability of 30 (out of a possible 100) indicates that 30% of the adult, English-speaking population will be able to read the piece comfortably.

The content under each heading in the abstract should be as follows:

Background: This should be one or two sentences to explain the context or elaborate on the purpose and rationale of the review.

Objectives: This should be a precise statement of the primary objective of the review, ideally in a single sentence. Where possible the style should be of the form 'To assess the effects of *[intervention or comparison]* for *[health problem]* for/in *[types of people, disease or problem and setting if specified]*'.

Search strategy: This should list the sources and the date of the last search, using the active form 'We searched….' or, if there is only one reviewer, the passive form can be used, e.g. 'Database X, Y, Z were searched'. If the CRG's specialised register was used, this should be listed first in the form 'Cochrane X Group specialised register'. The order for listing any other databases should be the Cochrane Central Register of Controlled Trials, MEDLINE, EMBASE, other databases. The date range of the search for each database should be given. For the Cochrane Central Register of Controlled Trials this should be in the form 'Cochrane Central Register of Controlled Trials (*The Cochrane Library* issue 1, 2003)'. For most other databases such as MEDLINE, it should be in the form 'MEDLINE (January 1966 to December 2002)'. Searching of bibliographies for relevant citations can be covered in a generic phrase 'reference lists of articles'. If there were any constraints based on language or publication status, these should be listed. If individuals or organisations were contacted to locate studies this should be noted and it is preferable to use 'We contacted pharmaceutical companies' rather than a listing of all the pharmaceutical companies contacted. If journals were specifically handsearched for the review, this should be noted but handsearching done by the reviewers to help build the specialised register of the CRG should not be listed.

Selection criteria: This should be given as '*[type of study]* of *[type of intervention or comparison]* in *[disease, problem or type of people]*'. The outcomes should only be included if the review was restricted to specific outcomes.

Data collection and analysis: This should be restricted to how data were extracted and assessed, and not include details of what data were extracted. This section should cover whether extraction and quality assessment of studies were done by more than one person. If the reviewers contacted investigators to obtain missing information, this should be noted here. What steps, if any, were taken to identify adverse effects should be noted.

Main results: This section should begin with the total number of trials and participants included in the review, and brief details pertinent to the interpretation of the results (e.g. the quality of the studies overall or a comment on the comparability of the studies, if appropriate). It should address the primary objective and be restricted to the main qualitative and quantitative results (generally including not more than six key results). The outcomes included should be selected on the basis of which are most likely to help someone making a decision about whether or not to use a particular intervention. Adverse effects should be included if these are covered in the review. If necessary, the number of

studies and participants contributing to the separate outcomes should be noted. The results should be expressed narratively as well as quantitatively if the numerical results are not clear or intuitive (such as those from a standardised mean differences analysis). The summary statistics in the abstract should be the same as those selected as the defaults for the review, and should be presented in a standard way, such as 'odds ratio 2.31 (95% confidence interval 1.13 to 3.45)'. Ideally, rates of events (percentage) or averages (for continuous data) should be reported for both comparison groups. If overall results are not calculated in the review, a qualitative assessment or a description of the range and pattern of the results can be given. However, 'vote counts' in which the numbers of 'positive' and 'negative' studies are reported should be avoided.

Reviewers' conclusions: The primary purpose of the review should be to present information, rather than to offer advice. The Reviewers' conclusions should be succinct and drawn directly from the findings of the review so that they directly and obviously reflect the main results. Assumptions should not be made about practice circumstances, values, preferences, tradeoffs; and the giving of advice or recommendations should generally be avoided. Any important limitations of data and analyses should be noted. Important conclusions about the implications for research should be included if these are not obvious.

HYPOTHETICAL EXAMPLE OF AN ABSTRACT
Almonds and raisins in the treatment of influenza in adults
Peach A, Apricot D, Plum P

Background
Almonds and raisins both have antiviral properties, but they are not widely used due to incomplete knowledge of their properties and concerns about possible adverse effects.

Objectives
The objective of this review was to assess the effects of almonds and raisins in adults with influenza.

Search strategy
We searched the Cochrane Acute Respiratory Infections Group trials register (searched 15 December 2002), the Cochrane Central Register of Controlled Trials (The Cochrane Library Issue 1, 2003), MEDLINE (January 1966 to December 2002), EMBASE (January 1985 to December 2002) and reference lists of articles. We also contacted manufacturers and researchers in the field.

Selection criteria
Randomised and quasi-randomised studies comparing almonds and/or raisins with placebo, or comparing doses or schedules of almonds and /or raisins in adults with influenza.

Data collection
Two reviewers independently assessed trial quality and extracted data. Study authors were contacted for additional information. Adverse effects information was collected from the trials.

Main results
Seventeen trials involving 689 people were included. Five trials involving 234 people compared almonds with placebo. Compared to placebo, almonds significantly shortened duration of fever by 23% (by 1.00 days, 95% confidence interval 0.73 to 1.29). Six trials involving 256 people compared raisins with placebo. Raisins significantly shortened duration of fever by 33% compared to placebo (by 1.27 days, 95% confidence interval 0.77 to 1.77). The little data available directly comparing almonds and raisins (two trials involving 53 people) indicated that the efficacy of the two drugs was comparable, although the confidence intervals were very wide. Based on four trials of 73 people, central nervous system effects were significantly more common with almonds than raisins (relative risk 2.58, 95% confidence interval 1.54 to 4.33).

Reviewers' conclusions
Almonds and raisins appear to be equally effective in the treatment of influenza. Both drugs appear to be relatively well tolerated, although raisins may be safer.

2a.4 Text

The text of the review should be as succinct as possible. It should be written so that someone who is not an expert in the area can understand it, in light of the following policy statement (taken from Cochrane News 1999; 15: 14):

'The target audience for Cochrane Reviews is people making decisions about healthcare. This includes healthcare professionals, consumers and policy makers with a basic understanding of the underlying disease or problem.

It is a part of the mission and a basic principle of the Cochrane Collaboration to promote the accessibility of systematic reviews of the effects of healthcare interventions to anyone wanting to make a decision about healthcare. However, this does not mean that Cochrane Reviews must be understandable to anyone, regardless of their background. This is not possible, any more than it would be possible for Cochrane Reviews to be written in a single language that is understandable to everyone in the world. It is important to translate the content, or elements of the content, of reviews into different languages and formats targeted at different audiences including healthcare professionals, consumers and policy makers in a variety of circumstances.

Cochrane Reviews should be written so that they are easy to read and understand by someone with a basic sense of the topic who may not necessarily be an expert in the area. Some explanation of terms and concepts is likely to be helpful, and perhaps even essential. However, too much explanation can detract from the readability of a review. Simplicity and clarity are also vital to readability.

The readability of Cochrane Reviews should be comparable to that of a well-written article in a general medical journal.'

Background: The review should begin with a brief synthesis of the underlying biology and healthcare of the topic being reviewed. This background should make clear the motivation and rationale for the review. It should be presented in a fashion that is understandable to the consumers of that healthcare.

Objectives: This should begin with a precise statement of the primary objective of the review, including the intervention(s) reviewed and the targeted problem. It might also mention why this review was undertaken and how it might relate to a wider review of a general problem. Any prior hypotheses should be stated and the comparisons that are made in the review should be consistent with these. If a review addresses more than one hypothesis and includes several comparisons, the comparisons should be grouped for each hypothesis or question.

Criteria for considering studies for this review: The criteria used to select studies for inclusion in the review should be stated. **Types of studies** (e.g. 'all randomised controlled comparisons' or 'all double blind randomised controlled trials'), **types of**

participants, **types of interventions** and **types of outcome measures** are subheadings in this section.

Search strategy for identification of studies: The data sources used to identify studies should be summarised, including bibliographic databases, reference lists from pertinent articles and books, conference proceedings and personal contact with experts or organisations active in the area. The databases searched, the date they were searched and the terms used should be stated, including any constraints, such as language. If a CRG has developed a specialised register of studies and this was searched for the review, a standard description of this register can be referred to but information should be included on when and how the specialised register was last searched. If journals were specifically handsearched for the review, this should be noted but handsearching done by the reviewers to help build the specialised register of the CRG should not be listed. Any additional data sources used should be listed, including any contacts made with individuals or organisations (including pharmaceutical companies) to identify studies. The search should be as up-to-date as possible.

Methods of the review: This should include the method used to apply the selection criteria (e.g. if they were applied independently by more than one reviewer), the criteria used to assess the quality of studies and how they were applied, how data were obtained (e.g. if individual patient data were sought ,or if the number of events was calculated from published survival curves), how the data were synthesised, and any statistical techniques used and sensitivity analyses performed. If a CRG uses a standard approach for all of their reviews, the methods section can reference a description of those methods in the CRG's module. Similarly, if a Methods Group (MG) has recommended a standard approach and a review uses that approach, the methods section can reference the relevant report or MG module (e.g. for the method of collecting data or assessing the quality of studies).

Description of studies: This should refer to the information contained in the 'Characteristics of Included Studies' and the 'Characteristics of Excluded Studies' tables. It should describe key characteristics of the study participants, interventions and outcome measures in the included studies and any important differences among the studies. The sex and age range of participants should be stated here unless it is obvious (e.g. if all the participants are pregnant). Reviewers should note any other characteristics of the studies that they regard as important for readers of the review to know.

Methodological quality of included studies: This should describe the general quality of the included studies and any important flaws in individual studies. If the quality of each study was assessed using explicit criteria, the criteria that were used should be described or referenced under 'Methods'. How each trial scored on each criterion can be summarised in this section or, preferably, included in the 'Characteristics of Included Studies'.

Results: This should be a summary of the main findings of the review and any sensitivity analyses that were undertaken. Subheadings can be used if they make reading easier (e.g. for each prior hypothesis if a review addresses more than one). The results of individual trials, and any statistical summary of these, should be included in Data tables. Reviewers should avoid making inferences in this section. A common mistake to avoid (both in describing the results and in drawing conclusions) is the confusion of 'no evidence of an

effect' with 'evidence of no effect'. When there is inconclusive evidence, it is wrong to claim that it shows that an intervention has 'no effect' or is 'no different' from the control intervention. It is safer to report the data, with a confidence interval, as being compatible with either a reduction or an increase in the outcome.

Discussion: This should include brief comments on any methodological limitations of the included studies and the review that are important for decisions about practice or future research. Comments on how the included studies fit into the context of other evidence might be included here, stating clearly whether the other evidence was systematically reviewed. Comments on how the results of the review fit into the context of current clinical practice might be included here, although reviewers should bear in mind that current clinical practice might vary internationally.

Reviewers' conclusions: The primary purpose of the review should be to present information, rather than to offer advice. **Implications for practice** and **Implications for research** are subheadings in this section. The implications for practice should be as practical and unambiguous as possible. They should not go beyond the evidence that was reviewed. 'No evidence of effect' should not be confused with 'evidence of no effect'. The implications for research should not include vague statements such as 'more research is needed'. Reviewers should state exactly what research is needed, why and how urgently. Opinions on how the review might be improved with additional data or resources might also be included here.

Acknowledgements: This section should be used to acknowledge any individuals or organisations who the reviewers wish to acknowledge but who have not made a sufficient contribution to the review to be included in the Contributions section.

2a.5 Conflict of interest

Any conflict of interest capable of influencing the judgements of any of the reviewers should be reported, including financial, personal, political or academic conflicts (see section 2.2). If there are no conflicts of interest, this should be stated explicitly, e.g. by reporting 'None known'.

2a.6 References

References to studies are organised under four standard headings: **included studies**, **excluded studies**, **studies awaiting assessment**, and **ongoing studies.** Other references include **additional references** that are cited in the review and **other published versions of the review**; e.g. if the review has been published in a journal. Reviewers should check their references for accuracy (Dickersin 1986, Eichorn 1987).

Studies awaiting assessment: Potentially relevant studies that have been identified, but cannot be assessed for inclusion until additional data or information are obtained, should be listed here. These need not be cited in the text of the review.

Ongoing studies: Studies which are ongoing but meet the inclusion criteria should be listed here.

Additional references: Other references cited in the text should be listed here. If a report of a study is cited in the text for some reason other than referring to the study (e.g. because of some background or methodological information in the report), it should be listed here as well as under the relevant study.

Other published versions: References to other published versions of the review in a journal, textbook or *CDSR* should be listed here.

2a.7 Tables and figures

Characteristics of included studies: This is a standard table with seven columns: study ID, methods, participants, interventions, outcomes, notes and allocation concealment. Reviewers must decide what characteristics of the included studies are likely to interest users of the review. It is possible to use codes so that each column can include several subcategories of information; e.g. a reviewer could include country, setting and sex under 'participants'. Information on the funding of a study could be included under 'notes'. Footnotes should be used for explanations of any abbreviations used (these will be published in the *CDSR*). Reviewers must also include information about the 'Data source' for all included studies to indicate whether published data only, unpublished data only or a mixture were used, or if unpublished data were sought but have not been used (e.g. because they have not been obtained).

Characteristics of excluded studies: Any studies meeting the inclusion criteria, or appearing to meet the inclusion criteria, that were excluded should be identified and the reason for exclusion should be given (e.g. inappropriate control group).

Characteristics of ongoing studies: This is a standard table with seven columns: Study ID, Trial name or title, Participants, Interventions, Outcomes, Starting date, Contact information and Notes. Footnotes should be used for explanations of any abbreviations used in the table (these will be published in the *CDSR*).

Comparisons and data: A review can include more than one comparison and a study can be included in more than one of these. The comparisons should correspond to the questions or hypotheses under 'Objectives'. Data for each comparison must be entered in a standardised format from which tables and figures for each comparison can be generated. Reviewers should try to avoid listing many comparisons or outcomes for which there are no data in the review since each comparison generates a graph even if it contains no data and analysis. Instead, reviewers should note these comparisons in the text of their review. Five types of tables are possible: dichotomous data, continuous data, individual patient data, generic inverse variance and other data.

Additional figures: From RevMan 4.2 onwards, Cochrane reviews can contain additional figures. Figures showing statistical analyses should follow the relevant guidance prepared by the Statistical Methods Group. Decisions about the suitability of pictures for inclusion in Cochrane reviews are the responsibility of the reviewers and the editors of their CRG. It is the responsibility of the reviewer(s) to obtain permission to include any figures or pictures for which the copyright is owned by someone else.

2a.8 Comments and criticisms

Summary, **Reply** and **Contributors** are subheadings in this section. The summary should be prepared by the criticisms editor for the CRG in consultation, if necessary, with the person submitting the comment. A reply to this should then be prepared by the reviewer(s). Details of the people who contributed to this process should be given. Further information on the comments and criticisms and the updating of reviews is given in section 10.11.

2a.9 Elements of Cochrane protocols and reviews that should be published:

The following elements of a protocol or a full review for a Cochrane Review in RevMan should be published when the protocol appears in *The Cochrane Database of Systematic Reviews*. If any of the sections marked below with a * are empty, the protocol or review should not be published until something has been added to the section.

PROTOCOL
*Title of review
*Name of contact reviewer
*Contact details of contact reviewer
*List of reviewers for citation
Contributions
Sources of support - intramural
Sources of support - extramural
What's new - text
What's new - issue protocol first published
*What's new - date of last substantive update
Notes - published
*Text of Review (These sections only to be published: Background, Objectives, Criteria for considering studies for this review, Types of studies, Types of participants, Types of interventions, Types of outcome measures, Search strategy for identification of studies, Methods of the review, Acknowledgements, Potential conflict of interest)
Other references - additional references
Additional tables
Additional figures
Comments and criticisms - title
Comments and criticisms - summary
Comments and criticisms - reply
Comments and criticisms - contributors

FULL REVIEW
*Title of review
*Name of contact reviewer
*Contact details of contact reviewer
*List of reviewers for citation
Contributions
Sources of support - intramural
Sources of support - extramural
What's new - text

What's new - issue protocol first published
What's new - issue review first published
*What's new - date of last substantive update
What's new - date new studies sought but none found
What's new - date new studies found but not yet included/excluded
What's new - date new studies found and included or excluded
What's new - date reviewers' conclusions section amended
Notes - published
Synopsis
*Abstract
*Text of Review (All sections to be published)
References to studies - included studies
References to studies - excluded studies
References to studies - studies awaiting assessment
References to studies - ongoing studies
Other references - additional references
Other references - other published versions of this review
Tables - characteristics of included studies
Tables - characteristics of excluded studies
Tables - characteristics of ongoing studies
Comparisons and data
Meta-analysis graphs
Additional tables
Additional figures
Comments and criticisms - title
Comments and criticisms - summary
Comments and criticisms - reply
Comments and criticisms – contributors

2a.10 References

Dickersin 1986. Dickersin K, Hewitt P. Look before you quote. BMJ 1986; 293:1000-2.

Eichorn 1987. Eichorn P, Yankauer A. Do authors check their references? A survey of accuracy of references in three public health journals. Am J Public Health 1987; 77:1011-2.

Flanagin 1998. Flanagin A, Carey LA, Fontarosa PB, Philips SG, Pace BP, Lundberg GD, Rennie D. Prevalence of articles with honorary articles and ghost authors in peer-reviewed medical journals. JAMA 1998; 280: 222-4.

ICMJE 1997. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. Canadian Medical Association Journal 1997; 156: 270-85.

Rennie 1997. Rennie D, Emanuel L, Yank V. When authorship fails: a proposal to make contributors accountable. JAMA 1997;278:579-85.

Rennie 1998. Rennie D, Yank V. If authors become contributors, everyone would gain, especially the reader. Amer J Public Health 1998;88:828-30.

Yank 1999. Yank V, Rennie D. Disclosure of researcher contributions: a study of original research articles in the Lancet. Annals of InternalMedicine 1999; 130: 661-70.

APPENDIX 2b. CODE OF CONDUCT FOR AVOIDING POTENTIAL FINANCIAL CONFLICTS OF INTEREST

The Cochrane Collaboration Code of Conduct for Avoiding Potential Financial Conflicts of Interest

1. General Principle

The essential activity of the Cochrane Collaboration is co-ordinating the preparation and maintenance of systematic reviews of the effects of healthcare interventions performed by individual reviewers according to procedures specified by the Collaboration. The performance of the review must be free of any real or perceived bias introduced by receipt of any benefit in cash or kind, any hospitality, or any subsidy derived from any source that may have or be perceived to have an interest in the outcome of the review. All entities that constitute the Cochrane Collaboration must accept this General Principle as condition of participation in the Collaboration.

2. Recommendations

2.1 Receipt of benefits from any source of sponsored research must be acknowledged and conflicts of interest must be disclosed in the *Cochrane Database of Systematic Reviews* and other publications that emanate from the Collaboration.

2.2 If a proposal raises a question of serious conflict of interest, this should be forwarded to the local Cochrane Centre for review (and the Steering Group notified accordingly). If the issue involves a Cochrane Centre, the issue should be referred to the Steering Group.

2.3 It is not mandatory to send funding proposals to the local Cochrane Centre or Steering Group prior to accepting them. However, such reviews would be desirable in cases of restricted donations, or any donation that appears to conflict with the General Principle.

2.4 The Steering Group should receive (and review at least annually) information about all external funds accepted by Cochrane entities. The Steering Group will use this information to prepare and distribute an annual report on the potential conflicts of interests attendant on the Collaboration's solicitation and use of external funds.

2.5 The Steering Group should constitute a subcommittee to view potential conflicts of interests, to offer recommendations for their resolution, and to consider appropriate sanctions to redress violations of the General Principle.

APPENDIX 3a. LOGISTICS OF DOING A REVIEW

3a.1 RESOURCES FOR A SYSTEMATIC REVIEW

Individual Cochrane Reviews are prepared by reviewers working in Collaborative Review Groups (CRGs). Each CRG has an editorial team responsible for producing a module of edited reviews for dissemination through the *Cochrane Database of Systematic Reviews*.

Because the Cochrane Collaboration is built around CRGs, it is important that each reviewer is linked with one. Besides ensuring that Cochrane Reviews are appropriately edited, this structure reduces the burden placed on individual reviewers since the editorial teams are responsible for providing most or all of the following types of support:

- conducting systematic searches for relevant studies and coordinating the distribution of potentially relevant studies to reviewers
- establishing specific standards and procedures for the CRG
- ensuring that reviewers receive the methodological support they need

The main resource required by reviewers is their own time. The majority of reviewers will contribute their time 'freely' because it will be viewed as part of their existing efforts to keep up-to-date in their areas of interest. In some cases, reviewers may need additional resources or, at least, be able to justify the amount of time required for a systematic review to colleagues who do not yet understand either what systematic reviews entail, or their importance.

The amount of time required will vary, depending on the topic, the number of studies, the methods used (e.g., the extent of efforts to obtain unpublished information), the experience of the reviewers, and the types of support provided by the editorial team. The workload associated with undertaking a review is thus very variable. However, consideration of the tasks involved and the time required for each of these might help a reviewer to estimate the amount of her or his time that will be required. These tasks include:

- Training
- Meetings
- Protocol development
- Searching for studies
- Assessing citations and full text reports of studies for inclusion in the review
- Assessing the quality of included studies and obtaining data
- Pursuing missing data and unpublished studies
- Analysing the data
- Interpreting the results and preparing a report
- Keeping the review up-to-date

Resources that might be required for these tasks, in addition to the reviewers' time, include:

- searching (identifying studies is primarily the responsibility of those involved in developing registers of studies, usually the editorial teams of the CRG. However, reviewers may share this responsibility and it may be appropriate to search additional databases for a specific review.)
- help for library work and photocopying
- a second reviewer, possibly a student or research assistant, to assess studies for inclusion, assess the quality of included studies, obtain data and conduct analyses
- statistical support for synthesizing (if appropriate) the results of the included studies
- equipment (e.g. computing hardware and software)
- supplies and services (long distance telephone charges, facsimiles, paper, printing, photocopying, audio-visual and computer supplies)
- office space for support staff
- travel funds

Many organisations currently provide funding for systematic reviews and additional agencies are likely to recognise the importance of supporting this type of work in the future. These include research funding agencies, those organisations that provide or fund healthcare services, those responsible for health technology assessment and those involved in the development of clinical practice guidelines. Although applications for funding need to adhere to the requirements of the funding organisation to which one is applying, a general outline of an application for funding for a systematic review should contain the following elements:

- Objectives
- Rationale
- Design of the review
- General approach
- Identification of studies
- Selection of studies for inclusion
- Assessments of the validity of included studies
- Obtaining data for the included studies
- Analysis
- Inferences and presentation of results
- Time-chart for major activities
- Budget

The objectives and design of a review are addressed in sections 3 to 8. Describing the rationale for a systematic review is analogous to describing the 'present state of knowledge' in a grant application for a primary study and may include a review of prior reviews on the same topic. The same scientific principles that apply to a review of studies can be applied to a review of reviews. The fundamental difference is the unit of analysis, which for a review of reviews is the review article.

Methodological issues that might need to be considered in reviewing reviews include how they will be identified, selected for detailed review and assessed (Sacks 1987, L'Abbe 1987, Oxman 1991). Reasons for undertaking a review of prior reviews, in addition to providing the rationale for an application for funding, include learning from what earlier reviewers have done, avoiding unnecessary duplication of efforts, identifying potentially relevant studies (including those that are unpublished), and collecting background

information that may be important in interpreting the results of individual studies and drawing conclusions from the results of the review.

A time chart with target dates for accomplishing key tasks can help with scheduling the time needed to complete a review. Such targets may vary widely from review to review. Reviewers, together with the editorial team for the CRG, must determine an appropriate time frame for a specific review. An example of a time chart with target dates is:

Month	
1 – 6	Additional searches for published and unpublished studies
1	Pilot test of inclusion criteria
1 – 6	Relevance assessments
1	Pilot test of validity criteria
1 – 8	Validity assessments
1	Pilot test of data collection
1 – 8	Data collection
1 – 8	Data entry
2 – 8	Missing information
6 – 8	Analysis
1 – 9	Preparation of report
10 -	Keeping the review up-to-date

3a.2 REGISTERING A PROTOCOL

Once a protocol has been completed it should be sent to the CRG editors to consider. When the editors are satisfied with the protocol they will include it in the CRG's module for incorporation in the Parent Database. Protocols are published and disseminated in *CDSR*. This will raise expectations and may discourage others from undertaking a review on the same topic. Editors and reviewers should not include a protocol in a module unless there is a firm commitment to complete the review within a reasonable time frame and to keep it up-to-date once it is completed.

3a.3 METHODS OF A COLLABORATIVE REVIEW GROUP

The editorial team of each CRG is responsible for documenting the methods used by the team for editing their module. The editorial team is also responsible for documenting any specific methods used by the CRG beyond the standard methods specified in the Handbook, including:

- methods used to review protocols
- any standard methodological criteria for including studies in reviews

- the search strategies used to develop and maintain the specialised register used by the CRG and method of distributing potentially relevant citations or full-text reports to reviewers
- any additional search strategies that reviewers are instructed to use routinely
- any standard methods used to select studies for reviews
- any standard criteria or methods used to assess the methodological quality of included studies
- any standard methods used for data collection
- any standard methods used for synthesising data
- any standard methods used for deriving conclusions or indicating the strength of the evidence on which the conclusions are based
- any decision rules used to categorise interventions (see section 9.6)
- any specific rules used for preparing the standard tables and figures
- the methods used to keep reviews up-to-date and respond to criticisms

Descriptions of specific methods used by each CRG are published as part of the Group's module in *The Cochrane Library*.

3a.4 REFERENCES

L'Abbe 1987. L'Abbe KA, Detsky AS, and O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987; 107: 24-33.

Oxman 1991. Oxman AD and Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidem 1991; 11:1271-8.

Sacks 1987. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450-5.

APPENDIX 5a. COCHRANE AND NATIONAL LIBRARY OF MEDICINE RANDOMIZED CONTROLLED TRIAL AND CONTROLLED CLINICAL TRIAL CRITERIA

5a.1 Cochrane criteria for randomized controlled trials (RCTs) and controlled clinical trials (CCTs)

Records identified for inclusion should meet the eligibility criteria devised and agreed in November 1992, which were first published, in 1994, in Section 5 of the Cochrane Reviewer's Handbook (previously called the Cochrane Handbook). According to these eligibility criteria:

A trial is eligible if, on the basis of the best available information (usually from one or more published reports), it is judged that:

- the individuals (or other units) followed in the trial were definitely or possibly assigned prospectively to one of two (or more) alternative forms of health care using
    - random allocation or
    - some quasi-random method of allocation (such as alternation, date of birth, or

case record number)

Trials eligible for inclusion are classified according to the reader's degree of certainty that random allocation was used to form the comparison groups in the trial. If the author(s) state explicitly (usually by some variant of the term 'random' to describe the allocation procedure used) that the groups compared in the trial were established by random allocation, then the trial is classified as an 'RCT' (randomized controlled trial). If the author(s) do not state explicitly that the trial was randomized, but randomization cannot be ruled out, the report is classified as a 'CCT' (controlled clinical trial). The classification 'CCT' is also applied to quasi-randomized studies, where the method of allocation is known but is not considered strictly random, and possibly quasi-randomized trials. Examples of quasi-random methods of assignment include alternation, date of birth, and medical record number.

The classification as RCT or CCT is based solely on what the author has written, not on the reader's interpretation; thus, it is not meant to reflect an assessment of the true nature or quality of the allocation procedure. For example, although double-blind trials are nearly always randomized, many trial reports fail to mention random allocation explicitly and should therefore be classified as 'CCT'.

Relevant reports are reports published in any year, of studies comparing at least two forms of health care (healthcare treatment, healthcare education, diagnostic tests or techniques, a preventive intervention, etc.) where the study is on either living humans or parts of their body or human parts that will be replaced in living humans (e.g., donor kidneys). Studies on cadavers, extracted teeth, cell lines, etc. are not relevant. *Searchers should identify all controlled trials meeting these criteria regardless of relevance to the entity with which they are affiliated.*

The highest possible proportion of all reports of controlled trials of health care should be included in CENTRAL. Thus, those searching the literature to identify trials should give reports the benefit of any doubts. Reviewers will decide whether to include a particular report in a review.

5a.2 National Library of Medicine definitions for Publication Type terms: RANDOMIZED CONTROLLED TRIAL, CONTROLLED CLINICAL TRIAL

RANDOMIZED CONTROLLED TRIAL:

A clinical trial that involves at least one test treatment and one control treatment, concurrent enrollment and follow-up of the test- and control-treated groups, and in which the treatments to be administered are selected by a random process, such as the use of a random numbers table. Treatment allocations using coin flips, odd-even numbers, patient social security numbers, days of the week, medical record numbers, or other such pseudo- or quasi-random processes, are not truly randomized and a trial employing any of these techniques for patient assignment is designated simply a CONTROLLED CLINICAL TRIAL.

CONTROLLED CLINICAL TRIAL:

A clinical trial involving one or more test treatments, at least one control treatment, specified outcome measures for evaluating the studied intervention, and [an intended to be bias-free] method of assigning patients to the test treatment. The treatment may be drugs, devices, or procedures studied for diagnostic, therapeutic, or prophylactic effectiveness. Control measures include placebos, active medicine, no-treatment, dosage forms and regimens, historical comparisons, etc. When randomization using mathematical techniques, such as the use of a random numbers table, is employed to assign patients to test or control treatments, the trial is characterized as a RANDOMIZED CONTROLLED TRIAL. However, trials employing treatment allocation methods such as coin flips, odd-even numbers, patient social security numbers, days of the week, medical record numbers, or other such pseudo- or quasi-random processes are simply designated as controlled clinical trials.

APPENDIX 5B: MEDLINE HIGHLY SENSITIVE SEARCH STRATEGY FOR B.1) SILVERPLATTER-MEDLINE, B.2) OVID-MEDLINE, AND B.3) PUBMED.

Upper case denotes controlled vocabulary. Lower case denotes free-text terms. Those wishing to run this search strategy are recommended to seek the advice of a trained medical librarian.

5b.1 Format for SilverPlatter version 3.10:

phase 1:
#1 RANDOMIZED-CONTROLLED-TRIAL in PT
#2 CONTROLLED-CLINICAL-TRIAL in PT
#3 RANDOMIZED-CONTROLLED-TRIALS
#4 RANDOM-ALLOCATION
#5 DOUBLE-BLIND-METHOD
#6 SINGLE-BLIND-METHOD
#7 #1 or #2 or #3 or #4 or #5
#8 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#9 #6 not #7

phase 2:
#10 CLINICAL-TRIAL in PT
#11 explode CLINICAL-TRIALS
#12 (clin* near trial*) in TI
#13 (clin* near trial*) in AB
#14 (singl* or doubl* or trebl* or tripl*) near (blind* or mask*)
#15 (#13 in TI) or (#13 in AB)
#16 PLACEBOS
#17 placebo* in TI
#18 placebo* in AB
#19 random* in TI
#20 random* in AB
#21 RESEARCH-DESIGN

#22 #9 or #10 or #11 or #12 or #14 or #15 or #16 or #17 or #18 or #19 or #20
#23 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#24 #21 not #22
#25 #23 not #8

phase 3:
#26 TG=COMPARATIVE-STUDY
#27 explode EVALUATION-STUDIES
#28 FOLLOW-UP-STUDIES
#29 PROSPECTIVE-STUDIES
#30 control* or prospectiv* or volunteer*
#31 (#29 in TI) or (#29 in AB)
#32 #25 or #26 or #27 or #28 or #30
#33 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#34 #31 not #32
#35 #33 not (#8 or #24)

#36 #9 or #25 or #35

5b.2 Format for OVID:

phase 1:
1 RANDOMIZED CONTROLLED TRIAL.pt.
2 CONTROLLED CLINICAL TRIAL.pt.
3 RANDOMIZED CONTROLLED TRIALS.sh.
4 RANDOM ALLOCATION.sh.
5 DOUBLE BLIND METHOD.sh.
6 SINGLE BLIND METHOD.sh.
7 or/1-6
8 ANIMAL.sh. not HUMAN.sh.
9 7 not 8

phase 2:
10 CLINICAL TRIAL.pt.
11 exp CLINICAL TRIALS.sh.
12 (clin$ adj25 trial$).ti,ab.
13 ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab.
14 PLACEBOS.sh.
15 placebo$.ti,ab.
16 random$.ti,ab.
17 RESEARCH DESIGN.sh.
18 or/10-17
19 18 not 8
20 19 not 9

phase 3:
21 COMPARATIVE STUDY.sh.
22 exp EVALUATION STUDIES.sh.
23 FOLLOW UP STUDIES.sh.

24 PROSPECTIVE STUDIES.sh.
25 (control$ or prospectiv$ or volunteer$).ti,ab.
26 or/21-25
27 26 not 8
28 27 not (9 or 20)

29 9 or 20 or 28

5b.3 Format for PubMed:

Phase 1
(randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double-blind method [mh] OR single-blind method [mh]) NOT (animal [mh] NOT human [mh])

Phases 1 and 2
(randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double-blind method [mh] OR single-blind method [mh] OR clinical trial [pt] OR clinical trials [mh] OR ("clinical trial" [tw]) OR ((singl* [tw] OR doubl* [tw] OR trebl* [tw] OR tripl* [tw]) AND (mask* [tw] OR blind* [tw])) OR (placebos [mh] OR placebo* [tw] OR random* [tw] OR research design [mh:noexp]) NOT (animal [mh] NOT human [mh])

All Phases
(randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double-blind method [mh] OR single-blind method [mh] OR clinical trial [pt] OR clinical trials [mh] OR ("clinical trial" [tw]) OR ((singl* [tw] OR doubl* [tw] OR trebl* [tw] OR tripl* [tw]) AND (mask* [tw] OR blind* [tw])) OR (-placebos [mh] OR placebo* [tw] OR random* [tw] OR research design [mh:noexp] OR comparative study [mh] OR evaluation studies [mh] OR follow-up studies [mh] OR prospective studies [mh] OR control* [tw] OR prospectiv* [tw] OR volunteer* [tw]) NOT (animal [mh] NOT human [mh])

Note: Subject specific terms (MeSH and textwords) should be ORed together, enclosed within parentheses, then ANDed with the appropriate version of the Cochrane highly sensitive search strategy.

APPENDIX 5C. EXAMPLE OF A SEARCH STRATEGY FOR ELECTRONIC DATABASES

(from the following Cochrane Review: Wilkinson C. Interventions for asymptomatic retinal breaks and lattice degeneration for preventing retinal detachment (Cochrane Review). In: *The Cochrane Library*, Issue 1, 2003. Oxford: Update Software.)

Search strategy for identification of studies

See: Collaborative Review Group search strategy

Trials were identified by electronic searches of the Cochrane Controlled Trials Register - CENTRAL (which includes the Cochrane Eyes and Vision Group specialized register), MEDLINE and EMBASE.

The following strategy was used to search CENTRAL Issue 1 2001 [search conducted January 5, 2001]:

#1 RETINAL-DETACHMENT:ME
#2 (RETINA* near ((((DETACH* or BREAK*) or PERFORATION*) or TEAR*) or HOLE*))
#3 (LATTICE near DEGENERAT*)
#4 RETINAL-PERFORATIONS:ME
#5 ((VITREO* near DETACH*) and POSTERIOR)
#6 ((VITREORETINAL or VITREO-RETINAL) near DEGENERAT*)
#7 (((((#1 or #2) or #3) or #4) or #5) or #6)
#8 LASER-COAGULATION*:ME
#9 LIGHT-COAGULATION:ME
#10 CRYOTHERAPY*1:ME
#11 ((LASER or LIGHT) near COAGULAT*)
#12 (LASER near PHOTOCOAGULAT*)
#13 CRYOPTHERAP*
#14 (((((#8 or #9) or #10) or #11) or #12) or #13)
#15 PROPHYLA*
#16 (#7 and (#14 or #15))

The following strategy was used to search MEDLINE to December 2000 [search conducted January 5, 2001]:

SilverPlatterASCII 3.0DOSN
#1 "RETINAL-DETACHMENT"/ all subheadings
#2 "RETINAL-PERFORATIONS"/ all subheadings
#3 "VITREOUS-DETACHMENT"/ all subheadings
#4 RETINA* near (DETACH* or BREAK* or PERFORATION* or TEAR* or HOLE*)
#5 (LATTICE near DEGENERAT*)
#6 VITREO?RETINAL next DEGENERAT*
#7(VITREO* near DETACH*) and POSTERIOR
#8 (#4 or #5 or #6 or #7) in TI,AB
#9 #1 or #2 or #3 or #8
#10 explode "LIGHT-COAGULATION"/ all subheadings
#11 explode "CRYOTHERAPY"/ all subheadings
#12(LASER or LIGHT) near COAGULAT*
#13 LASER near PHOTOCOAGULAT*
#14 CRYOTHERAP*
#15 (#12 or #13 or #14) in TI,AB
#16 #10 or #11 or #15
#17 PROPHYLA* in TI,AB
#18 #9 and (#16 or #17)

To identify randomized controlled trials, this search was combined with the Cochrane Highly Sensitive Search Strategy phases one and two as contained in the Cochrane Reviewer's Handbook (Clarke 2000).

The following strategy was used to search EMBASE to February 2001 [search conducted February 2, 2001]:

SilverPlatterASCII 3.0DOSN
#1 explode "RETINA-DETACHMENT"/ all subheadings
#2 "VITREOUS-BODY-DETACHMENT"/ all subheadings
#3 "VITREORETINAL-DEGENERATION"/ all subheadings
#4 RETINA* near (DETACH* or BREAK* or PERFORATION* or TEAR* or HOLE*)
#5 (LATTICE near DEGENERAT*)
#6 VITREO?RETINAL near DEGENERAT*
#7 (VITREO* near DETACH*) and POSTERIOR
#8 #4 or #5 or #6 or #7
#9 #1 or #2 or #3 or #8
#10 explode "LASER-COAGULATION"/ all subheadings
#11 "CRYOTHERAPY"/ all subheadings
#12 (LASER or LIGHT) near COAGULAT*
#13 LASER near PHOTOCOAGULAT*
#14 CRYOTHERAP*
#15 (#12 or #13 or #14) in TI,AB
#16 #10 or #11 or #15
#17 "PROPHYLAXIS"/ all subheadings
#18 PROPHYLA* in TI,AB
#19 #9 and (#16 or #17 or #18)

To identify randomized controlled trials, this search was combined with the following search:

SilverPlatterASCII 3.0DOSNEMBASE (R) 1998/07-1998/12
#1 "RANDOMIZED-CONTROLLED-TRIAL"/ all subheadings
#2 "RANDOMIZATION"/ all subheadings
#3 "CONTROLLED-STUDY"/ all subheadings
#4 "MULTICENTER-STUDY"/ all subheadings
#5 "PHASE-3-CLINICAL-TRIAL"/ all subheadings
#6 "PHASE-4-CLINICAL-TRIAL"/ all subheadings
#7 "DOUBLE-BLIND-PROCEDURE"/ all subheadings
#8 "SINGLE-BLIND-PROCEDURE"/ all subheadings
#9 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
#10 (RANDOM* or CROSS?OVER* or FACTORIAL* or PLACEBO* or VOLUNTEER*) in TI,AB
#11 (SINGL* or DOUBL* or TREBL* or TRIPL*) near (BLIND* or MASK*) in TI,AB
#12 #9 or #10 or #11
#13 HUMAN in DER
#14 (ANIMAL or NONHUMAN) in DER
#15 #13 and #14
#16 #14 not #15
#17 #12 not #16

APPENDIX 6. REVIEWS INCLUDING NON-RANDOMISED STUDIES

1 Rationale

The Cochrane Collaboration builds on ten principles, two of which are to minimise bias and to ensure relevance. In order to minimise bias, reviewers may choose to include only randomised controlled trials (RCTs) in their reviews. While this approach minimises bias it may not always ensure relevance. The challenge facing reviewers is this: How far is it possible to achieve a higher level of relevance by including evidence other than that derived from RCTs without violating the central principle: minimising bias?

2 What might be the advantages and dangers of including non-randomised studies in systematic reviews?

If a systematic review relies solely on data from randomised trials, it is open to a number of problems. The most obvious of these is that certain important health care problems have not been studied, or are impossible or very difficult to study in randomised trials. But randomised trials may be inadequate for other reasons also. For example, there may be insufficient information on the types of participant or outcome which are of relevance to the review (e.g. rare side effects), or the data may only contain short term follow-up when important findings depends on longer follow-up. Inclusion of evidence from non-randomised studies may resolve some of these problems, but it also poses problems and threats to validity as unexpected biases may creep in and invalidate the conclusions.

Some examples already exist where inclusion of non-randomised evidence in systematic reviews have been helpful. For example the possible causal relationship between prone sleeping position and cot death which was strongly supported by meta-analyses of observational studies (Beal 1991) was subsequently corroborated by national intervention programmes leading to a reduced rate for cot deaths (Wennergren 1997). A recent example of the opposite might be the many systematic reviews of observational studies of hormone replacement therapy in postmenopausal women showing a dramatic and highly significant decrease in mortality but contradicted by an ensuing large randomised trial showing no significant difference with a fairly narrow confidence interval (Petitti 1998).

Several empirical studies of the possible biases in non-randomised studies have been published recently (Britton 1998; Reeves 1998; Kunz 1998; Benson 2000; Concato 2000). The foci, the quality assessments and the conclusions of these studies vary and have led to some confusion and discussion. High quality research projects with prespecificied protocols are needed.

3 Guidelines for inclusion of non-randomised studies in Cochrane reviews

The Cochrane Non-Randomised Studies Methods Group (NRSMG) was registered in November 1999 and is currently developing guidelines for the inclusion of non-randomised studies in Cochrane reviews. The following guideline chapters are planned and under development:

1. Introduction
2. Types of study design
2.1. Scope and terminology of the NRSMG guidelines
2.2. What types of study designs should be included in a Cochrane review?
2.3. What types of research questions are expected to benefit from the inclusion of non-randomised evidence?
3. Searching for non-randomised studies
4. Quality assessment
5. Data extraction
6. Analysis
7. Interpretation

The draft chapters will be made available at www.cochrane.dk/nrsmg/ as they reach a useable form (during 2000 and 2001). The chapters will be approved by the NRSMG as they reach their final form. The full set of guidelines is expected to be ready by the end of 2001.

4 Further information

This appendix was prepared by Ole Olsen on behalf of the Cochrane Non-Randomised Studies Methods Group. Further information can be found in the NRSMG module in The Cochrane Library or at www.cochrane.dk/nrsmg/.

5 References

Beal 1991. Beal SM, Finch CF. An overview of retrospective case-control studies investigating the relationship between prone sleeping position and SIDS. J Paediatr Child Health 1991;27(6):334-9.

Benson 2000. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000;342:1878-86.

Britton 1998. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. Health Technol Assessment 1998;2(13).

Concato 2000. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000;342:1887-92.

Kunz 1998. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 1998; 317:1185-90.

Petitti 1998. Petitti DB. Hormone replacement therapy and heart disease prevention: experimentation trumps observation. JAMA 1998;280:650-2.

Reeves 1998. Reeves BC, MacLehose RR, IM Harvey, TA Sheldon, IT Russell, AMS Black. Comparisons of effect size estimates derived from randomised and non-randomised studies. In: Black N, et al, editors. Health Services Research Methods: a guide to best practice. London: BMJ Publishing Group, 1998; 73-85.

Wennergren 1997. Wennergren G, Alm B, Oyen N, Helweg-Larsen K, Milerad J, Skjaerven R, Norvenius SG, Lagercrantz H, Wennborg M, Daltveit AK, Markestad T, Irgens LM. The decline in the incidence of SIDS in Scandinavia and its relation to risk-intervention campaigns. Nordic Epidemiological SIDS Study. Acta Paediatr 1997;86:963-8.

APPENDIX 8a. EFFECT MEASURES FOR DICHOTOMOUS DATA

Outcome

	yes	no
Treatment	A	B
Control	C	D

Odds ratio = (A/B)/(C/D)
Relative risk (RR) = [A/(A+B)]/[C/(C+D)]
Relative risk reduction = 1 - RR
Risk difference (or Absolute risk reduction, ARR) (RD) = A/(A+B) - C/(C+D)

When the outcome is undesirable, a relative risk or odds ratio less than one represents a beneficial treatment (zero would represent 100% effectiveness). A risk difference (or absolute risk reduction) of less than zero represents a benefit (a risk difference that is equal to the risk in the control group represents 100% effectiveness).

The odds ratio can also be expressed as (A/C)/(B/D) (i.e. the odds of a case having been exposed relative to the odds of a control having been exposed). This and the expression given above are equivalent to A×D/B×C. Thus, if A is small relative to B and C is small relative to D, the odds ratio and the relative risk are approximately the same.

APPENDIX 8b. CALCULATING THE NUMBER NEEDED TO TREAT (NNT)

NNTs are a useful way to re-express the results of a study but some caution is needed when they are used in reviews. NNTs are specific to a particular length of follow-up since they are based on the number of people who will benefit within a certain period of time who otherwise would not benefit. Systematic reviews tend to combine trials of varying follow-up periods, which could make an NNT difficult to interpret (Smeeth 1999). NNTs should only be calculated when the follow-up periods are similar.

When summarising results, the 'control event rate' (the rate of events in the control group) can be substituted for the 'patient expected event rate' (the baseline risk). In practice, individual patients' expected event rate might differ importantly from the control event rate in the studies in a review.

The following abbreviations are used in this appendix:

CER = control event rate
EER = experimental event rate
PEER = patient expected event rate
NNT = Number needed to treat
RD = risk difference (or absolute risk reduction, ARR)
RR = relative risk
RRR = relative risk reduction
OR = odds ratio

Then:

RD = CER – EER
RR = EER/CER
RRR = RD/CER = 1 – RR

The RRR can be calculated from the OR using

$$RRR = CER - \frac{OR \times CER/(1 + CER)}{[OR \times CER/(1 + CER)]}$$

The NNT can then be calculated with either
NNT = 1/RD
NNT = 1/(CER - RR x CER)
NNT = 1/(RRR x CER)

If the CER is very small, say less than 5%, the OR is approximately equal to the RR and the RRR is approximately equal to (1 – OR). However, as the CER (or PEER) increases, the difference between the OR and the RR increases.

If the average CER across studies is used in the above formulae, the NNT will be for the average baseline risk observed across the included studies. Since the PEER (baseline risk) often varies across studies and is likely to vary across patient groups, it is general important to specify the baseline risk for which an NNT is reported and to report NNTs for a range of PEERs. For example, the range of CERs in the included studies can be used, giving NNTs based on the lowest, the average and the highest of these. However, this assumes that the RRR is the same for different baseline risks. Although this assumption is often correct, it is not always (Sharp 1996, Ioannidis 1997, Smith 1997, Thompson 1997, Smeeth 1999).

Confidence limits for NNTs should be calculated by using the upper and lower confidence limits for the summary statistic that is used to calculate the NNT (RR, OR or RD). For further discussion about NNTs and their calculation see (Sackett 1996, Senn 1998, Altman 1998).

References

Altman 1998. Altman DG. Confidence intervals for the number needed to treat. BMJ. 1998; 317:1309-12.

Ioannidis 1997. Ioannidis JP, Lau J. The impact of high-risk patients on the results of clinical trials. J Clin Epidemiol 1997; 50: 1089-98.

Sackett 1996. Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! Evidence Based Medicine, 1996 Sept-Oct; 1:164.

Senn 1998. Senn S, Walter S, Olkin I, Altman D, Deeks J, Sackett DL. Odds ratios revisited. Evidence-Based Medicine. 1998 May-June;71.

Smeeth 1999. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses – sometimes informative, usually misleading. Brit Med J 1999;318:1548-51.

Sharp 1996. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. BMJ 1996;313:735-8.

Smith 1997. Smith GD, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? BMJ 1997;315: 1610-4.

Thompson 1997. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. Stat Med 1997; 16: 2741-58.

APPENDIX 9. INCORPORATING ECONOMIC EVALUATION INTO THE COCHRANE REVIEW PROCESS

The Cochrane Collaboration's main role of 'preparing, maintaining and making accessible reviews of the effects of healthcare' is motivated by an underlying aim to help people make decisions about healthcare. However, in the face of limited resources, decision-makers need to consider further evidence when deciding how to act on the evidence from Cochrane reviews. Nearly every healthcare decision has an impact, not only on health and social welfare, but also on the use of resources. Therefore, to make the best decisions about alternative interventions, information is needed on resource use and costs as well as health effects.

The process of incorporating economic evaluation into Cochrane Reviews is not straight forward. As with many areas of scientific inquiry, the methodology is still developing. A particular challenge in the context of Cochrane Reviews is ensuring that economic information and analyses contained in reviews is relevant to people working in widely varying circumstances. For those who are considering addressing economic questions as part of their review, or along side of a Review, advice can be found in the module for the Cochrane Health Economics Methods Group in *The Cochrane Library*.

APPENDIX 11A. PRACTICAL METHODOLOGY OF META-ANALYSES USING UPDATED INDIVIDUAL PATIENT DATA

11a.1 Front page

PRACTICAL METHODOLOGY OF META-ANALYSES (OVERVIEWS) USING UPDATED INDIVIDUAL PATIENT DATA

LESLEY A. STEWART
MRC Cancer Trials Office, 5 Shaftesbury Road, Cambridge CB2 2BW, U.K.

AND

MICHAEL J. CLARKE
University of Oxford, Clinical Trial Service Unit and ICRF Cancer Studies, Radcliffe Infirmary, Oxford OX2 6HE, U.K.

on behalf of the
COCHRANE WORKING GROUP ON META-ANALYSIS USING INDIVIDUAL PATIENT DATA

(Originally published in Statistics in Medicine, Vol. 14, 2057-2079, 1995)

11a.2 Further information

For further information on the Cochrane Working Group on meta-analysis using individual patient data, please contact one of the authors:

Lesley A Stewart
MRC Cancer Trials Office
5 Shaftesbury Road
Cambridge CB2 2BW
UK
Phone: +44-1223-311110
Fax: +44-1865-58817
e-mail: LS @ cto.mrc.ac.uk

Michael J Clarke
UK Cochrane Centre
Summertown Pavilion
Middle Way
Oxford OX2 7LG
United Kingdom
Phone: +44-1865-516300
Fax: +44-1865-516311
e-mail: mclarke@cochrane.co.uk

11a.3 Workshop participation

Doug Altman, Colin Baigent, Marc Buyse, Iain Chalmers, Mike Clarke, Rory Collins, Carl Counsell, Jack Cuzick, Rob Edwards, Tricia Elphinstone, Vaughan Evans, Richard Gray, Liz Greaves, Francois Gueyffier, Heather Halls, Rob Henderson, Jini Hetherington, Sally Hunt, Peter Langhorne, Carol Lefebvre, David Machin, Silvia Marsoni, Veronique Mosseri, Lennarth Nyström, Mandy Ogier, Andy Oxman, Max Parmar, Richard Peto, Jean-Pierre Pignon, Sue Richards, Carmen Ruiz, Paul Seed, Michael Sextro, Lena Specht, Sally Stenning, Lesley Stewart, Annet te Velde, Jayne Tierney, Harm van Tinteren, Valter Torri, Paul Weston, Keith Wheatley, Chris Williams.

11a.4 Summary

Meta-analyses using updated individual patient data may provide the most reliable means of combining data from similar randomised controlled trials and the benefits of this approach to systematic review are described. Guidance, based on the experience of several groups who have undertaken such projects is given. This includes practical advice on initiating and maintaining collaboration, the time and resource required to undertake these usually international projects and methods of data checking and validation. Example proforma are included.

11a.5 Introduction

Systematic reviews using meta-analysis to combine the results of related randomised controlled trials are increasingly common, and the number of associated publications has

mushroomed. Although there is a burgeoning literature on the statistical methods of meta-analysis, less has been published on the practical methods of carrying out such projects. These can include calculations based solely on information presented in a few published papers, more detailed analysis of aggregate data supplied by individual trialists, and time-to-event analysis of thoroughly checked and updated individual patient data. The last of these has been described as the 'yardstick' against which all systematic reviews should be measured (1), and current limited empirical evidence shows that meta-analyses which rely solely on data extracted from published reports can give estimates of treatment effects, and of their significance, which are not confirmed when all of the relevant evidence is analysed (2 , 3 , 4). Given that the central collection, checking and analysis of individual patient data from all relevant trials can require a considerable amount of time, personnel and financial resource, further research is needed to determine when it is most appropriate to adopt this approach and what the most appropriate alternatives are if sufficient resources are not available. Irrespective of this, the additional benefits of meta-analyses based on individual patient data (IPD) when compared with meta-analyses based on published aggregate date include the ability to:

- Undertake survival and other time-to-event analyses
- Undertake subgroup analyses for important hypotheses about differences in effect
- Carry out detailed data checking and ensure the quality of randomisation and follow-up
- Ensure the appropriateness of analyses
- Update follow-up information

Further, as IPD meta-analyses require the collaboration of the investigators who conducted the trials, other benefits (which may also be found if trialists are approached for aggregate data) may include:

- More complete identification of relevant trials
- Better compliance with providing missing data
- More balanced interpretation of the results
- Wider endorsement and dissemination of the results
- Better clarification of further research
- Collaboration on further research

This paper provides guidance on the conduct of IPD meta-analyses, which aim to collect data on each randomised patient entered in all randomised trials addressing a particular question. The patient data are checked, collated and analysed centrally by a secretariat. Subsequent publication is generally made by the collaborative group of trialists, often following a meeting of this group at which the results and their implications are discussed. Until now, almost no information on either the techniques or the resources needed for such a project has been readily available. Thus each of the groups who have undertaken them has generally had to develop their own means of data collection, checking and analysis.

In the hope that this situation could be improved, a workshop (under the auspices of the Cochrane Collaboration) was convened in April 1994 to discuss the practicalities of meta-analyses based on individual patient data. This was attended by nearly 40 participants (Appendix A), all of whom had been involved in the planning or conduct of this form of meta-analysis. The aim was to discuss all practical aspects of such projects; to identify

areas of agreement and disagreement on the methods used; and to prepare published guidance available to anyone contemplating using this technique in a systematic review. Participants did not discuss whether or not meta-analyses using individual patient data are indeed a 'gold standard' or statistical methodology.

11a.5 Running a meta-analysis based on individual patient data

The steps involved in a meta-analysis of individual patient data are shown in figure 1 along with some very approximate guidance on the time required for these. The majority of effort is required to plan, initiate, set up and manage the study and, although much has been written about the statistical methodology of meta-analysis, this can often represent the least time consuming and difficult aspect of the project. Nurturing collaboration and careful checking of incoming data generally consume much more time and resource, since the ultimate aim is to obtain accurate, up to date and complete data from all patients in all relevant randomised trials.

Figure 1. Stages of an Individual Patient Based Systematic Review

NB: All estimates of time are necessarily very approximate and will depend on the size of the meta-analysis and the complexity of the data requested

(1) Development

Identify need for IPD meta-analysis

Devise questions

Identify trials (continues throughout project)

Refine questions

Meta-analysis of published data (if appropriate)

Write Protocol

Initial contact with trialists

Typically requires approximately 3-6 months minimum (3-4 person months minimum effort)

(2) Data Collection and Checking

Assess feasibility

Set up database

Request data

Check data

Analyse trials individually

Finalise data

Requires approximately one year (15 person months for 50 trials, 4-5 person months for 5 trials)

(3) Analysis and Dissemination of Results

Analyse data

Present results to trialists

Discuss results and implications with trialists

Draft manuscript

Requires approximately 6-9 months (10 -12 person months for 50 trials, 5-6 person months for 5 trials)

(4) Future Projects

Future updates

New projects

- extend scope of meta-analysis

- initiate new trials

The total time required for the meta-analysis is approximately 2 - 3 years (approximately 30 person months for a meta-analysis of 50 trials and 15 person months for 5 trials).

11a.6 Resource requirements

It is perhaps not generally appreciated just how much time and effort is involved in performing an IPD meta-analysis. It is not something to be undertaken lightly, and since a variety of clinical, scientific, statistical, computing and data management skills are required, it is generally not something to be undertaken by a single individual. Of necessity, projects usually take a few years from initiation to first publication. Although some of this time can be saved by involving more personnel, the project duration will be constrained by the time taken to secure the full involvement of the collaborating trialists. This collaboration is the main way of ensuring that the data to be analysed are as complete, accurate and reliable as possible.

Financial

Based on estimates provided by those attending the workshop, the average cost of running an IPD meta-analysis was approximately £1,000 per trial or £5-£10 per patient (£ Sterling, 1994), whichever was the less. However, these estimates, which did not include the costs associated with a Collaborative Group Meeting, were very approximate and retrospective and varied greatly depending on the size and complexity of the project. In addition, most estimates did not include the hidden costs associated with administration. Interestingly, those meta-analyses funded by direct grants, where presumably a more detailed record of costs was required, were considerably more expensive. Previous projects have been financed by both core and grant-based funding. The first cycle of project initiation, data collection and analysis is well suited to one-off grant applications because of its structure and timescale, although many IPD meta-analyses will require subsequent updating which may at first seem less attractive to some funders. Funding could be sought from a variety of sources: Government bodies, research organisations, charities and industry.

Staff
Most of the estimated costs were associated with staff, typically representing around 80% of the total budget. As discussed above, a range of skills are required and the involvement of the various personnel will vary over time. It is therefore usual for some groups co-ordinating IPD meta-analyses to be simultaneously involved in several projects, scheduling them so that the workload of the clinical, scientific, statistical, computing, data management, administrative and secretarial staff is evenly distributed.

Time
Figure 1 includes very approximate estimates of the minimum time required to complete the various stages of a meta-analysis, both on an absolute time scale and in terms of person months. It should be noted, though, that the actual time taken may vary considerably depending on the circumstances of each project. In most circumstances it is unlikely that an IPD meta-analysis could reach first publication in much less than three years.

11a.7 Planning the meta-analysis

As with a clinical trial, a good deal of planning and organisation is required before a meta-analysis can be launched and trialists are asked to provide data. After the identification of a suitable question, the first step is to identify all relevant randomised trials and to plan the conduct of the meta-analysis. In most cases this will involve developing a protocol or written plan of the proposed investigation. A good deal of resource is involved in this pre-data collection planning stage, which may take several months. There is therefore a potential problem in that several groups may independently embark upon the same or similar projects, representing both a duplication of effort and an annoyance to the trialist who receives multiple requests for the same data. One way to help avoid this is through the prospective registration of these meta-analyses with the Cochrane Collaboration, in the same way that systematic reviews using other techniques can be registered.

Establishing a Secretariat
At the earliest stages of the meta-analysis a secretariat to co-ordinate the project should be established. It is likely that this will consist of the scientific, statistical and data management staff who will do most of the work on the project, and also appropriate clinical experts. A larger Steering Group may also be formed to advise the secretariat on

strategic issues and analyses. This is likely to be made up of members of the secretariat, trialists and independent experts.

Methods of Identifying Trials

It is of the utmost importance that as high a proportion as possible of all relevant trials are identified, regardless of their results or publication status. Any trials that are missing should not be too numerous or unrepresentative to affect the results of the meta-analysis in any important way. This is true of any systematic review, irrespective of the analytical methods to be adopted, and searching for trials should continue throughout the duration of the project.

The first step towards identifying trials is usually to perform a computerised bibliographic search. However, such searches may miss a significant proportion of published trials. For example, it has been shown that electronic searching for randomised clinical trials using the US National Library of Medicine's database MEDLINE, might yield only around half of the relevant studies that are actually contained in the database (5). Further, MEDLINE indexes only 3,700 out of around 16,000 medical journals published worldwide (5). The coding of articles within MEDLINE is currently being revised to improve the retrieval of future RCTs and the Cochrane Collaboration is working with the National Library of Medicine (NLM) in the retrospective tagging of all previously published randomised trials. Other databases, for example, CancerLit, Current Contents, Excerpta Medica, The Index of Scientific and Technical Proceedings, Dissertation Abstracts and the Index to UK Theses, may be useful additions or alternatives to MEDLINE, but further research is required to determine which are most efficient in the various areas of medicine.

In order to make full use of the current computerised databases, it is important that efficient search strategies are used. An inexperienced searcher should seek as much help as possible. Optimal strategies for searching MEDLINE are currently under development (5) and these should be adopted as part of any systematic review. The latest version is shown in Appendix B. This strategy does not include subject specific searching so that individual searchers will need to add further steps, for example, adding terms such as the disease and therapy in question.

At present, problems will remain even with the best computer search strategy. Some relevant articles in the databases will be missed because of lack of clarity in the published reports or indexing errors, and the majority of medical journals are not covered by any literature database. Until all published randomised trials are accessible through MEDLINE, it is essential that electronic searches are supplemented by some hand-searching. This will need to include those journals that are most likely to contain relevant reports which cannot be identified in the existing databases, and also those meeting abstracts which are not available in any electronic form.

This aspect of any meta-analysis can be both time-consuming and labour intensive. Even a refined literature search strategy is likely to yield many more articles than will eventually prove relevant to the meta-analysis. A fair number of the unnecessary articles will have to be obtained as full papers in order to determine whether or not they are relevant. In addition, the thorough handsearching of journals and meeting abstract books requires a substantial amount of care, time and effort. The Cochrane Collaboration is attempting to coordinate such searching and it would be worthwhile for anyone planning

to do such a search to communicate first with the Collaboration to avoid duplication of effort.

An additional problem is that trials with positive results are more likely to be published than those with negative or inconclusive results (6 , 7 , 8, 9), thus skewing the published literature in favour of the positive. It is therefore extremely important that, whenever possible, unpublished trials are sought and included in meta-analyses (especially where the results of a trial might have influenced the decision on whether it would be published). Although data from unpublished trials have not been subject to peer review, obtaining the trial protocol and individual patient data enables thorough checking both of the data supplied and the trial design, allowing, in fact, a much more detailed review than is generally possible prior to the publication of a trial. Moreover, even if a trial has been published in a prestigious journal, this cannot be taken as guarantee of the quality of the actual data. All trials, both published and unpublished, should be subject to the same degree of careful checking prior to inclusion in an IPD meta-analysis.

The main reason for non-publication of a trial is failure by the authors to prepare a report (6 , 7 , 8), and these trials are usually small single institution studies. Finding such trials can therefore be difficult. Trial registers, which prospectively register trials at inception, are the best solution to this problem (10) and the conduct of all systematic reviews should be much simplified when the use of such registers becomes widespread (11). However, while it is to be hoped that increasing numbers of new trials will be registered, many existing trials will still not be included and the identification of these trials will continue to be a major part of most meta-analyses. As the collaborative group is likely to consist of international experts with a good knowledge of potentially relevant or otherwise unidentified trials, the direct contact with trialists that is an integral part of a meta-analysis based on individual patient data can be a rich source of information. In addition, the circulation of a list of all identified trials at appropriate clinical meetings may bring to light trials, as well as trialists, previously unknown to the secretariat and collaborative group. Other potential sources of information include pharmaceutical companies and regulatory authorities.

Developing a written plan or protocol

As with any formal research, some form of written plan or protocol should be produced for the meta-analysis. Examples of formats that have been used successfully in previous projects include a two page summary sheet and a longer document similar to the protocol for a clinical trial.

Table I. Possible items to include in a written plan or protocol for an individual patient data based meta-analysis

RATIONALE
Underlying biology
Review of trials
Preliminary meta-analysis

OBJECTIVES
Inclusion or eligibility criteria
Search strategies
Data to be collected
Brief description of data checking procedures

Main analyses to be performed
Publication policy
Suggested timetable for the meta-analysis
Provisional list of trials to be included

Table 1 shows some of the items that might be considered for inclusion in such a document. As a minimum, trialists being asked to participate in the project should be provided with some guidance on the proposed analyses along with a statement on publication policy and the confidentiality of data. The most difficult item is perhaps the inclusion of a meta-analysis based on data other than individual patient data which may have been performed as part of the planning stage of the IPD meta-analysis, as this may give the impression to some potential collaborators that the review has already been done and that they need not go to the trouble of supplying individual patient data. It is important, therefore, that if a preliminary meta-analysis is included it is accompanied by a suitable explanation of why it is not felt to be adequate and why individual patient data are being sought. For example, if the meta-analysis simply relied on data that could be easily abstracted from publications which had been identified by an inadequate MEDLINE search, it should be noted that such an analysis might be biased by a failure to include trials whose data could not be abstracted from the identified publications, published trials which were not found in the MEDLINE search, and trials which had not been published. In such a case it should be noted that, as well as helping to rectify these potential problems, collecting IPD allows the published data to be updated. The reasons for requesting individual, rather than aggregate, data should also be given. If the IPD meta-analysis has been preceded by a thorough meta-analyses using aggregate data then just this information, to indicate why individual patient data was now felt to be necessary, would be required (12).

Developing a written plan or protocol makes setting up a meta-analysis more rigorous by helping to identify problems and clarify issues early in the project. Specifying inclusion criteria means that trials can be evaluated for suitability at an early stage. Although there may be a temptation to request data from all trials at the outset of the meta-analysis, a more measured approach makes it less likely that trials will have to be withdrawn or excluded after the trialists have started to prepare and provide data. Time spent at this stage more than makes up for itself later, although it does mean that initiating collaboration may be delayed.

11a.8 Initiating collaboration

Having decided on the therapeutic questions to be addressed, identified the relevant trials and done the appropriate planning, the trialists need to be contacted and persuaded to participate. Generally this will involve inviting them to join the collaborative group and to provide the data required for the analysis. Occasionally it will also involve seeking the advice of the trialists on the data to be collected. Establishing collaboration can take some time, especially if a trial was done many years ago and the appropriate personnel have moved since their trial was published or registered. In this case it pays to be persistent and to write to all authors.

In the initial correspondence the secretariat should emphasise the collaborative nature of the project and state that publication of the meta-analysis results will be made in the name of the collaborative group and stress that any data supplied will be held securely and

treated as confidential. It is also useful to reassure trialists that data collection will be as simple and flexible as possible. Including a written plan or protocol in this initial mailing may help in explaining the project to trialists, and also demonstrate the seriousness with which it is being tackled. Enclosing a reply form may help in getting a prompt reply containing the basic trial information and ascertaining what data items the trialists would be able to provide. A trial protocol and other documentation including information on the method of treatment assignment (including details on stratification factors and block size) should also be requested at this stage. An example of an initial form inviting collaboration is given in Appendix C. However, it may take several letters or telephone calls and even, in a few extreme cases, meetings with the trialists to secure their participation in the meta-analysis.

11a.9 Data collection

Once a decision has been taken that the meta-analysis is indeed feasible, what is often the most labour intensive aspect of the project, both for the secretariat and the trialists supplying data, can begin. On average a minimum of one or two person weeks of secretariat time is required to collect the data, convert it to a standard format, check, query and rectify the inevitable problems for any one trial. However, this may vary considerably depending on the complexity of the data collected. Thus, depending on the size of the meta-analysis, completing this stage can take several months. Fewer trials will, of course, mean less work at this stage and increasing the number of staff working on the project can speed the checking process. However, the absolute amount of time taken will ultimately be determined by how long it takes trialists to provide the data and respond to queries. In most instances, therefore, it is unlikely that this stage can be completed in less than a year.

Deciding which data items to collect
The minimum data that can be collected for an IPD meta-analysis are the patient identifier, treatment allocated and outcome(s), together with the date of randomisation and date of outcome if time to event is to be calculated. It is, however, often important to collect additional baseline variables, even when subgroup analyses are not planned, because these data are extremely useful in checking the integrity of the randomisation process. The collection of additional outcome data might also be advisable.

The decision on which data items to collect can be made by the secretariat, steering group or by the collaborative group. Obviously this last option will be time consuming and may lead to potential disagreements if suggestions are conflicting or if some are rejected. Whichever approach is adopted, it is essential that clinical as well as statistical input is sought. The final list of suggested variables should be sent to trialists early in the project to check that each variable will be available from a large enough proportion of trials to justify its request and collection.

Data Collection
Specifying the desired format for data, suggesting codes where appropriate and providing data collection forms may help trialists. However, it is important that trialists should be allowed to supply data in whatever way is most convenient to them, whereupon the secretariat take responsibility for converting the data to the required format. In such instances, it is very important that there is a clear understanding between the secretariat and trialists as to the content of their non-standard data. At this stage it may be useful to

identify a single individual (generally the person responsible for preparing the data) to whom all queries can be addressed, as this can simplify and speed the process considerably. Examples of forms and formats for data that have been used in the past are given in Appendix D1, D2, D3 and D4.

Unavailable data

It must be appreciated that provision of data may entail considerable work for the trialist and so good communication is essential both to persuade them of the worth of the project and to explain what is required of them. Every effort should be made to reduce the burden on the trialist or data centre providing the information. On initial contact, some trialists may report that the data from their study are not available. Although in instances where data have been destroyed or lost, the trial may not be recoverable, it is often worth pursuing negative replies in case an alternative source of data can be found. For example, other people within a trial group may be more willing or able to supply the data. More usually the problem is one of insufficient resource, so that offers of assistance (usually in the form of sending someone to retrieve the data) are often effective. An invitation to the collaborators' meeting has often acted as an incentive to collaborate.

The aim of the meta-analysis should be to obtain individual data from all randomised patients in all relevant trials. If, despite all efforts to secure collaboration, data from one or more trials are not available, the question of how to deal with this arises. When a large proportion of the total randomised evidence (perhaps 90-95%) has been collected, the missing data may be considered unlikely to alter importantly the meta-analysis results. Nonetheless the unavailability of trials should be made clear in the published report of any meta-analysis.

If individual patient data are not available, aggregate data provided by trialists or data extracted from publications could be used. However, it is not clear whether or not the use of data extracted from published reports is desirable, given the potential problems with such data compared to data (aggregate or individual) supplied directly by the trialist. In addition, an explanation of why this was deemed acceptable for some trials would have to be given to those trialists who had put a great deal of effort into supplying individual patient data. The use of published data might therefore discourage some trialists from providing any data. Where a trialist is unable to supply individual patient data but can provide aggregate data, this would be more acceptable than published data alone, but, again, such data will preclude the specific advantages of individual patient data and this should be noted. However, completely excluding trials from the meta-analysis because individual patient data were not obtainable might cause problems through the omission of randomised evidence. Whenever the IPD meta-analysis is supplemented with trial results that are not based on the provision of individual patient data, this should be made clear. One option might be to conduct sensitivity analyses comparing a purely individual patient data based meta-analysis with one that incorporates whatever data are available on all relevant trials.

Data Checking

The main aims of data checking procedures should be to ensure the accuracy of data, integrity of randomisation and completeness of follow up. For any one trial, it is important that the results of all the data checks should be considered together to build up an overall picture of that trial and any associated problems. Where there are concerns, these should be brought to the attention of the trialist and sympathetic attempts made to

resolve them. This can often be done by letter or phone but may, occasionally, involve a visit to the trialist to help clarify and if necessary to rectify matters. The vast majority of cases will be resolved satisfactorily - often by the insertion of data that were not supplied initially. Although errors in data are common, having seen the patient data from hundreds of trials, the experience of the groups represented at the workshop is that fraud is very rare.

Checking data accuracy
All data supplied should be subject to the sort of range and consistency checks that would be used in a prospective trial. This should be irrespective of whether data were supplied electronically or had to be entered manually into the meta-analysis database (when it is vitally important to check the accuracy of data input). Any missing data, obvious errors, inconsistencies between variables or extreme values should be queried and rectified as necessary by the trialist. If details of the trial have been published these also should be checked against the raw data and any inconsistencies queried. All of the changes made to the data originally supplied by the trialists, and the reasons for these changes, should be recorded.

Checking the integrity of randomisation and follow up procedures
It is very important that the analysis should be based on the 'intention-to-treat' principle and therefore that data should be collected, and analyses based, on **all** randomised patients. Any randomised patients that have been excluded from the trial should, wherever possible, be reintroduced to the analyses.

As part of the checking process prognostic variables should be checked for balance across treatment arms. It is, however, important to remember that imbalances may occur by chance alone especially for non-stratified variables and when trials are small. Other checks that can be done include looking at the weekday of randomisation. For example, in the UK we would expect very few non-acute randomisations at the weekend (although, in studies from other countries it is important to appreciate cultural differences in working patterns). Similarly, randomisations in trials of acute disease would be expected to spread throughout the week. A visual display of the chronological sequence of randomisations can be illuminating. For example, figure 2, which is included with the trialist's permission, shows such a curve from an unpublished trial of radiotherapy versus chemotherapy in multiple myeloma. In this trial the radiotherapy equipment was unavailable for six months during the trial but patients continued to enter the chemotherapy arm. It was only when the individual patient data were provided for a meta-analysis that this problem was brought to the attention of the trialist who agreed that the appropriate solution was to exclude this small number of non-randomised chemotherapy patients from the analysis. Similarly, looking at chronological accrual may reveal a period at the beginning or end of a trial when full randomisation was not taking place.

Follow up
Where survival (or other time dependent variable) is the primary outcome it may be important that trial follow up is as up to date as possible since an increased follow-up may see a reduction in the treatment effect if the survival curves are converging (2 , 13) or an increased treatment effect if the curves are diverging (14). Thus, where appropriate, data should be checked to ensure that follow up is up to date and to ensure that it is balanced across treatment arms. Balance in follow up can be checked by selecting all patients

outcome-free and using the date of censoring as the event to carry out a 'reverse Kaplan-Meier' analysis producing censoring curves which should be the same for all arms of the trial. Any imbalance should be brought to the attention of the trialist and updated information should be sought. However, the trialist might not be able to provide updated follow up on all their patients. In such cases it may be possible for the secretariat to take responsibility for obtaining the additional follow up. For example, if death is a primary outcome, mortality information might be available from national death registers, provided that sufficient information is available to identify the patient. Some sources of this information are shown in Appendix E.

However, not all countries run such schemes and tracing the fate of patients especially those from older trials is not necessarily straightforward (16). In addition the cause of death information available from these sources might not be sufficiently accurate to use for analysis of cause-specific mortality (in those relatively few cases where such analyses are done as a supplement to the more usual analyses of death by all causes).

Figure 2. Entry of patients to randomized trial showing accrual of patients to chemotherapy (and radiotherapy) treatment group.
Not currently available

Analysis of individual trials
Trials should be analysed individually and the trialists should be sent a copy of any such analyses as well as a printout of their data as included in the meta-analysis database. This allows verification and also provides the trialist with an updated analysis of their own study which they may find useful for other purposes including further reports of their trial.

What to do if a trial cannot be used
If a trial fails the checking procedures and the responsible trialist is unable to rectify the data or to explain the observed anomalies, the question arises of what to do next. Ultimately the decision on whether or not a particular aspect of a trial indicates a serious bias is a subjective one and the best solution may be to bring the problems to the attention of the trialist, and then to make a joint decision on whether to include or exclude it from the meta-analysis. If it is decided that a trial has to be excluded, this should be reported when the results of the meta-analysis are published. This is best done sympathetically, for example by noting simply that the trial had not been randomised properly. It is not the role of a meta-analysis group to oversee or to police the conduct of clinical trials and to be too explicit in the rejection of a trial could endanger the goodwill and collaborative spirit necessary for future meta-analyses.

11a.10 The collaborators' meeting

A collaborators' meeting is an important and integral part of the meta-analysis. It ensures that collaborators are the first to see the results of the meta-analysis and that they have a chance to question and discuss these results and their implications before they become available to a wider audience. These discussions and any conclusions that arise may lead to further analyses and they can then be incorporated into the published report of the meta-analysis. In addition, having the meta-analysis debated and endorsed by an internationally recognised group of experts may help with dissemination of results, which is a vital part of any systematic review. Finally the assembly of this international group also provides an excellent opportunity for discussing and possibly deciding the areas of

treatment which require clarification or further research. In particular it can provide a good opportunity to discuss and propose future trials. The goodwill engendered is invaluable in completing, updating and publishing the analysis and the existence of the meeting may serve as an incentive to collaborate. Such meetings are also valuable in setting a deadline to which the secretariat and trialists supplying data have to work.

The planning and organisation of such a meeting requires considerable resource and its date must be planned well in advance to fit with the overall timetable for the meta-analysis. The meeting can be scheduled for various stages of the project. If held at a reasonably early stage, when a good deal of data may be outstanding, it acts as a good incentive for trialists who have not supplied data to do so as soon as possible. Alternatively, if it is held at a later point in time, after the majority of data has been assembled and analysed, the results presented are very similar to those that will be used finally, and the time between the meeting and publication will be minimised.

The main purpose of the meeting should be to present the results of the meta-analysis and to discuss the methods, results and implications with the trialists so that they can take a full and active role in this process. The meeting should probably have a structured format and there should be ample time for discussion. Equal proportions of presentation and discussion time might be a good balance. The meeting is also the appropriate place to discuss the future of the Collaborative Group, for example whether to update the IPD meta-analysis in the future.

All those present at the Oxford workshop who had organised such Collaborative Group meetings had provided accommodation free of charge to participating trialists. Some had provided either travel funds for all participants or for those who would otherwise be unable to attend. The provision of such funds obviously depends largely on circumstance: the number of people involved and whether it would be possible to generate sufficient sponsorship to pay for expenses. One possible approach is to secure full funding for the first collaborators' meeting but for trialists to pay for their own travel to subsequent meetings. The cost of holding a one-day meeting without the provision of travel funds was approximately £100 per delegate increasing to around £600 per delegate when travel was provided, although of course this is very dependent on how far participants had to travel to the meeting.

11a.11 Publication

IPD meta-analyses should aim to publish the results as soon as possible after the Collaborative Group Meeting. Primary publications should be in the name of the collaborative group responsible for the meta-analysis rather than individual authors, the secretariat or steering group. This emphasises the collaborative nature of the project and engenders continued collaboration. As IPD meta-analyses are usually international projects and since trialists may wish to place varying emphasis on the interpretation of the results, it is wise for the publication to concentrate on the presentation of the results leaving detailed interpretation to separate commentaries by independent experts.

11a.12 Research agenda

The methodology described in this report stems from the collective experience of many groups who have already conducted meta-analyses using individual patient data. Such

projects provide the most reliable and informative type of systematic review by collecting and analysing all of the relevant randomised evidence. Although some aspects of IPD meta-analyses cannot be done in any other way, for example time to event analyses, they are also particularly time and resource consuming. It is therefore important that additional empirical evidence of the relative values of the different techniques involved in such reviews should be sought and published. With this in mind, the Cochrane Working Group on meta-analysis using individual patient data has initiated a research agenda (Appendix F). In addition to questions directly related to the conduct of meta-analyses, trial information collected as an integral part of these projects is a useful resource which would allow research into randomised controlled trials generally. Some of the topics listed have already been investigated (2 , 17, 18, 19, 20, 21) and we would be interested to learn of any other relevant past, current or planned research.

11a.13 Conclusions

Meta-analyses based on updated individual patient data provide the most comprehensive and reliable means of assessing the results of existing randomised clinical trials. It is the only reasonable way of performing time to event analyses, the best way of performing subgroup analyses and allows the review to use common prognostic and outcome variables. The detailed checking of data possible with this approach also improves the accuracy of the data included in the meta-analysis, allowing the integrity of the randomisation and follow up procedures to be assessed centrally. However, considerable expertise, time, effort and resource are required to carry out meta-analyses using individual patient data. They should not be undertaken lightly and might best be carried out by a secretariat on behalf of an international collaborative group. We hope that the guidance contained in this report will prove useful to such people.

11a.14 Appendix A: Participants at the Cochrane Collaboration workshop on Meta-Analysis Using Individual Patient Data, Oxford, 1994

BELGIUM
International Institute for Drug Development, Brussels
Marc Buyse

DENMARK
Danish National Study Group, Herlev
Lena Specht

FRANCE
Clinical Trials and Meta-Analysis - Clinical Pharmacology Unit, Lyon
Francois Gueyffier

Institut Curie, Paris
Veronique Mosseri

Institut Gustave-Roussy, Villejuif
Jean-Pierre Pignon

GERMANY
German Hodgkin's Disease Study Group, Köln
Michael Sextro

ITALY
Mario Negri Institute, Milano

Silvia Marsoni, Valter Torri

NETHERLANDS
Antoni van Leeuwenhoek huis Institute, Amsterdam
Harm van Tinteren, Annet te Velde

SWEDEN
Umeå University Hospital , Umeå
Lennart Nyström

UK
Academic Section of Geriatric Medicine, University of Glasgow
Peter Langhorne

Clinical Trial Service Unit, Oxford
Colin Baigent, Mike Clarke, Rory Collins, Tricia Elphinstone, Vaughan Evans, Richard Gray, Liz Greaves, Heather Halls, Mandy Ogier, Richard Peto, Sue Richards, Keith Wheatley

CRC Wessex Medical Oncology Unit, Southampton
Chris Williams

Department of Clinical Neurosciences, University of Edinburgh
Carl Counsell

Imperial Cancer Research Fund, London
Doug Altman, Jack Cuzick, Rob Edwards

London School of Hygiene and Tropical Medicine, London
Paul Seed

MRC Cancer Trials Office, Cambridge
David Machin, Max Parmar, Sally Stenning, Lesley Stewart, Jayne Tierney, Paul Weston

University College Hospital, London
Carmen Ruiz

UK Cochrane Centre, Oxford (observers)
Iain Chalmers, Jini Hetherington, Sally Hunt, Carol Lefebvre, Andy Oxman

Witthenshawe Hospital, Manchester
Rob Henderson

11a.15 Appendix B: Medline search strategies for optimal sensitivity in identifying randomised clinical trials

Format shown is for SilverPlatter version 3.10. Upper case denotes controlled vocabulary. Lower case denotes free-text terms. Those wishing to run this search strategy are recommended to seek the advice of a trained medical librarian.

#1 RANDOMIZED-CONTROLLED-TRIAL in PT
#2 RANDOMIZED-CONTROLLED-TRIALS
#3 RANDOM-ALLOCATION
#4 DOUBLE-BLIND-METHOD
#5 SINGLE-BLIND-METHOD
#6 #1 or #2 or #3 or #4 or #5
#7 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#8 #6 not #7

#9 CLINICAL-TRIAL in PT

#10 explode CLINICAL-TRIALS
#11 (clin* near trial*) in TI
#12 (clin* near trial*) in AB
#13 (singl* or doubl* or trebl* or tripl*) near (blind* or mask*)
#14 (#13 in TI) or (#13 in AB)
#15 PLACEBOS
#16 placebo* in TI
#17 placebo* in AB
#18 random* in TI
#19 random* in AB
#20 RESEARCH-DESIGN
#21 #9 or #10 or #11 or #12 or #14 or #15 or #16 or #17 or #18 or #19 or #20
#22 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#23 #21 not #22
#24 #23 not #8

#25 TG=COMPARATIVE-STUDY
#26 explode EVALUATION-STUDIES
#27 FOLLOW-UP-STUDIES
#28 PROSPECTIVE-STUDIES
#29 control* or prospectiv* or volunteer*
#30 (#29 in TI) or (#29 in AB)
#31 #25 or #26 or #27 or #28 or #30
#32 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#33 #31 not #32
#34 #33 not (#8 or #24)

Reproduced with kind permission from Carol Lefebvre, UK Cochrane Centre

11a.16 Appendix C: Form supplied with invitation to collaborate in an individual patient-based meta-analysis

LOCALISED SOFT TISSUE SARCOMA META-ANALYSIS

Name:

Did we get your title, affiliation and address correct? If not please give correct details:

Telephone:
(*area code and number*)

Fax:
E-mail:

Please give your own reference or protocol number for this study.

Are the details concerning your study correct? Yes No

Is the most recent publication cited in the protocol reference list?

If no please give details:

Are you willing to take part in this overview? Yes No

If yes please confirm that you would be able to supply survival information for each patient randomised

Yes No Patient identifier Yes No Date of randomisation

Date of birth or age at randomisation	Survival status
Sex	Cause of death
Disease status	Date of death/last follow up
Disease site	Local recurrence status
Histology	Date of local recurrence
Histologic Grade	Distant recurrence status
Tumour size	Date of distant recurrence
Primary treatment	Whether excluded from own analysis
Treatment allocated	Reason for exclusion
Extent of resection	

How will you supply data?

Floppy disk: E-mail: Computer print-out: Sealed envelope:

Please give the method of randomisation used in this study

Central telephone call Other (please specify): Sealed envelope:

Please state stratification factors used (if any):

What proportions was this study designed to have in each arm? (eg 1:1

Please give the name and address of the appropriate contact for collection of data:

Please five details of any relevant publications or trials you may know of not listed in the tables or Appendix A of the protocol:

Signed Date

Please note that any information supplied will be treated in strict confidence and used only for the purpose of the overview

11a.17 Appendix D1: Example coding and formatting instructions for data supplied electronically

LOCALISED SOFT TISSUE SARCOMA META-ANALYSIS
Suggested Coding: Individual Patient Data

- Disks should be formatted for the DOS operating system.
- Files should be in DBASE, FoxPro (.dbf files) or ASCII format with fields separated by spaces. However, it would be preferable if you did not use spaces to denote unknown values (see below).

- You may code the data in whichever way is most convenient to you, although it would be helpful if you adopted the coding suggested on this sheet. If you are unable to do this, please supply full details of the coding system used.

Please list fields in the following order using the suggested coding:

Patient identifier	Type	Character	**Treatment Allocated**	Type	numeric
	Width	15		Width	1
	Any alphanumeric string up to15 characters			Code	1=treatment 2=control
Date of birth (DOB)	Type	date	**Extent of Resection**	Type	numeric
	Width	-			1
	Code	date in dd/mm/yy format unknown day=15/mm/y unknown month=157067yy unknown date=0/01/01			1=well clear 2=close/marginal 3=macroscopically involved 9=unknown
Age	Type	numeric	**Date of Randomisation (DOR)**	Type	date
	Width	3		Width	-
	Code	age in years unknown=999		Code	date in dd/mm/yy format
Sex	Type	numeric	**Survival Status**	Type	numeric
	Width	1		Width	1
	Code	1=female 2=male 9=unknown		Code	0=alive 1=dead If survival status is unknown code as 0, the patient being censored at the date of the last follow up
Disease status (at randomisation)	Type	numeric	**Cause of death**	Type	numeric
	Width	1		Width	1
	Code	1=primary 2=recurrent 3=metastatic 9=unknown		Code	1=soft tissue sarcoma 2=chemotherapy related 3=other 8=not applicabe 9=unknown
Disease site	Type	numeric	**Date of death/Last follow up**	Type	date
	Width	1		Width	-
	Code	1=extremity 2=trunk 3=head and neck 4=breast 5=uterus 6=retroperitoneum		Code	date in dd/mm/yy format unknown day=15/mm/yy unknown month=15/06/yy unknown date=01/01/01
			Local Recurrence Status	Type	numeric

		7=viscera/abdomen 9=unknown		Width Code	1 0=no recurrence 2=leiomyosarcoma 9=unknown
Histology	Type Width Code	numeric 1 1=MFH 1=recurrence 3=liposarcoma 4=synovial 5=malignant schwannoma 6=alveolar or embryonal rhabdomyosarcoma/Ewing's/PNET 7=AIDS-related sarcoma 8=other 9=unknown	**Date of Local Recurrence**	Type Width Code	date - date in dd/mm/yy format unknown day=15/mm/yy unknown month=15/06/yy unknown date=01/01/01
			Distant Recurrence Status	Type Width Code	numeric 1 0=no recurrence 1=recurrence 9=unknown
Grade	Code as convenient, but please supply full details of the coding system used				
			Date of Distant Recurrence	Type Width	date -
Tumour size	Type Width Code	numeric 2 Give the size of the largest single dimension in centimetres unknown=99		Code	date in dd/mm/yy format unknown day=15/mm/yy unknown month=15/06/yy unknown date=01/01/01
Primary Treatment	Type Width Code	numeric 3 *1st digit (pre-op treatment)* 0=non 1=radiotherapy 2=induction chemotherapy 3=radiotherapy + induction chemotherapy 9=unknown	**Excluded**	Type Width Code	- numeric 1 0=included in analysis 1=excluded from analysis 9=unknown
			Reason for Exclusion	Type Width Code	character 15 short string giving reason for exclusion ornumeric codes with code meanings provided

2nd digit (surgery)
1=amputation
2=excision
3=biopsy only
9=unknown

3rd digit (post-op treatment)
0=no radiotherapy
1=radiotherapy
2=unknown

11a.18 Appendix D2: Example of a form that could be used to supply data manually

Colectoral Cancer Collaboration (CCC) overview of mortality by randomly-allocated treatment in resectable colorectal cancer trials: provision of one line line of CONFIDENTIAL data for each patient ever randomised (INCLUDING any ineligible, withdrawn, inevaluable, lost or 'protocol deviant' patients).

Name of trialist of trial group:
Data Sheet No.:
Name of trial:
Staging system used: Dukes: |_| Astler-Coller: |_| TNM: |_| Other: |_| (please specify)
Treatment group 1 = ; Trt. gp. 2 = ; Trt. gp. 3= ; Trt. gp. 4=

Patient Identifier	Date randomi-sed	Trf. gp.	Date of surgery	Tum.site	Tumour stage	Gender	Entry age	Rec-ur?	Approx. date of 1st rec.	Site rec.	Dead/ other	Date died/last traced	Cause of death if died without recurrence
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|	|_|	|_|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|
|_|_|_|_|_|_|_|_|_|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|_|_|	|_|	|_|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|_|	|_|	|_|	|_|_|_|_|_|_|

Leave blanks if not applicable or if data not (or not yet) conveniently available

11a.19 Appendix D3: Coding scheme that was used with the form for supplying data manually

GUARANTEE OF CONFIDENTIALITY OF DATA: ANY INFORMATION PROVIDED OVERLEAF TO THE
CC SECRETARIAT WILL BE HELD SECURELY AND IN STRICT CONFIDENCE

NOTES OF FORMAT OF DATA REQUESTED OVERLEAF:

- Special coding conventions:
  Please accompany these forms by an explanatory letter about any special coding conventions (e.g. on tumour site, tumour staging or cause of death) you have used, plus notes on any special features of the study(s) to which you wish to draw attention.

- Dates that are not (or not yet) known exactly:
  either leave DAY blank and give (approximate or provisional) month and year;
  or leave DAY and MONTH blank, and just give approximate year.

BASELINE DATA:

Patient identifier:
Any convenient convention you wish, in case any correspondence becomes necessary. (If reporting several trials, please try to use a system that implicitly specifies both the trial and the patient.)

Date randomised:
Please describe ALL patients EVER randomised, including even lost, ineligible or withdrawn patients, and ignore all non-randomised patients.

Trt. gp. allocated:
Treatment group number: 1 or 2 only, for 2-group trials, or a wider range for trials with more arms, as defined by you at the top of the form. N.B: even if, in reality, some quite different (or even opposite!) treatment was inadvertently given, what is wanted is the originally-allocated treatment. (For patients erroneously entered more than once, give only the first allocation.)

Date of surgery:
See note above on approximate dates.

Tumour site:
0 = unspecified; 1 = colon; 2 = rectum; 3 = colon and rectum. If you prefer to use your own classification of tumour site (e.g. in order to code sigmoid tumours separately) please do so, and send us details of it.

Tumour stage:
Please use your own classification and send us details of it, or use the Dukes classification (A = lesion confined to muscularis propria; B = lesion extends through muscularis propria with negative nodes; C = positive nodes), or any other standard system (e.g. Astler-Coller modification, TNM etc). Extra codes: D = metastatic disease; X = benign tumour (eg adenoma); and Y = inoperable disease.

Gender:
1 = male; 2 = female.

Entry age:
Age at randomisation.

FOLLOW-UP DATA:

Recur?:

Any recurrence? 1 = none recorded; 2 = some recurrence (local or distant or both).

Approx. date of 1st recur.:
Give the best estimate you can: see note above on approximate dates.

Site of 1st recurrence:
0 = unknown; 1 = local only; 2 = local and distant; 3 = distant only.

Dead/other:
1 = alive when last traced; 2 = known to be dead; 3 = lost despite extensive inquires, but still alive when last traced.

Date died/last traced:
Date of death, or date last known to be alive, as accurately as possible: see note above on approximate dates.

Death cause:
If the patient died without reported recurrence, give underlying cause of death. Either state the cause in words, use an ICD code or use your own classification and send us details of it.

11a.20 Appendix D4: Example of instructions that could be used to create a formated electronic file

MACH-NC
Meta-analysis of Chemotherapy in Head and Neck Cancer

	Column	
Patient Identifier	2-11	10 characters
Date of birth	13-18	dd/mm/yy, 999999=Unknown
or age	17-18	2 digits (13-16 blanks) 99=Unknown
Sex	20	1=Male, 2=Female, 9=Unknown
Site of primary	22	1=Oral cavity, 2=Oropharynx, 3=Nasopharynx, 4=Larynx, 5=Hypopharynx, 6=Cervical node(s) without primary, 7=Others, 9=Unknown
T	24	O=TO, X=TX, S=Tis, 1=T1, 2=T2, 3=N3, 9=Unknown
N	25	O=NO, X=NX, 1=N1, 2=N2, 3=N3, 9=Unknown
M	26	O=MO, 1=M1, 9=Unknown
or stage	26	1 digit (24-25 blanks), 9=Unknown

(The aim of the next four questions is to identify presenting characteristics at the time of randomisation)

Recurrence at randomisation	28	0=No, 1=Yes
Second primary at randomisation	30	0=No, 1=Yes
Squamous cell	32	0=No, 1=Yes
Type of histology if not squamous cell	34-45	12 characters (*blanks for squamous cell)*
Treatment allocated	47	1=No chemotherapy, 2=Chemotherapy
Date of randomisation	49-54	dd/mm/yy, 999999=Unknown
Received at least one cycle of chemotherapy	56	0=No, 1=Yes, 9=Unknown
Date of last follow-up	58-63	dd/mm/yy, 999999=Unknown
Survival status	65	0=Alive, 1=Dead
Death related to treatment	67	0=No, 1=Yes
Complete response at the end	69	0=No, 1=Yes

of treatment (including salvage treatment)		*(collected for computation of disease-free survival)*
Recurrence of second primary	71	0=No, 1=Yes *(only for complete responders)*
Date of first event	73-78	dd/mm/yy, 999999=Unknown
Type of first event	80	1=locoregional, 2=metastasis, 3=locoregional + metastasis, 4=second primary without recurrence, 9=Unknown
Excluded from your analysis	82	0=No, 1=Yes
Reasons for exclusion	84-95	12 characters

11a.21 Appendix E: Sources of mortality information for individual patients

England and Wales
The Chief Medical Statistician (Dept MR) Health Statistics
OPCS
St Catherine House
10 Kingsway
London WC2B 6JP

France
INSEE
Département de Démographie
Division Répertoire et Mouvement de la Population
18, Bd Adolphe Pinard
75675 PARIS
Cedex 14

Service d'information sur les causes médicale de décès
INSERM SC8
55, Chemin de Rorde
BP 34
78100 LE VESINET

Isle of Man
Isle of Man Health Services Board
Registration Department
Markwell House
Market Street
Douglas
Isle of Man

Northern Ireland
The Central Services Agency
27 Adelaide Street
Belfast BT2 8SH

Norway
Statistisk Sentralbyraa
Skippergt. 15
PB 8131 Dep
N-0033 Oslo
Norway

Scotland
Departmental Record Officer
General Register Office for Scotland
New Register House
Edinburgh EH1 3YT

USA
National Death Index
Division of Vital Statistics
National Centre for Health Statistics
6525 Belcrest Road
Hyattsville, MD 20782
USA

11a.22 Appendix F: Research agenda proposed by Cochrane Working Group on Individual Patient Based Meta-Analyses

Although some aspects of IPD meta-analyses cannot be done in any other way, for example time to event analyses, these projects also particularly time and resource consuming. It is therefore important that additional empirical evidence of the relative values of the different techniques involved in such reviews should be sought and published.

A Research relating to individual patient-based meta-analysis

1. ***Comparison of individual patient data with summary data supplied by trialists*:** At least two individual patient based meta-analyses have been conducted following the collection of summary data from the same set of trials. These are in Hodgkin's disease and in antiplatelet therapy

2. ***Comparison of individual patient data with published data*:** This has been done for cisplatin-based therapy in ovarian cancer but most of the individual patient data meta-analyses could repeat those analyses. This would allow the evidence to be extended to other disease and therapy areas.

3. ***Comparison of individual patient data after extensive data-checking with individual patient data supplied initially*:** There are different levels of data-checking - from finding and querying missing or inconsistent data variables, to detailed investigation of the integrity of the randomisation and follow-up procedures. Detailed data-checking is resource-intensive and time-consuming and may delay the publication of the meta-analysis results, so empirical evidence of its value would be useful.

4. ***Comparison of trial quality as assessed using the individual patient data with quality as assessed from the published report:*** Does the individual patient data reveal problems in the randomisation or follow-up procedures that were not mentioned in the published report?

B Research relating to all types of meta-analysis and to RCTs

5. **Method of randomisation:** Sensitivity analyses could be performed using the method of randomisation (eg envelope, central computer, 'blinded' date of birth) to distinguish between RCTs. Stratification, minimisation and block size could also be investigated.

6. ***Size of RCTs*:** Sensitivity analyses could be performed to take into account the size of RCTs. This could also investigate whether there are important differences in the results from multi-centre or single institute trials.

7. ***Chronology of RCTs:*** Sensitivity analyses could be performed distinguishing between RCTs by their place in time - perhaps the early RCTs have the more striking results. A RCT's place in time could be defined in various ways (start date, finish date, publication date) and cumulative meta-analyses could be done ordered in these ways. Sensitivity analyses could also be performed distinguishing between RCTs published before the systematic review was conducted and those published afterwards.

8. ***Place of publication:*** Sensitivity analyses could be performed distinguishing between RCTs which have been published as full papers, as abstracts or are unpublished. This will also investigate whether there are any important differences between RCTs published in journals indexed by medical literature databases; between RCTs in those databases which would or would not have been found by a simple search strategy; between RCTs in or not in the 'major' journals identified by these databases;and between RCTs published in different languages.

9. ***Speed of publication:*** The variation in the speed of publication among trials with differing results could be investigated, especially with regard to changes in their results with further follow-up.

10. ***Repeated publications*****:** RCTs may be reported several times and it is often difficult to know that reports are of the same trial, and so may be included more than once in a meta-analysis. It has been suggested that positive trials are more likely to be published repeatedly. This could be investigated.

11. ***Fate of RCTs published as abstracts:*** Sensitivity analyses could be performed distinguishing between RCTs which were published as abstracts and then did or did not publish as full papers.

12. ***Citation bias:*** To investigate whether the RCTs in the meta-analyses selectively cite other RCTs with similar results. This could also investigate (using the Science Citation Index) which RCT publications are cited most often to see if their results are representative of the overall conclusion as shown by the meta-analysis or are they at an extreme?

13. ***Source of trial funding:*** Sensitivity analyses could be performed using the source of funding (eg drug company, government, charity, local) to distinguish between RCTs.

11a.23 Acknowledgements

We are grateful to the UK Cochrane Centre, in particular to Iain Chalmers for suggesting the Oxford Workshop and Caroline Caldicott for helping to organise it. We also thank Linda Baulk for typing this manuscript.

11a.24 References

1. Chalmers I. The Cochrane Collaboration: preparing, maintaining and disseminating systematic reviews of the effects of healthcare. Ann N Y Acad Sci 1993; **703**: 156-65.

2. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data:is there a difference? Lancet 1993; **341**: 418-22.

3. Pignon JP, Arriagada R. Meta-analysis. Lancet 1993; **341**: 964-5.

4. Pignon JP, Ducreux M, Rougier P. Meta-analysis of adjuvant chemotherapy in gastric cancer: a critical reappraisal. J Clin Oncol 1994; **12:** 877-9.

5. Dickersin K, Scherer R, Lefebvre C. Identification of relevant studies for systematic review. Br Med J 1994; **309**: 1286-91.

6. Cook DJ, Guyatt GH, Ryan et. Should unpublished data be included in meta-analyses. JAMA 1993; **269:** 2749-53.

7. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. JAMA 1992; **267**: 374-8.

8. Dickersin K, Min YI. NIH Clinical trials and publication bias. Online J Curr Clin Trials (serial online) 1993. Doc No 50.

9. Easterbrook PJ, Berlin JE, Gopalan R, Matthews DR. Publication bias in clinical trials. Lancet 1991; **337**: 865-72.

10. Dickersin K. Keeping posted. Why register clinical trials? - Revisited. Cont Clin Trials 1992; **13**: 170-7.

11. Easterbrook PJ. Directory of registries of clinical trials. Stats in Med 1992; **11:** 345-423.

12. Clarke MJ, Stewart LA. Obtaining data from randomised controlled trials: how much do we need in order to perform reliable and informative meta-analyses? Br Med J 1994; **309:**1007-10.

13. Advanced Ovarian Cancer Trialists Group. Chemotherapy in advanced ovarian cancer: an overview of randomised clinical trials. Br Med J 1991; **303:** 884-93.

14. Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormal, cytotoxic or immune therapy. Lancet 1992; **339:** 1-15, 71`-85.

15. Early Breast Cancer Trialists' Collaborative Group. Treatment of early breast cancer: Vol 1, Worldwide evidence 1990 Oxford: Oxford University Press, 1990.

16. Haugh MC, Cornu C, Boissel JP. Long-term survival follow-up in international clinical trials 1993; **14:** 416.

17. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. JAMA 1994; **272**: 158-62.

18. Rochon PA, Gurwitz JH, Cheung CM, Hayes HA, Chalmers TC. Evaluating the quality of articles published in journal supplements compared with the quality of those published in the parent journal. JAMA 1994; **272**: 108-13.

19. Gotzsche PC. Multiple publication of reports of drug trials. Eur J Clin Pharmacol 1989; **36**: 429-32.

20. Gotzsche PC. Reference bias in reports of drug trials. Br Med J 1987; **295:** 654-9.

21. Antman EM, Lau J, Kupelnick B, Mostellar F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. JAMA 1992; **268:** 240-8. (Impact of meta-analyses).

22. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trials results: effect on physician' willingness to prescribe. Lancet 1994; **343:** 1209-11. (Absolute versus relative difference for the report of treatment effect).

APPENDIX 11B. PROSPECTIVE META-ANALYSIS

A systematic review should, ideally, define the question to be addressed in advance of the identification of potentially eligible studies. However, these projects are by their nature, retrospective, since the studies included are usually identified after they have been completed and reported (Pogue 1998, Zanchetti 1998). The reviewer's knowledge of the results of the study may influence:

- the criteria for study selection
- the definition of a systematic review question
- the interventions and participant groups evaluated
- the outcomes to be assessed in the review

In contrast, a systematic review which is conducted as a prospective meta-analysis includes studies that were identified, evaluated and determined to be eligible for inclusion before their results became known. It is a method that has been used in recent years in cardiovascular disease (Simes 1995, CTTC 1995, WHO-ISHBPL 1998) and childhood leukaemia. (Shuster 1996, Valsecchi 1996) and can help to overcome some of the problems of traditional systematic reviews by enabling:

- hypotheses to be specified *a priori*, ignorant to the results of individual studies
- prospective application of selection criteria

- *a priori* statements of intended analyses, including subgroup analyses, to be made before the results of individual studies are known. This avoids potentially unreliable data-dependent emphasis on particular subgroups.

A Methods Group has been established to investigate methodological issues around such projects and to offer guidance on their conduct. For example, because studies should not be included in a prospective meta-analyses if their results are known before the decision is taken to include them, PMA will not always include all studies of a particular question. Research is needed to investigate the impact of this on systematic reviews.

To register a PMA as a Cochrane review, investigators need to submit a protocol to the relevant Collaborative Review Group (CRG). The protocol will then undergo the same peer review process as any Cochrane review. The decision as to whether or not a PMA should be a Cochrane Review rests with the CRG. If a CRG decides it does not have the expertise necessary to determine whether or not the submitted protocol meets the requirements of a PMA, members of the PMA Methods Group will be available to review the protocol.

References

CTTC 1995. Protocol for a prospective collaborative overview of all current and planned randomized trials of cholesterol treatment regimens. Cholesterol Treatment Trialists' (CTT) Collaboration. American Journal of Cardiology 1995; 75: 1130-4.

Pogue 1998. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet. 1998; 351: 47-52.

Shuster 1996. Shuster JJ. Gieser PW. Meta-analysis and prospective meta-analysis in childhood leukemia clinical research. Annals of Oncology 1996; 7: 1009-14.

Simes 1995. Simes RJ. Prospective meta-analysis of cholesterol-lowering studies: the Prospective Pravastatin Pooling (PPP) Project and the Cholesterol Treatment Trialists (CTT) Collaboration. American Journal of Cardiology 1995; 76: 122C-126C.

Valsecchi 1996. Valsecchi MG. Masera G. A new challenge in clinical research in childhood ALL: the prospective meta-analysis strategy for intergroup collaboration. Annals of Oncology 1996 ; 7: 1005-8.

WHO-ISHBPL 1998. Protocol for prospective collaborative overviews of major randomized trials of blood-pressure-lowering treatments. World Health Organization-International Society of Hypertension Blood Pressure Lowering Treatment Trialists' Collaboration . Journal of Hypertension 1998; 6: 127-37.

Zanchetti 1998. Zanchetti A, Mancia G. Searching for information from unreported trials--amnesty for the past and prospective meta-analyses for the future. Journal of Hypertension 1998; 16: 125.