THE COCHRANE COLLABORATION

HANDBOOK

Version 3.0 (compiled November 1996)

CONTENTS

The Cochrane Collaboration Handbook
Version 3.0 (compiled November 1996)

ABOUT THE HANDBOOK

Editors

Cynthia Mulrow and Andrew Oxman

Date of most recent amendment

21 October 1996

Date of most recent substantive amendment

21 October 1996

Handbook versions and related resources

Handbook	Ver	Release date	Description	Contact
Word Viewer Handbook	3.0	21 Oct 1996	Core text & appendices published with CDSR + on CC FTP servers	Cynthia Mulrow, SACC & Andy Oxman, NCC
Hypertext Handbook	3.0	21 Oct 1996	Windows help file published with CDSR + on CC FTP servers	Cynthia Mulrow, SACC & Andy Oxman, NCC
Related resources	Ver	Release date	Description	Contact
Cochrane Review Methodology Database (CRMD)	3.0	21 Oct 1996	Reference list + additional references + abstracts + full text of some references	Andy Oxman, NCC
Training Manual	3.0	21 Oct 1996	Printed document published by SACC includes additional objectives, summaries, examples & exercises	Cynthia Mulrow, SACC
RevMan help + Manual	3.0	21 Oct 1996	Help screens + software manual	Monica Fischer, NCC
FAQ list NCC	3.0	21 Oct 1996	Frequently asked questions	Rasmus Moustgaard,

Related resources	Ver	Release date	Description	Contact
Examples library	3.0	21 Oct 1996	Examples + exercises for each step in training manual	Cynthia Mulrow, SACC
Slides library	3.0	21 Oct 1996	Power Point files	Cynthia Mulrow, SACC
Empirical Methodological Studies (EMS) module		08 Oct 1997	Register of studies + systematic reviews	Andy Oxman, NCC
MWG modules		02 Dec 1997	Includes scope,	MWG convenors

How to cite the handbook

Cochrane Library (hypertext or Word Viewer) version:

> Mulrow CD, Oxman AD (eds). Cochrane Collaboration Handbook [updated 21 October 1996]. Available in The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 3. Oxford: Update Software; 1996. Updated quarterly. Available from BMJ Publishing Group, London.
>
> When referring to a specific section or subsection refer to it by the title and section number, NOT page numbers, e.g.:
>
> Mulrow CD, Oxman AD (eds). Formulating the problem. Cochrane Collaboration Handbook [updated 21 October 1996]; Section 4. Available in The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 3. Oxford: UpdateSoftware; 1996. Updated quarterly. Available from BMJ Publishing Group, London.

Contact adresses

Dr Cynthia Mulrow
San Antonio Cochrane Center
Audie L. Murphy Memorial Veterans Hospital
VA ACOS/AC (11C) 6
7400 Merton Minter Boulevard
San Antonio, Texas 78284
USA
Telephone: 1-210-617-5190
Facsimile: 1-210-617-5234
E-mail: cochrane@merece.uthscsa.edu

Dr Andy Oxman
Health Services Research Unit
National Institute of Public Health
P.O. Box 4404 Torshov
0403 Oslo
NORWAY
Telephone: +47 22 04 22 00
Fax: +47 22 04 25 95
E-mail: andrew.oxman@labmed.uio.no

Advisory Group

Douglas Altman, Mildred Cho, Mike Clarke, Lelia Duley, David Moher, Dave Sackett, Ken Schulz, Mark Starr, Lesley Stewart.

Sources of support

Health Services Research and Development Service and the University of Texas Health Science Center, San Antonio (USA)
National Institute of Public Health (Norway)
Norwegian Ministry of Health and Social Affairs (Norway)
Norwegian Research Council (Norway)
US Veterans Health Administration (USA)

Acknowledgements

Many people have contributed to this and preceding editions of the Cochrane Collaboration 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 their own time to making it a reality.

We would particularly like to thank the following people for their contributions to this edition of the Handbook: Claire Glenton, Joe Cavellero, Michael Brand, Valerie Lawrence, Christina Aguilar, and Bob Badgett.

The first edition of the Handbook (1994) was developed principally with funds provided by the National Health Service Research & Development Programme, England, Oxford Regional Health Authority, and the Nuffield Provincial Hospitals Trust, UK; and the LW Frohlich Fund, USA. Section VI, from which this version of the Handbook has evolved, 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.

Support for editorial work on the second edition of the Handbook (1995) was provided by the National Institute of Public Health, the Norwegian Research Council and the Norwegian Department of Health and Social Affairs, Norway; the Swedish Council on Technology Assessment in Health and the Swedish Planning and Rationalization Institute for the Health and Social Services, Sweden; and the National Health Service Research & Development Programme, England, UK.

WHAT'S NEW?

Last version of the Handbook

14th July 1995

Editorial responsibility

Responsibility for maintaining material formerly contained in Sections I to V of The Cochrane Collaboration Handbook has been devolved as described below. The Handbook now consists solely of what was formerly Section VI: Preparing and Maintaining Systematic Reviews. Cynthia Mulrow, director of the San Antonio Cochrane Center, has joined Andy Oxman as co-editor. The entire Handbook has been 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 over the past three years 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, MWGs 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
MWGs	MWG rep on Steering Group	Andy Oxman

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.)

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.

Reviews of non-experimental evidence

Some CRGs, Fields and Methods Working Groups (MWGs) 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".

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 MWGs for inclusion in CDSR.

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 2.2 and section 2a.6 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.

- Subheadings for ongoing studies and bibliographic information

  Implementation of subheadings for ongoing studies has been postponed pending recommendations from the newly formed Trials Registers Development Group. It is anticipated that a structured format for bibliographic information (using specified fields for authors, titles, etc.) will also be used in the future. These changes, once they are implemented will facilitate linkages between trials registers, other databases such as MEDLINE and Cochrane Reviews, and facilitate maintenance of reviews.

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 {188}. 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 random errors (simple mistakes), thus providing more reliable results upon which to draw conclusions and make decisions {103, 233}. 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 {335, 177, 70, 439}.

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 {43, 24, 440, 37, 441}. It was not until the late 1980s that people drew attention to the poor scientific quality of healthcare review articles {189, 197, 361}. 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 {442}, the number of systematic reviews published in healthcare journals {443}, and the rapid growth of the Cochrane Collaboration {444}. Recognition of the importance of systematic reviews has also stimulated a growing number of empirical studies of the methods used in reviews {445}.

This Handbook builds on the work of many people, including those represented in the Cochrane Review Methodology Database {442}, input from Cochrane Methods Working Groups {444}, practical experience and feedback from Collaborative Review Groups {444} which have taken on the daunting task of systematically reviewing the effects of healthcare within their areas of interest, and Cochrane Centres {444} which provide training for reviewers. Whenever possible recommendations made here are based on empirical evidence and advice from Cochrane Methods Working Groups.

Our aim is to help reviewers make good decisions about the methods they use relative to the specific healthcare problems that they address, rather than dictate arbitrary standards. 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 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. 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.

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
- Critical appraisal 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.

References

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

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

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

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

103. 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.

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

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

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

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

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

335. 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.

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

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

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

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

442. Glenton C, Oxman AD. Cochrane Review Methodology Database. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996; Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

443. NHS Centre for Reviews and Dissemination. Database of Abstracts of Reviews of Effectiveness. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996; Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

444. Cochrane Collaboration. Cochrane Database of Systematic Reviews. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996; Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

445. Kunz R, for the Empirical Methodological Studies MWG. Register of Empirical Methodological Studies. Oxford: Update Software, 1996; Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

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 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 by collaborative trialists' groups 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*, 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 to *The Cochrane Database of Systematic Reviews* (CDSR).

Each review consists of:

- a cover sheet - giving the title, citation details and contact addresses
- 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 conclusions.
- standard tables and figures - showing characteristics of the included studies, specification of the interventions that were compared, the results of the included studies, and a log of the studies that were excluded.
- references

Standard headings and tables guide reviewers preparing a 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 is described in the appendix attached to this section (Guide to the format of a Cochrane Review).

2.1 Outline of a Cochrane Review

Cover sheet:

Title
Short title
Reviewer(s)
Contact address
    Name
    Organisation and address
    Telephone number

- Facsimile number
- E-mail
- Date last edited
- Date of last substantive update
- Sources of support to the review

Abstract

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

Text

- Background
- Objectives
- Selection criteria
  - Types of studies
  - Types of participants
  - Types of interventions
  - Types of outcome measures
- Search strategy
- Methods
- Description of studies
- Methodological quality
- Results
- Discussion
- Conclusions
  - Implications for practice
  - Implications for research
- Acknowledgements
- Conflicts of interest

Tables and figures:

- Table of comparisons
- Table of included studies
- Table of excluded studies
- Data tables and graphs

References:

- References to studies
  - Studies included in this review
  - Studies excluded from this review
  - Studies awaiting assessment

Ongoing studies
Other references
Additional references
Previously published versions of this review

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.

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 2c):

- 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, such reviews 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. Financial conflicts of interest cause the most concern and must be reported if there are any. However, 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.

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 at all (505). However, financial conflicts of interest, which cause the most concern, can and should be avoided. Conflicts of interest that cannot be avoided should be disclosed.

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 Collaborative Review Group editors know of potential conflicts even when they are confident that their judgements will not be or were not 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 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 following two statements in addition to disclosing conflicts of interest (see appendix 2d):

1) "I have participated sufficiently in the conception and design of this Review and the analysis of the data, as well as the writing of the Review to take public responsibility for it. I believe the manuscript represents valid work. I have reviewed the final version of the Review and approve its public dissemination."

2) "I certify that any past or present affiliations with or involvement in any organisation or entity with a direct financial interest in the subject matter or materials discussed in the Review (e.g., employment, consultancies, stock ownership, honoraria, expert testimony) are listed below."

2.3 Publication of Cochrane Reviews in print journals and books

Authors of Cochrane Reviews may wish to seek co-publication in peer-reviewed medical journals, particularly in those journals that have expressed enthusiasm for co-publication of Cochrane Reviews. A co-publication policy is appended that has been worked out with the editors of the BMJ and The Lancet, and subsequently agreed to by the editors of most speciality journals affiliated with the BMJ, the Canadian Medical Association Journal and Annals of Internal Medicine and agreed to in principle by the International Committee of Journal Editors (Vancouver Group), although it is not binding on the members of the latter.

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 are granting the Collaboration world-wide licenses for these activities, and do not sign over exclusive copyright to any journal or other publisher. Although journals are free to request a non-exclusive copyright that permits it to publish and re-publish the review, but this cannot restrict the publication of the review by the CC in whatever form then CC feels appropriate.

Journals can insist that the publication of the review in the Cochrane Database of Systematic Reviews (CDSR) should not precede publication in print, but publication in print should not be subject to lengthy production times. Because of editors' desires for priority, reviewers should submit their articles after agreement from the editor for their CRG, but before publication in the CDSR. Journals can also request revision of the review for editorial or content reasons. External peer review provided by journals may enhance the value of the review and should be welcomed.

Also, journals generally may require shorter reviews. Selective shortening of reviews may be appropriate, but there should not be any substantive differences between the review in the journal and in the Cochrane Database.

We suggest that the passage below and a copy of the publication policy be provided to editors upon submission of articles for publication, and that the letter of submission be copied to the CRG editor for information. This policy and procedure will be new to most journal editors and may require direct discussion with the journal editor. Problems encountered in this process should be noted to the CRG editors.

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, a charitable international organisation whose purpose is preparing, maintaining, and disseminating systematic reviews of the effects of health care. As outlined in the appended policy, Cochrane reviews cannot be subject to the exclusive copyright requested by some journals. This policy permits journals to publish reviews, with priority if required, but permits the Cochrane Collaboration to also publish and disseminate such reviews.

References

505. 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 in any scientific endeavour, the methods to be used should be established beforehand. Since reviews are retrospective analyses of data collected by others, it is important to make the process as rigorous and well defined as possible {448}. However, it is also necessary to maintain a practical perspective.

Just as protocols for RCTs 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 based on the results. *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 8.8) 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 drafts of the appropriate sections of the final report. The sections of the report that should be (at least partially) completed to register a planned review are:

- Cover sheet
- Background
- Objectives
- Selection criteria
- Search strategy
- Methods

The content of these sections for a completed report 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

The protocol should begin with a brief synthesis of the underlying biology, psychology or sociology and healthcare issues that provide the motivation and rationale for the review.

Systematic reviews can, in general, be motivated by several 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, they are intended first and foremost to provide unbiased, up-to-date summaries of what we know and do not know about the effects of different forms of healthcare. They should help people make practical decisions about healthcare. This has important implications for deciding whether 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 a 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.
- 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 outcomes that are likely to be important to decision makers.
- It is not helpful to include in a review evidence where the risk of bias is high, 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.
- Results should be presented in such a way that their applicability in different contexts can be assessed by decision makers.
- Reviewers should bear in mind that different people might make different decisions based on the same evidence (for good reasons). The primary aim of a Cochrane Review 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.

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 questions (objectives) along with the proposed search strategy and plan 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. For examples of protocols see *The Cochrane Database of Systematic Reviews* {444}.

References

444. Cochrane Collaboration. Cochrane Database of Systematic Reviews. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996, Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

448. 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 {448}. This is best done by asking clearly framed questions. Such questions are essential for determining the structure of a review {449, 37, 450}. 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 {451, 452}. 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 chemotherapeutic agent given in a particular way for any stage of breast cancer. In the former example the stage and severity of the disease in question would be defined very precisely, whereas the treatment formulation would be defined very precisely in the latter example.

An overview of the key components of questions follows with examples of useful issues to consider for each component. Reviewers need to be careful that they understand the terms and language that are used in various countries to describe these components.

4.2.1 What types of people?

It is often helpful to consider the types of people that are of interest in two steps. First, define the disease or condition that are of interest. Explicit criteria sufficient for establishing the presence of that 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 based on factors such as age, sex, race, educational status, or presence of a particular condition such as angina or shortness of breath. One might also be interested in a particular setting based on 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 solid ground. For example, focusing a review of the effectiveness of mammographic screening specifically 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 on a particular subgroup of people on the basis of age, sex or astrological sign simply because of personal interests when there is no underlying biologic or sociological justification should be avoided. When there is uncertainty about whether there are important differences in effects among different subgroups of people with the disease or condition in which one is interested, it is probably best to include all the relevant subgroups and then test for important and plausible differences in effect in the analysis (see section 4.5 below and section 8.3).

4.2.2 What types of comparisons?

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 types of control groups that are acceptable for 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 will be discussed further in the section on assessing the validity of studies (section 6).

4.2.3 What types of outcomes?

The third key component of a well-formulated question is the delineation of outcomes that are of interest. Are explicit criteria necessary for establishing the presence of those outcomes? If so, what criteria are sufficient? Will combinations of outcomes be considered? For example, should a study that only has data on nonfatal and fatal strokes combined be accepted for review if the question specifically relates to stroke mortality?

In general, Cochrane Reviews should include all reported outcomes that are likely to be meaningful to people deciding 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 to use chemotherapy for non-small-cell lung cancer, even if the available trials only report survival data.

While all important outcomes should be included in Reviews, trivial outcomes should not be included. Reviewers need to avoid overwhelming people with data that is of little or no importance while they must be careful not to leave out important data. For example, a wide range of adverse outcomes is often reported in drug trials. 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 in the text of the review.

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

4.2.4 What types of study designs?

Certain study designs are superior to others when answering questions. Randomised trials 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. Although questions related to diagnostic or screening strategies may be optimally addressed by randomised trials, questions about the accuracy of diagnostic tests can be addressed by cross-sectional 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 one is in matching questions to aspects of design, the less likely one is 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 and silly 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 health care {256, 123, 62}. However, the extent and even the direction of bias in non-randomised studies is often impossible to predict.

Second, it is often difficult to locate RCTs {20} and there is a risk of publication bias {19}. Trials that do not show an intervention to be effective are sometimes less likely to be published than trials that show an effect. This means that systematic reviews that fail to include unpublished trials may be biased towards overestimating the effectiveness of an intervention. Extensive efforts are going into building registers of controlled trials {444} to make it easier for reviewers to locate RCTs and protect against bias in the identification of trials. Similar efforts have not been put into identifying 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 of 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 trial registers. However, the extent to which trial registers are developed varies and it may be necessary for reviewers to conduct supplemental searches. It may also be necessary to search one or more trial registers.

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 data. 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 under "type II diabetes", "insulin", "renal function" and "long-term", relevant data is likely to be missed. On the other hand, if "insulin" or "type II diabetes" is used alone as a search term, hundreds of studies that are not germane to the question are likely to be identified.

In general, useful key components to use when searching include the condition or disease of interest and the intervention, exposure, or diagnostic test being evaluated. Although one may be specifically interested in a particular setting, studies are often not indexed by setting type in electronic data bases. Also, multiple outcomes may be evaluated in studies, some of which may be relevant yet were not considered in the indexing of the article. More detail on this topic will be provided in the next section on locating and selecting studies (section 5).

Whatever search strategies are used, it will be necessary to go through several articles 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 based on 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 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: a) determination of final criteria that will be used to select appropriate studies for review, and b) 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 will be explicated 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, one might ask a broad question regarding whether antiplatelet agents are effective in preventing thrombotic events in humans. Alternatively, another reviewer might ask whether a particular

antiplatelet agent, aspirin, is effective in decreasing the risks of a particular thrombotic event, stroke, in elderly persons with a previous history of stroke.

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 8.3). 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 conflict with the reviewer's beliefs are excluded.

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 interventions behave very differently or that various definitions of the condition of interest or the setting are associated with markedly different courses and outcomes. It is fine to mix apples and oranges, if your question is about fruit, but not if your question is, for example, 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, synthesis and interpretation of data may be particularly challenging. Broadly focused reviews can also become unwieldy to maintain and present.

One option that has been found useful is to build a broadly focused review based on a series of more narrowly focused reviews. For example, health care 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 a reviewer and a review group (e.g. how well developed the review group's trials register is, the availability of resources, time and interest).

4.6 Changing questions

While questions should be posed before initiating the actual review, it is important that these questions do not become a straight jacket which prevents exploration of unexpected issues {453}. 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. 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 study results or was it simply that you hadn't initially considered alternate but acceptable ways of defining the study populations, interventions or outcomes of interest?
- Are your search strategies appropriate for the refined question?
- Is your data collection tailored to the refined question?

References

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

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

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

62. 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.

123. 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.

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

444. Cochrane Collaboration. Cochrane Database of Systematic Reviews. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996, Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

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

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

450. Cooper HM. The problem formulation stage. In: Cooper HM, ed. Integrating Research. A

Guide for Literature Reviews. Newbury Park: Sage Publications, 1984; 19-37.

451. 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.

452. Counsell C. Formulating the questions and locating the studies for inclusion in systematic reviews. Ann Intern Med 1996; In press.

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

5. LOCATING AND SELECTING STUDIES

A comprehensive, unbiased search is one of the key differences between a systematic review and a traditional review. While electronic databases such as MEDLINE are powerful tools for locating studies, only 30 - 80% of all known published randomised controlled trials are identifiable using MEDLINE, depending on the area or specific question {20}. Non-English-language references are underrepresented in MEDLINE and only published articles are included, so there is the potential for publication bias {4, 18, 19, 20, 105, 142, 145, 169, 217, 218} and language bias (478, 479). To protect against bias and ensure that all relevant data are included in a review it is important to use multiple sources to identify studies and a systematic approach to selecting studies.

In this section we first discuss sources for identifying studies, including Collaborative Review Group trial registers, checking reference lists, personal communication, electronic databases and hand searching. We then discuss developing a search strategy, selecting studies, documenting a search strategy, keeping track of studies and, finally, contributing to trial registers.

5.1 Collaborative Review Group trial registers

In keeping with the principles guiding the development of the Cochrane Collaboration, each review group should aim to meet its needs to identify studies within its scope in a way that will ensure comprehensiveness, protect against bias and minimise needless duplication of efforts. In general, this is done through developing and maintaining registers of studies. It is up to each review group to decide on and implement effective mechanisms for ensuring that studies in its trials register find their way into reviews. Typically, the editorial team will assume at least some, if not all, responsibility for examining new studies and forwarding them to appropriate reviewers who then must assess whether the study meets the inclusion criteria for their reviews. In some cases, reviewers themselves may be asked to routinely check the review group's trials register, or some combination of approaches might be used. The approach that is used should be reported in a review, including the last date when the review group's trials register was checked.

There are important differences in how review groups have developed their registers and how well developed they are (see Group details for CRGs in CDSR) {444}. Irrespective of this, reviewers share responsibility with their editorial team for ensuring that as many relevant studies as can be found are included in their reviews. The primary source of studies for Cochrane Reviews is a review group's trials register. However, in addition to ensuring that all relevant studies in the register are included in a review, reviewers should consider using appropriate additional strategies to further ensure that all relevant studies are included. In developing a search strategy, it is important to consult with the coordinator of the review group to avoid unnecessary duplication of effort and build upon (and contribute to) familiarity and efficient use of alternative databases and other search strategies.

5.2 Checking reference lists

Reviewers should check the reference lists of all relevant articles that are obtained. This is generally an efficient means of identifying studies, although it can be time consuming. Additional, potentially relevant, articles that are identified should be retrieved and assessed for possible inclusion in the review. The potential for reference bias (a tendency to preferentially cite studies supporting one's own views) when doing this should be kept in mind and guarded

against by using multiple other strategies {160, 205}.

Some of the most convenient and obvious sources of trial references are existing reviews. Copies of previously published reviews on the topic of interest should be obtained and checked for references to the original studies. The process of following up references from one article to another is sometimes referred to as pearling, the ancestry approach, or citation chasing. The Cochrane Library includes the York Database of Abstracts of Reviews of Effectiveness (DARE). DARE provides information on previously published reviews of the effects of healthcare. It includes four main types of records:

- Structured abstracts - Abstracts assessing and summarising previously published systematic reviews judged to be of good quality.
- ACP Journal Club abstracts of reviews - Abstracts of reviews produced by the American College of Physicians Journal Club.
- INAHTA records - References to reports prepared by members of the International Network of Agencies for Health Technology Assessment (INAHTA), many with structured abstracts.
- Source records - references to other published systematic reviews. These records in general have not been quality assessed by the Centre for Reviews and dissemination.

MEDLINE, EMBASE and other bibliographic databases can also be used to identify review articles. In MEDLINE review articles are indexed under Publication Type as "Review". If many reviews are found, it is often helpful to select out those reviews that are coded as "Meta-analysis" or "Review, Academic" under Publication Type.

5.3 Personal communication

Colleagues can be an important source of information on recent trials that have not yet been published or on older trials which were never published. These informal channels of communication can sometimes be the only means of identifying unpublished data and "grey literature" such as technical reports, conference proceedings or dissertations. Formal letters of request for information can also be used to identify studies. One way of doing this is to construct a comprehensive list of relevant articles along with the inclusion criteria for the review and send a letter to the first author of each paper, asking the authors if they know of any published or unpublished studies relevant to the area not included in the list. It may be desirable to send copies of the same letter to other experts and pharmaceutical companies or others with an interest in the area. Examples of letters used as part of a search for studies are included in the Library of Examples.

5.4 Electronic databases

5.4.1 Cochrane Controlled Trials Register

The Cochrane Controlled Trials Register (CCTR) is part of the Cochrane Library and contains reference information to randomised controlled trials and controlled clinical trials in health care. These references are contributed by review groups and other entities that have identified the trials through various search mechanisms. Both the content and the indexing of CCTR are evolving to enhance its usefulness to reviewers as a source for identifying trials.

5.4.2 MEDLINE and EMBASE

Index Medicus and Excerpta Medica are indexes of healthcare journals that are available in electronic form as MEDLINE and EMBASE, respectively. These reference databases can be searched by subject terms assigned by indexers employed by the publishing organisation. In theory this can be a great advantage. It standardises terminology applied to a specific concept and allows articles referring to the same concept to be identified by searching for a single subject term. Using the appropriate search 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 article. However, when conducting a systematic review, the precision with which subject terms are applied to references should be viewed with healthy scepticism. Authors may not describe their methods or objectives well, indexers are generally not experts in the field of the article that they are indexing, and the indexers are subject to the normal mistakes people make when working rapidly at a rather tedious job. Strategies for getting around this limitation are discussed below. Another limitation of subject indexes is that they are incomplete. Not all journals are covered by either index, and articles from a portion of the included journals are included on a selective basis.

The overlap in journals covered by MEDLINE and EMBASE is approximately 34% {468}. The actual degree of reference overlap depends on the topic with reported overlap values in particular areas ranging from 10% to 75% {175, 468, 469, 470, 471}. Studies comparing searches of the two databases have generally concluded that a comprehensive search requires that both databases be searched. The same studies found that although MEDLINE and EMBASE did not return the same sets of references, they both returned roughly similar numbers of relevant references. Given that search results from each database can be expected to be of approximately equal value, the choice of which to search first may often be a matter of cost with MEDLINE usually the less costly option.

Controlled-vocabulary search terms are not standardised between databases so the search strategy will need to be customised for each database. One method of quickly identifying the appropriate controlled-vocabulary terms for a database is to retrieve from the database articles that have met inclusion criteria and note common text words and the terms the indexers applied to the articles.

5.4.3 SCISEARCH

Science Citation Index, and its electronic counterpart Scisearch, list where a published article was subsequently cited. Use of this technique requires that at least one key article exists on a topic and that those reporting subsequent research cite the key article or articles. It is a way of searching forward in time from the publication of an important article.

5.4.4 Registers of clinical trials

Registers containing references to planned, ongoing or completed trials which were registered prospectively can be a valuable source of trials. Identifying trials at inception and not on the basis of publication status reduces the risk of publication bias that may affect the results of a review. A list of registers of clinical trials is included as an appendix (appendix 5a).

5.4.5 Other bibliographic databases

Occasionally other bibliographic databases may provide a useful additional source for identifying trials for a specific review. *Online Databases in the Medical and Life Sciences*, a

directory of 795 databases, can be used to locate and assess the relevance of additional databases {94}.

5.5 Hand searching

To facilitate the location of all published trials the Cochrane Collaboration has organised extensive hand searching efforts, i.e., manually examining each issue of a journal and reading each title/abstract/body of an article sufficiently to determine whether the article is a randomised controlled trial (RCT), a controlled clinical trial (CCT) or a meta-analysis. The Baltimore Cochrane Center is co-ordinating the development of the International Register of Clinical Trials for the Collaboration, which includes co-ordination of the hand searching efforts as well as collection of the results of electronic searches. By December 1995, the Baltimore Cochrane Center had submitted approximately 43,000 references to the United States' National Library of Medicine for coding in MEDLINE as either RCT or CCT, thereby tripling the number of trials that can be identified using those terms. (See appendix 5b for NLM and Cochrane definitions of RCT and CCT.)

Hand searching efforts are being co-ordinated to avoid duplication of work. Each journal should be hand searched only once and the results of the search submitted to a common database or databases, such as MEDLINE and the Cochrane Controlled Trials Register.

5.6 Developing a search strategy

The process by which studies will be selected for inclusion in a review should be described in the protocol. More rigorous procedures reduce the risks of error and bias, but also increase the resources and time involved in preparing a review. The selection of studies for inclusion in a review is a funnelling process that involves multiple stages. When reviewing the results of an electronic search, the first stage involves assessing titles and abstracts to determine whether each article might meet predetermined eligibility criteria. If, given the information available, it can be determined that an article definitely does not meet inclusion criteria, it can be excluded. If the title or abstract leave room for doubt in the reviewer's mind that the article cannot definitely be excluded, the full text of the article should be retrieved. Review of the full text may lead the reviewer to exclude the study because it does not meet inclusion criteria. If the article is not excluded, it may then be formally abstracted as described in section 7. At each stage of the selection process, it is important to err on the side of over-inclusion because once a trial has been excluded from the selection process it is unlikely to be reconsidered. Questionable articles which are included at one stage can be excluded at a latter stage when more information on the study is available.

There are diminishing marginal returns for search efforts, i.e., after a certain stage has been reached, each additional unit of time invested in searching returns fewer references that are relevant to the study. 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 review group's trials register is developed and the resources that are available.

Librarians who have specialised in electronic searching can often be helpful in developing and executing a search strategy. Librarians may lack knowledge in the subject area of a review, and they may be trained to run searches that are precise (i.e. include few non-relevant

citations) but not necessarily comprehensive. Therefore, it is important to work closely with a librarian in deciding which databases to search and what search terms to use in each database. It is often helpful to provide one or more published reviews on the topic and synonyms for key concepts. Ideally reviewers should be present when the search is conducted. There are often costs associated with each database and each downloaded record and judgements about what to download often need to be made while the search is being conducted. Increasing the comprehensiveness of a search entails reducing its precision and retrieving more non-relevant articles. Although it is important to be as comprehensive as possible, it is always necessary to strike a balance between comprehensiveness and precision. Developing a search strategy is an iterative process in which the terms that are used are modified based on what has already been retrieved.

A good approach to developing a comprehensive search strategy is to begin with multiple terms that describe the condition or disease of interest and join these together with the Boolean "OR" operator. The results then can be narrowed down by using terms (joined together with "OR") that describe the interventions being evaluated and pulling out only articles that address both the condition AND interventions of interest. Although a research question may address particular populations, settings or end points, these concepts are often not well indexed by the electronic databases and therefore generally do not lend themselves well to searching. To gain an understanding of what is and is not well indexed by a particular database it is useful to review the controlled-vocabulary terms applied to a group of relevant studies. In general, both index terms and text words (found in the title or abstract) should be used.

Methodological terms can also be used to focus a search. The UK Cochrane Centre has designed a search strategy for RCTs that includes controlled-vocabulary and free text terms such as "randomized-controlled-trial," "clinical trial" and "placebo" (Appendix 5c). These terms can be combined (using the Boolean "AND" operator) with terms for the condition of interest alone or after focusing in on articles that address the condition AND interventions of interest (see figure below). insert Venn diagrams showing 3 options: condition/disease + intervention, condition/disease + RCT/CCT, all three).

While MEDLINE and EMBASE include citations from 1966 and 1974, respectively, Index Medicus and Excerpta Medica, the print versions of these databases, include citations from 1879 and 1948, respectively. Manually searching the earlier printed subject indexes may occasionally be worthwhile if relevant studies were published before the onset of the electronic databases.

5.7 Selecting studies

All the articles that are identified must then be assessed to see whether the inclusion criteria for the review have been met. Reviewers must decide:

- whether more than one reviewer will assess the relevance of each article
- 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 judgement. To help ensure that these judgements are reproducible, it is desirable for more than one reviewer to apply the inclusion criteria to all the potentially relevant articles that are retrieved. However, the amount of judgement needed varies from review to review. Some reviewers may decide that the judgements that must be made about which studies to include in a review are quite straightforward. The use of multiple reviewers to assess relevance may appear frivolous in these cases. Whatever the case, the number of people assessing the relevance of each trial should be stated in the Methods section of the review (if it is not stated in a description of the methods used by all the reviewers in a particular collaborative review group).

Experts in a particular area frequently have pre-formed opinions that can bias their assessments of both the relevance and validity of articles {15, 233}. 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 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 (480). However, this takes much time, and may not be warranted given the resources required and the uncertain benefit in terms of protecting against bias.

Disagreements about whether a trial 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 reference the study 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.

5.8 Documenting a search strategy

The search strategy should be described in sufficient detail that the process could be replicated. Documentation of the search strategy should include the main sources that were used to locate studies, the search strategy for each database that was searched; temporal constraints (e.g., 1966 to 1996); and language constraints, if any (although these language constraints should, so far as possible, be avoided). As noted above, how and when the review group's trials register was last checked should be reported. Examples of descriptions of search strategies can be found in CDSR. There is wide variation in the extent to which reviewers have conducted searches beyond what is included in the review groups trials register and in how detailed the description of the search strategy is. While the ideal may be a register that is sufficiently developed that additional searches by reviewers are not necessary, most review

groups have not reached this stage of development. Because Cochrane Reviews are kept up-to-date, searches are ongoing and both the search strategies that are used and the reporting of these strategies can continue to be improved. To facilitate the documentation of search strategies it is suggested that a print-out of the search strategy (i.e., not re-typed) should be used to describe what was done, together with the title of the database, the date when the search was run and the years covered by the search. For example:

> MEDLINE on SilverPlatter was searched January 1996, period 1966 to date, using the following search strategy: (followed by a print-out of the search strategy).

The search strategies used to develop a trial register are described in the review group module. These should not be reported in the text of Cochrane Reviews.

5.9 Keeping track of studies

When retrieving references from multiple sources or when running multiple search strategies on the same database it is common to retrieve the same reference more than once. A reference management system eases the work of identifying these duplicate titles and assists in tracking the retrieval status of each article. Interlibrary loan and document retrieval services can lose requests or may not deliver an article for months after it was requested. It is also a good idea to document the reason for exclusion of each article that was identified but ultimately not included in the review. When searching multiple databases, it can be helpful to know which references were identified in each database. This can help guide decisions about which databases to search routinely for maintaining the review group's trials register and, if appropriate, individual reviews.

Specially designed reference management databases such as ProCite and Reference Manager are useful and relatively easy to use. General database packages such as Access and FoxPro include powerful query capabilities and lend themselves well to customisation but require at least minimal programming and database design skills to set up. Currently the Baltimore Cochrane Center is using ProCite and requires that hand search results are submitted in ProCite format or as an ASCII text file. Review groups are using several different software programs to manage their trials registers and there is not a single best choice. Specifications are being developed for ("TrakMan") software that will address the specific needs of Collaborative Review Groups and should ease the task of managing references for reviewers, coordinators and editors in review groups.

5.10 Contributing to trial registers

To ensure the long-term viability of review groups and the Cochrane Collaboration it is important that efforts to develop and maintain trials registers are co-ordinated effectively and efficiently. This is necessary to maximise access to studies by reviewers, thereby helping to ensure that reviews are comprehensive, up-to-date, and unbiased. It is also necessary to minimise duplication of effort and the burden that is placed on reviewers to conduct searches, both initially and on an ongoing basis.

To obtain the maximum benefit and minimise wasted efforts it is important that reviewers work closely with review group coordinators in developing and implementing the searches they conduct. Specifically:

- Developing a search strategy should begin with consulting the review group's trials register and familiarising oneself with the methods used to develop and maintain the register. The editorial team of a review group should be able to offer guidance and sometimes practical help with developing and implementing a search strategy.
- Studies identified by reviewers that are within the scope of the review group should be added to the review group's trials register. Each review group should have in place appropriate mechanisms for facilitating the contribution of studies from reviewers to its trials register.
- Reviewers may wish to contribute in specific ways to developing and maintaining a trials register beyond what is required for a specific review; e.g. by hand searching journals or conference proceedings. Any hand searching that is undertaken should be co-ordinated with the coordinator for the review group.
- Personal communication with experts in a field might contribute to the development of a review group's trials register as well as to a specific review. It may be useful to discuss specific plans to contact experts with the coordinator or editors of a review group to help co-ordinate these efforts.
- Reviewers can help to inform and revise the strategies used to maintain a trials register based on their experience conducting searches for specific reviews.

References

4. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. J R Statist Soc A 1988; 151:445-63.

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

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

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

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

94. Online Databases in the Medical and Life Sciences. New York: Cuadra/Elsevier, 1987.

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

142. 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.

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

160. Gotzsche, P.C. Reference bias in reports of drug trials. BMJ 1987; 295:654-6, 1987.

169. 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.

175. 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. Pharm Weekbl Sci 1992; 14:316-20.

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

217. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. Stat Med 1987; 6:11-29.

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

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

468. Smith BJ, Darzins PJ, Quinn M, Heller RF. Modern methods of searching the medical literature. Med J Aust 1992; 157(9):603-11.

469. Ramos-Remus C, Suarez-Almazor M, Dorgan M, Gomez-Vargas A, Russell AS. Performance of online biomedical databases in rheumatology. J Rheumatol 1994; 21(10):1912-21.

470. Rovers JP, Janosik JE, Souney PF. Crossover comparison of drug information online database vendors: Dialog and MEDLARS. Annals of Pharmacotherapy 1993; 27(5):634-9.

471. Odaka T, Nakayama A, Akazawa K, et al. The effect of a multiple literature database search a numerical evaluation in the domain of Japanese life science. J Med Syst 1992; 16:77-81.

6. CRITICAL APPRAISAL OF STUDIES

Critical appraisal 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. Parameters that warrant appraisal are those related to applicability of findings, validity of individual studies, and certain design characteristics that affect interpretation of results. Applicability, 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 studies as well as the reviewers define the people, interventions, and outcomes of interest.

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 anti-ischemic agents on symptoms of chest pain in adults with prior myocardial infarction. However, the trials may examine different preparations and doses of anti-ischemic 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 what and how to abstract data related to applicability and design factors likely to affect the interpretation of findings will be given in the next section (section 7). The remainder of this section will focus on critically appraising the validity of individual studies included in a systematic review. As most Cochrane Reviews focus on randomised trial data, we will concentrate on how to appraise the validity of such data.

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 {240}. An important issue that should not be confused with validity is precision. Precision is a measure of the likelihood of 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 using meta-analysis. More precise results are given more weight.

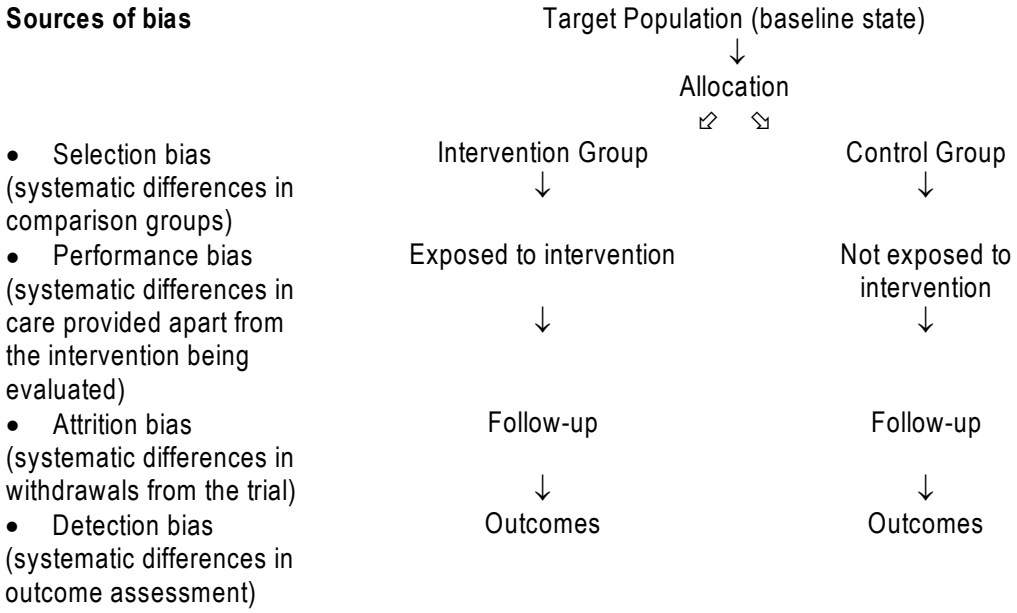
Variation in validity can explain variation in the results of the studies included in a systematic review. More rigorous studies may be more likely to yield results that are closer to the "truth". Quantitative analysis of results from trials of variable validity can result in "false positive" conclusions (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 provide less precise or biased estimates of an intervention's effect {135}.

It is important to systematically complete critical appraisal of all studies in a review even if there is no variability in either 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 systematic review's conclusions would not be nearly 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 errors 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 {240, 491}. There is, however, a logical basis for suspecting such relationships and good reason to assess these four potential biases {81}.



6.3 Selection bias

One of the most important biases that may distort treatment comparisons is that which can result from the way that comparison groups are assembled {486}. 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 random assignment by those involved in recruitment.

The process of concealing assignment until treatment has been allocated has sometimes been referred to as "randomization blinding" {123}. 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 allocation of treatments) reduces performance and detection biases. Second, from a practical standpoint, concealing treatment assignment 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 whatever outcomes are being compared. In contrast, control of detection bias is often outcome-specific and may be accomplished successfully for some outcomes in a trial but not

others. Thus, blinding up to allocation and blinding after allocation of treatment are addressing different sources of bias, are inherently different in their practicability and may apply to different parts of a trial. 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 treatment has been assigned.

Empirical research has shown that lack of adequate allocation concealment is associated with bias {123, 62}. 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, trials can be judged on the reported method of allocation concealment. Information should be presented that provides some assurance that allocations were not known until the point of allocation, at least. The method for assigning participants to treatments should be robust against patient and clinician bias and its description should be clear. The following are some approaches that can be used to assure adequate concealment schemes.

- centralised (e.g., group assignment 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 group assignments in a locked unreadable computer file that can be accessed only after entering characteristics of an enrolled subject
- sequentially numbered, sealed, opaque envelopes

Other approaches may include statements that imply an approach like 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 seem to provide adequate concealment.

Approaches to allocation concealment that should be considered clearly inadequate include alternation, case record numbers, dates of birth, day of the week, and any allocation procedure that is entirely transparent before assignment, such as an open list of random numbers of assignments. When trials 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 Review Manager (RevMan) they are required to 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 care provided to comparison groups other than the intervention of interest. 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 indeed important in protecting against bias {62, 487, 488}. 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 {489, 490}. Furthermore, there is evidence that participants who are aware of their assignment status report more symptoms, leading to biased results {488}. 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 treatment? Were those providing care unaware of the assigned therapy?

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

Blinding is likely to be particularly important in research with subjective outcome measures such as pain {62, 487, 488}. 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 binding. In some areas it may be desirable to use the same criterion across reviews, in which case a review group might want to agree to a standard approach for assessing blinding {89, 62, 480, 491}.

6.5 Attrition bias

Attrition bias refers to systematic differences between groups in losses of participants from the study. It has sometimes been referred to as exclusion bias but we call it attrition bias to prevent confusion with pre-allocation exclusion and inclusion criteria for enrolling people. 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 trial reports on attrition after allocation has not been found to be consistently related to bias {62}. 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 authors.

6.6 Detection bias

Detection bias refers to systematic differences in outcome assessment. Trials that blind outcome assessors regarding treatment allocation should logically be less likely to be biased than trials that do not. However, at least two studies have failed to demonstrate empirically a relationship between blinding of outcome assessment and study results, possibly due to inadequacies in trial reports {62, 245}.

Somewhat different from bias in outcome assessment is bias due to selective reporting of results. This source of bias may be important in areas where multiple outcome measures are used, as in evaluations of treatments for rheumatoid arthritis {157}. Therefore, reviewers may want to consider specification of predefined primary outcomes and analyses by the investigators' indicators of validity. Alternatively, selective reporting of results 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 8.8). However, having used several explicit criteria to assess validity, it is desirable to summarise these somehow 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:

	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 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 have identified 25 scales and 9 checklists that have been used to assess the validity and "quality" of randomised controlled trials {240, 491}. These scales and checklists include anywhere from 3 to 57 items and take from 10 to 45 minutes to complete. Almost all 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. 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 controlled trials, the relationship between such a score and the degree to which a trial 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 review groups 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 weights) for each item. While this approach offers appealing simplicity, it is not supported by empirical evidence {62, 146}.

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. They may carry a greater risk of confusing the quality of reporting with the validity of the trial. 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 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 {492}. 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 {492}.

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 {493, 494, 495, 496}. 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 trials, cohort studies and case-control studies and relate to the extent that those entered into a study are appropriately accounted for in the

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 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 exposure to other factors that could affect outcomes. In addition to considerations of blinding, which are like those in 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 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 advised.

6.9 Application of critical appraisal criteria

Several basic decisions must be made regarding the critical appraisal studies, like 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 reviewers' backgrounds 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 and minimise errors and improve reliability of findings, but more than one creates the potential for disagreement among reviewers. When multiple reviewers are planned, there should be an explicit procedure or decision rule identified *a priori* for identifying and resolving disagreement. As a 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). The next section (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 test any assessment criteria that are planned on a pilot sample of articles to ensure that the appraisal criteria can be applied consistently. A suggested sample would be three to six papers that span a range of low to high-risk bias.

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 {233}, they may nonetheless give more consistent assessments of the validity of trials than persons without content expertise {480} and they may 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.

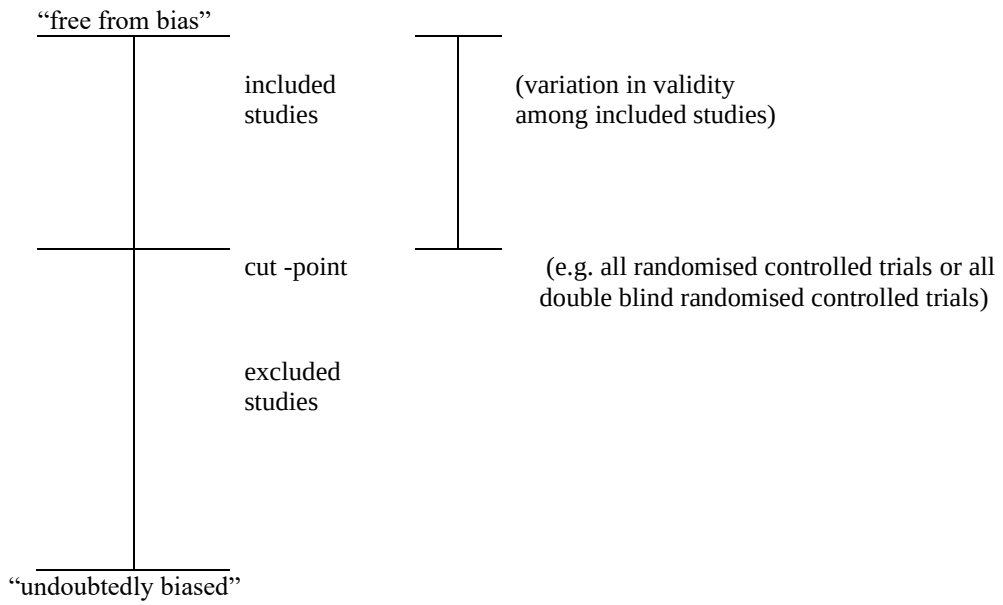
Reviewers must also decide whether those assessing study validity will be blinded to the names of the authors, institutions, journal, and results of a trial when they apply critical appraisal criteria to the methods. Some empirical evidence suggests that blind assessment of reports might produce lower and more consistent scores than open assessments {480}. However, such assessments are very time consuming. Reviewers must weigh the potential benefits of blind assessments against the costs involved in deciding whether to blind the reviewers. Further research is underway comparing blind and open assessments of trial 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 trials
- in sensitivity analyses
- as weights in statistical analysis (meta-analysis) of the 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 trials 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 cut-point for inclusion. With a sufficiently high cut-point, variation in validity among included reports may be moot.

There are several methods to examine whether validity may explain differences among study results {135}. 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, preferably one specified *a priori*. This approach can be used whether the results are heterogeneous, as 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 (a type of "cumulative meta-analysis"), examining the impact on the overall results as trials of decreasing validity are included (see section 8.6.2).

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 8.6.1). 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 any of the available "quality" scales to differences in the risk of bias.

It is possible using RevMan 3.0 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. A function to facilitate sensitivity analyses also has not yet been implemented, but it is possible to do these by deleting or adding studies. RevMan does not include an option for weighting studies by methodological validity and neither cumulative meta-analysis or meta-regression is possible using RevMan 3.0.

6.11 Limitations of critical appraisal

There are two major difficulties with critically appraising the validity of studies. The first is inadequate reporting of trials {67, 213, 497}. 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 clarifying data from investigators, but this may be difficult. The application of standards for reporting trials {67, 497} can facilitate critical appraisal.

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 trial 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 also 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.

References

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

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

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

89. 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 (eds). Effective Care in Pregnancy and Childbirth. Oxford: Oxford University Press, 1989; 39-65.

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

135. 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.

146. 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.

157. 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

213. 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

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

240. 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.

245. 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.

486. Kleijnen J, Gotzsche P, Kunz RA, Oxman AD, Chalmers I. So what's so special about randomization? In: Chalmers, I, Maynard, A (eds): Non-Random Reflections on Health

Services Research. London: BMJ. In press.

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

488. 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.

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

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

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

491. 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.

492. 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.

493. Freemantle N, Bero L, Grilli R, Grimshaw J, Oxman A (eds.) Effective Professional Practice Module of The Cochrane Database of Systematic Reviews. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996; Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

494. 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.

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

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

497. 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

The data collection form is a bridge between what has been reported by primary investigators (journal articles, project reports, personal communications, etc.) and what is ultimately reported by a reviewer. The data collection form serves at least three important functions. First, as alluded to in earlier sections, the data collection form is directly linked to and is therefore a visual representation of the formulated review question and planned critical appraisal of included studies. 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 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 appraisals 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 with 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 files from electronic data collection forms to RevMan, technical expertise beyond the scope of most reviewers is needed.

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 a pilot 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. As mentioned in section 5, review groups currently use several different reference management programs, such as Reference Manager and ProCite, to manage their trials registers. Software that will address the specific searching and reference management needs of Collaborative Review Groups is under development. 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 8.6).

The software that is essential for every Cochrane Review is RevMan. This software facilitates data management, quantitative analyses, and report generation. RevMan is provided at no cost to reviewers in Collaborative Review Groups. RevMan workshops are offered by Cochrane Centres. Technical support for RevMan is provided through coordinators of review groups, who are supported by Cochrane Centres and the Software Development Group.

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, 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 included studies table, as footnotes 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 must be revised. Coding the form with a revision date or version number reduces the chances of erroneously using an outdated form.

Each included study must be given an identifier which is entered into RevMan when entering information about references. The same identifier should be used for all reports referring to the same study. The identifier is used in the included and excluded studies tables and in data tables in RevMan. Reviewers may choose to use the last name of the first author and the year of publication, the country or location where a study was done, or other informative identifiers. Studies can be sorted alphabetically by these codes.

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 should be 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 (otherwise, data may be abstracted from studies that will be excluded and this will be a waste of resources).

Cochrane Reviews include an excluded studies table. This table is not meant to include information on all studies that were considered in the search and selection process. Rather it pertains to studies that appear to meet the inclusion criteria and which others might believe to be relevant, but upon closer inspection must be 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 critically appraising each study, it is necessary to code specific study characteristics. These characteristics can be categorized into groups that match information that will be entered into RevMan: methods, participants, interventions, and outcomes. Information under participants often includes data relevant to the study setting and diagnostic criteria for the condition of interest, if pertinent. Remember, it is easy to get carried away wanting to ensure sufficient information is collected. 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 the previous 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 to collect 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 the validity criteria that are used.

7.5.2 Participants

Characteristics of participants may vary substantially across studies; some review groups 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 different ethnic populations, then this information should be collected. If treatment effects are thought constant over ethnic groups, and if such information won't be useful to help apply results, it should not be collected. Typical items that are useful for assessing applicability include age and gender. 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 health care intervention studies are: acute care hospitals, emergency facilities, offices or clinics, extended care facilities such as boarding and nursing homes, and communities. Sometimes trials are conducted in different geographical regions that have important differences in cultural characteristics that could affect delivery of an intervention and 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 code trials regarding definition and severity of heart failure (e.g., systolic or diastolic dysfunction, severe systolic dysfunction with ejection fractions of < 20%, etc.) and 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 twenty-four 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 health care 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; 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 abstraction form to design. Researchers of primary studies often report more than one outcome (mortality, morbidity, quality of life, etc.), may report the same outcome using different measures, may include outcomes for sub-groups 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 can be anticipated 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 8.9.5 and section 8.6.4).

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 (n), the mean and the standard deviation (SD). 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 formats of the coding that is used 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 to or near 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 training if the form was designed by a different reviewer.

7.7 Pilot testing and form revisions

All forms should be pilot tested using a representative sample of the studies to be reviewed. The pilot test is likely to identify data that are not needed or are missing. Reviewers may provide feedback that certain coding instructions are confusing or incomplete (do not adequately describe all of the types of responses). When multiple reviewers are participating on a project, there may need to be a consensus among reviewers 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 must 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 persons 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 a spreadsheet or data base program. Keeping the "consensus" information separate is essential for assessing the reliability of coding.

It may not be important to formally examine reliability for all 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.

8. ANALYSING AND PRESENTING RESULTS

8.1 Do not start here!

Please read sections 3 - 6 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), identifying, selecting and critically appraising studies (section 4 and section 5) and collecting data (section 6).

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 bottom line or the diamond used at the bottom of a graph to sum up the results of a meta-analysis. Sometimes reviewers and users of reviews in their rush to get to the diamond 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.

8.2.1 Why use statistics in a review?

To prepare a review, investigators must collect data from individual studies, just as they must collect data from individual patients in primary studies. In either case, statistical methods can be used to analyse and summarise data. If used appropriately, they 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 reviews is to compare the number of "positive" studies with the number of "negative" studies. With such a "vote counting" approach, a study may be counted as "positive" in one review and "negative" in another, depending on how the results are interpreted by the reviewers. 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" {131, 103}.

Another error that can easily occur when statistical methods are not used is inappropriate weighting of the results of the individual studies. For example, vote counting gives studies equal weight, ignoring differences in the size and quality of the studies. 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 meta-analysis are valid, any more than it does for a primary study. Moreover, like any tool, statistical methods can be misused.

8.2.2 When not to use statistics in a review

A systematic review that does not include a statistical analysis can be just as valuable as a meta-analysis. Unfortunately, reviewers sometimes do not believe this and they feel they will have failed if they cannot do a meta-analysis. This can lead people to do silly things.

As noted in section 3, Cochrane Reviews are intended first and foremost to provide unbiased, up-to-date summaries of what we know and do not know about the effects of different forms of healthcare. The aim is to help people make practical decisions about healthcare. The results of statistical analyses (meta-analysis) contribute to this in an important way, but this is only one element of a summary, often not the most important and can be misleading if done inappropriately.

There are two important reasons for not using statistics in a review. The most important reason for not using meta-analysis is the lack of relevant, valid data. This should be obvious, but by focusing undue attention on meta-analysis reviewers sometimes are led to include marginally relevant or poor-quality studies, use highly questionable data, and make tenuous assumptions. This is based on the notion that some data (whatever the quality) is better than none. This is not always the case. Misleading data is worse than no data. Moreover, it can be an important help to people making decisions to know that there are no reliable data upon which to base a decision. This can result in people undertaking studies to produce the data that are needed, and it can enable people to take decisions that they might not otherwise have taken.

The second important reason for not doing a meta-analysis is that it sometimes does not make sense, i.e., some meta-analyses are difficult to understand and may combine study results in misleading ways. For example, a summary of controlled trials to improve prescribing could be very helpful to many people with an interest in this problem. However, an estimate of the average effect of interventions to improve prescribing would likely be meaningless and might be misleading.

The second reason for not doing a meta-analysis can sometimes be anticipated in advance and other times may only become clear until after data are collected and important differences in study participants, interventions, outcomes, or designs are recognised. The latter has unfortunately been referred to as "failed meta-analysis" {485}. There is no reason to consider an appropriate decision not to use meta-analysis as a failure!

In general, if it is not clear how the results of a meta-analysis will help people making decisions, it should not be done.

8.2.3 What does an 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 what 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 investigate differences among studies, to derive estimates of effect across studies, and to conclude how much confidence should be placed in those estimates.

8.2.4 What 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 3). 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, or 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. Considerations about how to formulate the questions that a review addresses are addressed in section 3. Essentially the same considerations apply to deciding what comparisons to make, what outcomes to combine and what key characteristics (of participants, interventions, outcomes, and study design) to consider when investigating heterogeneity (variation in effects). These considerations must be addressed when setting up a Table of Comparisons using the Review Manager software (RevMan) and in deciding what information to put in the Table of Included Studies.

8.2.5 What data 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 CDSR, but the default display for each review will be determined by the reviewers.

Because concealment of allocation is a validity criterion that applies to all outcomes, and there is good evidence to support the logical arguments for using this criterion, it is assigned by reviewers when study references are entered in RevMan and automatically assigned to studies in tables of results as one characteristic by which studies can be sorted. 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 weight) 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 8.9.1.

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 useful to people making decisions. So far as possible, results should be transformed, or data should be sought from investigators so that all of the study results can then be presented in that 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 8.8 below).

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). Often information such as this is easier to understand and more helpful to people making decisions when it is dichotomised. However, deciding on a cut point can be quite 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 the data in a table for "other data". Another option is to present all the data as standardised mean differences. Alternatively, the log-odds-ratio can be used as a measure of treatment difference {455}.

Several 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 what we know about the effects of audit and feedback on professional practice, it can be helpful to summarise all valid studies comparing the effects of audit and feedback with no intervention on objective measures of performance, even though the types of professionals, settings and outcome measures vary substantially from study to study. Although it might be possible to transform results in this case to standardised mean differences, this 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.

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 either dichotomous or continuous 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 and are not 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 possible explanations of heterogeneity.

8.3 Investigating heterogeneity and deciding on subgroup analyses

Having decided what comparisons to make and what data are needed for each comparison, the next question to ask is 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 play 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" {228}. Tests of homogeneity are formal statistical analyses for estimating the probability of observed differences in results having occurred because of chance alone. The more significant (closer to zero) the test, the less likely it is that the observed differences were caused by chance alone.

When there is "statistically significant" heterogeneity (a low probability, say less than 1 out of 10 or 1 out of 20), it suggests that the observed differences in results are likely caused by factors other than chance. When looking for possible explanations for the differences, however, such as differences in the interventions that were studied, it is important to be cautious. Even if a meta-analysis included only well-designed, randomised controlled trials, patients were not randomised to one study compared with another. Because studies usually differ in many ways, it is risky to attribute differences in results to any one factor.

RevMan will automatically test the homogeneity of the results of the individual studies being combined for each comparison of dichotomous or continuous data. For other data, as well as for dichotomous and continuous data, thoughtful consideration of differences between studies based on an understanding of the underlying biology, psychology and sociology is essential.

Irrespective of the extent of heterogeneity in results, it is frequently of interest to examine a particular category of participants in a review; for example, women, a certain age group, or those with a specific pattern of disease. These examinations, or subgroup analyses, are exceedingly common, but they are also often misleading {232, 196}. Conclusions based on subgroup analyses can do harm both when a particular category of people is denied effective treatment (a "false-negative" conclusion), and when ineffective or even harmful treatment is given to a subgroup of people (a "false-positive" conclusion). Subgroup analyses can also generate misleading recommendations about directions for future research that, if they are followed, can 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 in subgroup response is real and, consequently, when to include a conclusion based on a subgroup analysis in a review {196}. They can also be used to guide decisions about whether apparent differences in response due to differences in the intervention (such as the specific drug or dose that was used) are real.

- Is there indirect evidence in support of the difference?

Another way of asking this question is: Is the difference plausible, 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 *a priori* (in the protocol for the review) and the number of planned subgroup analyses should be kept to a minimum to avoid spurious findings. The greater the number of hypotheses tested, the greater the number of subgroup differences one will find, by chance. Decisions about which analyses to do and which ones 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 the most salient difference 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, different patients (varying in risk of adverse outcomes, for example), varying degrees of cointervention, or varying criteria for the outcome, could all explain differences in the results of drug trials. Stated simply, inferences based on between-study comparisons are based on comparisons between non-comparable groups. Even when all the individual studies are RCTs, patients were not randomized 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 3.0 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, in general, subgroup analyses based on differences between studies are discouraged as noted above.

Commonly, several comparisons for different subgroups are simply conducted without formally testing for interactions. This practice, together with only reporting subgroups within which sizeable treatment differences are found, can lead to an overestimate of the significance, as

well as of the size of the difference. It should be avoided.

Regression models, such as logistic regression {381} can also be used for analysis of interactions if the interactions are modelled by product terms. This approach 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 with this approach 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 consider it is important to formally test the significance of an apparently important difference in response in a subgroup should consider asking a statistician for help. It is not important for reviewers to understand the details of the appropriate 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 it is important that appropriate statistical methods are used.

8.4 Summarising effects across studies

The central aim of most Cochrane Reviews is to provide a reliable estimate of the effects of an intervention or type of intervention, based on a weighted average of the results of all the available good quality studies. Typically, the weight given to each study is the inverse of the variance, i.e., more precise estimates (from larger studies with more events) are given more weight {178}. It is possible to give studies more or less weight based on their methodological quality as well as their precision, but this is rarely done {135}. More often "sensitivity analyses" (see section 8.8) in which the robustness of the results is tested by excluding subgroups of studies, such as those of poorer methodological quality, are done.

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 people in the same study.

Occasionally a meta-analysis is published where an "effect size" is calculated for each study group based on the difference in outcomes before and after treatment and groups from different studies are directly compared with each other. When this is done, the results of a meta-analysis must be interpreted with a great deal of caution. Even if all of the data come from RCTs, the power of randomisation is completely lost when this approach is taken, and the data have been reduced to the equivalent of before-after studies. This approach to synthesising data is not supported by RevMan and should be avoided.

The confidence interval around an estimate derived from an appropriate statistical synthesis indicates how precise that estimate is. It is, in principle, no different than a confidence interval around an estimate derived from primary research {334}. However, one important consideration that arises in meta-analyses is whether to incorporate between study variation in estimating the confidence interval (a "random effects" approach). If there is little between study variation (i.e. the test for homogeneity results in a large *p* value, say greater than 0.10), this will make little difference. If there is significant between study variation, an analysis that ignores this (a "fixed effects" approach) will give a narrower confidence interval than one that does not. If there is significant heterogeneity reviewers should be cautious about interpreting the overall estimate derived from a meta-analysis, as well as attributing between-study differences to any one factor.

There is disagreement on how to summarise the results of studies, and whether they should be combined when there is substantial heterogeneity {80}. However, some common ground exists 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. Reviewers should attempt to explore the reasons for the heterogeneity and to explain it, but the interpretations must be cautious because the analyses are usually *post hoc*. Subgroup analyses in general are often useful but difficult to interpret (see section 8.3) {232, 196}.

One potential explanation for heterogeneity that can be investigated in an *a priori* fashion is methodological quality. Differing methodological quality in the component studies could produce both bias and accompanying heterogeneity. If there are important differences in the quality of studies, reviewers should consider excluding poorer quality studies.

Secondly, it is generally agreed that the fixed effects approach for a test of significance for the overall null hypothesis ("no effect in all studies") is statistically valid. A statistically significant result indicates that a treatment effect exists in at least one of the studies.

Thirdly, whether heterogeneity is present or not, the overall typical effect measure from a fixed effects approach is an informative average measure of treatment effect.

Fourthly, the random effects methods rely 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" methods give more weight to the smaller studies than the "fixed effects" methods and smaller studies are often of poorer quality and may be more susceptible to publication bias.

That leaves as the primary dispute the different approaches to the calculation of a confidence interval around a typical effect measure when there is clear heterogeneity. If the test of heterogeneity is statistically significant and the differences in results are practically important:

- Reviewers should consider the sources of the heterogeneity (e.g., differences in dose, timing of treatment, length of treatment, participant characteristics). In particular, reviewers should explore whether differences in the control of bias among the studies explain the heterogeneity (e.g., concealment of randomisation).

- If the heterogeneity can be explained, reviewers should consider (with due caution) the option of presenting the results of subgroup analyses.

If the heterogeneity cannot be explained, reviewers should consider the following options:

- Do not aggregate the studies at all.
- Use the fixed effects model with appropriate cautious interpretation.
- Use the random effects model with appropriate cautious interpretation.
- Use both the fixed and random effects models as an additional aid in explaining the uncertainty around an analysis with heterogeneous studies.

- Whatever is done, reviewers should clearly explain what was done and why.

Various methods are available for analyses using fixed and random effects models. Generally, not much difference exists between the results from the various fixed effect approaches {224}. Less is known about the various random effects approaches.

8.5 Statistical methods available in Review Manager and CDSR

Meta-analyses use a variety of techniques. Unfortunately, there is not one "correct" technique. The choice of technique depends on the nature of the data being analysed. Fortunately, as with other uses of statistics, a conceptual understanding of the principles is more important than detailed knowledge of the specific techniques. The statistical methods available in RevMan 3.0 include:

Type data	Summary statistic	Model	Method

Dichotomous	odds ratio (O-E)	fixed effect	Peto
		random effects	Mantel-Haenszel 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

For dichotomous data there are pros and cons for each of the summary statistics: odds ratio, relative risk, risk difference (see appendix 8a on effect measures for dichotomous data). Although the odds ratio has been commonly used in meta-analysis, there are concerns about the potential for invalid clinical interpretation of the odds ratio {64}. The odds ratio has statistical advantages relating to its sampling distribution and its suitability for modelling {456}, 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 {64}.

The statistical method that is used and the presentation of results can be considered as separate issues. The number needed to treat (NNT), {180} for example, is a sensible way to describe results, but this does not imply that the analysis needs to be done using the NNT or the risk difference (from which the NNT can be derived directly).

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). 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). An NNT calculator is not available in RevMan 3.0 but will likely be available in subsequent versions of the software.

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. 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.

In 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 it is recommended that:

- Each review group should develop a policy regarding which statistical method(s) to use. This should be done in consultation with a statistician.
- If a reviewer elects to deviate from the review group policy, the reason for this should be reported in the review.

8.6 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 text of the review. Reviewers may want to make additional tables or graphs available (e.g., on an FTP server), but there currently is not a mechanism to publish these in CDSR.

8.6.1 Multivariate meta-analysis

"Meta-regression" can be used to assess statistically whether specific factors (co-variates) influence the magnitude of effect across studies {381}. Such analyses should best be used as hypothesis generating tools. As noted above, 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 co-variates 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 {457, 458}. The co-variates that are investigated can also be either general (e.g., the event rate in the control group) or problem specific. A variety of statistical methods including weighted least squares, logistic and hierarchical models can be used for meta-regression analyses {381, 383, 459}.

8.6.2 Cumulative meta-analysis

In cumulative meta-analysis studies are added one at a time in a specified order, e.g., according to date of publication or quality {179}. This approach can be used in sensitivity analyses (see section 8.8), for example to determine whether results have been robust over time or across studies of differing quality. Cumulative meta-analysis (and other sensitivity analyses) cannot easily be done using RevMan 3.0. Future modifications of the software will facilitate this.

8.6.3 Bayesian meta-analysis

Bayesian approaches to meta-analysis, such as the Confidence Profile Method {22, 341}, are based on the principle that each observation or set of observations should be viewed in conjunction with a prior probability describing our prior knowledge about the phenomenon of interest. The new observations alter this prior probability generate a posterior probability. Traditional statistical methods used in meta-analysis assume that nothing is known about the

magnitude of the treatment effect before a study. In Bayesian terms, the prior probability distribution is a uniform one, 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 frequently they will vary considerably. To implement credible Bayesian analyses additional development in eliciting prior probability distributions and conducting robust analyses is needed {464}.

8.6.4 Time to event (survival) analysis

Several approaches are available for extracting and analysing data from reports of studies with survival time (or time to an event) as the endpoint of interest. RevMan 3.0 does not support this type of analysis. However, it is possible to conduct and display the results of analyses for different follow-up periods, thereby approximating a survival analysis (see, for example {466}). For reviews where survival time or time to an event is particularly important, individual patient data is highly desirable (see section 10). 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 (see, for example {460}).

The analysis should be based on full survival data whenever possible. Not only is this approach more powerful {115}, it also provides insight into the course of the disease and treatment effect over time. When full survival data is not obtainable, it is still desirable to examine the distributions of survival times in those studies for which full data are available.

8.6.5 Exact calculations

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 as is done in RevMan 3.0. A better approach might be to use exact calculations. Computational methods are available for computing an exact confidence interval for the common odds ratio {50} and these are available in commercial software packages (e.g., Egret {467}).

8.7 Dose response analyses

The same principles described above for subgroup analyses can be applied to dose response and duration response analyses. Meta-regression methods can be used to estimate dose-response parameters {28, 108}. 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 trial and a consistent relationship is found across studies. While reviewers should attend to dose effects and consider this as a possible explanation for heterogeneity of results, they should be cautious about drawing conclusions about dose or duration effects based on between study differences.

8.8 Sensitivity analyses and publication bias

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 the key decisions and assumptions that are open to question, and

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 (including the types of participants, interventions and outcome measures, and methodological cut-points)

- including or excluding studies where there is some ambiguity as to whether they meet the inclusion criteria

- excluding unpublished studies

- excluding studies of lower methodological quality

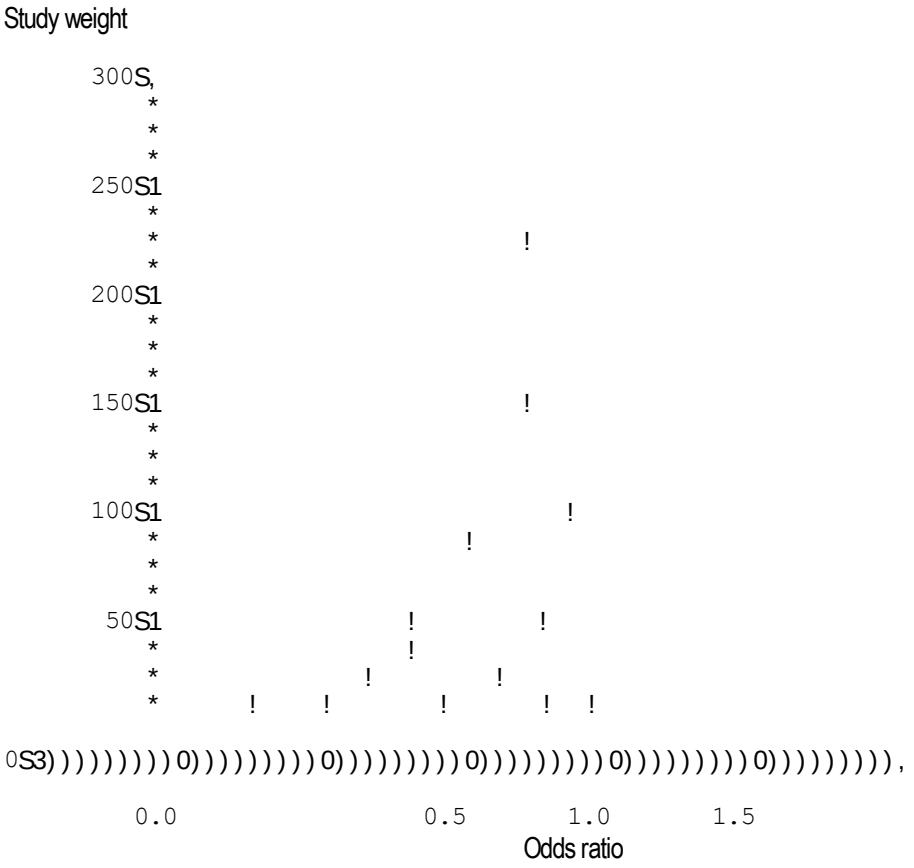
- reanalysing the data using a reasonable range of results for studies where there may be some uncertainty about the results (eg. because of inconsistencies in how the results are reported that cannot be resolved by contacting the investigators or because of differences in how outcomes are defined or measured)

- reanalysing the data imputing a reasonable range of values for missing data

- reanalysing the data using different statistical approaches (eg. using a random effects model)

Although several ways of examining the potential impact of publication bias on the results of a systematic review have been described, all of them are problematic {4, 5, 33, 36, 105, 145}. There is a growing consensus that the only satisfactory way to address this bias is through prospective registration of trials {41, 136, 138}. Nonetheless, reviewers may want to consider using a "funnel plot" to look for publication bias if there is enough studies. This can be done by plotting the study weight (one divided by the variance of the odds ratio) or sample size (on the "y" axis) against the odds ratios (on the "x" axis). Typically, this will give the appearance of a funnel with larger studies (with greater weights) at the top in the middle and smaller studies spread out at the bottom. A gap at the bottom of the funnel on the right side (assuming smaller odds ratios (less than one) represent beneficial effects) indicates that small "positive studies were identified, whereas equal numbers of small studies in the opposite direction were not, as illustrated below:

Figure. Funnel graph of study weight against odds ratio



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

8.9 Problems

8.9.1 Missing data

Three types of data may be missing in 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 8.8 in relationship to publication bias. Methods for handling missing information in the studies that are identified include excluding studies for which data are missing, simple imputations and complex strategies for imputing values for the missing data {149, 461}. No simple solutions exist for the problem of missing data. In general:

- Whenever possible, missing data should be obtained from the investigators.
- The assumptions of whatever method is used 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 8.8).

8.9.2 Multiple comparisons and the play of chance

For reviews and studies with multiple subgroup analyses and many different outcomes if enough analyses are done some are likely to produce "statistically significant" results just by chance alone. For example, at $p = 0.05$, one in 20 comparisons is likely to be "statistically significant." The more comparisons that are made, the more likely that one of them will be found statistically significant.

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 study results are presented, it is not always possible to know how many comparisons were really made. Often interesting findings are selected 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 collected data that were not reported. This is particularly important for studies reporting multiple outcomes, because of the potential for biased reporting {157, 161}.

There is no simple nor totally satisfactory solution to the problem of multiple comparisons, but the following advice can be offered:

- Keep subgroup analyses to a minimum.
- Decide in advance (and state in the protocol) 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 specified in the protocol for a review.

- Interpret "chance" findings (ones that were not hypothesised in advance) cautiously, even when they are "statistically significant".

8.9.3 Unit of analysis errors

In some studies people are allocated by group (e.g., by practice, by hospital or by community) to avoid contamination or for convenience. Often when this is done the unit of allocation is different from the unit of analysis, i.e., people are allocated by groups and analysed as though they had been allocated individually. This is sometimes called a unit of analysis error {483}. 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 result in studies having narrower confidence intervals and receiving more weight than is appropriate.

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. If an estimate of the intra-cluster variation can be obtained, it may be possible to adjust the results {484, 462}.

8.9.4 Ordinal data

Little work has been done to date 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 {463}. (More to come, pending publication of the cited papers and further input from the Statistical Methods Working Group.)

8.9.5 Cross-over trials

Cross-over trials can present a problem, particularly if both cross-over and parallel group trials are included together. One approach to this problem is to treat the cross-over studies in the same fashion as parallel group studies. That is, results from the treatment period can be handled as if they came from one group of patients and results from the control period can be treated as if they came from an independent group of patients. However, this approach ignores the fact that it is the same patients in both arms of the study, and they are not independent of each other. An alternative, more robust approach is to use techniques specifically developed for paired designs and for combining paired designs with unpaired designs {143}. If both approaches yield similar results, the results of the first approach (which can be done using RevMan) are on much stronger ground. (For an example of this approach see {465}.

8.10 Displaying and describing analyses

The analytic methods that are used in reviews should be described in the Methods section. The main results should be described and clarified in the text. Reviewers should be careful to ensure that the focus in the Results section is consistent with the reviews objectives and the comparisons specified in the protocol of the review. *Post hoc* analyses should be reported as such.

Cochrane Reviews contain the actual data entered into the Data Tables by reviewers. Tables and graphs that are displayed in RevMan and CDSR are generated "on the fly" using MetaView. This allows both reviewers and those using CDSR to select alternative statistical methods and view the results in ways other than those specified in the Methods section. To

avoid confusion, the default in CDSR will be set for each review to the methods specified by the reviewers. This will ensure that the display of results is consistent with what is described in the text, unless someone using CDSR actively chooses to override this default.

There is a convention in CDSR of displaying summary statistics for unfavourable outcomes so that for dichotomous data odds ratios and relative risks less than one (and risk differences less than zero) indicate that treatment is better. In graphical displays this is represented by estimates indicating that treatment is better than control on the left side of the vertical line indicating no difference. Both reviewers and users of the reviews can be confused when this convention is not maintained and there is a greater risk of errors being made. However, there are circumstances where it may not be appropriate to adhere to this convention. 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'. In some analyses two different interventions might be compared, so that deciding which is "treatment" and which is "control" is arbitrary. 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. To accommodate this, RevMan 3.0 allows reviewers to edit the labels used for "treatment" and "control" groups and these changes will be reflected in future editions of CDSR.

8.11 Where to go for help

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

Additional help can be found in the references cited in this section of the Handbook, other references in the Cochrane Review Methodology Database, and in the frequently asked question (FAQ) list for RevMan and the Handbook. If you discover that there are important gaps in the help that can be found in the Handbook, specific problems that are not addressed, please let us (Cindy Mulrow or Andy Oxman) know! If you have solutions to problems that others might encounter, make sure you 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.

References

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

5.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.

22.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.

28.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.

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

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

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

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

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

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

103.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.

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

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

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

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

135.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.

136.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.

138.Dickersin K. Keeping posted. Why register clinical trials? - revisited. Controlled Clin Trials

1992; 13:170-7.

143.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.

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

149.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.

157.Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal anti-inflammatory drugs in rheumatoid arthritis [published erratum appears in Controlled Clin.Trials 1989 Sep;50(9):356]. Controlled Clin Trials 1989; 10:31-56.

161.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(6):613]. J Clin Epidemiol 1990; 43:1313-8.

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

179.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.

180.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.

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

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

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

232.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.

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

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

381. 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 1994; Doc No 134:[8425

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

455. Whitehead A, Bailey AJ, Elbourne D. A meta-analysis of clinical trials involving both continuous and binary outcomes. Stat Med. In press.

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

457. Morgenstern H. Uses of ecological analysis in epidemiologic research. Am J Public Health 1982; 72:1336-4.

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

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

466. Mari JJ, Streiner D. Family intervention for those with schizophrenia. In: Adams C, Anderson J, de Jesus Mari J (eds.) Schizophrenia Module of The Cochrane Database of Systematic Reviews , [updated 06 June 1996]. Available in The Cochrane Library [database on disk and CDROM]. The Cochrane Collaboration; Issue 2. Oxford: Update Software; 1996. Updated quarterly. Available from: BMJ Publishing Group, London.

460. 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.

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

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

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

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

462. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. Am J Epidemiol 1981; 114:906-914.

463. Whitehead A, Jones NMB. A meta-analysis of clinical trials involving different classifications of response into ordered categories. Stat Med 1994; 13:2503-2515.

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

465. 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 Med1990; 88:279-86.

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

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 conclusions of a review 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 deciding
- Clarification of any important trade-offs between the expected benefits, harms, and costs of the intervention of interest

Because Cochrane Reviews have an international audience, the discussion and 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 aim of 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, and 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 how strong the evidence is supporting a conclusion that there is 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 {34}. 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?

Explicit approaches to grading the strength of evidence underlying a conclusion are available {498, 499, 129, 454, 500}, 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 may elect to use a standard approach to grading the strength of evidence across reviews included in its module and, over time, it may be possible for the Collaboration 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 Group, 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."* {501}

Users of Cochrane Reviews must decide, either implicitly or explicitly, how applicable the evidence is to their 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 like 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 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 to why the evidence should not be applied under certain circumstances {379, 502}. 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 (for example, 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; for example, in developing countries {502}.

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 import consideration in determining the applicability of results. However, it is important to distinguish between two issues. The first is 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 impact of an intervention (e.g., the number needed to treat) by applying the estimated relative effect (e.g. relative risk) 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 a review.

9.2.4 Variation in the results of the included studies

In addition to identifying limitations of the applicability of the results of a 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 8.3, differences between subgroups, particularly between study differences, 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 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 deciding.

9.4 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 are made. Health care 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 {21}. It is beyond the scope of Cochrane Reviews to incorporate formal economic analyses (although they might well be used for such analyses) {251, 252}. 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.

9.5 Implications

The above cautions about drawing conclusions not withstanding, review groups (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 {253}, 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 (e.g., continuity of caregiver during labour)

B) Forms of care that should be abandoned in light of the available evidence (e.g., routine weight measurement throughout pregnancy or elective caesarean for 'gestational diabetes')

C) Forms of care that involve important trade-offs between known benefits and known adverse effects (e.g., epidural analgesia in labour versus other forms of pharmacological relief)

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 (e.g., support for caregivers during at-risk pregnancy)

B) Forms of care that have not been shown to have the effects expected from them, but which may require further attention (e.g., routine iron supplementation during pregnancy)

C) Forms of care with reasonable evidence that they are not effective for the purpose for which they have been used (e.g., dietary regulation for 'gestational diabetes')

9.6 Common errors in reaching conclusions

A common mistake that occurs, both in describing results and in reaching conclusions, when there is inconclusive evidence is to confuse 'no evidence of effect' with 'evidence of no effect'. Reviewers should be careful not to do this. A second common mistake is to reach conclusions which go beyond the evidence that is reviewed. Reviewers should be cautious to avoid errors such as these. Statements 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.

In summary, it is not possible to provide firm guidelines regarding how to prepare a discussion and reach conclusions about the implications of a review for practice and future research. However, systematic consideration of the issues outlined in this section can guide the interpretation of the results of a review and help reviewers to avoid making incorrect or misleading inferences.

References

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

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

499. US Preventive Services Task Force. Guide to Clinical Preventive Services. 2nd ed.

Baltimore: Williams & Wilkins, 1996; xxxix-lv.

129. 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.

454. Guyatt GH, Sackett DL, Sinclair JC et al; for the Evidence-Based Medicine Working Group. JAMA 1995; 274:1800-4.

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

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

379. 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.

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

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

251. 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.

252. 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.

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

10. IMPROVING AND UPDATING REVIEWS

If Cochrane Reviews are to be useful to those who want to take more informed decisions in health care 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 most 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 under the aegis of collaborative trialists' groups, 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), but they also depend on the success of negotiations among the trialists. 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. In general, these people have selected themselves during the early phases of the Collaboration's development. 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 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 must also disclose any potential conflict of interest that they might have.

10.2 Consumer involvement

Healthcare consumers and other users of Cochrane Reviews must be involved in developing reviews to help ensure that Reviews:

- Are targeted at problems that are important to the 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

Relatively little is known about the effectiveness of various means of involving consumers in the review process or, more generally, in healthcare research all the way from how questions are formulated to how results are used in decisions. The Collaboration is dedicated to consumer involvement in principle. This is based on our values and good logic. Researchers and funders have 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. 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.

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 trials registers by CRGs
- Providing training and support to those undertaking searching activities
- Developing the Cochrane Controlled Trials Register (CCTR) to facilitate the transfer of trials between CRGs and other Cochrane entities, and to facilitate access to trials in other registers contributed to CCTR
- Working with the US National Library of Medicine to improve the coding of trials in MEDLINE and to develop an ancillary database of trials not included in MEDLINE
- Developing and evaluating strategies to improve the coding and classification of trials

This work involves many people engaged in a variety of activities through CRGs, Cochrane Centres, Fields and Methods Working Groups. The Trials Registers Development Group, the Baltimore Cochrane Center and the Coding and Classification Methods Working Group have key responsibilities for co-ordinating and guiding these activities, as described in CDSR.

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 Collaborative Review Groups, 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 Working 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 CDSR.

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 centralised efforts to assist with this and ensure the thoroughness and efficiency with which RCTs are identified
- Use of explicit criteria for selecting trials for inclusion in a review and for assessing the quality of included trials
- Application of these criteria 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 availability of resources and data
- Collection of individual patient data from trialists 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 sub-group analyses that are undertaken
- Carefully drawn conclusions, including implications for practice and future research, based on cautious interpretation of results - taking into account the limitations of the review and variability in the values and conditions of those making decisions
- 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 the validity of the results of a review.

10.7 Software and informatics support

Review Manager (RevMan) is designed to assist reviewers in constructing reviews in the structured format described in section 2. This software will continue to be developed and will 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 of *a priori* protocols for planned reviews, as described in section 3.

The development of RevMan, 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 advisory groups for each of the three programs. Development of specialised software to facilitate the management of trials by CRGs (TrakMan) is currently being explored. 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 Working 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 CDSR.

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, CRG 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 collaborative review group 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 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. A library of training materials, including examples, illustrations and frequently asked questions is being developed in conjunction with the Handbook to facilitate access to these materials. The San Antonio Cochrane Center is responsible for co-ordinating training efforts and producing a Training Manual based on the Handbook.

10.8.2 Training for editors

CRG editors need skills related to the area covered by the review group, skill, and experience in critically appraising studies of the effects of healthcare and an ability to edit scientific material for publication. An CRG editor may have to:

- Identify and define topics to be reviewed
- Identify potential reviewers who are capable of preparing a systematic review, and brief 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
- Arrange that existing reviews are updated, by organising the retrieval of newly registered studies within the scope of the CRG, sending them to reviewers, and helping them to update their reviews

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 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, 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 specialised aspects of editing - such as the correct use of meta-analysis, how to plan the work of a collaborative review group and how to help and support the reviewers.

Within each CRG, the senior editor or editors should take responsibility for the systematic development of the editorial skills of the other editors. Since this task will be new for most senior editors, it may be useful for the Cochrane Collaboration to hold introductory workshops.

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.

10.9 Peer review

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, 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 aims to create an iterative system through which successive versions of each review will reflect not only the emergence of new data, but also valid criticisms, solicited or unsolicited, from whatever source. This process is facilitated by the contact details both for reviewers and for editorial offices in each review. Successive versions of a particular review, together with any intervening criticisms, will be archived electronically.

10.9.1 Refereeing

There are no standard methods for refereeing systematic reviews. However, several general principles merit mention. First, peer review can be useful for several stages of the review process: question formulation, protocol development, completion of the review and updating. Second, peer review should include multiple referees or editors with both methodological and topical expertise, and with differing viewpoints. Some of the peer reviewers or referees should be external to the CRG from which the review originates. Referees should include people without direct financial, intellectual, or personal conflicts of interest concerning the topic being addressed. Statements regarding conflict of interest are required in Cochrane Reviews and CRGs may want to require these from referees. Third, explicit standardised methods and checklists aimed at ensuring comprehensiveness and limiting bias should be encouraged among peer- reviewers. Fourth, peer review should be constructive, courteous, and timely.

Specific areas to address at each stage of peer review vary. The main issues to consider at question formulation are whether: a) there is any overlap or potential duplication of effort with another reviewer either within or outside the originating review group; b) objectives are clearly phrased and include all of the components of well-formulated questions; and c) the review is likely to be feasible. Review at this stage can often be accomplished quickly by a CRG's editorial team.

Reviewing protocols is more time-consuming and 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. Peer review at review completion includes a second critique of the review's methods as well as a critique of the actual results, presentation of results, discussion, and conclusion. Critiques of completed reviews are best done by multiple individuals, some of whom are external and independent of the CRG. The importance of external peer review of protocols may vary from CRG to CRG and within a CRG from review to review.

Differences among referees' critiques should be elucidated and reconciled whenever possible. Potential mechanisms to use for reconciliation of different critiques are arbitration by one or more of the editors or use of an additional independent referee. Examples depicting flow diagrams of editorial processes that are utilized by different review groups are given in the Library of Examples.

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. One such checklist, adapted from multiple citations {37, 55, 177, 189, 195, 197, 335, 348, 438, 441}, is shown below. Other checklists, including ones used by CRGs, are included in the Library of Examples.

Problem Formulation
- Are review questions well formulated with specified key components?
- Were major changes in review questions avoided during the review process?

Study Identification
- Is there a thorough search for relevant data using appropriate sources?
- Are there unbiased explicit searching strategies that are appropriately matched to the question?

Study Selection
- Are appropriate inclusion and exclusion criteria used to select articles?
- Are selection criteria applied in a manner that limits bias?
- Are major changes in selection criteria avoided during the review process?

Appraisal 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 important parameters, such as study designs, considered in the synthesis?
- Are reasonable decisions made concerning whether and how to combine data?
- 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?

Conclusions
- Are conclusions supported by the data reviewed?
- Are plausible competing explanations of observed effects addressed?
- Is evidence appropriately interpreted as inconclusive (no evidence of effect) or as showing a particular strategy did not work (evidence of no effect)?
- Are important considerations for decision makers identified, including values and contextual factors that might influence decisions?

10.10 Updating reviews

When registering a systematic review with the Cochrane Collaboration, reviewers agree to keep it current. Keeping a review up-to-date entails repeating, at periodic intervals, the steps

involved in the original review. Some of the steps will require minimal effort (for example, reviewing your 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 group's coordinator or search co-ordinator. 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 trials register, CCTR and MEDLINE. Soon, the Cochrane Collaboration will have a Criticism Management System (see below) whereby users of Cochrane Reviews can provide comments and criticisms of reviews. This database will provide an additional source of pertinent studies cited by users of CDSR.

Original data collection forms should be used to abstract new research evidence. If new research evidence addresses important variables that were not included in original collection form, forms may be modified. For example, if reviewers had originally only abstracted morbidity and mortality outcomes in trials addressing treatment of oat cell carcinoma of the lung, and recent trials 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 probably represent substantive changes that merit editorial critique through a 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 on the cover page of the review. At a minimum, updates should be considered on an annual basis. Even if no substantive new evidence is found on annual review and no major amendment is indicated, the date of the latest search for evidence should be made clear in the search strategy section of the review.

10.11 Responding to criticisms

The San Francisco Cochrane Center is spearheading the development of a Criticism Management System for CDSR to facilitate criticism and response to criticisms of Cochrane Reviews. This system will include a Criticism Database which should be operational by 1997. Guidelines to help ensure that CRGs and reviewers take appropriate account of criticisms that are received will be incorporated in future editions of the Handbook as the Criticism Management System evolves.

References

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

55. 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.

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

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

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

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

335. 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.

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

438. 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.

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

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 trialists or collection of individual patient data (IPD) from the trialists. 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 {474} and in more restricted reviews in which data from a relatively small number of trials assessing a specific health care intervention are collected and combined {475}. Systematic reviews based on IPD have been described as the yardstick against which all reviews should be measured {476}. 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 systematic review when deciding whether to embark on such a project.

11.2 Methods Working Group on Individual Patient Data Reviews

To try to help with this decision and with the logistics of such projects, a Cochrane Collaboration Methods Working 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 {477}. This report is included as an appendix.

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 aim should be that all randomised patients, and no non-randomised patients, from all relevant trials are included and that they are analysed using the intention-to-treat principle. In this way, systematic biases and chances 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 trialists to supply their data since this should increase the likelihood that data will be received for all relevant trials. In addition, trialists should know that any data supplied for the review will be held in confidence and will not be used for any other purpose without the permission of the trialists, and that the reports of the review will be published in the names of the collaborating trialists 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 trial results come from a central re-analysis of the raw data from each trial. The necessary data items are sought, and, after central processing, any inconsistencies or problems are discussed and hopefully resolved by communication with the responsible trialists.

The finalised data for each trial are then analysed separately to obtain summary statistics, which are combined to give an overall estimate of the effect of treatment. In this way, patients are only directly compared with others in the same trial and the entire dataset is not pooled as though it came from a single, homogeneous randomised trial.

11.4 How can an IPD meta-analysis help?

If a systematic review relies solely on data from published trials, it is open to several problems. The most obvious of these is that unpublished trials 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 relevance to 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, the fact that it has not been might not be clear from the published report). Collection of either aggregate or individual patient data from trialists will resolve some of these problems: unpublished trials can be included, updated data on specific types of patient and outcome can be requested, and whether the data are based on the randomised allocations can be clarified.

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 trialists; and to update follow-up information through patient record systems (such as mortality registers) where available. In addition, it might be easier for a trialist 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 patients, the trialist can simply send these details instead of preparing new tables.

Furthermore, as IPD meta-analyses involve the collaboration of the trialists other benefits (that may also be found if the trialists are contacted for aggregate data) may include more complete identification and understanding of the trials; 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 Working Group on IPD based meta-analysis was to establish, and encourage the tackling of, a research agenda to investigate this approach to systematic reviews (see appendix). 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 and have shown the importance of the former to help control publication bias, to ensure the use of the intention-to-treat principle in the analysis, and to obtain a fuller picture of the effects of different treatments over time {95, 222, 475}.

11.6 Further information

This chapter was prepared by Mike Clarke and Lesley Stewart on behalf of the Cochrane Collaboration Working Group on meta-analyses using individual patient data.

Many of the topics discussed here are expanded on in {477}. 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 Collaboration reviewers would like further information, they are welcome to contact the Methods Working Group.

References

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

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

474. 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.

475. 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.

476. 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.

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

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

447. Jepson R, Hetherington J. Glossary of terms in the Cochrane Collaboration. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996, Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

446. Last JM (ed). A Dictionary of Epidemiology. 2nd ed. New York: Oxford University Press, 1988.

503. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1991.

504. Goodman C. Literature searching and evidence interpretation for assessing health care practices. Stockholm: SBU, 1993.

APPENDICES

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.

Short title: This can be up to 36 characters.

Reviewers: This can be the name of an individual (surname followed by initials), several individuals or a collaborative group (e.g. Antiplatelet Trialists' Collaboration).

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

Primary contact for this review: This should be the address and other contact details for the person to whom correspondence about the review should be addressed. The contact information must be entered using specified fields.

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

Date of last substantive update: The reviewer(s) or editors of a Collaborative Review Group (CRG) must decide whether an amendment is substantive or not. In general, substantive amendments are ones which require editorial review. Non-substantive amendments include corrections of typographical errors.

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.

This review should be cited as: This is generated automatically based on other information that is entered by reviewers.

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 "intramural" (from within the institution or organisation where the reviewers are employed) and "extramural" (external funding).

2a.2 Abstract

All reviews must include an abstract of not more than 250 words. The abstract should fit onto a computer screen and should provide a concise summary of the review in a structured format. Abstracts are made freely accessible on the Internet. The content under each heading in the abstract should be as follows.

Objective: This should be a precise statement of the primary objective of the review. It should

include information about the specific population, intervention and outcomes that are reviewed.

Search strategy: A succinct summary of data sources (the search strategy used to identify trials) should be given including any time restrictions or other constraints.

Selection criteria: The inclusion for the review should be summarised.

Data collection & analysis: This should outline the methods used to obtain, critically appraise and analyse data from the included studies.

Main results: The main results of the review, whether qualitative or quantitative, should be stated.

Conclusions: The primary implications for practice should be clearly stated, avoiding over-generalization. Important implications for further research can also be stated.

2a.3 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.

Background: The review should begin with a brief synthesis of the underlying biology and health care 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 health care.

Objective: 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. The comparisons made should be listed in the comparisons table. If a review addresses more than one hypothesis and includes several comparisons, the comparisons should be grouped for each hypothesis or question.

Selection criteria: The criteria used to select trials 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 intervention** and **types of outcome measures** are subheadings in this section.

Search strategy: The data sources used to identify trials should be summarised, including bibliographic databases, reference lists from pertinent articles and books, conference proceedings and personal contact with experts or institutions active in the area. The databases searched and the terms used should be stated, including any constraints, such as language. Any hand-searched journals should be listed. If a collaborative review group has developed a specialised register of trials, reviews contributed by that group can include or refer to a standard description of how trials were identified, noting when and how the trials register was last searched and any additional data sources used.

Methods: 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, if the number of events was calculated from published survival curves), how the data were synthesised, any statistical techniques used and sensitivity analyses performed. If a collaborative review group uses a standard approach for all the reviews prepared by that group, the methods section can reference a description of those methods in the group's module. Similarly, if a Methods Working Group (MWG) has recommended a standard approach and a review uses that approach, the methods section can reference the relevant report or MWG module, e.g., for assessing the quality of RCTs or for statistical analysis.

Description of studies: This should refer to the information contained in the **included studies table** and the **excluded studies table** and should describe or explain 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 must decide what other characteristics of the trials it is important for readers of the review to know.

Methodological quality: 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 **table 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 be cautious about reporting subgroup analyses. They 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'.

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.

Conclusions: Implications for practice and implications for research are subheadings. The implications for practice should be as practical and unambiguous as possible. They should not go beyond the evidence that was reviewed.

The implications for research should not include statements like "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: Credit should be given where credit is due. This could, for example, include all the members of a collaborative trialists' group for individual patient data reviews.

Reviewers should ensure that anyone personally acknowledged is informed.

2a.4 Tables and figures

Table of comparisons: The comparisons made in the **data tables** must be listed here. A review can include more than one comparison and a study can be included in more than one of these. Currently it is not possible to view this table in CDSR. In RevMan it can be viewed and edited as a hierarchical list. The comparisons should correspond to the questions or hypotheses in the text under **objectives**.

Table of included studies: This is a standard table. Its design represents a compromise between ease of use and flexibility. Currently it is only possible to view one "row" of the column at a time (displayed as a column with each cell appearing as a separate box). Software that will facilitate viewing this and the **excluded studies** tables as tables is being developed.

The table has six columns (in addition to the identifier for the study): **study, methods, participants, interventions, outcomes,** and **notes**. The last column is for personal notes and is not published with the review. It is possible to use codes so that each column can include several sub-categories of information, e.g., a reviewer could include country, setting and sex under **participants**. Reviewers must decide what characteristics of the included studies are likely to interest users of the review.

In addition to including appropriate citations, reviewers must include information about **concealment of allocation** and **data sources** for all included studies, using specified categories for each. For concealment of allocation, each study must be marked as: adequate (A), unclear (B) or inadequate (C); or it must be indicated that allocation concealment was not used (D) as a criterion to assess validity. For data sources it must be indicated 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).

Table 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., data not available).

Data tables: Data for each comparison specified in the **table of comparisons** must be entered in a standardised format from which tables and figures for each comparison can be generated. Three types of tables are possible: dichotomous data, continuous data, and other data. The table used for "other data" is like the included studies table and enables the reviewer to specify appropriate headings for the columns.

2a.5 References

References to studies are organised under four standard headings: **included studies**, **excluded studies**, **studies awaiting assessment**, **ongoing studies.** Other references include **additional references** that are cited in the review and **previously published versions of the review**, e.g., if the review has been published in a journal. Reviewers should check their references for accuracy {4004, 4005}. Examples of the correct forms of references for a journal article and book, respectively, are:

Liggins GC, Howie RN. A controlled trial of antepartum glucocorticoid treatment for prevention of the respiratory distress syndrome in premature infants.

Pediatrics 1972; 50:515-25.

Crowley P. Promoting pulmonary maturity. In: Chalmers I, Enkin MW, Keirse MJNC, eds. Effective Care in Pregnancy and Childbirth. Oxford: Oxford University Press, 1989: 746-64.

Examples of the correct form for other references can be found in Appendix 2b. The titles of journals should be abbreviated according to the style used in MEDLINE. If the number of authors exceeds six, list six followed by 'et al'.

Included studies: Each study should be given a short name, e.g., the last name of the first author and the year of publication or the location of the study. The studies should be listed alphabetically by short name with the references (if there are more than one for a trial) arranged chronologically or in some other logical order for each study.

Excluded studies: These should be listed alphabetically by a short name with the reference(s). If there are more than one reference for a study, they should be listed chronologically or in some other logical order.

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 have been cited in the text of the review.

Ongoing studies: Studies which are ongoing that meet the inclusion criteria should be listed here with contact addresses, and information about whether recruitment is still open if this information is known and whether the investigators are seeking other centres to join the trial.

Additional references: Other references cited in the text should be listed here. If a report of a trial is cited in the text for some reason other than referring to the trial (e.g., because of some background or methodological information in the report), it should also be listed here.
Trials should be identified by their short names in square brackets in the text, e.g. [Liggins 1972]. Other references should be identified by the last name of the author and the year of publication in parenthesis, e.g. (Crowley 1989). The last name of the first author followed by 'et al' should be used if there are multiple authors, e.g. (Antman et al 1992). If no author is given, the title of the journal or publication should be used, e.g. (Lancet 1993) or (International Register of Perinatal Trials 1991). A semicolon should be used between references, e.g. (Crowley 1989; Liggins et al 1977).

Previously published versions: References to any previously published versions of the review in a journal or textbook should be listed here.

2a.6 Conflict of interest

Any conflict of interest capable of influencing the judgements of any of the reviewers shouls 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".

Appendix 2b. Format for references

Reviewers should format references according to the uniform requirements for manuscripts submitted to biomedical journals published by the Committee of Medical Journal Editors:

1) Standard journal article (List all authors, but if the number exceeds six, give six followed by et al.)

- You CH, Lee KY, Chey RY, Menguy R. Elec-trogastrographic study of patients with unexplained nausea, bloating and vomiting. Gastroenterology 1980 Aug;79(2):311-4.

As an option, if a journal carries continuous pagination throughout a volume, the month and issue number may be omitted.

- You CH, Lee KY, Chey RY, Menguy R. Elec-trogastrographic study of patients with unexplained nausea, bloating and vomiting. Gastroenterology 1980; 79:311-4.

- Goate AM, Haynes AR, Owen MJ, Farrall M, James LA, Lai LY, et al. Predisposing locus for Alzheimer's disease on chromosome 21. Lancet 1989; 1:352-5.

2) Organisation as author

- The Royal Marsden Hospital Bone-Marrow Transplantation Team. Failure of syngeneic bone-marrow graft without preconditioning in post-hepatitis marrow aplasia. Lancet 1977;2:742-4.

3) No author given

- Coffee drinking and cancer of the pancreas [editorial]. BMJ 1981; 283:628.

4) Article not in English

- Massone L, Borghi S, Pestarino A, Piccini R, Gambini C. Localisations palmaires purpuriques de la derma-tite herpetiforme. Ann Dermatol Venereol 1987; 114:1545-7.

5) Volume with supplement

- Magni F, Rossoni G, Berti F. BN-52021 protects guineapig from heart anaphylaxis. Pharmacol Res Commun 1988;20 Suppl 5:75-8.

6) Issue with supplement

- Gardos G, Cole JO, Haskell D, Marby D, Paine SS, Moore P. The natural history of tardive dyskinesia. J Clin Psychopharmacol 1988;8(4 Suppl):31S-37S.

7) Volume with part

- Hanly C. Metaphysics and innateness: a psycho-analytic perspective. Int J Psychoanal 1988;69(Pt 3):389-99.

8) Issue with part

- Edwards L, Meyskens F, Levine N. Effect of oral iso-tretinoin on dysplastic nevi. J Am Acad Dermatol 1989;20(2 Pt 1):257-60.

9) Issue with no volume

- Baumeister AA. Origins and control of stereotyped movements. Monogr Am Assoc Ment Defic 1978;(3):353-84.

10) No issue or volume

- Danoek K. Skiing in and through the history of medicine. Nord Medicinhist Arsb 1982:86-100.

11) Pagination in Roman numerals

- Ronne Y. Ansvarsfall. Blodtransfusion till fel patient. Vardfacket 1989;13: XXVI-XXVII.

12) Type of article indicated as needed

- Spargo PM, Manners JM. DDAVP and open-heart surgery [letter]. Anaesthesia 1989; 44:363-4.

- Fuhrman SA, Joiner KA. Binding of the third component of complement C3 by Toxoplasma gondii [abstract]. Clin Res 1987; 35:475A.

13) Article containing retraction

- Shishido A. Retraction notice: Effect of platinum compounds on murine lymphocyte mitogenesis [Retraction of Alsabti EA, Ghalib ON, Salem MH. In: Jpn J Med Sci Biol 1979;32:53-65]. Jpn J Med Sci Biol 1980; 33:235-7.

14) Article retracted

- Alsabti EA, Ghalib ON, Salem MH. Effect of platinum compounds on murine lymphocyte mitogenesis [Retracted by Shishido A. In: Jpn J Med Sci Biol 1980; 33:235-7]. Jpn J Med Sci Biol 1979;32:53-65.

15) Article containing comment

- Piccoli A, Bossatti A. Early steroid therapy in IgA neuropathy: still an open question [comment]. Nephron 1989; 51:289-91. Comment on: Nephron 1988; 48:12-7.

16) Article commented on

- Kobayashi Y, Fujii K, Hiki Y, Tateno S, Kurokawa A, Kamiyama M. Steroid therapy in IgA nephropathy: a retrospective study in heavy proteinuric cases. Nephron 1988; 48:12-7. Comment in: Nephron 1989; 51:289-91.

17) Article with published erratum

- Schofield A. The CAGE questionnaire and psychological health [published erratum appears in Br J Addict 1989; 84:701]. Br J Addict 1988; 83:761-4.

Books and Other Monographs

18) Personal author(s)

- Colson JH, Armour WJ. Sports injuries and their treatment. 2nd rev. ed. London: S. Paul, 1986.

19) Editor(s), compiler as author

- Diener HC, Wilkinson M, editors. Drug-induced headache. New York: Springer-Verlag, 1988.

20) Organisation as author and publisher

- Virginia Law Foundation. The medical and legal implications of AIDS. Charlottesville: The Foundation, 1987.

21) Chapters in a book

- Weinstein L, Swartz MN. Pathologic properties of invading microorganisms. In: Sodeman WA Jr, Sodeman WA, editors. Pathologic physiology: mechanisms of disease. Philadelphia: Saunders, 1974:457-72.

22) Conference proceedings

- Vivian VL, editor. Child abuse and neglect: a medical community response. Proceedings of the First AMA National Conference on Child Abuse and Neglect; 1984 Mar 30-31; Chicago. Chicago: American Medical Association, 1985.

23) Conference paper

- Harley NH. Comparing radon daughter dosimetric and risk models. In: Gammage RB, Kaye SV, editors. Indoor air and human health. Proceedings of the Seventh Life Sciences Symposium; 1984 Oct 29-31; Knoxville (TN). Chelsea (MI): Lewis, 1985:69-78.

24) Scientific or technical report

- Akutsu T. Total heart replacement device. Bethesda (MD): National Institutes of Health, National Heart and Lung Institute; 1974 Apr. Report No.: NIH-NHLI-69-2185-4.

25) Dissertation

- Youssef NM. School adjustment of children with congenital heart disease [dissertation]. Pittsburgh (PA): Univ. of Pittsburgh, 1988.

26) Patent

- Harred JF, Knight AR, McIntyre JS, inventors. Dow Chemical Company, assignee. Epoxidation process. US patent 3,654,317. 1972 Apr 4.

Other Published Material

27) Newspaper article

- Rensberger B, Specter B. CFCs may be destroyed by natural process. The Washington Post 1989 Aug 7; Sect. A:2 (col. 5).

28) Audiovisual

- AIDS epidemic: the physician's role [videorecording]. Cleveland (OH): Academy of Medicine of Cleveland, 1987.

29) Computer file

- Renal system [computer program]. MS-DOS version. Edwardsville (KS): MediSim, 1988.

30) Legal material

- Toxic Substances Control Act: Hearing on S. 776 Before the Subcomm. on the Environment of the Senate Comm. on Commerce. 94th Cong., 1st Sess. 343 (1975).

31) Map

- Scotland [topographic map]. Washington: National Geographic Society (US), 1981.

32) Book of the Bible

- Ruth 3:1-18. The Holy Bible. Authorized King James version. New York: Oxford Univ. Press, 1972.

33) Dictionary and similar references

- Ectasia. Dorland's illustrated medical dictionary. 27th ed. Philadelphia: Saunders, 1988:527.

34) Classical material

- The Winter's Tale: act 5, scene 1, lines 13-16. The complete works of William Shakespeare. London: Rex, 1973.

Unpublished Material

35) In press

- Lillywhite HD, Donald JA. Pulmonary blood flow regulation in an aquatic snake. Science. In press.

Appendix 2c. 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 health care 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 2d. Conditions of publication

THE COCHRANE COLLABORATION
PO Box 726, OXFORD OX2 7UX, UK, Fax +44 (0)1865 516311

Conditions of Publication/Information on Authors

Cochrane Review ID:
First Author:
Title:

This form must be filled out, signed by all authors, and returned to the Review Group Co-ordinator before your review can be published as a Cochrane Review.

My signature on this document affirms that I agree to the conditions listed below for publication of my review in the Cochrane Database of Systematic Reviews.

1. I have participated sufficiently in the conception and design of this Review and the analysis of the data, as well as the writing of the review to take public responsibility for it. I believe the manuscript represents valid work. I have reviewed the final version of the Review and approve its public dissemination.

2. I certify that any past or present affiliations with or involvement in any organisation or entity with a direct financial interest in the subject matter or materials discussed in the Review (e.g., employment, consultancies, stock ownership, honoraria, expert testimony) are listed below. (Write "none" if this is the case or specify below any affiliation or involvement that might be perceived as a possible conflict of interest.)

3. I grant to the Collaboration an irrevocable, paid up, nonexclusive, worldwide licence (with right to sublicense) to incorporate in The Cochrane Database of Systematic Reviews all material which I have provided to the Collaboration and all material which I shall in the future provide to the Collaboration (together, "the Material"); and as part of or in conjunction or connection with such Database, to publish, reproduce, print, copy, vend, market, disseminate via electronic networks, amend, translate, adapt, make derivative works based upon, distribute copies of, use and display the Material and any such derivative works.

Person who is taking principal responsibility for this review

Signature:
Name:
Mailing address:
Phone:
Fax:
E-Mail:

Other Authors (Give current mailing address.)

Signature:
Name:
Mailing address:
Phone:
Fax:
E-Mail:

Signature:
Name:
Mailing address:
Phone:
Fax:
E-Mail:

Signature:
Name:
Mailing address:
Phone:
Fax:
E-Mail:

Signature:
Name:
Mailing address:
Phone:
Fax:
E-Mail:

Signature:
Name:
Mailing address:
Phone:
Fax:
E-Mail:

Appendix 3a. The 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 a review group. 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 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 review group
- ensuring that reviewers receive the methodological support they need

The main resource required by reviewers is their own time. Most 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 trials, 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 trials
- Assessing citations and full text reports of trials for inclusion in the review
- Assessing the quality of included trials 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 trials is primarily the responsibility of those involved in developing registers of trials, 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 trials for inclusion, assess the quality of included trials, obtain data and conduct analyses
- statistical support for synthesizing the results of the included trials
- 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 trials
- Selection of trials for inclusion
- Assessments of the validity of included trials
- Obtaining data for the included trials
- Analysis
- Inferences and presentation of results
- Time-chart for major activities
- Budget

The objectives and design of a review are addressed in sections 4-8. 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 trials 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 {335, 177, 198}. 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 trials (including unpublished trials), and collecting background information that may be important in interpreting the results of individual trials 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 review group, 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 trials
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 Group'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 collaborative review group is responsible for documenting the methods used by the team for editing the module. The editorial team is also responsible for documenting any specific methods used by the Group beyond the standard methods specified in the Handbook, including:

- any standard methodological criteria for including studies in reviews
- the search strategies used to develop and maintain the register of trials used by the group 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 synthesizing 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 categorize interventions (see section 9.5)

- any specific rules used for preparing the standard tables and figures
- the methods used to keep reviews up to date and respond to criticisms
- methods used to review protocols

Descriptions of specific methods used by each CRG are published as part of the Group's module in CDSR.

3a.4 References

335. 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.

177. L'Abbe KA, Detsky AS, and O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987; 107(2): 24-33.

198. Oxman AD and Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidem 1991; 11:1271-8.

Appendix 5a. Registers of clinical trials

From Dickersin K, Larson K. Establishing and maintaining an international register of RCTs. In: Cochrane Collaboration Handbook. Oxford: The Cochrane Collaboration, 1995 [updated 14 July 1995]. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996; Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

AIDS CLINICAL TRIALS INFORMATION SERVICE (ACTIS)

Ms. Sharon Marsh
P. O. Box 6421
Bethesda, MD 20849-6421 USA
Telephone: (301) 251-5750
FAX: (301) 738-6616

Scope: All federally sponsored drug trials for AIDS and HIV infection and efficacy trials sponsored by pharmaceutical companies.

AIDS/HIV TREATMENT DIRECTORY

Joseph Guimento, MPA, Projects Coordinator
Treatment Information Services
American Foundation for AIDS Research (AmFAR)
733 Third Avenue, 12th Floor
New York, NY 10017 USA
Telephone: (212) (800) 39-AMFAR
FAX: (212) 682-9812

Scope: Active HIV-associated clinical trials.

AIDS/HIV TREATMENT DIRECTORY FOR NEW ENGLAND

Ms. Janine Grant, Educational Director
Community Research Initiative (CRI) of New England
320 Washington Street, 3rd Floor
Brookline, MA 02146 USA
Telephone: (617) 566-4004
FAX: (617) 566-8226

Scope: HIV/AIDS clinical trials open for enrollment in the New England states.

ARC DATABASE OF CLINICAL TRIALS IN RHEUMATOID ARTHRITIS

Mr. Peter S. Blair
ARC Epidemiology Research Unit
Stopford Building
Oxford Road
Manchester M13 9PT UK

Telephone: 44 061 275-5037
FAX: 44 061 275-5043

Scope: Completed, randomized, Phase III, controlled trials of treatments in rheumatoid arthritis.

CARDIOVASCULAR RANDOMIZED CLINICAL TRIAL (CVRCT) REGISTRY

Allan D. Kitching, MD, FRCPC
Department of Medicine
St. Joseph's Hospital
Fontbonne Hall, Room 515
50 Challton Avenue East
Hamilton, ON 18N 4A6 CANADA
Telephone: (905) 522-1155 Ext. 3736
FAX: (905) 521-6068

Scope: Completed, active and planned clinical trials of treatments of cardiovascular disease.

COLUMBIA REGISTRY OF CLINICAL TRIALS

E. Andrew Balas, MD, PhD
Health Services Management
University of Missouri - Columbia
324 Clark Hall
Columbus, MO 65211 USA
Telephone: (314) 882-8416
FAX: (314) 882-6158

Scope: Completed, prospective, contemporaneously controlled clinical trials in health services research, utilization and information management, where the effect of the intervention is measured on the quality of patient care (process and/or outcome).

CRITICAL PATH AIDS PROJECT

Mr. Kiyoshi Kuromiya
Critical Path Project, Inc.
2062 Lombard Street
Philadelphia, PA 19146 USA
Telephone: (215) 545-2212
FAX: (215) 735-2762

Scope: Philadelphia regional trials for persons with AIDS/HIV.

DENTALPROJ

Ms. Carla G. Flora
Management Information Section
National Institute of Dental Research
5333 Westbard Avenue, Room 707
Bethesda, MD 20892 USA
Telephone: (301) 594-7645
FAX: (301) 594-7655

Scope: Dental research funded by NIDR.

DEPARTMENT OF VETERANS AFFAIRS COOPERATIVE STUDIES REGISTRY

Ping C. Huang, PhD
Cooperative Studies Program 12A4
Department of Veteran Affairs Central Office
810 Vermont Avenue, NW
Washington, DC 20420 USA
Telephone: (202) 535-7154
FAX: (202) 535-7153

Scope: All clinical trial proposals submitted to the VA Cooperative Study Program.

DIRECTORY OF HIV CLINICAL TRIALS IN THE BAY AREA

Ms. Nicole Mandel
Community Consortium
3180 18th Street, Suite 201
San Francisco, CA 94110 USA
Telephone: (415) 476-9554
FAX: (415) 476-4734

Scope: All studies deal with some aspect of HIV disease and are IRB approved.

THE EXPERIMENTAL TREATMENT GUIDE

Mr. Kenneth Fornataro, Executive Director
Mr. Joel Beard, Information Manager
AIDS Treatment Data Network
259 West 30th Street, 9th Floor
New York, NY 10001 USA
Telephone: (212) 268-4196 (English) or 643-0870 (Spanish)
FAX: (212) 268-4199
E-mail: AIDSTreat@AOL.com

Scope: AIDS/HIV related clinical and educational research studies.

FNCLCC CANCER CLINICAL TRIALS REGISTRY

Françoise Bonichon, MD
Biostatistics Unit
FNCLCC Cancer Trials Registry
Fondation Bergonié
180, Rue de Saint Genes
33000 Bordeaux Cedex FRANCE
Telephone: 33 56.92.43.00 or 56.92.44.90 (direct)
FAX: 33 56.91.68.29

Scope: Phase III and randomized Phase II cancer clinical trials in which patients from French Cancer Centers are included.

INTERACT

Ms. Robyn Sussel
INTERact Secretariat
The Canadian HIV Trials Network
200-1033 Davie Street
Vancouver, BC V6E 1M7 CANADA
Telephone: (604) 631-5327
FAX: (604) 631-5210

Scope: Inaugurated in August 1994, this register incorporates the contents of at least six major national and international registries of clinical trials of HIV/AIDS treatments.

THE INTERNATIONAL COMMITTEE ON THROMBOSIS AND HAEMOSTATIS (ICTH) REGISTRY ON CLINICAL TRIALS

Jean-Pierre Boissel, MD
Nadine Bossard, MD
Unité de Pharmacologie Clinque
Hospices Civils de Lyon, Université Claude Bernard
162, Avenue Lacassagne BP 3041
69394 Lyon Cedex 03 FRANCE
Telephone: 33 78.53.12.61
FAX: 33 78.53.10.30

Scope: Controlled, multicentre trials of treatment of thrombotic and/or atherosclerotic diseases and bleeding disorders.

INTERNATIONAL REGISTRY OF CLINICAL TRIALS OF HIV THERAPIES AND VACCINES

Marcel Fortin, BA, MLIS

The Canadian HIV Trials Network
200-1033 Davie Street
Vancouver, BC V6E 1M7 CANADA
Telephone: (604) 631-5327
FAX: (604) 631-5210

Scope: Studies of HIV therapies and vaccines. This registry has been subsumed by INTERact.

INTERNATIONAL REGISTRY OF PERINATAL TRIALS

Roger Soll, MD
International Registry of Perinatal Trials
Department of Pediatrics
University of Vermont College of Medicine
Given Building
Burlington, VT 05405 USA
Telephone: (802) 656-2296
FAX: (802) 656-4844

Scope: Planned, ongoing, or completed but unpublished trials using random allocation of participants to alternative forms of care during pregnancy, childbirth, the puerperium, or the neonatal period.

INTERNATIONAL REGISTRY OF VISION TRIALS

Kay Dickersin, PhD
Department of Epidemiology and Preventive Medicine
University of Maryland School of Medicine
506 West Fayette Street
Baltimore, MD 21201 USA
Telephone: (410) 706-5295
FAX: (410) 328-0110
E-mail: kdickers@umabnet.umd.edu

Scope: Randomized or quasi-randomized clinical trials in vision research.

MUSCULOSKELETAL RESEARCH REGISTER

Mr. Chad B. Munger
Department of Health Policy and Research
American Academy of Orthopaedic Surgeons
6300 North River Road
Rosemont, IL 60018 USA
Telephone: (708) 823-7186
FAX: (708) 823-8027

Scope: All studies concerned with musculoskeletal conditions in humans.

NAM UK DIRECTORY OF HIV/AIDS TREATMENTS AND TRIALS

Mr. Edward King
NAM Charitable Trust
Unit 52, The Eurolink Centre
49 Effra Road, Brixton
London SW2 1B2 UK
Telephone: 44 71 737-2004
FAX: 44 71 737-6190

Scope: Open or closed clinical trials of treatments for HIV/AIDS; also, such trials terminated since November 1991. In addition, a separate section of the directory provides reviews of published data on drugs and on treatments for opportunistic infections.

NCI CANCER CONTROL INTERVENTION STUDIES

Brenda K. Edwards, PhD
Surveillance Program
Division of Cancer Prevention and Control
National Cancer Institute
9000 Rockville Pike, EPN Room 343
Bethesda, MD 20892 USA
Telephone: (301) 496-8506
FAX: (301) 402-0816

Scope: NCI funded intervention studies pertaining to cancer prevention and control.

NATIONAL CLINICAL TRIALS REGISTRY

R. John Simes, MD, FRCAP, Director
Davina Ghersi, Registrar
The National Clinical Trials Registry
NHMRC Clinical Trials Centre
University of Sydney
Edward Ford Building (A27)
Sydney, NSW 2006 AUSTRALIA
Telephone: 61 (02) 692-4561
FAX: 61 (02) 552-3881

Scope: Randomized treatment allocation evaluating interventions in cancer and cardiovascular disease in humans.

NATIONAL INSTITUTES OF HEALTH INVENTORY OF CLINICAL TRIALS AND STUDIES

John H. Ferguson, MD
Office of Medical Applications of Research
National Institutes of Health
Federal Building, Room 618
7550 Wisconsin Avenue
Bethesda, MD 20892 USA
Telephone: (301) 496-1143/44
FAX: (301) 402-0420

Scope: All clinical trials funded by the NIH. Note that while this register still exists, it is not currently being maintained.

NEUROSURGERY CLINICAL TRIALS REGISTRY

Stephen J. Haines, MD
Department of Neurosurgery
University of Minnesota Health Sciences Center
420 Delaware Street, SE, Box 96 UMHC
Minneapolis, MN 55455 USA
Telephone: (612) 624-8651
FAX: (612) 624-0644

Scope: Comparative clinical trials of topics relevant to the practice of neurosurgery.

OTTAWA STROKE TRIAL REGISTRY

David Moher, MSc
Clinical Epidemiology Unit
Loeb Medical Research Institute
1053 Carling Avenue
Ottawa, Ontario K1Y 4E9 CANADA
Telephone: (613) 798-5555 ext. 6055
FAX: (613) 761-5351

Scope: Randomized clinical trials in which the primary focus is stroke, including trials of stroke prevention, acute stroke treatment and long-term care of stroke patients. The definition of stroke used includes cerebral thrombosis, cerebral embolus and SAH.

PHYSICIAN DATA QUERY (PDQ)

Gisele Sarosy, MD
International Cancer Information Center
National Cancer Institute
Bldg 82, Rm 113
Bethesda, MD 20892 USA

Telephone: (301) 496-7406
FAX: (301) 480-8105

Scope: Protocols sponsored by NIH are included, along with protocols submitted from individual institutions or drug companies that have been approved by the PDQ Editorial Board.

PHILADELPHIA AIDS PROTOCOL TESTING DIRECTORY

Mr. Richard Scott, Director
Medical Affairs, Policy and Planning
AIDS Activities Coordinating Office
Department of Public Health
500 South Broad Street, 2nd Floor
Philadelphia, PA 19146 USA
Telephone: (215) 875-5688
FAX: (215) 875-5210

Scope: Studies concerning treatment for AIDS located within Philadelphia County boundaries.

REGISTER OF INVESTIGATIVE PROTOCOLS TO TREAT MALIGNANT BRAIN TUMORS IN NORTH AMERICA

Ms. Jennifer Ricks
International Cancer Information Center
National Cancer Institute
Bldg. 82, Room 103
Bethesda, MD 20892 USA
Telephone: (301) 496-7406
FAX: (301) 480-8105

Scope: Active Phase I-III trials employing surgery, radiation, chemotherapy, or other treatment modalities for treatment of malignant brain tumors in adults and children. No closed or discontinued trials are included.

REGISTRY OF HIV CLINICAL TRIALS IN CANADA

Mr. Marcel Fortin
The Canadian HIV Trials Network
200-1033 Davie Street.
Vancouver, BC V6E 1M7 CANADA
Telephone: (604) 631-5327
FAX: (604) 631-5210

Scope: Any clinical trials taking place in Canada concerning HIV and HIV-related illnesses.

SOUTHERN CALIFORNIA HIV TREATMENT DIRECTORY

Ms. Jill Glassbrenner
S.W. Community Based AIDS Treatment Group
1800 North Highland Avenue, Suite 610
Los Angeles, CA 90028 USA
Telephone: (213) 469-5888
FAX: (213) 464-4404

Scope: Open for enrollment phase I-III clinical studies on HIV+ patients.

SPANISH DATABASE OF CLINICAL TRIALS

Fernando Garcia Alonso, MD
Subdirección General de Evaluación de Medicamentos
Ministerio de Sanidad y Consumo
Paseo del Prado 18-20
Madrid 28014 Spain
Telephone: 34 (1) 420.2068
FAX: 34 (1) 420.3217

Scope: Drug clinical trials to be performed in Spain.

UKCCCR REGISTER

Peter Fayers, Senior Statistician
UK Coordinating Committee on Cancer Research (UKCCCR)
MRC Cancer Trials Office
5 Shaftsbury Road
Cambridge CB2 2BW UK
Telephone: 44 0223 311110
FAX: 44 0223 311844

Scope: Phase I, II or III cancer trials. Plans for a European register (ECCTR) which will subsume this one.

WASHINGTON-BALTIMORE HIV RESEARCH DIRECTORY

Peter Hawley, MD
Whitman-Walker Clinic
1407 S Street, NW
Washington, DC 20009 USA
Telephone: (202) 745-6159
FAX: (202) 745-0238

Scope: AIDS/HIV research protocols in the Washington, DC to Baltimore, MD area.

Appendix 5b. Definitions of RCT and CCT

From Dickersin K, Larson K. Establishing and maintaining an international register of RCTs. In: Cochrane Collaboration Handbook. Oxford: The Cochrane Collaboration, 1995 [updated 14 July 1995]. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996; Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

5b.1 Criteria for registering studies with the Cochrane Collaboration's International Register of Published RCTs of Health Care
5b.2 National Library of Medicine Definitions Publication type terms: RANDOMIZED CONTROLLED TRIAL and CONTROLLED CLINICAL TRIAL.

5b.1 Criteria for registering studies with the Cochrane Collaboration's International Register of Published RCTs of Health Care

Overarching principle: The highest possible proportion of all reports of RCTs of health care should be included in the International Register. Thus, those searching the literature to identify trials should give reports the benefit of any doubts. Publications which simply mention the possibility of undertaking an RCT should not be included, however.

Eligibility criteria: Reviewers will decide whether to include a particular report in a review. The aim of the Register is to provide reviewers with all possible trials they may wish to include in a review, not to decide whether a report is worthy or relevant for inclusion.

Relevant reports are reports published in any year, of studies comparing at least two forms of health care (medical treatment, medical 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.
Studies are eligible for inclusion in the Register if allocation to the intervention was random or intended-to-be-random (e.g., alternation), or if a concurrent control group was used in the trial and it is possible that a random or intended-to-be-random method was used to allocate participants to the study groups. Judgements as to the quality of the methods used or whether the authors actually did what they claimed should not be used to decide eligibility for inclusion in the Register.

A trial should thus be included in the Register if, based on the best available information, 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).

In addition:

- If one or more outcomes were assessed using "double blinding" or "double masking" such that neither the participant/patient nor the assessor was aware of the intervention received, but randomization is not mentioned explicitly in the text, a trial should be included.
- Crossover trials, in which patients have been assigned to the first intervention using

random or quasi-random allocation, should be included.

- Reports of trials dealing only with animals should not be included.
- Units of randomization may be individuals, groups (such as communities or hospitals), organs (such as eyes) or other parts of the body (such as teeth).
- A report of a randomized trial should be included even when no results are presented or when results are limited to the analyses of baseline variables.
- Articles describing an intended or ongoing trial or commenting on a trial (such as in an editorial) should be brought to the attention of the Baltimore Cochrane Center, but are not eligible for inclusion in the International Register of Published RCTs.

5b.2 National Library of Medicine Definitions Publication type terms: RANDOMIZED CONTROLLED TRIAL, CONTROLLED CLINICAL TRIAL

MH Randomized Controlled Trial
DC 2
IDXSH NULL LIST
MH_TH NLM (1991)
MHLEX NON
GM publication type: for reports of randomized controlled trials
AN publication type only; to designate a type of clin trial in which two or more groups are chosen at random, one receiving the service, the other not; for randomized controlled trials as a subject or of value as research, index under RANDOMIZED CONTROLLED TRIALS (MH); do not confuse with Publication type CONTROLLED CLINICAL TRIAL; do not interpret trial design: use term of author; if in doubt, read MeSH definitions; coord IM or NIM any other epidemiol or statist method of design present; Manual 17.36+, 26.26.3
MS 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.
OL search policy: Online Manual; iuse: main heading AND RANDOMIZED CONTROLLED TRIAL (PT)
MH Controlled Clinical Trial
DC 2
IDXSH NULL LIST
MH_TH NLM (1995)
MHLEX NON
GM publication type: for reports of controlled clinical trials
AN publication type only; to designate a type of clin trial of drugs, devises, procedures for diag, ther or prev eff under a tightly designed protocol; for controlled clin trials as a subject, index under CONTROLLED CLINICAL TRIALS (MH); do not confuse with Publication Type RANDOMIZED CONTROLLED TRIAL; do not interpret trial design: use term of author; if in doubt, read MeSH definitions; Manual 17.11+
MS A clinical trial involving one or more test treatments, at least one control treatment, specified outcome measures for evaluating the studied intervention, and a 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.
OL search policy: Online Manual; iuse: main heading AND CONTROLLED CLINICAL TRIAL (PT)

Appendix 5c. Optimal search strategy for RCTs

From Dickersin K, Larson K. Establishing and maintaining an international register of RCTs. In: Cochrane Collaboration Handbook. Oxford: The Cochrane Collaboration, 1995 [updated 14 July 1995]. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996; Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

5c.1 Silver Platter (version 3.10) format

#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 or #6
#8 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#9 #7 not #8
#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 (#14 in TI) or (#14 in AB)

#16 PLACEBOS
#17 placebo* in TI
#18 placebo* in AB
#19 random* in TI
#20 random* in AB
#21 RESEARCH-DESIGN
#22 #10 or #11 or #12 or #13or #15 or #16 or #17 or #18 or #19 or #20 or #21
#23 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#24 #22 not #23
#25 #24 not #9
#26 TG=COMPARATIVE-STUDY
#27 explode EVALUATION-STUDIES
#28 FOLLOW-UP-STUDIES
#29 PROSPECTIVE-STUDIES
#30 control*or prospectiv* or volunteer*
#31 (#30 in TI) or (#30 in AB)
#32 #26 or #27 or #28 or #29 or #31

#33 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#34 #32 not #33
#35 #34 not (#9 or #25)
#36 #9 or #25 or #35

- 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.

5c.2 CD-PLUS Ovid (version 3.0) format

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

10 CLINICAL TRIAL.pt.
11 exp CLINICAL TRIALS
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

21 COMPARATIVE STUDY.sh.
22 exp EVALUATION STUDIES
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 26 not (9 or 20)
29 9 or 20 or 28

- 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.

Appendix 8a. Effect measures for dichotomous data

		Adverse outcome	
		+	-
Treatment	+	A	B
	-	C	D

Odds ratio = (A/B)/(C/D)
Relative risk (RR) = [A/(A+B)]/[C/(C+D)]
Relative risk reduction = 1 - RR
Absolute risk reduction (ARR) = A/(A+B) - C/(C+D)
Number needed to treat = 1/ARR

a) When the outcome is undesirable, a relative risk or odds ratio less than one represents a beneficial treatment or exposure, with zero representing 100% effectiveness. An absolute risk reduction of less than zero representatives a benefit, and 100% effectiveness would be equivalent to the risk observed in the control group.

b) 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), and both expressions are equivalent to A×D/B×C. From the two expressions, you can see that 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.

GLOSSARY

Absolute risk reduction
See *risk difference*.

Additive model
A model in which the combined effect of several factors is the sum of the effects produced by each of the factors. For example, if one factor multiplies risk by *a* and a second factor by *b*, the combined effect of the two factors is *a* + *b*. See also *multiplicative model*.

Administrator (of a *Collaborative Review Group*)
See *Coordinator*

Adminors
The name of an *e-mail* discussion list for *coordinators* of *Collaborative Review Groups*, and the name of a sub-directory for this group of people on the UK Cochrane Centre's *FTP* Server.

Allocation concealment
See *concealment of allocation*.

Applicability (synonyms: external validity, generalisability, relevance, transferability)
The degree to which the results of an observation, study or review hold true in other settings.

Attrition bias
Systematic differences between comparison groups in withdrawals or exclusions of participants from the results of a study. For example, patients may drop out of a study because of side effects of the intervention. Excluding these patients from the analysis could result in an overestimate of the effectiveness of the intervention.

Bayesian analysis
An approach that can be used in single studies or meta-analysis which incorporates a prior *probability distribution* based on subjective opinion and objective evidence, such as the results of previous research. Bayesian analysis uses *Bayes' theorem* to update the prior distribution in light of the results of a study, producing a posterior distribution. Statistical inferences (*point estimates*, *confidence intervals*, etc.) are based on this posterior distribution. The posterior distribution also acts as the prior distribution for the next study. This approach has many attractive features, but is controversial because it depends on opinions, and frequently they will vary considerably.

Bayes' theorem
A probability theorem used to obtain the probability of a condition in a group of people with some characteristic (e.g. exposed to an intervention of interest, or with a specified result on a diagnostic test) on the basis of the overall rate of that condition (the prior probability) and the likelihoods of that characteristic in people with and without the condition.

Bias
A systematic error or deviation in results or inferences. In studies of the effects of healthcare bias can arise from systematic differences in the groups that are compared (*selection bias*), the care that is provided, or exposure to other factors apart from the intervention of interest (*performance bias*), withdrawals or exclusions of people entered into the study (*attrition bias*) or

how outcomes are assessed (*detection bias*). Bias does not necessarily carry an imputation of prejudice, such as the investigators' desire for particular results. This differs from conventional use of the word in which bias refers to a partisan point of view. Many varieties of bias have been described {489}. See also *methodological quality*, validity.

Blinding (synonym: masking)
Keeping secret group assignment (e.g., to treatment or control) from the study participants or investigators. Blinding is used to protect against the possibility that knowledge of assignment may affect patient response to treatment, provider behaviours (*performance bias*) or outcome assessment (*detection bias*). Blinding is not always practical (e.g., when comparing surgery to drug treatment). The importance of blinding depends on how objective the outcome measure is; blinding is more important for less objective outcome measures such as pain or quality of life. See also *single blind*, double blind, and *triple blind*.

Breslow-Day test
A statistical test for the *homogeneity* of *odds ratios*.

Case series
An uncontrolled observational study involving an intervention and outcome for more than one person.

Case study (synonyms: anecdote, case history, single case report)
An uncontrolled observational study involving an intervention and outcome for a single person.

Case-control study (synonyms: case referent study, retrospective study)
A study that starts with identification of people with the disease or outcome of interest (cases) and a suitable control group without the disease or outcome. The relationship of an attribute (intervention, exposure, or risk factor) to the outcome of interest is examined by comparing the frequency or level of the attribute in the cases and controls. For example, to determine whether thalidomide caused birth defects a group of children with birth defects (cases) could be compared to a group of children without birth defects (controls). The groups would then be compared with respect to the proportion exposed to thalidomide through their mothers taking the tablets. Case-control studies are sometimes described as being *retrospective* as they are always performed looking back in time.

CCTR
See *Cochrane Controlled Trials Register*.

CD-ROM (Compact Disc - Read Only Memory)
A computer storage medium. A CD-ROM can contain a database of information (e.g. *MEDLINE*, or the *Cochrane Controlled Trials Register*) that may be searched either on a personal computer or a computer linked to a network.

CDSR
See *Cochrane Database of Systematic Reviews*.

Chi-square test
Any statistical test based on comparison of a test statistic to a chi-square distribution. The *Mantel-Haenszel test* is a well-known chi-square test.

Client Manager
Client Manager is a DOS-based software package (written for the *Cochrane Collaboration*) that enables the creation of a *database* for storing names and contact information. Client Manager is used by some *Collaborative Review Groups*, Cochrane Centres, and other entities within the Collaboration to keep up-to-date contact information on people. It is being replaced by a Windows-based package. See HIREx.

CI
See *Confidence interval*

CINAHL (Cumulative Index of Nursing and Allied Health Literature)
Electronic *database* covering the major journals in nursing and allied health. Years of coverage: 1983 - present.

CL
See *Cochrane Library*

Clinical trial (synonyms: therapeutic trial, intervention study)
A trial that tests out a drug or other intervention to assess its effectiveness and safety. This general term encompasses *randomised controlled trials* and *controlled clinical trials*.

Cochrane Centres
An entity in the *Cochrane Collaboration* with responsibility for helping to co-ordinate and support the Collaboration. Responsibilities include maintaining a directory of people contributing to the *Cochrane Collaboration*; helping to establish *Collaborative Review Groups*; organising workshops, seminars and annual colloquia to support and guide the development of the *Cochrane Collaboration*. Each Centre is responsible for providing support within a specified geographic area. Details of Centre responsibilities and a list of the Centre responsible for any given country are available in *CDSR*.

Cochrane Collaboration
An international organisation that aims to help people make well informed decisions about health by preparing, maintaining, and ensuring the accessibility of systematic reviews of the benefits and risks of healthcare interventions.

Cochrane Collaboration Handbook
Guidelines for preparing and maintaining *Cochrane Reviews*. Information about the Collaboration that was previously contained in Sections I to V of the Handbook is now maintained and published as a module in *CDSR*. The Handbook is published in the *Cochrane Library* and it can be downloaded from *FTP* servers at the Australasian, Canadian and UK *Cochrane Centres*.

Cochrane Consumer Network
A registered *entity* in the *Cochrane Collaboration* responsible for co-ordinating and facilitating consumer involvement in the Collaboration.

Cochrane Controlled Trials Register (CCTR)
A database of references to controlled trials in health care. Cochrane groups and other organisations have been invited to contribute their specialised registers, and these registers, together with references to clinical trials identified on *MEDLINE*, form The Cochrane Controlled

Trials Register (CCTR).

Cochrane Database of Systematic Reviews (CDSR)
The major product of the *Cochrane Collaboration*. It brings together all the currently available *Cochrane Reviews* and is updated quarterly. It also contains information about the Collaboration. *Collaborative Review Groups* submit *modules* of edited *reviews* and other information to the *Parent Database* for inclusion in the *CDSR*. See *Cochrane Library*.

Cochrane Library (CL)
A collection of databases, published on disk and CD-ROM and updated quarterly, containing the *Cochrane Database of Systematic Reviews*, the *Cochrane Controlled Trials Register*, the *Database of Abstracts of Reviews of Effectiveness*, the *Cochrane Review Methodology Database*, and information about the *Cochrane Collaboration*.

Cochrane Review
A Cochrane Review is a systematic, up-to-date summary of reliable evidence of the benefits and risks of healthcare. Cochrane Reviews are intended to help people make practical decisions. For a review to be called a "Cochrane Review" it must be in the *Parent Database* maintained by the *Cochrane Collaboration*. The Parent Database is composed of *modules* of reviews submitted by *Collaborative Review Groups* (CRGs) registered with the Cochrane Collaboration. The reviews contributed to one of the modules making up the Parent Database are refereed by the *editorial team* of the CRG, as described in the CRG module. Reviewers adhere to guidelines published in the *Cochrane Handbook*. The specific methods used in a Review are described in the text of the review. Cochrane Reviews are prepared using *Review Manager* software provided by the Collaboration and adhere to a structured format that is described in the Handbook.

Cochrane Review Methodology Database (CRMD)
A bibliography of articles and books about methodological issues relevant to summarising evidence of the effects of healthcare. It is published in the *Cochrane Library*.

Cohort study (synonyms: follow-up, incidence, longitudinal, prospective study)
An observational study in which a defined group of people (the cohort) is followed over time and outcomes are compared in subsets of the cohort who were exposed or not exposed, or exposed at different levels, to an intervention or other factor of interest. Cohorts can be assembled in the present and followed into the future (a "concurrent cohort study"), or identified from past records and followed forward from that time up to the present (a "historical cohort study"). Because *random allocation* is not used, matching or statistical adjustment must be used to ensure that the comparison groups are as similar as possible.

Cointervention
In a *randomised controlled trial*, the application of additional diagnostic or therapeutic procedures to members of either or both the experimental and the control groups.

Collaborative Review Group (CRG)
The primary working entity (organisational unit) of the *Cochrane Collaboration*. CRGs are made up of individuals sharing an interest in a particular healthcare problem or type of problem. The main purpose of a CRG is to prepare and maintain *systematic reviews* of the effects of health care within the scope of the Group. Members participate in the Group not only by preparing *Cochrane Reviews* but also by *hand-searching* journals or other activities that help

the Group to fulfil its aim. Each CRG is co-ordinated by an *editorial team*, responsible for regularly updating and submitting to the *Parent Database* an edited *module* of Reviews and information about the Group.

Collaborative Trialists Group
Investigators who conducted similar *randomised controlled trials* independently and agree to contribute *individual patient data* from their trials to a *meta-analysis*.

Concealment of allocation
The process used to prevent foreknowledge of group assignment in a randomised controlled trial, which should be seen as distinct from *blinding*. The allocation process should be impervious to any influence by the individual making the allocation by having the *randomisation* process administered by someone who is not responsible for recruiting participants, for example, a hospital pharmacy, or a central office. Using methods of assignment such as date of birth and case record numbers (see *quasi random allocation*) are open to manipulation. Adequate methods of allocation concealment include centralized randomisation schemes; randomisation schemes controlled by a pharmacy; numbered or coded containers in which capsules from identical-looking, numbered bottles are administered sequentially; on-site computer systems, where allocations are in a locked unreadable file; and sequentially numbered opaque, sealed envelopes.

Conference abstracts
Short summaries of presentations at conferences. May be published as proceedings.

Confidence interval (CI)
The range within which the "true" value (e.g., size of effect of an intervention) is expected to lie with a given degree of certainty (e.g. 95% or 99%). Note: Confidence intervals represent the probability of *random errors*, but not systematic errors (*bias*).

Confounding
A situation in which a measure of the effect of an intervention or exposure is distorted because of the association of exposure with other factor(s) that influence the outcome under study.

Consumer (healthcare consumer)
Someone who uses, is affected by, or who is entitled or compelled to use a health-related service.

Consumer advocate or representative
Consumer who is actively involved with other consumers and able to represent the perspectives and concerns of that broader group of people. A consumer advocate or representative should be linked with other consumers, accountable to them, and should not have a conflict of interest in that role.

Contamination
In *clinical trials*, the inadvertent application of the intervention being evaluated to people in the *control* group or inadvertent failure to apply the intervention to people assigned to the intervention group.

Contingency table
A tabular cross-classification of data such that subcategories of one characteristic are indicated

horizontally (in rows) and subcategories of another characteristic are indicated vertically (in columns). *Tests of association* between the characteristics can be readily applied. The simplest contingency table is the fourfold, or 2x2 table, which is used in *clinical trials* to compare *dichotomous* outcomes, such as death, for an intervention and control group or two intervention groups.

Continuous data
Data with a potentially infinite number of possible values along a continuum. Height, weight, and blood pressure are examples of continuous variables.

Control
1. In *clinical trials* comparing two or more interventions, a control is a person in the comparison group that receives a *placebo*, no intervention, usual care or another form of care.
2. In *case-control studies* a control is a person in the comparison group without the disease or outcome of interest.
3. In statistics control means to adjust for or take into account extraneous influences or observations.
4. Control can also mean programs aimed at reducing or eliminating the disease when applied to communicable (infectious) diseases.

Controlled clinical trial
Refers to a study that compares one or more intervention groups to one or more comparison (*control*) groups. Whilst not all controlled studies are *randomised*, all *randomised trials* are *controlled*.

Co-ordinating editor (of a *Collaborative Review Group*)
A member of the *editorial team* located at the *editorial base* who is responsible for editing reviews and, supported by other members (particularly the *coordinator*) for fostering the smooth running of the Group and assuring the quality of the Group's *module*.

Coordinator (of a *Collaborative Review Group*) (Previously known as *Administrator*)
The key person in managing and supporting a *Collaborative Review Group* (CRG) on a day-to-day basis. Most CRGs have a full-time coordinator working from an *editorial base*. Responsibilities of a coordinator include co-ordinating the activities of the CRG; fostering liaison and communication between *editors* and *reviewers*; setting up and maintaining a *trials register*; producing newsletters; providing *reviewers* with the relevant software (*RevMan*), manuals and support to do their reviews; transferring reviews to the *Parent Database* via the *Module Manager* (*ModMan*) software, for inclusion in the *Cochrane Database of Systematic Reviews*. Coordinators come from a variety of backgrounds, and whilst some also prepare *systematic reviews* in addition to their work as coordinator, many do not.

Cost-benefit analysis
An *economic analysis* that converts effects into the same monetary terms as the costs and compares them.

Cost-effectiveness analysis
An *economic analysis* that converts effects into health terms and describes the costs for some additional health gain (e.g. cost per additional stroke prevented).

Cost-utility analysis
An *economic analysis* that converts effects into personal preferences (or *utilities*) and describes how much it costs for some additional quality gain (e.g., cost per additional quality-adjusted life-year).

CRG
See *Collaborative Review Group*.

CRMD
See *Cochrane Review Methodology Database*

Critical appraisal
The process of assessing and interpreting evidence by systematically considering its *validity*, results and relevance.

Cross-sectional study (synonym: prevalence study)
A study that examines the relationship between diseases (or other health related characteristics) and other variables of interest as they exist in a defined population at one particular time. The temporal sequence of cause and effect cannot necessarily be determined in a cross-sectional study.

Cross-over trial
A type of clinical trial comparing two or more interventions in which the participants, upon completion of the course of one treatment are switched to another. For example, for a comparison of treatments A and B, half the participants are randomly allocated to receive them in the order A, B and half to receive them in the order B, A. A problem with this design is that the effects of the first treatment may carry over into the period when the second is given.

Cumulative meta-analysis
In cumulative meta-analysis studies are added one at a time in a specified order (e.g., according to date of publication or quality) and the results are summarised as each new study is added. In a graph of a cumulative meta-analysis each horizontal line represents the summary of the results as each study is added, rather than the results of a single study.

Current Contents
Electronic *database* that provides access to the tables of contents and bibliographic data from current issues of the world's leading scholarly research journals in the sciences, social sciences, arts, and humanities. Over 6,600 journals covered.

DARE
See *Database of Abstracts of Reviews of Effectiveness*.

Database
A collection of organised information, usually held on a computer. In some ways a database is similar to a filing system, but with important advantages: the information can be revised and kept up to date easily, and the computer can retrieve information from it very quickly. Electronic databases such as *MEDLINE*, *EMBASE* and the *CDSR* can be distributed on disk, CD-ROM or via the *Internet*.

Database of Abstracts of Reviews of Effectiveness (DARE)

A collection of structured abstracts and bibliographic references of systematic reviews of the effects of healthcare. See the *Cochrane Library*.

Decision analysis
A technique used to aid decision-making under conditions of uncertainty by systematically representing and examining all of the relevant information for a decision and the uncertainty around that information. The available choices are plotted on a decision tree. At each branch, or decision node, the probabilities of each outcome that can be predicted are estimated. The relative worth or preferences of decision-makers for the various possible outcomes for a decision can also be estimated and incorporated in a decision analysis.

Degrees of freedom
The number of independent comparisons that can be made between the members of a sample. It refers to the number of independent contributions to a sampling distribution (such as *chi-square distribution*). In a *contingency table* it is one less than the number of row categories multiplied by one less than the number of column categories; e.g. a 2 x 2 table comparing two groups for a *dichotomous* outcome, such as death, has one degree of freedom.

Detection bias (synonym: ascertainment bias)
Systematic differences between comparison groups in how outcomes are ascertained, diagnosed or verified.

Dichotomous data (synonym: binary data)
Observations with two possible categories such as dead/alive, smoker/non-smoker, present/not present.

Double blind (synonym: double masked)
Neither the participants in a trial nor the investigators (outcome assessors) are aware of which intervention the participants are given. The purpose of blinding the participants (recipients and providers of care) is to prevent *performance bias*. The purpose of blinding the investigators (outcome assessors, who might also be the care providers) is to protect against *detection bias*. See also *blinding*, single blind, triple blind, concealment of allocation.

E-mail (electronic mail)
Allows users on the *Internet* and other networks (both local and global) to communicate electronically by sending messages to individuals, or groups of individuals. E-mail is, for most people, cheaper and faster than other communication channels such as fax and standard mail.

Economic analysis (synonym: economic evaluation)
Comparison of the costs and outcomes of alternative health care interventions. See *cost-benefit analysis*, cost-*effectiveness analysis* and *cost-utility analysis*.

Editor (of a *Collaborative Review Group*)
A member of the *editorial team*, often not located at the *editorial base*, who not only prepares and maintains one or more *systematic reviews* as a member of a *CRG*, but also has responsibilities to support the *co-ordinating editor* in editing *systematic reviews* prepared by others, and in fostering the smooth running of the Group.

Editorial base
Collaborative Review Groups have an editorial base where the co-ordinating editor, the

coordinator, secretarial support, and the Group's *trials register* are located, and to which *reviewers* are encouraged to come to work on their reviews.

Editorial process
The process by which each individual *CRG* decides on the criteria for editing and including *reviews* in its edited *module* for inclusion in the *Cochrane Database of Systematic Reviews*. *Protocols* may also be reviewed both by the editors (internal review) and by external peer reviewers. See also *referee process*.

Editorial team (of a *Collaborative Review Group*)
Normally consists of a *coordinator*, several *editors,* and a secretary.

Effect size
1. A generic term for the *estimate of effect* for a study.
2. A dimensionless measure of effect that is typically used for *continuous data* when different scales (e.g., for measuring pain) are used to measure an outcome and is usually defined as the difference in means between the intervention and *control* groups divided by the standard deviation of the control or both groups. See *standardised mean difference*.

Effectiveness
The extent to which a specific intervention, when used under ordinary circumstances, does what it is intended to do. Clinical trials that assess effectiveness are sometimes called management trials. See also *intention-to-treat*.

Efficacy
The extent to which an intervention produces a beneficial result under ideal conditions. Clinical trials that assess efficacy are sometimes called explanatory trials and are restricted to participants who fully co-operate.

EMBASE (Excerpta Medica database)
A European-based electronic *database* of pharmacological and biomedical literature covering 3,500 journals from 110 countries. Years of coverage - 1974 to present.

Empirical
Empirical results are based on experience (or observation) rather than on reasoning alone.

Entities
The term used for registered groups in the *Cochrane Collaboration* (*Collaborative Review Groups*, *Centres*, *Fields*, *Methods Working Groups,* and the *Cochrane Consumer Network*).

Epidemiology
The study of the distribution and determinants of health-related states or events in specified populations.

Estimate of effect (synonym: treatment effect)
In studies of the effects of healthcare, the observed relationship between an intervention and an outcome expressed as, for example, a *number needed to treat*, *odds ratio*, *risk difference*, *relative risk*, *standardised mean difference*, or *weighted mean difference*.

Event rate

The proportion of participants in a group in whom an event is observed. Thus, if out of 100 patients the event (e.g. a stroke) is observed in 32, the event rate is 0.32.

Expected date (of a *Cochrane Review*)
The time by which a user of *CDSR* can expect to have access to a completed review. It appears on the title page of a *protocol* in *CDSR*, and is the date by which the review is expected to have been completed and gone through the *editorial process* of the *CRG* responsible for the *module* in which the review is found.

External peer reviewer (of a *Cochrane Review*)
A person with relevant content, methodological or user expertise who critically examines reviews in her/his area of expertise.

External validity (synonyms: external validity, generalisability, relevance, transferability)
The degree to which the results of an observation hold true in other settings. See also *validity*.

Extramural
Outside (the walls or boundaries of) a place or institution. Refers to "external" sources of support (such as funding) as opposed to "internal" (*intramural*) support.

F-test (synonym: variance ratio test)
A statistical test of the hypothesis that two population *variances* are the same. The *t test* assumes that this is the case.

Factorial design
Most trials only consider a single factor, where an intervention is compared with one or more alternatives, or a *placebo*. In a trial using a 2x2 factorial design, participants are allocated to one of four possible combinations. For example, in a 2x2 factorial, *RCT* of nicotine replacement and counselling, participants would be allocated to: nicotine replacement alone, counselling alone, both, or neither. In this way it is possible to test the independent effect of each intervention on smoking cessation and the combined effect of (interaction between) the two interventions.

Fields (Field co-ordination)
Fields are Cochrane *entities* that focus on dimensions of health care other than health problems such as the setting of care (e.g., primary care), the type of consumer (e.g. older people), the type of provider (e.g. nursing) or the type of intervention (e.g. physical therapies). Among their tasks, people working in Fields *hand search* specialist journals, help to ensure that priorities and perspectives in their field of interest are reflected in the work of *Collaborative Review Groups*, compile specialist *databases*, co-ordinate activities with relevant agencies outside the Collaboration and comment on *systematic reviews* relating to their particular area.

Fixed effect model
A statistical model that stipulates that the units under analysis (e.g., people in a trial or study in a *meta-analysis*) are the ones of interest, and thus constitute the entire population of units. Only within-study variation is taken to influence the uncertainty of results (as reflected in the *confidence interval*) of a meta-analysis using a fixed effect model. Variation between the estimates of effect from each study (heterogeneity) does not affect the confidence interval in a fixed effect model. See *random effects model*.

Fourfold table (synonym: 2x2 table)
A *contingency table* with two rows and two columns used in *clinical trials* to compare *dichotomous* outcomes, such as death, for an intervention and control group or two intervention groups.

FTP (File Transfer Protocol) Server
Enables users to open a connection to a host computer and log in (often anonymously, by using 'anonymous' as the user's name). Once logged in, files can be transferred between the host computer and the remote computer (the computer to which the host has been connected).

Funnel plot
A graphical display of sample size plotted against effect size that can be used to investigate publication bias. When many studies have been located that estimate the same effect, the distribution of points should resemble a funnel shape with a widening in the spread of effect sizes as sample size decreases. A gap on one side of the wide part of the funnel indicates that some studies have not been published or located. Funnel plots are not presently available in the Cochrane software.

Generalisability (synonyms: applicability, external validity, relevance, transferability)
Generalisability is the degree to which the results of a study or *systematic review* can be extrapolated to other circumstances, particularly to routine health care situations.

Gold standard
The method, procedure or measurement that is widely accepted as being the best available against which new interventions should be compared. It is particularly important in studies of the accuracy of diagnostic tests. For example, *handsearching* is sometimes used as the gold standard for identifying trials against which electronic searches of *databases* such as *MEDLINE* are compared.

Handbook
See *Cochrane Collaboration Handbook*

Handsearching
Handsearching within the *Cochrane Collaboration* refers to the planned searching of a journal page by page (i.e., by hand), including editorials, letters, etc., to identify all reports of *randomised controlled trials*. Normally a person would start by handsearching the current year of a journal and work backwards to 1948 (or volume 1 if after 1948). Once a trial is found, it is coded appropriately using definitions agreed within the *Cochrane Collaboration* and published in *CDSR*. All the identified trials, regardless of the topic, are sent to the Baltimore Cochrane Center, for forwarding to the US National Library of Medicine for re-tagging on *MEDLINE*, and trials that are within the scope of a *Collaborative Review Group* or *Field* go into their specialised *register of trials*. A handsearching manual is available through the Baltimore Cochrane Center, and should be read before handsearching is commenced. A journal hand search registration form must be completed for each journal title and sent to Baltimore to avoid duplication of effort.

Heterogeneity
In *systematic reviews* heterogeneity refers to variability or differences between studies in the estimates of effects. A distinction is sometimes made between "statistical heterogeneity" (differences in the reported effects), "methodological heterogeneity" (differences in study

design) and "clinical heterogeneity" (differences between studies in key characteristics of the participants, interventions, or outcome measures). Statistical tests of heterogeneity are used to assess whether the observed variability in study results (effect sizes) is greater than that expected to occur by chance. However, these tests have low *statistical power*. See also *homogeneity*.

HIREx
Windows-based *database* for storing names, addresses and other information. Currently under development at the Canadian *Cochrane Centre* for use by Cochrane *entities*.

Historical control
Person or group for whom data were collected earlier than for the group being studied. Because of changes over time in risks, prognosis, healthcare, etc. there is a large risk of *bias* (in studies that use historical controls) due to systematic differences between the comparison groups.

Homogeneity
In *systematic reviews* homogeneity refers to the degree to which the results of studies included in a review are similar. "Clinical homogeneity" means that, in trials included in a review, the participants, interventions and outcome measures are similar or comparable. Studies are considered "statistically homogeneous" if their results vary no more than might be expected by the play of chance. See *heterogeneity*.

Incidence
The number of new cases of a disease or event in a population during a specific period.

Index Medicus
Catalogue of the United States National Library of Medicine (NLM), and a periodical index to the medical literature. Available in printed form, or electronically as *MEDLINE*.

Individual patient data
In *systematic reviews* this term refers to the availability of raw data for each study participant in each included trial, as opposed to aggregate data (summary data for the comparison groups in each study). Reviews using individual patient data require collaboration of the investigators who conducted the original trials, who must provide the necessary data.

Intention-to-treat
An intention-to-treat analysis is one in which all the participants in a trial are analysed according to the intervention to which they were allocated, whether they received it or not. Intention-to-treat analyses are favoured in assessments of *effectiveness* as they mirror the noncompliance and treatment changes that are likely to occur when the intervention is used in practice, and because of the risk of *attrition bias* when participants are excluded from the analysis.

Internal validity
See *validity*.

Internet
Network of millions of computers world-wide. Computers on the Internet use compatible communication standards and share the ability to contact each other and share data. Users of

the Internet communicate via *electronic mail (e-mail)*, via Telnet (a process which allows a person to log in to a remote host), and via *FTP*. See also *World Wide Web*.

Intervention study
See *Clinical trial*.

Intramural
Within (the walls or boundaries of) a community or institution (e.g. a university). Used to distinguish from "external" (*extramural*) sources of support (such as funding).

LILACS (Latin American and Caribbean Health Sciences Literature)
An electronic *database* based on a regional database of medical and science literature. It is compiled by the Latin American and Caribbean Center for Health Science Information, a unit of the Pan American Health Organisation.

Logistic model
A statistical model of an individual's risk (probability of disease or some other outcome) as a function of a risk factor or intervention. This model has attractive statistical features and is widely used as a *regression model* for *dichotomous* outcomes. In *meta-analysis* (or *meta-regression*) the logistic model can be used to explore the relationship between study characteristics and study results.

Logistic regression
Logistic regression is used to investigate the relationship between an event rate or proportion and a set of independent variables. In *systematic reviews* it can be used to explore the relationship between key characteristics of the included studies and the results (observed effects) for each study.

Log-odds ratio
The (natural) log of the *odds ratio*. It is used in statistical calculations and in graphical displays of odds ratios in *systematic reviews*.

Mantel-Haenszel test
A summary *chi-square test* for stratified data and used when collecting for *confounding*. In meta-analyses the Mantel-Haenszel test is used to analyse data stratified (grouped) by study.

Masking
See *blinding*.

Mean (synonyms: arithmetic mean, average)
The average value, calculated by adding all the observations and dividing by the number of observations.

MEDLINE (MEDlars onLINE)
An electronic *database* produced by the United States National Library of Medicine. It indexes millions of articles in selected (about 3,700) journals. It is available through most medical libraries, and can be accessed on *CD-ROM*, the *Internet* and by other means. Years of coverage - 1966 to present.

MeSH headings (Medical Subject Headings)

Terms used by the United States National Library of Medicine to index articles in *Index Medicus* and *MEDLINE*. Designed to reduce problems that arise from, for example, differences in British and American spelling. The MeSH system has a tree structure in which broad subject terms branch into a series of progressively narrower subject terms.

Meta-analysis
The use of statistical techniques in a *systematic review* to integrate the results of the included studies. Also used to refer to systematic reviews that use meta-analysis.

Meta-regression
Multivariate *meta-analytic* techniques, such as *logistic regression*, used to explore the relationship between study characteristics (e.g. *allocation concealment*, baseline risk, timing of the intervention) and study results (the magnitude of effect observed in each study) in a systematic review.

MetaView
Software incorporated in *RevMan* and *CDSR* that does statistical analyses and prepares tabular and graphical displays of the results of the studies included in a review.

Methodological quality (synonyms: validity, internal validity)
The extent to which the design and conduct of a trial are likely to have prevented *systematic errors* (bias). Variation in quality can explain variation in the results of trials included in a *systematic review*. More rigorously designed (better 'quality') trials are more likely to yield results that are closer to the 'truth'. See also *external validity*, *validity*.

Methods Working Group (MWG)
An entity in the *Cochrane Collaboration* made up of individuals who are interested in the judgements that lead to selection, appraisal, synthesis, interpretation, and dissemination of health care information, such as statistical methods and informatics. Each MWG is responsible for preparing and maintaining a module that is published in *CDSR* and includes a description of the groups scope and activities.

Minimisation
A method of allocation used, particularly in small trials, to provide comparison groups that are closely similar for several variables. It can be done with or without a component of randomisation. It is best performed centrally with the aid of a computer program to ensure *allocation concealment*.

ModMan (Module Manager)
Software developed by the *Cochrane Collaboration* to allow *Collaborative Review Groups* to assemble and manage their edited *protocols* and *reviews*. ModMan also contains information. about the *Collaborative Review Group*. ModMan is used by *Collaborative Review Group coordinators* to edit and update modules that are sent electronically, at quarterly intervals, to the *Parent Database* for inclusion in the *CDSR*. A variation of the ModMan software is also used by other Cochrane *entities* to prepare *modules* for CDSR.

Module
Edited protocols and reviews, and information about a *Collaborative Review Group,* is referred to as the Group's module. This module is transferred electronically using *ModMan* to the *Parent Database* at quarterly intervals, for inclusion in the *CDSR*. Other Cochrane *entities* also

produce modules for inclusion in the Parent Database and CDSR.

Multiplicative model
A model in which the joint effect of two or more factors is the product of their effects. For example, if one factor multiplies risk by *a* and a second factor by *b*, the combined effect of the two factors is *a* x *b*. See also *additive model*.

MWG
See *Methods Working Groups*.

Negative study
A term used to refer to a study that does not have "*statistically significant*" (positive) results indicating a beneficial effect of the intervention being studied. The term can generate confusion because it refers to both statistical significance and the direction of effect, studies often have multiple outcomes, the criteria for classifying studies as "negative" are not always clear and, in the case of studies of risk or undesirable effects, "negative" studies are ones that do not show a harmful effect.

Null hypothesis
The statistical hypothesis that one variable (e.g., whether a study participant was allocated to receive an intervention) has no association with another variable or set of variables (e.g., whether a study participant died), or that two or more population distributions do not differ from one another. In simplest terms, the null hypothesis states that the results observed in a study are no different from what might have occurred by the play of chance.

Number needed to treat (NNT)
The number of patients who need to be treated to prevent one bad outcome. It is the inverse of the *risk difference*.

Observational study (synonym: non-experimental study)
A study in which nature can take its course. Changes or differences in one characteristic (e.g., whether or not people received the intervention of interest) are studied in relation to changes or differences in other(s) (e.g., whether or not they died), without the intervention of the investigator. There is a greater risk of *selection bias* than in experimental studies (*randomised controlled trials*).

Odds ratio (OR)
The ratio of the odds of an event in the experimental (intervention) group to the odds of an event in the *control* group. Odds are the ratio of the number of people in a group with an event to the number without an event. Thus, if a group of 100 people had an *event rate* of 0.20, 20 people had the event and 80 did not, and the odds would be 20/80 or 0.25. An odds ratio of one indicates no difference between comparison groups. For undesirable outcomes, an OR that is less than one indicates that the intervention was effective in reducing the risk of that outcome. When the event rate is small, odds ratios are very similar to *relative risks*.

Open clinical trial
1. A *clinical trial* in which the investigator is aware which intervention is being given to which participant (random allocation may or may not be used).
2. A clinical trial in which the investigator decides which intervention is to be given (non-random allocation). Also called *open label design.*

3. A clinical trial with an open *sequential design*.

Open label design
A trial in which the investigator decides who receives which intervention rather than using *random allocation*. See also *open clinical trial*.

Ordinal data
Data that are classified into more than two categories where there is a natural order to the categories: for example, non-smokers, ex-smokers, light smokers, and heavy smokers. Ordinal data are often reduced to two categories to simplify analysis and presentation, which may result in a considerable loss of information.

Paired design
A study in which participants or groups of participants are matched (e.g., based on prognostic factors) and one member of each pair is allocated to the experimental (intervention) group and the other to the *control* group.

Parallel group trial (synonym: independent group design)
A trial that compares two groups of people, one of which receives the intervention of interest and one of which is a *control* group. Some parallel trials have more than two comparison groups and some compare different interventions without including a non-intervention control group.

Parent Database
The compilation of *modules* prepared by *Collaborative Review Groups* and other Cochane *entities* published as the *Cochrane Database of Systematic Reviews* (CDSR).

Performance bias
Systematic differences in care provided apart from the intervention being evaluated. For example, if patients know they are in the *control* group they may be more likely to use other forms of care, patients who know they are in the experimental (intervention) group may experience *placebo* effects, and care providers may treat patients differently according to what group they are in. *Blinding* of study participants (both the recipients and providers of care) is used to protect against performance bias.

Peto method
A way of combining *odds ratios* that has become widely used in *meta-analysis*. The calculations are straightforward and understandable, but thids method produces biased results in some circumstances. It is a *fixed effect model*.

Peto odds ratio
An approximation to the exact odds ratios which are used when doing a meta-analysis using the *Peto method*. In some circumstances the Peto odds ratio can differ substantially from the exact odds ratio.

Phase I studies
The first stage in testing a new drug in humans. Usually performed on healthy volunteers without a comparison group.

Phase II studies
Second stage in testing a new drug in humans. Often performed on healthy volunteers. These are sometimes *randomised controlled trials*.

Phase III studies
Studies that are a full-scale evaluation of treatment. After a drug has been shown to be reasonably effective, it is essential to compare it to the current standard treatments for the same condition. Phase III studies are often *randomised controlled trials*.

Phase IV studies
Studies that are concerned with post-marketing surveillance. They are often promotional exercises aimed at bringing a new drug to the attention of a large number of clinicians and may be of limited scientific value.

Placebo
An inactive substance or procedure administered to a patient, usually to compare its effects with those of a real drug or other intervention, but sometimes for the psychological benefit to the patient through a belief that s/he is receiving treatment. Placebos are used in *clinical trials* to *blind* people to their treatment allocation. Placebos should be indistinguishable from the active intervention to ensure adequate *blinding*.

Placebo effect
A favourable response to an intervention, regardless of whether it is the real thing or a *placebo*, attributable to the expectation of an effect, i.e. the power of suggestion. The effects of many healthcare interventions are attributable to a combination of both placebo and "active" (non-placebo) effects.

Point estimate
The results (e.g., mean, weighted difference, odds ratio, relative risk or risk difference) obtained in a sample (a study or a meta-analysis) which are used as the best estimate of what is true for the relevant population from which the sample is taken. A *confidence interval* is a measure of the uncertainty (due to the play of chance) associated with that estimate.

Positive study
A term used to refer to a study with results indicating a beneficial effect of the intervention being studied. The term can generate confusion because it can refer to both *statistical significance* and the direction of effect, studies often have multiple outcomes, the criteria for classifying studies as *negative* or positive are not always clear and, in the case of studies of risk or undesirable effects, "positive" studies are ones that show a harmful effect.

Precision
1. A measure of the likelihood of *random errors* in the results of a study, meta-analysis, or measurement. *Confidence intervals* around the estimate of effect from each study are a measure of precision, and the weight given to the results of each study in a *meta-analysis* (typically the inverse of the *variance* of the estimate of effect) is a measure of precision (i.e., the degree to which a study influences the overall estimate of effect in a meta-analysis is determined by the precision of its estimate of effect).
2. The proportion of relevant citations located using a specific search strategy, i.e., the number of relevant studies (meeting the inclusion criteria for a *trials register* or a review) divided by the total number of citations retrieved.

Primary study (synonyms: included study, original study)
"Original research" in which data are first collected; a study included in a *systematic review*. The term primary research is sometimes used to distinguish it from "secondary research" (reanalysis of previously collected data), *meta-analysis*, and other ways of combining studies (such as *economic analysis* and *decision analysis*).

Prevalence
The number of existing cases of a particular disease or condition in a given population at a designated time.

Prevalence study
See *cross-sectional study*.

Probability distribution
The function that gives the probabilities that a variable equals each of a sequence of possible values. Examples include the binomial, chi square, normal and Poisson distributions.

ProCite
A software package designed to manage bibliographic references. Needs to be used in conjunction with Biblio-links to download records from electronic databases such as *MEDLINE*. Examples of other similar packages are Papyrus and *Reference Manager*.

Proportional hazards model (synonym: Cox model)
A statistical model in survival analysis that asserts that the effect of the study factors (e.g., the intervention of interest) on the hazard rate (the risk of occurrence of an event, such as death, at a point in time) in the study population is *multiplicative* and does not change over time.

Prospective study
In evaluations of the effects of healthcare interventions, a study in which people are divided into groups that are exposed or not exposed to the intervention(s) of interest before the outcomes have occurred. *Randomised controlled trials* are always prospective studies and *case control studies* never are. Concurrent cohort studies are prospective studies, whereas historical cohort studies are not (see *cohort study*), although in *epidemiology* a prospective study is sometimes used as a synonym for cohort study. See *retrospective study*.

Protocol
The plan or set of steps to be followed in a study. A protocol for a *systematic review* should describe the rationale for the review; the objectives; and the methods that will be used to locate, select, and critically appraise studies, and to collect and analyse data from the included studies.

Publication bias
A bias in the published literature where the publication of research depends on the nature and direction of the study results. Studies in which an intervention is not found to be effective are sometimes not published. Because of this, *systematic reviews* that fail to include unpublished studies may overestimate the true effect of an intervention.

P-value
The probability (ranging from zero to one) that the observed results in a study, or results more

extreme, could have occurred by chance. In a *meta-analysis* the P-value for the overall effect assesses the overall *statistical significance* of the difference between the treatment and control groups, whilst the P-value for the *heterogeneity* statistic assesses the statistical significance of differences between the effects observed in each study.

Quality
See *methodological quality*.

Quality score
A value assigned to represent the *validity* of a study either for a specific criterion, such as *allocation concealment,* or overall. Quality scores can be use letters (A, B, C) or numbers. An advantage of using letters is that the order of best to worst may be more obvious than for numbers.

Quasi-random allocation
A method of allocating participants to different forms of care that is not truly *random*, for example, allocation by date of birth, day of the week, medical record number, month of the year, or the order in which participants are included in the study (e.g. alternation).

Quasi-randomised trial
A trial using a *quasi-random* method of allocating participants to different forms of care. There is a greater risk of *selection bias* in quasi-random trials where allocation is not adequately concealed compared with *randomised controlled trials* with adequate *allocation concealment*.

Random
Governed by chance. See *randomisation*.

Random allocation
A method that uses the play of chance to assign participants to comparison groups in a trial, e.g. by using a random numbers table or a computer-generated random sequence. Random allocation implies that each individual or unit being entered into a trial has the same chance of receiving each of the possible interventions. It also implies that the probability that an individual will receive a particular intervention is independent of the probability that any other individual will receive the same intervention. See also *concealment of allocation*, *quasi-random allocation*, randomisation.

Random effects model
A statistical model sometimes used in *meta-analysis* in which both within-study *sampling error* (*variance*) and between-studies variation are included in the assessment of the uncertainty (*confidence interval*) of the results of a meta-analysis. See *Fixed effect model*. If there is significant *heterogeneity* among the results of the included studies, random effects models will give wider confidence intervals than fixed effect models.

Random error (synonym: sampling error)
Error due to the play of chance. *Confidence intervals* and *P-values* represent the probability of random errors, but not systematic errors (*bias*).

Random permuted blocks
A method of *randomisation* that ensures that, at any point in a trial, roughly equal numbers of participants have been allocated to all the comparison groups. Permuted blocks are often used

in combination with *stratified randomisation*.

Random selection (synonym: random sampling)
A method of obtaining a representative, unbiased group of people from a larger population. Random selection which is not related to how participants are allocated to comparison groups is frequently used in *cross-sectional* and *cohort studies*, which are not *randomised controlled trials*, and it is frequently not used in randomised controlled trials. In older trial reports, however, the term is occasionally used instead of *random allocation* or *randomisation*.

Randomisation (spelled randomization in US English)
Method used to generate a *random allocation* sequence, such as using tables of random numbers or computer-generated random sequences. The method of randomisation should be distinguished from *concealment of allocation* because of the risk of *selection bias* despite the use of randomisation if there is not adequate *allocation concealment*. For instance, a list of random numbers may be used to randomise participants, but if the list is open to the individuals responsible for recruiting and allocating participants, those individuals can influence the allocation process, either knowingly or unknowingly.

Randomisation blinding
See *concealment of allocation*.

Randomised controlled trial (RCT) (Synonym: randomised clinical trial)
An experiment in which investigators randomly allocate eligible people into (e.g., treatment and control) groups to receive or not to receive one or more interventions that are being compared. The results are assessed by comparing outcomes in the treatment and control groups. NOTE: when using randomized controlled trial as a search term (publication type) in MEDLINE, the US spelling (randomized) must be used.

RCT
See *randomised controlled trial*.

Referee
See *referee process.*

Referee process
System by which a *review* goes out to editors and sometimes one or more external parties with content, methodological or user expertise. These people are sometimes called *external peer reviewers* or *referees*. See also *editorial process*.

Reference Manager
A software package designed to manage bibliographic references. Sometimes confusingly referred to as RefMan (see *RevMan*). Examples of other similar packages are Papyrus and *ProCite*.

Register of trials
See *trials register*.

Regression model
A mathematical representation of the relationship of a dependent variable (outcome) to a combination of explanatory variables (sometimes called predictor variables or covariates).

Relative Risk (RR) (synonym: risk ratio)
The ratio of risk in the intervention group to the risk in the control group. The risk (proportion, probability or rate) is the ratio of people with an event in a group to the total in the group. A relative risk of one indicates no difference between comparison groups. A relative risk of one indicates no difference between comparison groups. For undesirable outcomes, an RR that is less than one indicates that the intervention was effective in reducing the risk of that outcome.

Reliability
Refers to the degree to which results obtained by a measurement procedure can be replicated. Lack of reliability can arise from divergences between observers or measurement instruments, or instability in the attribute being measured.

Retrospective study
A study in which the outcomes have occurred to the participants before the study commenced. *Case control studies* are always retrospective, *cohort studies* sometimes are, *randomised controlled trials* never are. See *prospective study*.

Review
1. A *systematic review*.
2. A review article in the medical literature which summarises a number of different studies and may draw conclusions about a particular intervention. Review articles are often not systematic. Review articles are also sometimes called overviews.
3. To referee a paper. See *referee*, *referee process*, external peer reviewer.

Reviewer
Somebody responsible for preparing and, in the case of *Cochrane Reviews*, keeping up to date a *systematic review*. Reviewer is also sometimes used to refer to an *external peer reviewer*, or *referee*.

Review Manager (RevMan)
Software developed for the *Cochrane Collaboration* to assist *reviewers* in preparing *Cochrane Reviews*. Reviewers enter their *protocols* and reviews into RevMan, from which they can be imported into *ModMan* by a Collaborative Review Group *coordinator* for inclusion in the *Parent Database* and *CDSR* as part of the Group's edited *module*.

Review protocol
See *protocol*.

Risk difference (RD) (synonym: absolute risk reduction)
The absolute difference in the *event rate* between two comparison groups. A risk difference of zero indicates no difference between comparison groups. For undesirable outcomes, an RD that is less than zero indicates that the intervention was effective in reducing the risk of that outcome.

Risk factor
An aspect of a person's condition, lifestyle or environment that increases the probability of occurrence of a disease. For example, cigarette smoking is a risk factor for lung cancer.

Run-in period

A period before a trial is commenced when no treatment is given. The data from this stage of a trial are only occasionally of value but can serve a valuable role in screening out ineligible or non-compliant participants, in ensuring that participants are in a stable condition, and in taking baseline observations. A run-in period is sometimes called a *washout period* if treatments that participants were using before entering the trial are discontinued.

Sampling error
See *random error*.

Search strategy
1. The methods used by a *Collaborative Review Group* (CRG) to identify trials within the Group's scope. This includes *handsearching* relevant journals, searching electronic *databases*, contacting drug companies, other forms of personal contact and checking reference lists. CRGs must describe their search strategy in detail in the Group's *module*. *Reviewers* can refer to the Group's search strategy when preparing a *Cochrane Review*, and if necessary, supplement this with a description of their own additional searches.
2. The methods used by a reviewer to locate relevant studies, including the use of a CRG's *trials register*.
3. The combination of terms used to identify studies in an electronic *database* such as *MEDLINE*.

Selection bias
1. In assessments of the *validity* of studies of healthcare interventions, selection bias refers to systematic differences between comparison groups in prognosis or responsiveness to treatment. *Random allocation* with adequate *concealment of allocation* protects against selection bias. Other means of selecting who receives the intervention of interest, particularly leaving it up to the providers and recipients of care, are more prone to bias because decisions about care can be related to prognosis and responsiveness to treatment.
2. Selection bias is sometimes used to describe a systematic error in *reviews* due to how studies are selected for inclusion. *Publication bias* is an example of this type of selection bias.
3. Selection bias, confusingly, is also sometimes used to describe a systematic difference in characteristics between those who are selected for study and those who are not. This affects the generalisability (external validity) of a study but not its (internal) validity.

Sensitivity analysis
An analysis used to determine how sensitive the results of a study or *systematic review* are to changes in how it was done. Sensitivity analyses are used to assess how robust the results are to uncertain decisions or assumptions about the data and the methods that were used.

Sequential trial
A trial in which the data are analysed after each patient's results become available, and the trial continues until a clear benefit is seen in one of the comparison groups, or it is unlikely that any difference will emerge. The main advantage of sequential trials is that they will be shorter than fixed length trials when there is a large difference in the effectiveness of the interventions being compared. Their use is restricted to conditions where the outcome is known relatively quickly.

Single blind (synonym: single masked)
The investigator is aware of the treatment/intervention the participant is getting, but the participant is unaware. See also *blinding*, double blind, *triple blind*.

Specialised register
See *Register of trials*.

Standardised mean difference
The difference between two means divided by an estimate of the within-group standard deviation. When an outcome (such as pain) is measured in a variety of ways across studies (using different scales) it may not be possible directly to compare or combine study results in a *systematic review*. By expressing the effects as a standardised value, the results can be combined since they have no units. Standardised mean differences are sometimes referred to as a *d* index.

Statistical power
The probability that the *null hypothesis* will be rejected if it is indeed false. In studies of the effectiveness of healthcare interventions, power is a measure of the certainty of avoiding a false negative conclusion that an intervention is not effective when in truth it is effective. The power of a study is determined by how large it is (the number of participants), the number of events (e.g. strokes) or the degree of variation in a continuous outcome (such as weight), how small an effect one believes is important (i.e. the smallest difference in outcomes between the intervention and the control groups that is considered to be important), and how certain one wants to be of avoiding a false positive conclusion (i.e. the cut-off that is used for *statistical significance*).

Statistical significance
An estimate of the probability of an association (effect) as large or larger than what is observed in a study occurring by chance, usually expressed as a *P-value*. For example, a P-value of 0.049 for a *risk difference* of 10% means that there is less than a one in 20 (0.05) chance of an association that is as large or larger having occurred by chance and it could be said that the results are "statistically significant" at P = 0.05). The cut-off for statistical significance is usually taken at 0.05, but sometimes at 0.01 or 0.10. These cut-offs are arbitrary and have no specific importance. Although it is often done, it is inappropriate to interpret the results of a study differently according to whether the P-value is, say, 0.055 or 0.045 (which are quite similar values, not diametrically opposed ones).

Stratified randomisation
In any randomised trial it is desirable that the comparison groups should be as similar as possible as regards participant characteristics that might influence the response to the intervention. Stratified randomisation is used to ensure that equal numbers of participants with a characteristic thought to affect prognosis or response to the intervention will be allocated to each comparison group. For example, in a trial of women with breast cancer, it may be important to have similar numbers of pre-menopausal and post-menopausal women in each comparison group. Stratified randomisation could be used to allocate equal numbers of pre- and post-menopausal women to each treatment group. Stratified randomisation is performed either by performing separate *randomisation* (often using *random permuted blocks*) for each stratum, or by using *minimisation*.

Study validity
See *validity*.

Surrogate endpoints (synonym: intermediary outcomes; surrogate outcomes)
Outcome measures that are not of direct practical importance but are believed to reflect

outcomes that are important; for example, blood pressure is not directly important to patients, but it is often used as an outcome in clinical trials because it is a risk factor for stroke and heart attacks. Surrogate endpoints are often physiological or biochemical markers that can be relatively quickly and easily measured, and that are taken as being predictive of important clinical outcomes. They are often used when observation of clinical outcomes requires long follow-up.

Systematic error
See *bias*.

Systematic review (synonym: systematic overview)
A *review* of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review. Statistical methods *(meta-analysis)* may or may not be used to analyse and summarise the results of the included studies. See also *Cochrane Review*.

***t*-distribution, *t*-test (synonym: Student *t*-test)**
The *t*-distribution is the distribution of a quotient of independent random variables, the numerator of which is a standardised normal random variable and the denominator of which is the positive square root of the quotient of a *chi-square* distributed random variable and its number of degrees of freedom. The *t*-test uses the *t*-distribution to test whether two means differ significantly or to test linear regression or correlation coefficients.

Test of association
See *statistical significance*.

Therapeutic trial
See *clinical trial*.

Trials register
In the *Cochrane Collaboration*, this is a *database* of bibliographic references to *randomised controlled trials* and *controlled clinical trials* relevant to a *Collaborative Review Group* or *Field*, that is maintained at the *editorial base*. Software such as *ProCite* or *Reference Manager* is used to manage the database. Once a relevant report of a trial is identified, it is photocopied, coded and entered onto the register. Wherever possible, relevant trial reports are downloaded directly into the register from an electronic database such as *MEDLINE*. Information about unpublished and ongoing trials is also included in trials registers.

Triple blind (synonym: triple masked)
An expression that is sometimes used to indicate that knowledge of which study participants are in which comparison group is kept secret from the statistician doing the analysis as well as from the study participants and investigators (outcome assessors). See also *blinding,* single blind, double blind.

Unit of allocation
The entity that is assigned to different comparison groups in a trial. Most commonly, individuals are allocated, but in some trials, people are assigned to the intervention and *control* groups in groups to avoid *contamination* or for convenience; for example, practices, hospitals or communities can be allocated. See *unit of analysis error*.

Unit of analysis error
In some studies people are allocated in groups instead of individually (e.g., by practice, by hospital or by community). Often when this is done the *unit of allocation* is different from the unit of analysis, i.e., people are allocated by groups and analysed as though they had been allocated individually. This is sometimes called a unit of analysis error. 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 overly narrow *confidence intervals* and false positive conclusions that the intervention had an effect when in truth there is greater uncertainty than what is reflected by the *P-value*. In the context of a review, it can result in studies having narrower confidence intervals and receiving more weight than is appropriate.

Utility
In economic and decision analysis, the desirability of an outcome, usually expressed as being between zero and one (e.g., death typically has a utility value of zero and a full healthy life has a value of one).

Validity (synonym: internal validity)
Validity is the degree to which a result (of a measurement or study) is likely to be true and free of *bias* (systematic errors). Validity has several other meanings, usually accompanied by a qualifying word or phrase; for example, in the context of measurement, expressions such as "construct validity", "content validity" and "criterion validity" are used. The expression "internal validity" is sometimes used to distinguish validity (the extent to which the observed effects are true for the people in a study) from *external validity* or *generalisability* (the extent to which the effects observed in a study truly reflect what can be expected in a target population beyond the people included in the study). See also *methodological quality*, *random error*.

Variable
Any quantity that varies. A factor that can have different values.

Variance
A measure of the variation shown by a set of observations, defined by the sum of the squares of deviations from the mean, divided by the number of *degrees of freedom* in the set of observations.

Venn diagram
A pictorial presentation of the extent to which two or more quantities or concepts are mutually inclusive and mutually exclusive.

Weighted least squares regression (in meta-analysis)
A *meta-regression* technique for estimating the parameters of a multiple regression model, wherein each study's contribution to the sum of products of the measured variables (study characteristics) is weighted by the precision of that study's estimate of effect.

Washout period
The stage in a *cross-over trial* when treatment is withdrawn before the second treatment is given. Washout periods are usually necessary because of the possibility that the intervention administered first can affect the outcome variable for some time after treatment ceases. A *run-in period* before a trial starts is sometimes called a washout period if treatments that participants were using before entering the trial are discontinued.

Weighted mean difference (in meta-analysis)
A method of *meta-analysis* used to combine measures on continuous scales (such as weight), where the mean, standard deviation and sample size in each group are known. The weight given to each study (e.g., how much influence each study has on the overall results of the meta-analysis) is determined by the *precision* of its estimate of effect and, in the statistical software in *RevMan* and *CDSR*, is equal to the inverse of the *variance*. This method assumes that all of the trials have measured the outcome on the same scale. See also *standardised mean difference*.

World Wide Web (WWW)
A part of the *Internet* with a graphical interface. "Web pages" or "home pages" are HyperText Markup Language (HTML) documents on the WWW. Hypertext allows users to jump from one place in a document to another, from one document to another, and from one computer on the WWW to another. A connection through a cable or over the telephone and a Web browser (software program), such as Netscape, are needed to access and view WWW documents. The *Cochrane Collaboration* has several home pages, including: the main Collaboration home page at the Canadian Cochrane Centre (http://hiru.mcmaster.ca/cochrane/) and the Australasian Cochrane Centre home page (http://som.flinders.edu.au/cochrane/). Links to other Cochrane Collaboration and related sites can be found through those two locations.

Glossary references

339. Cooper H, Hedges LV. The Handbook of Research Synthesis. New York: Russell Sage Foundation, 1993.

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

447. Jepson R, Hetherington J. Glossary of terms in the Cochrane Collaboration. Available in: The Cochrane Library [database on disk and CDROM]. Oxford: Update Software, 1996, Issue 3. Updated quarterly. Available from: BMJ Publishing Group, London.

446. Last JM (ed). A Dictionary of Epidemiology. 2nd ed. New York: Oxford University Press, 1988.

503. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1991.

504. Goodman C. Literature searching and evidence interpretation for assessing health care practices. Stockholm: SBU, 1993.

Acknowledgements
This glossary was originally compiled by Ruth Jepson (Menstrual Disorders Group) and Jini Hetherington (Collaboration Secretariat) with help from Hilda Bastian (Consumer Network), Liz Chalker (Acute Respiratory Infections Group), Iain Chalmers (UK Cochrane Centre), Jon Deeks (Centre for Statistics in Medicine, Oxford, UK), Cindy Farquhar (Menstrual Disorders Group), Monica Fischer (Nordic Cochrane Centre), Gerry Fowkes (Peripheral Vascular Diseases Group), Helen Handoll (Musculoskeletal Injuries Sub-Group), Emma Harvey (Cochrane

Collaboration on Effective Professional Practice), Carol Lefebvre (UK Cochrane Centre), Reive Robb (Infectious Diseases Group), Rochelle Seifas (Schizophrenia Group), Bev Shea (Musculoskeletal Group), and Mark Starr (Update Software); and initial secretarial support from Hazel Dunn (Peripheral Vascular Diseases Group).

Please send suggestions for improving the glossary to:

Claire Glenton
Health Services Research Unit
National Institute of Public Health
P.O. Box 4404 Torshov
0403 Oslo
NORWAY

Fax: +47 22 04 25 95
E-mail: claire.glenton@labmed.uio.no