Doctorate in Health Psychology

Systematic Reviews of Health Behaviour Interventions: Workshop Manual

Dr Chris Bridle, CPsychol.

Associate Professor of Health Psychology
Institute of Clinical Education
Warwick Medical School
University of Warwick

Available at:

www.warwick.ac.uk/go/chrisbridle

THIS IS A DRAFT – JULY 2006

Acknowledgement

The information in this manual is based largely on the guidance issued by the Centre for Reviews and Dissemination at the University of York, and contains information taken from materials and resources issued by a number of other review groups, most notably the Cochrane Collaboration.

Contents

Introduction

Unit 1: Background Information 5

Unit 2: Resources Required 11

Unit 3: Developing a Protocol 15

Unit 4: Formulating a Review Question 19

Unit 5: Searching for Evidence 24

Unit 6: Selecting Studies for Inclusion 36

Unit 7: Data Extraction 38

Unit 8: Critical Appraisal 41

Unit 9: Synthesising the Evidence 46

Unit 10: Interpreting the Findings 57

Unit 11: Writing the Systematic Review 61

Appendices

A: Glossary of systematic review terminology 63

B: Design algorithm for health interventions 66

C: RCT quality criteria and explanation 67

Further information:

Dr Chris Bridle, CPsychol
Associate Professor of Health Psychology
Institute of Clinical Education
Warwick Medical School
University of Warwick

Tel: +44 (024) 76150222
Fax: +44 (024) 76523701
Email: C.Bridle@warwick.ac.uk

Introduction

This training handbook will take you through the process of conducting systematic reviews of health behaviour interventions. The purpose of this handbook is to describe the key stages of the systematic review process and to provide some working examples and exercises for you to practice before you start your systematic review.

The handbook is not intended to be used as a single resource for conducting reviews, and you are strongly advised to consult more detailed methodological guidelines, some useful examples of which are highlighted below.

Overall learning outcomes

Working through this handbook will enable you to:

- Identify the key stages involved in conducting a systematic review
- Recognise some of the key challenges of conducting systematic reviews of health behaviour interventions
- Develop a detailed protocol for conducting a systematic review
- Formulate an answerable question about the effects of health behaviour interventions
- Develop a comprehensive search strategy in order to locate relevant evidence
- Evaluate the methodological quality of health behaviour interventions
- Synthesise evidence from primary studies
- Formulate evidence-based conclusions and recommendations
- Report and disseminate the results of a systematic review
- Evaluate the methodological quality of a systematic review
- Feel smug and superior when pontificating in front of your ill-informed colleagues

Additional reading

There are many textbooks and online manuals that describe systematic review methodology. Although these sources may differ in terms of focus (e.g. medicine, public health, social science, etc.), there is little difference in terms of content and you should select a textbook or online manual that best meets your needs. Some examples are listed below:

Textbooks

Brownson, R., Baker, E., Leet, T. & Gillespie, K. (2003). *Evidence-based Public Health*. Oxford University Press: Oxford.

Egger, M., Smith, G. & Altman, D. (2001). *Systematic Reviews in Health Care: Meta-analysis in context (2nd Ed.)*. BMJ Books: London.

Khan, K.S., Kunz, R., Kleijnen, J. & Antes, G. (2003). *Systematic Reviews to Support Evidence-Based Medicine: How to apply findings of healthcare research*. Royal Society of Medical Press: London.

Petticrew, M. & Roberts, H. (2005). *Systematic Reviews in the Social Sciences*. Blackwell Publishing: Oxford.

OnLine Manuals / Handbooks

Cochrane Collaboration Open-Learning Materials for Reviewers Version 1.1, November 2002. http://www.cochrane-net.org/openlearning/

Cochrane Reviewers' Handbook 4.2.5. http://www.cochrane.org/resources/handbook/index.htm

Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD Report Number 4 (2nd Edition). NHS Centre for Reviews and Dissemination, University of York. 2001. http://www.york.ac.uk/inst/crd/report4.htm

Evidence for Policy and Practice Information and Co-ordinating Centre Review Group Manual. Version 1.1, Social Science Research Unit, Institute of Education, University of London. 2001. http://eppi.ioe.ac.uk/EPPIWebContent/downloads/RG_manual_version_1_1.pdf

Handbook for compilation of reviews on interventions in the field of public health (Part 2). National Institute of Public Health. 2004. http://www.fhi.se/shop/material_pdf/r200410Knowledgebased2.pdf

Unit 1: Background Information

Learning Objectives

- To understand why research synthesis is necessary
- To understand the terms 'systematic review' and 'meta-analysis'
- To be familiar with different types of reviews (advantages / disadvantages)
- To understand the complexities of reviews of health behaviour interventions
- To be familiar with international groups conducting systematic reviews of the effectiveness of health behaviour interventions

Why reviews are needed

- Health care decisions, whether about policy or practice, should be based upon the best available evidence
- The vast quantity of research makes it difficult / impossible to make evidence-based decisions concerning policy, practice and research
- Single trials rarely provide clear or definitive answers, and it is only when a body of evidence is examined as a whole that a clearer, more reliable answer emerges

Two types of review

Traditional narrative review: The authors of these reviews, who may be 'experts' in the field, use informal, unsystematic and subjective methods to collect and interpret information, which is often summarised subjectively and narratively:

- Processes such as searching, quality assessment and data synthesis are not usually described and are therefore very prone to bias
- Authors of these reviews may have preconceived notions or biases and may overestimate the value of some studies, particularly their own research and research that is consistent with their existing beliefs
- A narrative review is not to be confused with a narrative systematic review – the latter refers to the type of synthesis within a systematic review

Systematic review: A systematic review is defined as a review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyse data from the studies that are included in the review:

- Because systematic reviews use explicit methods they are less prone to bias and, like other types of research, can be replicated and critically appraised
- Well-conducted systematic reviews 'top' the hierarchy of evidence, and thus provide the most reliable basis for health care decision making

Table 1.1: Comparison of traditional and systematic reviews

Components of a review	Traditional, narrative reviews	Systematic reviews
Formulation of the question	Usually address broad questions	Usually address focused questions
Methods section	Usually not present, or not well-described	Clearly described with pre-stated criteria about participants, interventions and outcomes
Search strategy to identify studies	Usually not described; mostly limited by reviewers' abilities to retrieve relevant studies; prone to selective citation	Clearly described, comprehensive and less prone to selective publication biases
Quality assessment of identified studies	Studies included without explicit quality assessment	Studies assessed using pre-stated criteria; effects of quality on results are tested
Data extraction	Methods usually not described	Undertaken pre-planned data extraction forms; attempts often made to obtain missing data from authors of primary studies
Data synthesis	Qualitative description employing the vote counting approach, where each included study is given equal weight, irrespective of study size and quality	Greater weights given to effect measures from more precise studies; pooled, weighted effect measures with confidence limits provide power and precision to results
Heterogeneity	Usually dealt with in a narrative fashion	Heterogeneity dealt with by narratively, graphically and / or statistically; attempts made to identify sources of heterogeneity
Interpreting results	Prone to cumulative systematic biases and personal opinion	Less prone to systematic biases and personal opinion; reflects the evidence presented in review

What is meta-analysis?

Meta-analysis is the statistical combination of data from at least 2 studies in order to produce a single estimate of effect

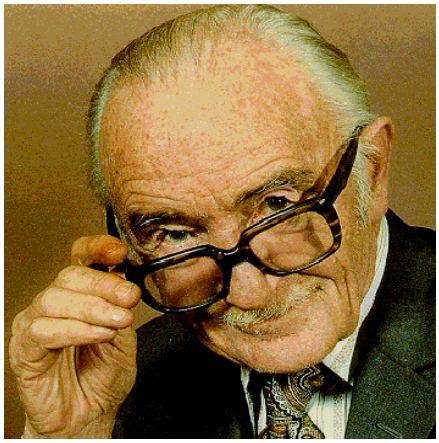
- Meta-analysis is NOT a type of review - meta-analysis IS a statistical procedure – that's all!
- A meta-analysis does not have to be conducted in the context of a systematic review, and a systematic review does not have to conduct a meta-analysis
- It is always desirable to systematically review a research literature but it may not be desirable, and may even be harmful, to combine statistically research data

Systematic reviews and evidence-based medicine

"It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials" (Archie Cochrane, 1979).

The Cochrane Collaboration is named in honour of the British epidemiologist Archie Cochrane. The Collaboration is an international non-profit organisation that prepares, maintains, and disseminates systematic up-to-date reviews of health care interventions.

Systematic reviews are the foundation upon which evidence-based practice, policy and decision making are built.



Archie Cochrane (1909-1988)

Who benefits from systematic review

Anyone who comes into contact with the healthcare system will benefit from systematic reviews

- Practitioners, who are provided with an up-to-date summary of the best available evidence to assist with decision making
- Policy makers, who are provided with an up-to-date summary of best available evidence to assist with policy formulation
- Public, who become recipients of evidence-based interventions
- Researchers, who are able to make a meaningful contribution to the evidence base by directing research to those areas where research gaps and weaknesses have been identified by systematic review
- Funders, who are able to identify research priorities and demonstrate the appropriate allocation of resources

Clinical vs. behavioural interventions

Systematic reviews have been central to evidence-based-medicine for more than two decades. Although review methodology was developed in the context of clinical (e.g. pharmacological) interventions, recently there has been an increasing use of systematic reviews to evaluate the effects of health behaviour interventions. Systematic reviews of health behaviour interventions present a number of methodological challenges, most of which derive from a focus or emphasis on:

- Individuals, communities and populations
- Multi-faceted interventions rather than single component interventions

- Integrity of intervention implementation – completeness and consistency
- Processes as well as outcomes
- Involvement of 'users' in intervention design and evaluation
- Competing theories about the relationship between health behaviour and health beliefs
- Use of qualitative as well as quantitative approaches to research and evaluation
- The complexity and long-term nature of health behaviour intervention outcomes

International review groups

The increasing demand for rigorous evaluations of health interventions has resulted in an international expansion of research groups / institutes who conduct systematic reviews. These groups often publish completed reviews, methodological guidelines and other review resources on their webpages, which can usually be freely downloaded. Some of the key groups conducting reviews in areas related to health behaviour include:

- Agency for Healthcare Research and Quality:
  http://www.ahrq.gov/
- Campbell Collaboration:
  http://www.campbellcollaboration.org/
- Centre for Outcomes Research and Effectiveness:
  http://www.psychol.ucl.ac.uk/CORE/
- Centre for Reviews and Dissemination:
  http://www.york.ac.uk/inst/crd/
- Cochrane Behavioral Medicine Field:
  http://www.cochranebehavmed.org
- Cochrane Collaboration – The Cochrane Library:
  http://www.thecochranelibrary.com
- Effective Public Health Practice Project:
  http://www.city.hamilton.on.ca/PHCS/EPHPP/EPHPPResearch.asp
- Guide to Community Preventive Services:
  http://www.thecommunityguide.org
- MRC Social and Public Health Sciences Unit:
  http://www.msoc-mrc.gla.ac.uk/
- National Institute for Health and Clinical Excellence:
  http://www.publichealth.nice.org.uk/page.aspx?o=home
- The Evidence for Practice Information and Co-ordinating Centre:
  http://eppi.ioe.ac.uk/

ONE TO READ

Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. Eval Health Prof 2002;25:12-37.

ONE TO REMEMBER

The major benefit of systematic review is that it offers the opportunity to limit the influence of bias, but only if conducted appropriately.

EXERCISE

1. In pairs, use the examples below to discuss some of the differences between reviews of clinical interventions *vs.* reviews of health behaviour interventions.

   Examples: a) Clinical, e.g. *effectiveness of antibiotics for sore throat*

   b) Health Behaviour, e.g. *effectiveness of interventions for smoking cessation*

Clinical	**Behavioural**
Study participants:	
……	……
Types of interventions:	
……	……
Types of outcomes (process, proxy outcomes, intermediate and / or long-term):	
……	……
Participants involved in design of intervention:	
……	……
Potential influences on intervention success / failure: external factors (e.g. social, political, cultural, etc.) and internal factors (e.g. training of those implementing intervention, literacy of population, access to services, etc.)	
……	……

Unit 2: Resources Required

Learning Objective

- To be familiar with the resources required to conduct a systematic review
- To know how to access key review resources

Types of resources

As Figure 1.1 suggests, conducting a systematic review is a demanding, resource-heavy endeavour. The following list outlines the main resources required to complete a systematic review:

- *Technological resources*: Access to electronic databases, the internet, and statistical, bibliographic and word processing software
- *Contextual resources*: A team of co-reviewers (to reduce bias), access to / understanding of the likely users of the review, funding and time
- *Personal resources*: Methodological skills / training, a topic in which you are interest, and bundles of patience, commitment and resilience

The Cochrane Collaboration software, Review Manager (RevMan), can be used for both the writing of the review and, if appropriate, the meta-analysis. The software, along with the user manual, can be downloaded for free: http://www.ccims.net/RevMan.

Unfortunately RevMan does not have a bibliographic capability, i.e. you can not download / save results from your internet / database literature searches. The bibliographic software to which the University subscribes is RefWorks: http://www.uwe.ac.uk/library/info/research/

Time considerations

The time it takes to complete a review will vary depending on many factors, including the review's topic and scope, and the skills and experience of the review team. However, an analysis of 37 medically-related systematic reviews demonstrated that the average time to completion was 1139 hours (approximately 6 months), but this ranged from 216 to 2518 hours (Allen & Olkin, 1999). The component mean times were:

342 hours	Protocol development
246 hours	Searching, study retrieval, data extraction, quality assessment, data entry
144 hours	Synthesis and statistical analysis
206 hours	Report and manuscript writing
201 hours	Other (administrative)

Not surprisingly, there was an observed association between the number of initial citations (before inclusion / exclusion criteria are applied) and the total time taken to complete the review. The time it takes to complete a health behaviour review, therefore, may be longer due to use of less standardised terminology in the psychology literature, resulting in a larger number of citations to be screened for inclusion / exclusion.

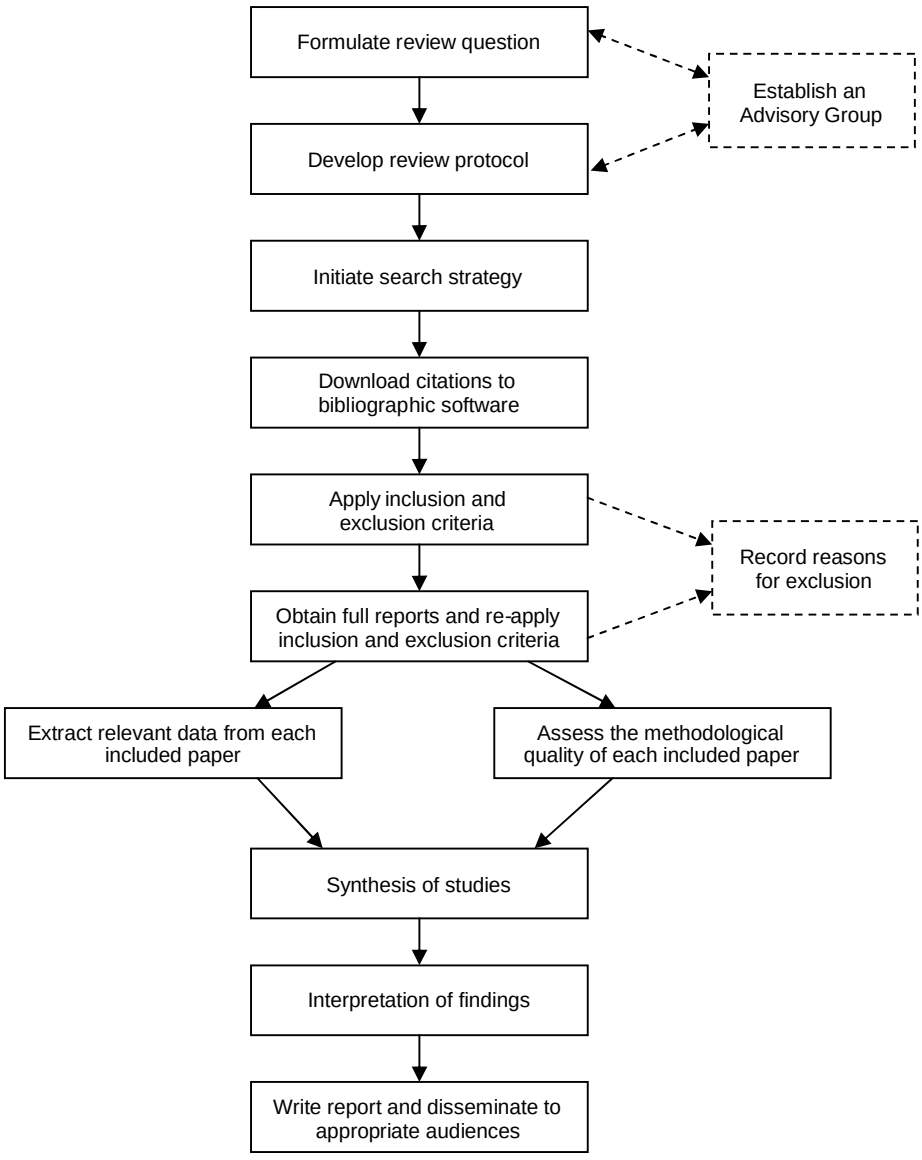
Example: Typical systematic review timeframe

Review Stage	Task	Project Days	Month
Protocol development	Specification of review objective, questions and methods in consultation with advisory group	20	1 - 2
Literature searches (electronic)	Develop search strategy, conduct searches, record search results - bibliographic database	15	2 – 3
Inclusion assessment 1	Search results screened for potentially relevant studies	5	3 – 4
Retrieval of primary studies	Download electronic copies, order library copies / inter-library loans, distribute papers to reviewers	15	3 – 5
Inclusion assessment 2	Full-text papers screened for inclusion – reasons for exclusion recorded	10	3 – 5
Validity assessment and data extraction	Independent validity assessment and data extraction checked for accuracy	15	4 – 6
Synthesis and interpretation	Tabulate data, synthesise evidence, investigate potential sources of heterogeneity	10	6 – 7
Draft report	Write draft report and submit to review team for comment	10	7 – 8
Submission and dissemination	Final draft for submission and dissemination	5	8 – 9
		105	**9**

In the above example the 'project days' are the minimum required to complete each stage. In most cases, therefore, completing a systematic review will take at least 105 project days spread across 9 months.

Targets for achieving particular review stages will vary from review to review. Trainees, together with their supervisors and other relevant members of the Health Psychology Research Group, must determine an appropriate time frame for the review at the earliest opportunity.

Fig 1: Flow chart of a systematic review



ONE TO READ

Allen IE, Olkin I. Estimating Time to Conduct a Meta-analysis From Number of Citations Retrieved. JAMA 1999;282(7):634-5.

ONE TO REMEBER

Good methodological guidance is one of the many resources needed to complete a systematic review, and whilst many guidelines are freely available online, perhaps the most useful are CRD's Report 4 and the Cochrane Reviewers' Handbook.

EXERCISE

1. In your own time, locate and download one complete set of guidelines and file with the workshop material.

2. In your own time, list the resources you are likely to need in order to complete your systematic review, and determine their availability to you.

Unit 3: Developing a Protocol

Learning Objectives

- To understand the rationale for developing a review protocol
- To recognise the importance of adhering to the review protocol
- To know what information should be reported in the review protocol
- To be familiar with the structure of the review protocol

Protocol: What and why?

A protocol is a written document containing the background information, the problem specification and the plan that reviewers follow in order to complete the systematic review.

- The first milestone of any review is the development and approval of the protocol before proceeding with the review itself.
- A systematic review is less likely to be biased if the review questions are well-formulated and the methods used to answer them are specified *a priori*.
- In the absence of a protocol, or failing to adhere to a protocol, it is very likely that the review questions, study selection, data analysis and reporting of outcomes will be unduly driven by (a presumption of) the findings.
- A clear and comprehensive protocol reduces the potential for bias, and saves time during both the conduct and reporting of the review, e.g. the introduction and methods sections are already written.

Protocol structure and content

The protocol needs to be comprehensive in scope, and provide details about the rationale, objectives and methods of the review. Most protocols report information that is structured around the following sections:

- *Background*: This section should address the importance of conducting the systematic review. This may include discussion of the importance or prevalence of the problem in the population, current practice, and an overview of the current evidence, including related systematic reviews, and highlighting gaps and weaknesses in the evidence base. The background should also describe why, theoretically, the interventions under review might have an impact on potential recipients.
- *Objectives*: You will need to determine the scope of your review, i.e. the precise question to be asked. The scope of the review should be based on how the results of the review will be used, and it is helpful to consult potential users of the review and / or an advisory group when determining the review's scope. In all cases, the question should be clearly

formulated around key components, e.g. Participants, Interventions, Comparison and Outcomes.

- *Search strategy*: Report the databases that are to be searched, search dates and search terms (e.g. subject headings and text words), and provide an example search strategy. Methods to identify unpublished literature should also be described, e.g. hand searching, contact with authors, scanning reference lists, internet searching, etc.

- *Inclusion criteria*: Components of the review question (e.g. Participants, Interventions, Comparisons and Outcomes) are the main criteria against which studies are assessed for inclusion in the review. All inclusion / exclusion criteria should be reported, including any other criteria that were used, e.g. study design. The process of study selection should be described, e.g. the number of reviewers involved, whether the process will be independent, and how disagreements will be resolved.

- *Data extraction*: Describe what data will be extracted from primary / included studies. It is often helpful to structure data extraction in terms of study details, participant characteristics, intervention details, results and conclusions. The data extraction process should be described, e.g. the number of reviewers involved, whether the process will be independent, and how disagreements will be resolved.

- *Critical appraisal / quality assessment*: The criteria / checklist to be used for appraising the methodological quality of included studies should be specified, as should the way in which the assessment will be used. The process of conducting quality assessment should be described, e.g. the number of reviewers involved, whether the process will be independent, and how disagreements will be resolved.

- *Method of synthesis*: Describe the methods to be used to present and synthesise the data. Reviews of health behaviour interventions often tabulate the included studies and perform a narrative synthesis due to expected heterogeneity. The protocol should identify *a priori* potential sources of effect heterogeneity and specify the strategy for their investigation.

Additional considerations

In addition to detailing the review's rationale, questions / objectives and methods, the protocol should ideally describe the strategy for disseminating the review findings, a timetable for completing review milestones, responsibilities of review team members, and role of the external advisory group.

- *Dissemination strategy*: Failing to disseminate research findings is unethical. The protocol should specify the relevant audiences to who the review results are to be disseminated, which may include academics, researchers, policy makers, practitioners and / or patients. The protocol should also describe the dissemination media to be used, e.g. journal publication, conference presentation, information sheet, online document, etc. The strategy should be precise, i.e. name the appropriate journal(s), conference(s), etc.

- *Timetable*: Identify review milestones and specify a timetable for their completion. Key milestones include: (1) protocol development and approval, (2) retrieval of study papers, (3) data extraction and quality assessment, (4) synthesis and analysis, (5) writing the draft review report, (5) submission of the final review report (i.e. your assessment requirement), and (6) a period for disseminating the review.

- *Review Team*: Your review team will consist of you as first reviewer, another trainee to act as second reviewer, and a staff member of the Health Psychology Research Group who will supervise the review. It is your responsibility to negotiate and clarify roles and responsibilities within the review team.

- *Advisory Group*: Systematic reviews are more likely to be relevant and of higher quality if they are informed by advice from people with a range of experiences and expertise. The Advisory Group should include potential users of the review (e.g. patients and providers), and those with methodological and subject area expertise. The size of the Advisory Group should be limited to no more than six, otherwise the group will become difficult to manage. Advisory Groups will be more effective / helpful if they are clear about the task(s) to which they should and shouldn't contribute, which may include:

    - Providing feedback (i.e. peer-review) on draft versions of the protocol and review report
    - Helping to make and / or refine aspects of the review question, e.g. PICO
    - Helping to identify potential sources of effect heterogeneity and sub-group analyses
    - Providing or suggesting important background material that elucidates the issues from different perspectives
    - Helping to interpret the findings of the review
    - Designing a dissemination plan and assisting with dissemination to relevant groups

ONE TO READ

Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: Comparing what was done to what was planned. JAMA 2002;287(21):2831-2834.

ONE TO REMEMBER

Do not start your systematic review without a fully-developed and approved protocol.

EXERCISE

1. Choose one of the review topics from the list below. Brainstorm, in groups, who you might want to include in an Advisory Group. After brainstorming all potential members, reduce the list to a maximum of 6 members.

   - Interventions for preventing tobacco sales to minors
   - Workplace interventions for smoking cessation
   - Primary prevention for alcohol misuse in young people
   - Interventions to improve immunisation rates

2. In your own time, search the Cochrane Library for protocols related to your area of interest and familiarise yourself with the structure and content.

Unit 4: Formulating a Question

Learning Objectives

- To understand the importance of formulating an answerable question
- To be able to identify and describe the key components of an answerable question
- To be able to formulate an answerable question

Importance of getting the question right

A well-formulated question will guide not only the reader in their initial assessment of the relevance of the review, but also the reviewer

- on how to develop a strategy for searching the literature
- the criteria by which studies will be included in the review
- the relevance of different types of evidence
- the analysis to be conducted

Post-hoc questions are more susceptible to bias than questions determined *a priori*, and it is thus important that questions are appropriately formulated before beginning the review.

Components of an answerable question (PICO)

An answerable, or well-formulated, question is one in which key components are adequately specified. Key components can be identified using the PICO acronym: Participants (or Problem), Intervention, Comparison, and Outcome. It is also worthwhile at this stage to consider the type of evidence most relevant to the review question, i.e. PICO-T.

- *Participants*: Who are the participants of interest? Participants can be identified by various characteristics, including demography (e.g. gender, ethnicity, S-E-S, etc.), condition (e.g. obesity, diabetes, asthma, etc.), behaviour (e.g. smoking, unsafe sex, physical activity, etc.) or, if meaningful, a combination of characteristics, e.g. female smokers.

- *Intervention*: What is the intervention to be evaluated? The choice of intervention can be topic-driven (e.g. [any] interventions for smoking cessation), approach-driven (e.g. peer-led interventions), theory-driven (e.g. stage-based interventions) or, if meaningful, a combination of characteristics, e.g. stage-based interventions for smoking cessation.

- *Comparison*: What comparator will be the basis for evaluation? Comparators may be no intervention, usual care or an alternative intervention. In practice, few review questions refer

explicitly to a named comparator, in which case the protocol should describe potential comparators and the strategy for investigating heterogeneity as a function of comparator.

- *Outcome*: What is the primary outcome of interest? The outcome that will be used as the primary basis for interpreting intervention effectiveness should be clearly identified and justified, usually in terms of its relationship to health status. For example, smoking cessation interventions often report cessation and motivation as outcome variables, and it is more meaningful to regard cessation as the primary outcome and motivation as a secondary outcome.

Using the PICO components

Well-formulated questions are a necessary pre-condition for clear meaningful answers. Not all questions components need to be explicitly specified, but using the PICO framework will help to formulate an answerable review question, as illustrated below.

Table 4.1: Question formulation using PICO components

Poorly formulated / Unfocussed	Well-formulated / Focussed
Effects of drugs on mental illness	Effects of cannabis on psychosis
Effectiveness of training for UWE staff	Effects of systematic review training on number of review publications among the Health Psychology Research Group
Effectiveness of smoking cessation interventions	Effects of stage-based smoking cessation interventions
Effectiveness of smoking cessation interventions	Effects of stage-based smoking cessation interventions in primary care for adolescents
Effectiveness of smoking cessation interventions	Effects of peer-led stage-based smoking cessation interventions in primary care for adolescents

Type of Evidence

A well-formulated question serves as a basis for identifying the relevant type of evidence required for a meaningful answer. This is because different types of evidence (i.e. design or methodology) are more or less relevant (i.e. valid or reliable) depending on the question being asked.

In health-related research, the key questions and the study designs offering the most relevant / reliable evidence are summarised below:

Type of Question		**Relevant (best) Evidence**
Intervention	-	Randomised controlled trial
Prognosis	-	Cohort
Aetiology	-	Cohort, case-control
Harm	-	Cohort, case-control
Diagnosis	-	Cross-sectional, case-control
Experience	-	Qualitative

- Because there is little standardisation of 'study design' terminology in the literature, an algorithm for identifying study designs of health interventions is presented in Appendix B.

Additional considerations

The PICO-T components provide a useful framework for formulating answerable review questions. However, there are additional issues that merit further consideration when conducting systematic reviews of health behaviour interventions, two key issues include:

- the use of qualitative research
- the role of heath inequalities.

Careful consideration of these issues may help in refining review questions, selecting methods of analysis (e.g. identifying heterogeneity and sub-groups), and interpreting review results.

Qualitative research

Several research endeavours, most notably the Cochrane Qualitative Research Methods Group (http://mysite.freeserve.com/Cochrane_Qual_Method/index.htm), are beginning to clarify the role / use and integration of qualitative research in systematic reviews. In particular, qualitative studies can contribute to reviews of effectiveness in the following ways:

- Helping to frame review questions, e.g. identifying relevant interventions and outcomes
- Identifying factors that enable / impede the implementation of the intervention
- Describing the experience of the participants receiving the intervention
- Providing participants' subjective evaluations of outcomes
- Providing a means of exploring the 'fit' between subjective needs and evaluated interventions to inform the development of new interventions or refinement of existing ones

Health inequalities

Health inequalities refer to the gap in health status and in access to health services, which exists between different social classes, ethnic groups, and populations in different geographical areas. Where possible, systematic reviews should consider health inequalities when evaluating intervention effects. This is because the beneficial effects of many interventions may be substantially lower for some population sub-groups. Many interventions may thus increase rather than reduce heath inequalities, since they primarily benefit those who are already advantaged.

Evans and Brown (2003) suggest that there are a number of factors that may be used in classifying health inequalities (captured by the acronym PROGRESS)

It may be useful for a review to evaluate intervention effects across different sub-groups, perhaps identified in terms of the PROGRESS factors.

Kristjansson et al (2004) provide a good example of a systematic review addressing health inequalities among disadvantaged (low S-E-S) school children.

- **P**lace of residence
- **R**ace / ethnicity
- **O**ccupation
- **G**ender
- **R**eligion
- **E**ducation
- **S**ocio-economic-status
- **S**ocial capital

ONE TO READ

Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 2003;327:1459–61 – this is a great example of how rigid adherence to the idea of 'best evidence' can sometimes be ludicrous!

ONE TO REMEMBER

A clear question is vital for developing a comprehensive search strategy, selecting relevant evidence for inclusion and drawing meaningful conclusions.

EXERCISE

1. Using the table below, formulate an answerable review question based on your presentation topic (this will be used in later exercises):

   P = ……

   I = ……

   C = ……

   O = ……

   Q = ……

   ……

   e.g. the effectiveness of (I) versus (C) for (0) in (P)

2. What type(s) of study design(s) should be included in the review?

   - Randomised controlled trial / cluster randomised controlled trial
   - Quasi-randomised controlled trial / pseudo-randomised trial
   - Cohort study with concurrent control / Controlled before-after study
   - Uncontrolled before-after study / cohort study without concurrent control
   - Qualitative research

Unit 5: Searching for Evidence

Learning Objectives

- To understand the importance of a comprehensive search
- To be able to develop a search strategy for locating relevant evidence
- To acquire basic skills to conduct a literature search

Potential for bias

Once an appropriate review question has been formulated, it is important to identify all evidence relevant to the question. An unrepresentative sample of included studies is a major threat to the validity of the review. The threat to validity arises from:

- *Reporting bias*: the selective reporting of research by researchers based on the strength and / or the direction of results
- *Publication bias*: the selective publishing of research (by editors) in peer-reviewed journals based on the strength and / or the direction of results
- *Language bias*: an increased potential for publication bias in English language journals
- *Geographical bias*: major databases (e.g. Medline) index a disproportionate amount of research conducted in North America and, by default, published in the English language

A good search

The Centre for Reviews and Dissemination has usefully produced a comprehensive checklist for finding studies for systematic reviews (http://www.york.ac.uk/inst/crd/revs.htm). Briefly, a good search strategy will

- be based on a clear research question
- attempt to locate up-to-date research, both published and unpublished, and without language restriction
- use a range of search media, including
  - electronic searching of research databases and general internet search engines
  - manual searching, including hand searching of relevant journals and screening the bibliographies of articles retrieved for the review
  - personal contact with key authors / research groups
- record all stages and results of the search strategy in sufficient detail for replication

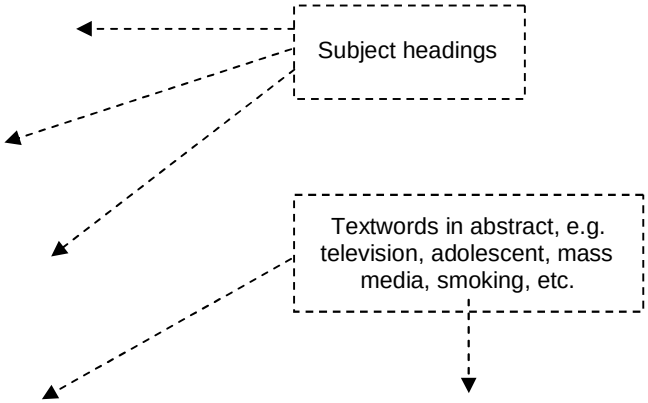
Components of database searching

Research databases do not search the full-text of the article for the search terms entered - only citation information is searched. Two distinct types of information are searched in the citation: subject headings, and textwords. The following complete reference shows the information that is available for each citation.

Example:

Unique Identifier: 2014859
Record Owner: NLM
Authors: Bauman KE. LaPrelle J. Brown JD. Koch GG. Padgett CA.
Institution: Department of Health Behavior and Health Education, School of Public Health, University of North Carolina, Chapel Hill 27599-7400.
Title: The influence of three mass media campaigns on variables related to adolescent cigarette smoking: results of a field experiment.
Source: American Journal of Public Health. 81(5):597-604, 1991 May.
Abbreviated Source: Am J Public Health. 81(5):597-604, 1991 May.
Publication Notes: The publication year is for the print issue of this journal.
NLM Journal Code: 1254074, 3xw
Journal Subset: AIM, IM
Local Messages: Held at RCH: 1985 onwards, Some years online fulltext - link from library journal list
Country of Publication: United States
MeSH Subject Headings

- **Adolescent**
- ***Adolescent Behavior**
- **Child**
- ***Health Education / mt [Methods]**
- **Human**
- ***Mass Media**
- **Pamphlets**
- **Peer Group**
- **Radio**
- **Regression Analysis**
- ***Smoking / pc [Prevention & Control]**
- **Southeastern United States**
- **Support, U.S. Gov't, P.H.S.**
- **Television**



Abstract
BACKGROUND: This paper reports findings from a field experiment that evaluated mass media campaigns designed to prevent cigarette smoking by adolescents. METHODS: The campaigns featured radio and television messages on expected consequences of smoking and a component to stimulate personal encouragement of peers not to smoke. Six Standard Metropolitan Statistical Areas in the Southeast United States received campaigns and four served as controls. Adolescents and mothers provided pretest and posttest data in their homes. RESULTS AND CONCLUSIONS: The radio campaign had a modest influence on the expected consequences of smoking and friend approval of smoking, the more expensive campaigns involving television were not more effective than those with radio alone, the peer- involvement component was not effective, and any potential smoking effects could not be detected.

ISSN: 0090- 0036

Publication Type: Journal Article.
Grant Number: CA38392 (NCI)
Language: English
Entry Date: 19910516
Revision Date: 20021101
Update Date: 20031209

Subject headings (or MeSH headings in Medline)

Subject headings are used in different databases to describe the subject of each article indexed in the database. For example, MeSH (Medical Subject Headings) are used in the Medline database, which uses more than 25,000 terms to describe studies and the headings are updated annually to reflect changes in terminology.

- Each database will have different controlled vocabulary (subject headings) meaning that search strategies will need to be adapted for each database that is searched
- Subject headings are assigned by error-prone human beings, e.g. the mass media article above was not assigned with the mass media subject heading in the PyscINFO database
- Search strategies should always include text words in addition to subject headings
- For many health behaviour topics there may be few subject headings available, in which case the search strategy may comprise mainly text words.

Text words

These are words that are used in the abstract of articles (and title) to assist with finding the relevant literature. Text words in a search strategy always end in .tw, e.g. adolescent.tw will find the word adolescent in the abstract and title of the article. A general rule is to duplicate all subject headings as text words, and add any other words may also describe the component of PICO.

- Truncation $: will pick up various forms of a text word

  e.g. teen$ will pick up teenage, teenagers, teens, teen

  e.g. Smok$ will pick up smoke, smoking, smokes, smoker, smokers

- Wildcards ? and #: these syntax commands pick up different spellings

  ? will substitute for one or no characters, so is useful for locating US and English spellings, e.g. colo?r.tw will pick up color and colour

  # will substitute for one character so is useful for picking up plural or singular versions of words, e.g. wom#n will pick up women and woman

- Adjacent ADJn - this command retrieves two or more query terms within *n* words of each other, and in any order. This syntax is important when the correct phraseology is unknown

  e.g. sport ADJ1 policy will pick up sport policy and policy for sport

  e.g. mental ADJ2 health will pick up mental health and mental and physical health

You will need to be become familiar with database idiosyncrasies, including:

- Use of different syntax to retrieve records, e.g. $ or * are used in different databases
- Use of different subject headings between databases, meaning that search strategies will need to be adapted for each database that is searched reviewers – this applies only to subject headings, not text words

Developing a database search strategy

- Identify relevant databases
- Identify primary concept for each PICO component
- Find synonyms / search terms for each primary concept
    - MeSH / Subject Headings / Descriptors, and Textwords
- Add other PICO components to limit search, e.g. study design filter

Study design filters

Study design filters can be added to search strategies in order to filter-out study designs not relevant to the review question. The sensitivity and specificity of study design filters depends on both the study design and database being searched. The use of such filters should be considered carefully.

- Study design filters appear reliable for identifying systematic reviews, studies conducting meta-analyses, and randomised controlled trials
- Use of study design filters is not generally recommended for non-randomised trials, resulting from poor and inconsistent use of non-standardised terminology
- Qualitative research: A CINAHL database filter is available from the Edward Miner Library http://www.urmc.rochester.edu/hslt/miner/digital_library/tip_sheets/Cinahl_eb_filters.pdf

CRD has a collection of study design filters for a range of databases, which can be downloaded: http://www.york.ac.uk/inst/crd/intertasc/index.htm

Research databases

Some examples of electronic databases that may be useful to identify health behaviour research include (websites listed for free access databases):

Psychology:	PsycINFO / PscyLIT
Biomedicine:	CINAHL, LILACS (Latin American Caribbean Health Sciences Literature: http://www.bireme.br/bvs/I/ibd.htm), Web of Science, Medline, EMBASE, CENTRAL (http://www.update-software.com/clibng/cliblogon.htm), CHID (Combined Health Information Database: http://chid.nih.gov/), CDP (Chronic Disease Prevention: http://www.cdc.gov/cdp/), SportsDiscus
Sociology:	Sociofile, Sociological Abstracts, Social Science Citation Index
Education:	ERIC (Educational Resources Information Center), C2-SPECTR (Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register: http://www.campbellcollaboration.org), REEL (Research Evidence in Education Library, EPPI-Centre: http://eppi.ioe.ac.uk)

Public Health:	BiblioMap (EPPI-Centre: http://eppi.ioe.ac.uk), HealthPromis (Health Development Agency Evidence: http://www.hda-online.org.uk/evidence/ - now held at NICE: http://www.publichealth.nice.org.uk), Popline (Population health and family planning: http://db.jhuccp.org/popinform/basic.html), Global Health
Qualitative:	ESRC Qualitative Data Archival Resource Centre (QUALIDATA) (http://www.qualidata.essex.ac.uk), Database of Interviews on Patient Experience (DIPEX) (http://www.dipex.org).
Ongoing:	National Research Register (http://www.update-software.com/national/), MRC Research Register (http://fundedresearch.cos.com/MRC/), Meta-Register of Controlled Trials (http://controlled-trials.com), Health Services Research Project (http://www.nlm.nih.gov/hsrproj/), CRISP (http://crisp.cit.nih.gov/).
Grey literature:	Conference Proceedings Index (http://www.bl.uk/services/current/inside.html), Conference Papers Index (http://www.cas.org/ONLINE/DBSS/confsciss.html), Theses (http://www.theses.org/), SIGLE, Dissertation Abstracts (http://wwwlib.umi.com/dissertations/), British Library Grey Literature Collection (http://www.bl.uk/services/document/greylit.html), Biomed Central (http://www.biomedcentral.com/)

Additional searching

Only about 50% of all known published trails are identifiable through Medline, and thus electronic searching should be supplemented

- Hand searching of key journals and conference proceedings
- Scanning bibliographies / reference lists of primary studies and reviews
- Contacting individuals / agencies / research groups / academic institutions / specialist libraries

Record, save and export search results

Always keep an accurate record of your searching. Below is an example of one way to record searches as they are carried out. It helps the searcher to keep track of what has been searched, and will also be useful when searches need to be updated.

It is essential to have bibliographic software (e.g. RefWorks) into which database search results (i.e. the retrieved citations) can be exported before being screened for inclusion / exclusion.

Citations from unpublished literature may need to be manually entered into the bibliographic software. Saving search results will assist with the referencing when writing the final review.

Example: Search record sheet

Review: _____

Searcher: _____ **Date:** _____

Database	Dates Covered	Date of search	Hits	Full record/ Titles only	Strategy Filename	Results Filename
MEDLINE	1966-2003/12	20/01/04	237	Full Records	medline1.txt	medres1.txt
EMBASE	1985-2003/12	20/01/04	371	Titles	embase1.txt	embres1.txt
PsychInfo						
CINAHL						
Brit Nursing Index						
HealthStar						

ONE TO READ

Harden A, Peersman G, Oliver S, Oakley A. Identifying primary research on electronic databases to inform decision-making in health promotion: the case of sexual health promotion. Health Education Journal 1999;58:290-301.

ONE TO REMEMBER

The search strategy must be comprehensive, thorough and accurately recorded – a poor search is a major threat to the validity of the review.

EXERCISE

1. Go through the worked example searching exercise.

2. Go back to PICO question developed in Unit Five.

   A). Find Medical Subject Headings (MeSH)/descriptors and text words that would help describe each of the PICO components of the review question.

	MeSH/descriptors e.g. Adolescent (Medline) e.g. High School Students (PsycINFO)	**Text words** student, school, teenage
P =	……	……
	……	……
	……	……
	……	……
	……	……
I =	……	……
	……	……
	……	……
	……	……
	……	……
C = *May not be required*		
	……	……
	……	……
	……	……
	……	……
O =	……	……
	……	……
	……	……
	……	……
	……	……

   B). Which databases would be most useful to locate studies on this topic? Do the descriptors differ between the databases?

   ……

   ……

   ……

WORKED EXAMPLE

We will work through the process of finding primary studies for a systematic review, using the review below as an example:

Sowden A, Arblaster L, Stead L. Community interventions for preventing smoking in young people (Cochrane Review). In: The Cochrane Library, Issue 3, 2004. Chichester, UK: Wiley & Sons, Ltd.

1 adolescent/
2 child/
3 Minors/
4 young people.tw.
5 (child$ or juvenile$ or girl$ or boy$ or teen$ or adolescen$).tw.
6 minor$.tw
7 or/1-6

All the subject headings and textwords for P

8 exp smoking/
9 tobacco/
10 "tobacco use disorder"/
11 (smok$ or tobacco or cigarette$).tw.
12 or/8-11

All the subject headings and textwords for O

13 (community or communities).tw.
14 (nationwide or statewide or countrywide or citywide).tw.
15 (nation adj wide).tw.
16 (state adj wide).tw.
17 ((country or city) adj wide).tw.
18 outreach.tw.
19 (multi adj (component or facet or faceted or disciplinary)).tw.
20 (inter adj disciplinary).tw.
21 (field adj based).tw.
22 local.tw.
23 citizen$.tw.
24 (multi adj community).tw.
25 or/13-24

All the subject headings (none found) and textwords for I

26 mass media/
27 audiovisual aids/
28 exp television/
29 motion pictures/
30 radio/
31 exp telecommunications/
32 videotape recording/
33 newspapers/
34 advertising/
35 (tv or televis$).tw.
36 (advertis$ adj4 (prevent or prevention)).tw.
37 (mass adj media).tw.
38 (radio or motion pictures or newspaper$ or video$ or audiovisual).tw.
39 or/26-38

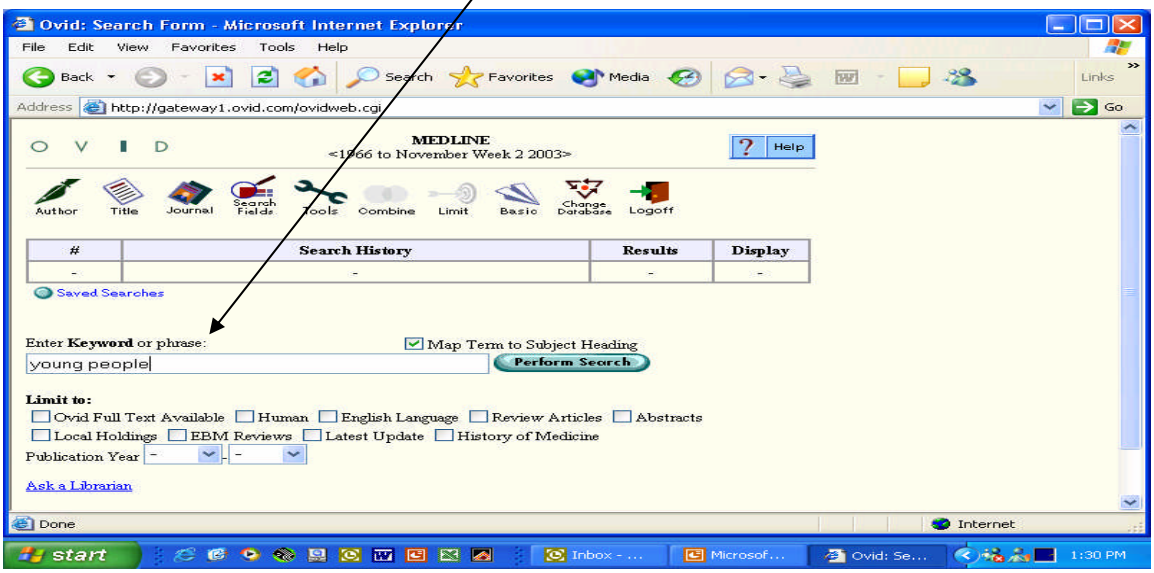
Mass media intervention excluded as not a community-based intervention (see search line 42)

40 7 and 12 and 25
41 7 and 12 and 39
42 40 not 41

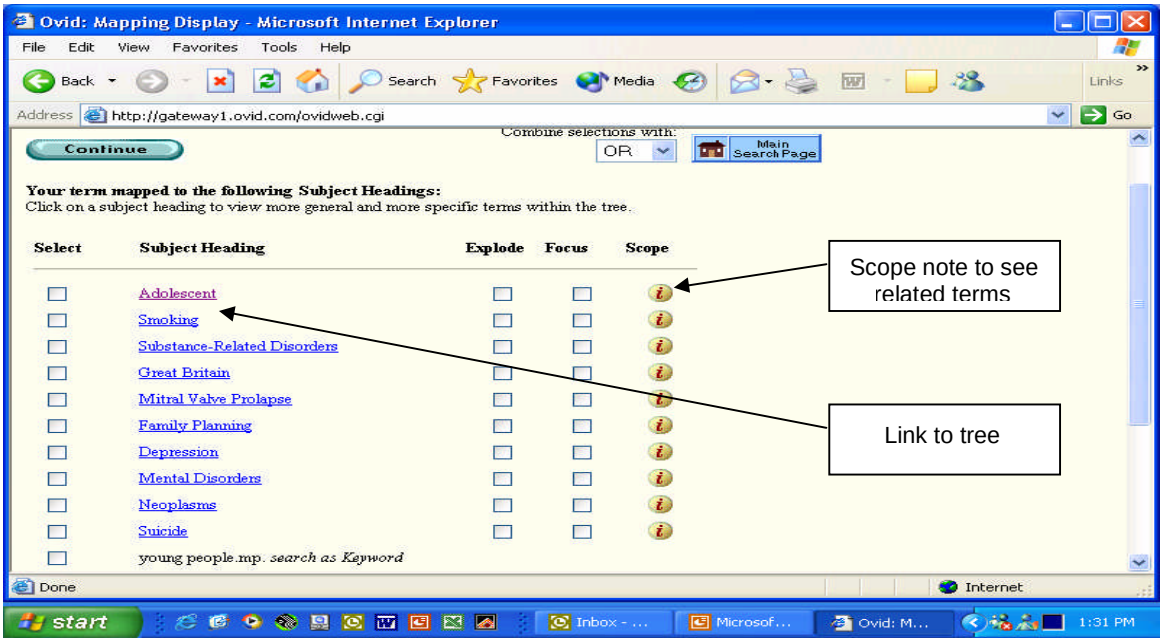
40 = young people & smoking & community-based interventions
41 = young people & smoking & mass media interventions
42 = community interventions not including mass media interventions

1. Start with the primary concept, i.e. young people.

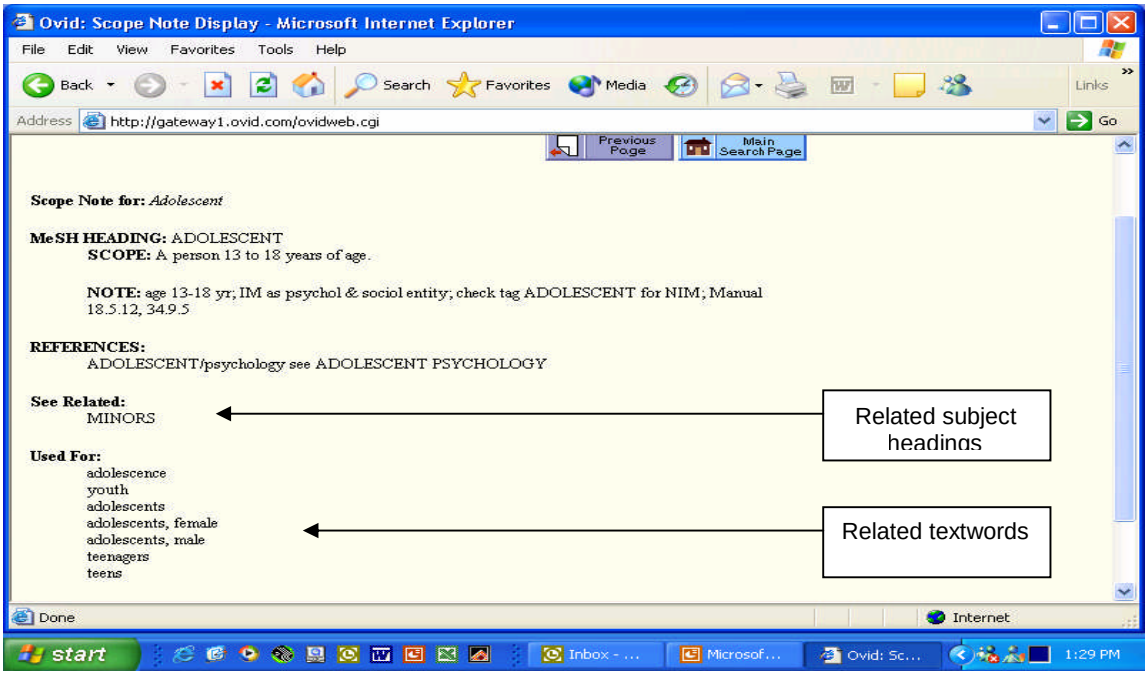
2. The Ovid search interface allows plain language to be 'mapped' to related subject headings, terms from a controlled indexing list (called controlled vocabulary) or thesaurus (e.g. MeSH in MEDLINE). **Map the term 'young people'**



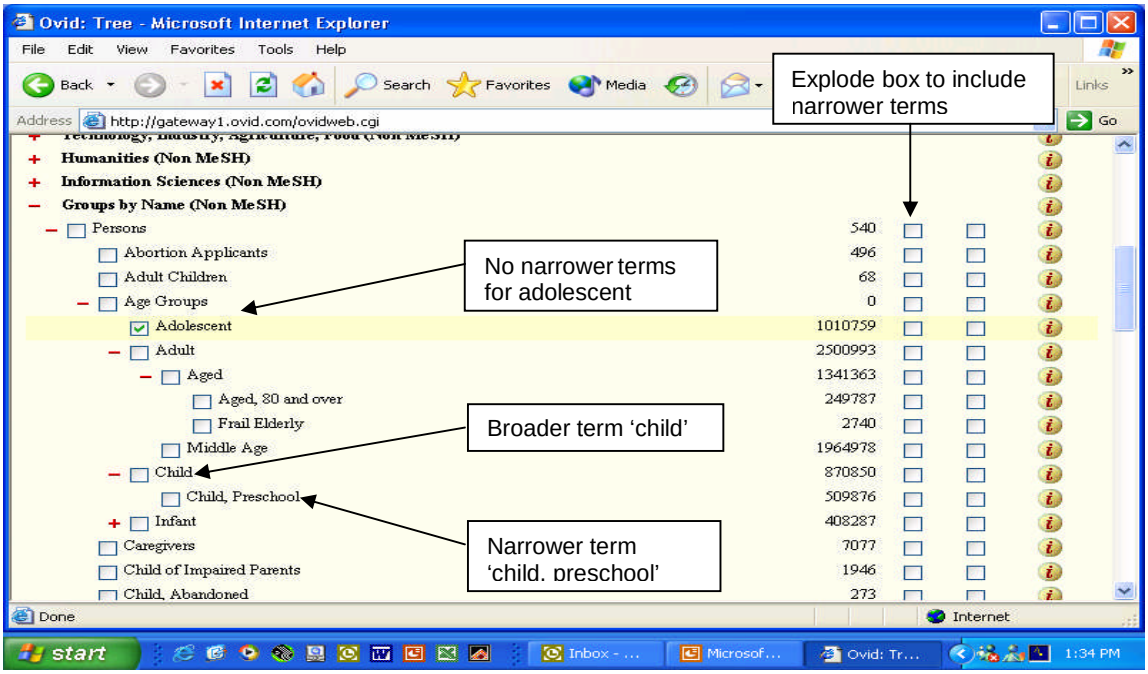
3. The result should look like this:



4. **Click on the scope note for the Adolescent** term (i symbol) to find the definition of adolescent and terms related to adolescent that can also be used in the search strategy. Note that Minors can also be used for the term adolescent.



5. **Click on Previous page and then Adolescent** to view the tree (the numbers will be different).

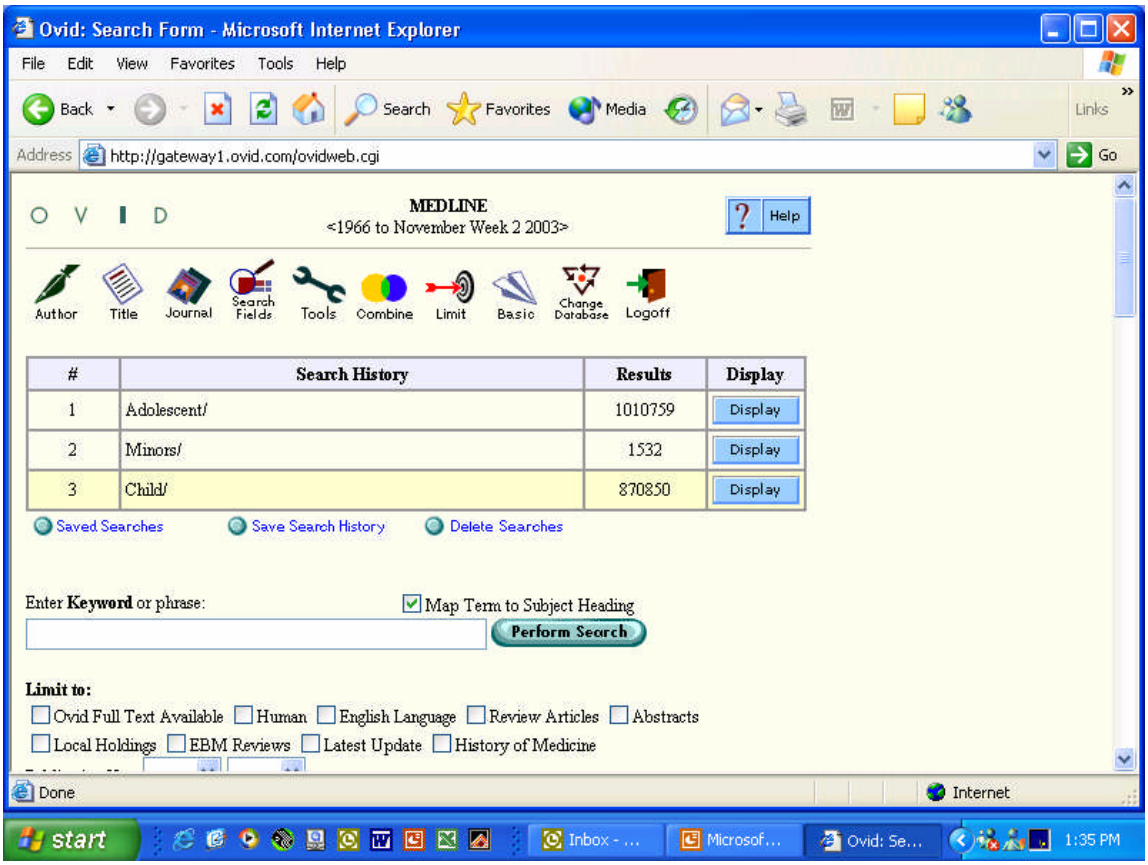


6. Because adolescent has no narrower terms click 'continue' at the top of the screen. This will produce a list of all subheadings. (If adolescent had narrower terms that are important to include the explode box would be checked).

7. **Press continue** (it is not recommended to select any of the subheadings for public health reviews).

8. The screen will now show all citations that have adolescent as a MeSH heading.

9. Repeat this strategy using the terms **child** and **minors**.



10. Using freetext or text-words to identify articles.
    Truncation - $ - Unlimited truncation is used to retrieve all possible suffix variations of a root word. Type the desired root word or phrase followed by either of the truncation characters '$' (dollar sign). Another wild card character is '?' (question mark). It can be used within or at the end of a query word to substitute for one or no characters. This wild card is useful for retrieving documents with British and American word variants.

11. Freetext words for searching – **type in young people.tw.**
    You can also combine all text words in one line by using the operator OR - this combines two or more query terms, creating a set that contains all the documents containing any of the query terms (with duplicates eliminated). For example, **type in (child$ or juvenile$ or girl$ or boy$ or teen$ or adolescen$).tw.**

12. Combine all young people related terms by **typing or/1-6**

13. Complete searches 8-12 and 13-25 in the worked example. Combine the three searches (7, 12, 25) by using the command AND.

Well done!

Now try a search using the PICO question you developed in Unit Five. A good start is to look at citations that are known to be relevant and see what terms have been used to index the article, or what relevant words appear in the abstract that can be used as text words.

Good luck!

Unit 6: Selecting Studies for Inclusion

Learning Objectives

- To be familiar with the process required to select papers for inclusion
- To understand the importance of independent application of inclusion / exclusion criteria
- To know why and how to record inclusion / exclusion decisions

Selection process

Once literature searches have been completed and saved in suitable bibliographic software, the records need to be screened for relevance in relation to inclusion / exclusion criteria, i.e. PICO-T. Individuals may make systematic errors (i.e. bias) when applying criteria, and thus each stage of the selection process should seek to minimise the potential for bias

- At least 2 reviewers should independently screen all references before decisions are compared and discrepancies resolved
- Reasons for exclusion should be recorded

First, all records identified in the search need to be screened for potential relevance

- If a paper does not satisfy one or more of the inclusion criteria it should be excluded, i.e. ruled-out
- For papers that can not be ruled-out, full-text copies should be ordered / obtained
- Decisions at this stage may be difficult, since the available information is limited to an abstract or, in some cases, a title only - if in doubt, a full-text copy of the paper should be obtained

Second, re-apply the inclusion criteria to the full-text version of papers identified during the first round of screening

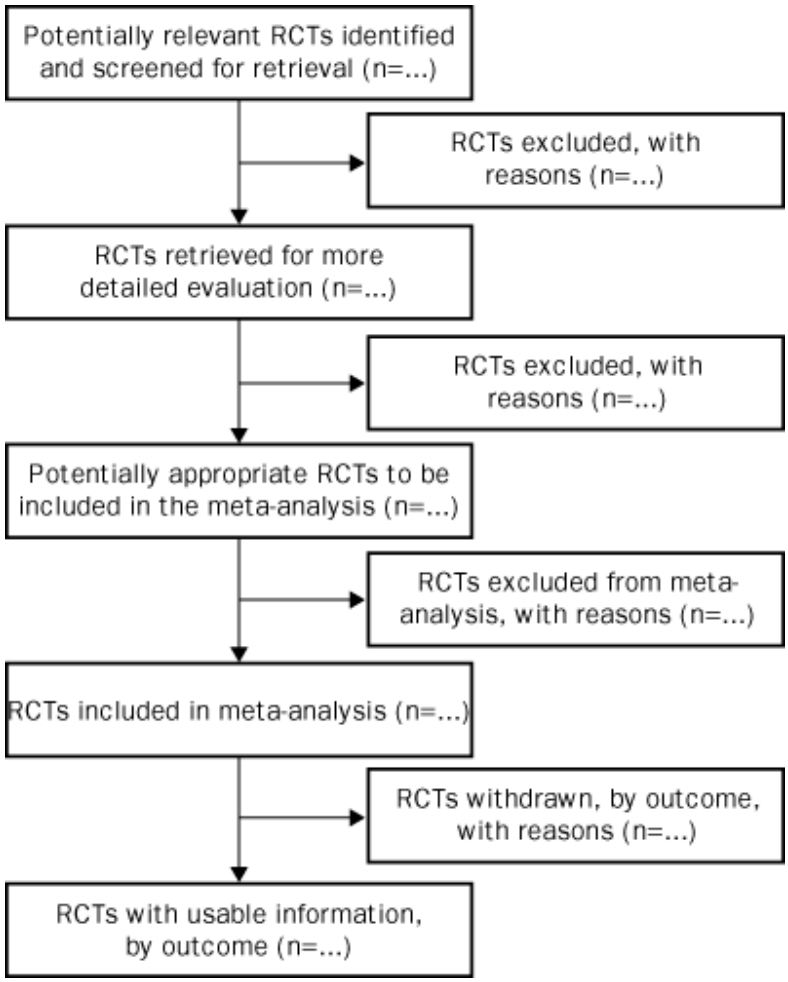
- If a paper does not satisfy one or more of the inclusion criteria it should be excluded, i.e. ruled-out
- Papers that satisfy ALL inclusion criteria are retained – all other papers are excluded
- The remaining papers are those of most relevance to the review question

Record your decisions

In a RCT, or any other primary study, it is important to be able to account for all participants recruited to the study, and a systematic review is no different, other than in this context our participants are study papers, and thus far better behaved. Recording selection decisions is important:

- Some reviews include hundreds of papers, making it difficult to keep track of all papers
- It will help deal with accusations of bias, e.g. '…you didn't include my paper …'
- Many journals require decision-data to be published as part of the review, often in the form of a flow chart, as in the example below

Figure 6.1: Flow of studies through a systematic review



Unit 7: Data Extraction

Learning Objectives

- To understand the importance of a well-designed, unambiguous data extraction form
- To know where to find examples of data extraction forms
- To identify the necessary data to extract from primary studies

Data extraction: What and why?

Data extraction refers to the systematic recording and structured presentation of data from primary studies. Clear presentation of important data from primary studies:

- Synthesis of findings becomes much easier
- A record to refer back to during the latter stages of the review process
- A great, comprehensive resource for anyone in the area, e.g. researchers and practitioners

Useful data to extract

It is important to strike the right balance between too much data and too few data, and this will vary from one review to the next. Common data include:

- Publication details: Author(s), year of publication, study design, target behaviour.
- Participants: n recruited, key characteristics (i.e. potential prognostic factors),
- Intervention details, e.g. full description of interventions given to all conditions, including controls, and stating whether controls
- Intervention context, e.g. who provided the intervention, where and for how long.
- Process measures, e.g. adherence, exposure, training, etc
- Results, e.g. attrition, N analysed, for each primary outcome (summary, contrast, precision)
- Comment, e.g. author's conclusion, as well as the conclusion / comment of the reviewer

Table 7.1: Example data extraction table for smoking cessation trial

Study	Participants	Intervention	Results	Conclusion / Comment
Smith, et al., (2003)	N Randomised: 290 (I=150, C=140) Age: m=43. Gender: 30% female Type: UK Community (Patient) Recruitment: Non-smoking related attendance at GP surgery	I: 3 x 30 min weekly stage-based, group MI with take-home intervention pack. C: GP advice Provider: Practice Nurse Setting: GP Surgery Follow-up: 2 months Outcome: Abstinence (3 wks), self-report questionnaire	Dropout: 82 (I=53, C=29) N Analysed: 208 (I=97, C=111) Abstinence: 31 (I=19, C=12) (p<0.05) **Reviewer analysis:** ITT OR=1.54 (95% CI, 0.63 to 4.29)	**Author:** Brief, stage-based MI with take-home material is an effective smoking cessation intervention. **Reviewer:** High attrition (I, OR = 2.09) and *ns* difference with ITT analysis. Tailoring unclear, re: group-level MI. Authors' conclusions are inconsistent with data.

Data extraction process

A template for entering data should be designed (using WORD, ACCESS, or similar) for capturing data identified for extraction in the protocol.

- Pilot the extraction form on a few papers among the review group
- Ensure extraction form captures all relevant data
- Ensure there is consistency among reviewers in the data being extraction and how it is being entered
- Data extracted by one reviewer and checked for accuracy by another

ONE TO READ

Clarke MJ, Stewart LA. Obtaining data from randomised controlled trials: How much do we need for reliable and informative meta-analysis? BMJ 1994;309:1007-1010.

ONE TO REMEMBER

Data to be extracted should be determined by the review question at the planning stage, not at the conduct stage by data reported in included studies – adhere to the protocol.

EXERCISE

1. In your own time, compare the style and content of the example data extraction templates in two or more of the following publications:

    - CRD Report Number 4. http://www.york.ac.uk/inst/crd/crd4_app3.pdf
    - Hedin A, and Kallestal C. Knowledge-based public health work. Part 2: Handbook for compilation of reviews on interventions in the field of public health. National Institute of Public Health. 2004. http://www.fhi.se/shop/material_pdf/r200410Knowledgebased2.pdf
    - The Community Guide http://www.thecommunityguide.org/methods/abstractionform.pdf
    - The Effective Public Health Practice Project reviews – (data extraction templates can be found in the appendices of reviews) http://www.city.hamilton.on.ca/phcs/EPHPP/default.asp

Unit 8: Critical Appraisal

Learning Objectives

- To know the benefits and limitations of quality assessment of primary studies
- To identify quality-related methodological criteria for a quantitative and qualitative study
- To understand the term 'bias' and distinguish between types of bias
- To gain experience in appraising health-related research, both qualitative and quantitative

Validity

Validity refers to prevention of systematic errors (bias) not precision (random errors). The interpretation of results depends on study validity, both internal and external validity:

- *Internal validity*: The extent to which the design, conduct and analysis of the study eliminate the possibility of bias. In systematic reviews, critical appraisal (or quality assessment) assesses internal validity, i.e. the reliability of results based on the potential for bias.
- *External validity*: The extent to which the results of a trial provide a correct basis for generalisations to other circumstances, i.e. the 'generalisability' or 'applicability' of results. Only results from internally valid studies should be considered for generalisability.

Bias

Bias refers to the systematic distortion of the estimated intervention effect away from the 'truth', caused by inadequacies in the design, conduct, or analysis of a trial. In other words, bias is the extent to which the observed *effect* may be due to factors other than the named intervention. There are four key types of bias that can systematically distort trial results:

- *Ascertainment bias*: Systematic distortion of the results of a randomised trial as a result of knowledge of the group assignment by the person assessing outcome, whether an investigator or the participant themselves.
- *Attrition bias*: Systematic differences between the comparison groups in the loss of participants from the study. Non-random differences in attrition after allocation may reflect dissatisfaction, usually with the treatment intervention, e.g. unpleasant, inconvenient, ineffective, etc.
- *Performance bias*: Systematic differences in the care provided to the participants in the comparison groups other than the intervention under investigation.
- *Selection bias*: Systematic error in creating intervention groups, such that they differ with respect to prognosis. That is, the groups differ in measured or unmeasured baseline characteristics because of the way participants were selected or assigned.

Critical appraisal criteria

Criteria used to critically appraise methodological quality relate to aspects of study design, conduct and analysis that reduce / remove the potential for one or more of the main sources of bias (see Appendix C). For example, the potential for ascertainment bias can be significantly reduced by blinding outcome assessors.

Poor reporting in primary studies makes it difficult to determine whether the criterion has been satisfied. For example, there are many ways in which researchers can randomise participants to treatment conditions, but study papers may merely report that participants *were* randomised without reporting *how*. This is important because some methods of randomisation are appropriate (e.g. computer generated random number tables) and some are flawed (e.g. alternation). This may seem pedantic, but there are very real effects associated with these seemingly unimportant distinctions.

As Table 8.1 illustrates, dimensions of methodology (i.e. criteria) are associated with large distortions in estimates of intervention effects.

Distortions have both qualitative and quantitative implications. In a study with an unclear / unreported method of randomisation, for example, a true effect of an odds ration of 1.2 (i.e. harmful effect) will – based on a 30% overestimation – translate into a beneficial effect of 0.84!

Quality of reporting does not account for these distortions, i.e. failing to report criterion-specific information is more likely to reflect poor methodology than poor reporting.

Table 8.1: Criteria and biased intervention effects

Quality Criteria	Mean% overestimation of intervention effect
Flawed randomisation	41
Unclear randomisation	30
Open allocation	25
Unblinded outcome assessment	35
Lack of blinding	17
No *a priori* sample size calculation	30
Failure to use ITT analysis	25
Poor quality of reporting	20

Khan et al, 1995; Moher et al, 1998

The relationship between criteria and bias is not always exclusive, and some criteria (e.g. method of randomisation) are related to more than one type of bias and the magnitude of effect may be mediated by other criteria. For example, in some situations the benefit of using an adequate method of randomisation may be undermined by a failure to conceal allocation, whereas in other situations the bias associated with use of a flawed method of randomisation may have little effect if allocation to conditions is concealed. This makes the interpretation of critical appraisal difficult.

The role of critical appraisal

The need to critically appraise the methodological quality of studies included in a review arises because studies of lower methodological quality tend to report different (usually more beneficial) intervention effects than studies of higher quality. However, there is much ongoing debate about the advantages and disadvantages of quality assessing studies included in a systematic review.

i Quality assessment may be beneficial when used:

- As threshold for study inclusion
- As explanation for differences in results between studies, e.g. in sensitivity analyses
- For making specific methodological recommendations for improving future research
- To guide an 'evidence-based' interpretation of review findings

ii Quality assessment of included studies may introduce bias into the review:

- Incorrect to assume that if something wasn't reported, it wasn't done
- Lack of evidence for relationship between some assessment criteria and study outcomes
- Simple vote counting (e.g. 3/10) ignores inherent limitations of 'assessing quality'

iii Variations in methodological rigour should not be ignored, but the potential benefits of quality assessment are dependent on an interpretation of quality based on:

- Sensible application of relevant criteria
- Broader potential for bias, not individual criteria, e.g. ascertainment bias not just blinding of outcome assessor
- Likely impact of any 'potential bias' on outcomes, e.g. little potential for bias from unblinded outcome assessors if assessment is objective / verifiable – death!

Critical appraisal tools

Numerous critical appraisal scales and checklists are available, many of which are reviewed in CRD Report 4. The choice as to which appraisal tool to use should be determined by the review topic and, in particular, the design of study being appraised. For quantitative research, examples include:

- CASP Checklist for Randomised Controlled Trials: http://www.phru.nhs.uk/casp/rct/
- Effective Public Health Practice Project : The Quality Assessment Tool for Quantitative Studies (http://www.city.hamilton.on.ca/phcs/EPHPP/).
- Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. J Epidemiol Community Health 2000;56:119-27.
- 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. A. Are the Results of the Study Valid? Evidence-Based Medicine Working Group. JAMA 1993;270(21):2598-2601.

If results from qualitative research are to contribute to the evidence-based interpretation of the review results, the quality of that evidence must be assessed. There are a number of checklists available to assess qualitative research, including:

- CASP Checklist tool for Qualitative Research: http://www.phru.nhs.uk/casp/qualitat.htm
- Greenhalgh T, Taylor R. Papers that go beyond numbers: Qualitative research. BMJ 1997;315:740-3.
- Health Care Practice Research and Development Unit, University of Salford, UK. Evaluation Tool for Qualitative Studies: http://www.fhsc.salford.ac.uk/hcprdu/tools/qualitative.htm
- Spencer L, Ritchie J, Lewis J, Dillon L. Quality in Qualitative Evaluation: A framework for assessing research evidence. Government Chief Social Researcher's Office. Crown Copyright, 2003. www.strategy.gov.uk/files/pdf/Quality_framework.pdf

ONE TO READ

Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001;323:42–6.

ONE TO REMEMBER

Critical appraisal of methodological quality requires careful consideration, and should be interpreted in relation to the broader context of the study.

EXERCISE

1. In groups, use the checklist provided to appraise the methodological quality of one of the following studies:

   i. Sahota P, Rudolf MCJ, Dixey R, Hill AJ, Barth JH, Cade J. Randomised controlled trial of primary school based intervention to reduce risk factors for obesity. BMJ 2001;323:1029i-1032.

   ii. Gortmaker S, Cheung S, Peterson K, Chomitz G, Cradle J, Dart H, Fox M, Bullock R, Sobol A, Colditz G, Field A, Laird N. Impact of a school-based interdisciplinary intervention on diet and physical activity among urban primary school children. Arch Pediatr Adolsc Med 1999;153:975-983.

   iii. Cass A, Lowell A, Christie M, Snelling PL, Flack M, Marrnganyin B, Brown I. Sharing the true stories: Improving communication between Aboriginal patients and healthcare workers. Med J Aust 2002; 176:466-70

Unit 9: Synthesising the Evidence

Learning Objectives

- To understand the different methods available for synthesising evidence
- To understand the terms: meta-analysis, confidence interval, heterogeneity, odds ratio, relative risk, narrative synthesis

Two general methods of synthesis

- Qualitative: narrative summary and synthesis of data
- Quantitative: data combined statistically to produce a single numeric estimate of effect, i.e. meta-analysis

The decision about which method of synthesis to use depends on the diversity of studies included in the review, i.e. heterogeneity.

Heterogeneity

Heterogeneity refers to differences between studies in terms of key characteristics. Studies will differ in an almost infinite number of ways, so it is helpful to think of these differences as falling under the rubric of one of three broader types of heterogeneity.

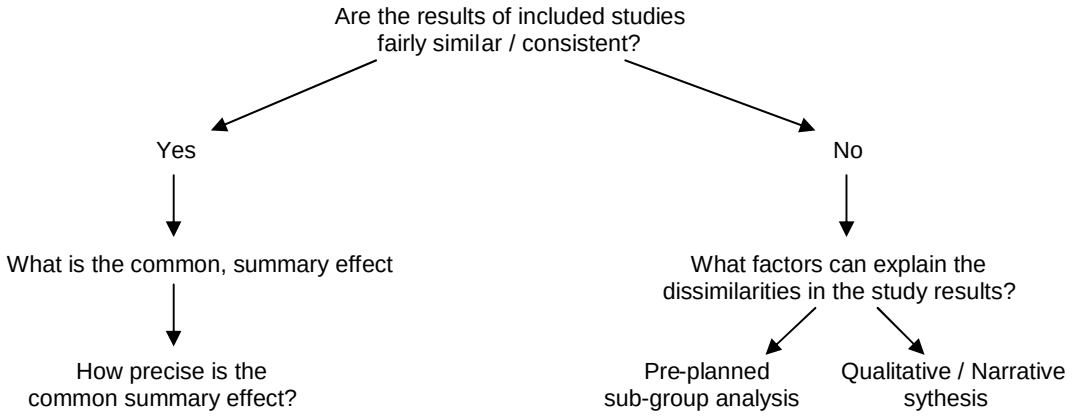
- *Clinical heterogeneity* refers to differences in the studies concerning the participants, interventions and outcomes, e.g. age, context, intervention intensity, outcome definition, etc.
- *Methodological heterogeneity* refers to differences between how the studies were conducted, e.g. study design, unit of randomisation, study quality, method of analysis, etc.
- *Statistical heterogeneity* refers to variation between studies in the measured intervention effect

Studies should only be combined statistically if they are sufficiently similar so as to produce a meaningful average effect.

- If there is reason to believe that any clinical or methodological differences may influence the size or direction of the intervention effect, it may not be appropriate to pool studies
- It is inappropriate to calculate an average effect if there is a large amount of statistical heterogeneity between studies

Central questions of interest

The purpose of synthesising evidence is to assess homogeneity of effect and, where necessary, identify the source or sources of effect heterogeneity



Key steps in synthesising evidence

The process of synthesising data should be explicit and rigorous. The following steps are recommended:

- Tabulate summary data
- Graph data (where possible) – forest plot
- Check for heterogeneity
  - No – meta-analysis
  - Yes – subgroup analysis, or qualitative synthesis
- Evaluate the influence of study quality on review results, e.g. sensitivity analysis
- Explore potential for publication bias, e.g. funnel plot

Tabulate summary data

Tabulating the findings from the studies helps

- the reviewer in assessing whether studies are likely to be homogenous or heterogeneous
- the reader in eyeballing the types of studies that were included in the review

Because health behaviour interventions differ in numerous ways, data tabulation needs to be selective and focussed on characteristics that may influence the effectiveness of the intervention.

Table 9.1: Example of data tabulation

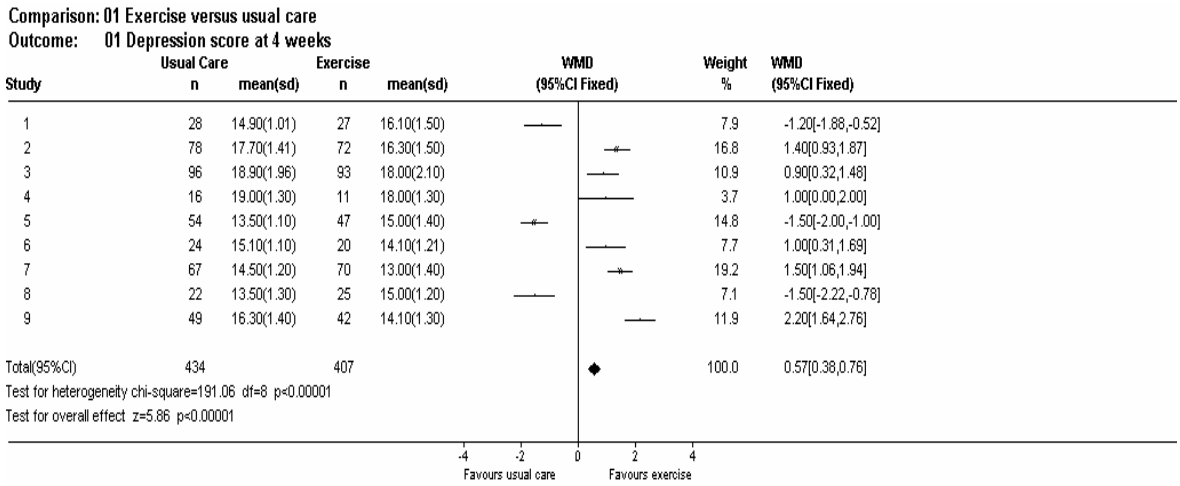
Study	Participants	Intervention	Context	Comparison	Outcome (Abstinence)	Summary effect OR (95%CI)	Validity
Smith, et al (2003)	290, UK GP patients	Group MI + written advice	Nurse, GP surgery, 3 pw	Usual care	Self-report at 2 months	1.54 (0.63,4.29)	Poor
Jones, et al (2004)	600, UK community	Group MI	Researcher, community centre, 2 pw	No intervention	Biochemical validation at 12 months	1.03 (0.33,1.22)	Good
Davis, et al (2005)	100, UK students	Stage-based	Written material	No intervention	Self-report at 2 months	2.54 (1.33,4.89)	Poor
McScott, (2006)	60, UK GP patients	Individual MI	Counsellor, home visit, 1pw	No intervention	Self-report at 1 month	1.87 (1.12,3.19)	Poor

Graph data

Where sufficient data are available, graphically present data using a Forest Plot

- Presents the point estimate and CI of each trial
- Also presents the overall, summary estimate

Graph 9.1: Workplace exercise interventions for mild depression



Check for heterogeneity

Use the tabulated data and graphical representation to check for heterogeneity

- Tabulated data should be used to check for heterogeneity among potential determinants of intervention effectiveness, i.e. clinical and methodological heterogeneity, as well as the direction of study results
- Graphical data can be used to assess statistical heterogeneity, such as point estimates on different sides of the line of unity, and CIs that do not overlap between some studies
- Statistical assessment of heterogeneity is provided by the chi-square statistic, which is produced by default in the Forest Plot. Significance is set at $p<.1$ for the chi-square, with non-significance indicating non-heterogeneous data

Caution: If the chi-square heterogeneity test reveals no statistical heterogeneity it should not be assumed that a meta-analysis is appropriate

- Chi-square has limited power to detect significant differences
- Health behaviour interventions have numerous potential sources of variation, which, for individual studies, may cause important but non-significant variations in intervention effectiveness
- Similar effect sizes may be obtained from studies that are conceptually very different and which merit separate assessment and interpretation

Best advice: In reviews of health behaviour interventions the reviewer needs to make the case for meta-analysis before proceeding.

If significant heterogeneity is found, or suspected:

- Investigate statistically what factors might explain the heterogeneity, e.g. subgroup analysis
- Investigate qualitatively what factors might explain the heterogeneity, i.e. narrative synthesis

If no heterogeneity is found or suspected:

- Perform meta-analysis

Qualitative synthesis of quantitative studies

If the studies included in the review are heterogeneous then it is preferable to perform a qualitative or narrative synthesis. Explicit guidelines for narrative synthesis are not available, but the central issues are the same

- Explore included studies to identify factors that may explain variations in study results
- Ideally, narrative synthesis should stratify results (e.g. favourable or unfavourable) and discuss in relation to factors identified *a priori* as potential sources of effect heterogeneity

- Important sources of heterogeneity are likely to be aspects related to participant characteristics, features of the intervention, outcome assessment and validity

Meta-Analysis: Process

If studies are sufficiently similar, meta-analysis may be appropriate. Meta-analysis essentially computes a weighted average of effect sizes, usually weighted to study size

- Calculate summary measure of effect for each included study
- Compute the weighted average effect
- Measure how well individual study results agree with the weighted average and, where necessary, investigate sources of statistical (i.e. effect) heterogeneity

Meta-analysis: Summary measures of effect

Effect size refers to the magnitude of effect observed in a study, which may be the size of a relationship between variables or the degree of difference between group means / proportions. Calculate summary effect measure for chosen comparison

- Dichotomous data: Relative Risk (aka Risk Ratio), Attributable Risk (aka Risk Difference), Odds Ratio, and Number Needed to Treat. These effects measures are calculated from a 2x2 contingency table depicting participants with or without the event in each condition.
- Continuous data: weighted mean difference or, especially when different measurement scales have been used, standardised mean difference, e.g. Glass's Δ, Cohen's *d*, Hedge's *g*. These effect measures can be calculated from a range of data presented in primary studies including Pearson's *r*, t-tests, F-tests, chi-square, and z-scores.
- Effect measures are estimates, the precision of which should be reported, i.e. confidence interval (CI). CIs indicate the precision of the estimated effect by providing the range within which the true effect lies, within a given degree of assurance, e.g. 95%.

There is no consensus regarding which effect measure should be used for either dichotomous or continuous data, but two issues should guide selection of summary effect measure:

- Communication (i.e. a straightforward and clinically useful interpretation)
- Consistency of the statistic across different studies

Meta-analysis: Models

Fixed effects model

- Assumes the true treatment effect is the same value in each study (fixed); difference between studies is due to random error

Random effects model

- Assumes treatment effects for individual studies vary around some overall average effect

- Allows for random error plus inter-study variability, resulting in wider confidence intervals
- Studies weighted more equally, i.e. relatively more weight is given to smaller studies

Which model to use

Most meta-analyses published in the psychology literature have used a fixed effects model. This is wrong. The random effects model should always be the preferred option because

- it offers a more realistic representation of reality
- real-world data from health behaviour interventions will have heterogeneous population effect sizes, even in the absence of known moderator variables
- it permits unconditional inferences, i.e. inferences that generalise beyond the studies included in the meta-analysis

Dealing with statistical heterogeneity

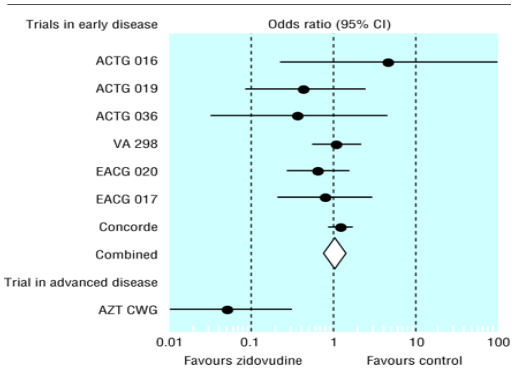
Studies should not be combined statistically if there is significant variation in reported intervention effects. If variation is confined to a selection of clearly distinct studies, it may be appropriate to perform subgroup analyses, i.e. conduct and compare separate meta-analyses based on subgroups of studies.

Trials involving patients with early stage HIV show no benefit for ZDT, i.e. people with early stage HIV do not live longer if they take ZDT.

The one trial involving patients with advanced stage HIV (AZT CWG), however, does show a significant benefit, i.e. people with advanced stage HIV do live longer if they take ZDT.

This relatively small but clinically important finding would be masked in a combined meta-analysis, which would suggest that ZDT has no effect on mortality.

Graph 9.2: HIV mortality results in ZDT trials, stratified by infection stage (early vs late)



Subgroup analyses must be interpreted with caution because the protection of randomisation is removed. For example, even where primary studies are well-conducted randomised controlled trials the results from subgroup analyses nevertheless

- reflect indirect comparisons, e.g. the effects of ZDT were not compared directly (i.e. in the same study) between people with early and late stage HIV, but indirectly, i.e. across different studies

- have greater potential for bias and confounding because they are *observational* in nature, e.g. the apparent benefits of ZDT in the AZT CWG trial may reflect any number of differences between trials other than infection stage, such as study quality, use of co-interventions, age, etc.

Subgroup analyses should be specified *a priori* in the review protocol, kept to a minimum and thought of as hypothesis generating rather than conclusion generation, e.g. infection stage may be a determinant of ZDT effectiveness.

Influence of quality of results

There is evidence that studies of lower methodological quality tend to report different (usually more beneficial) intervention effects than studies of higher quality. The influence of quality on the review results needs to be assessed. The impact of quality on results can be discussed narratively as well as being presented graphically, e.g. display study quality and results in a tabular format.

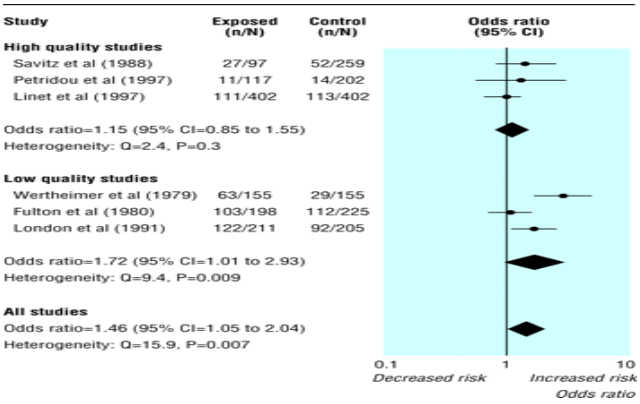
Where studies have been combined statistically, sensitivity analysis is often used to explore the influence of quality of results. Sensitivity analysis involves conducting repeated meta-analyses with amended inclusion criteria to determine the robustness of review findings.

The combined meta-analysis suggests that exposure to residential EMG is associated with a significantly greater risk of childhood leukaemia (OR = 1.46, 95% CI 1.05, 2.04).

The size of the effect in low quality studies is larger (OR = 1.72, 95% CI 1.01, 2.93), whereas the effect is not only smaller but non-significant in high quality studies (OR = 1.15, 95% CI 0.85, 1.55).

This suggests that study quality is influencing review results.

Graph 9.3: Case-control studies relating residential EMG exposure to childhood leukaemia, stratified by quality



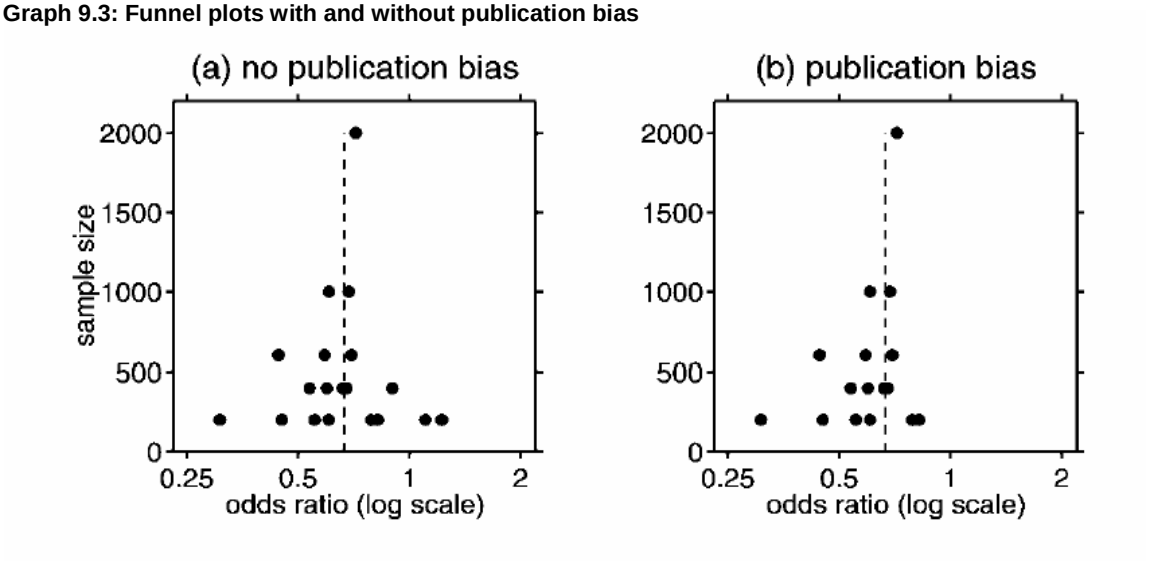
Potential for publication bias

Publication biases exists because research with statistically significant or interesting results is potentially more likely to be submitted, published and published more rapidly, especially in English language journals, than research with null or non-significant results.

Although a comprehensive search that includes attempts to locate unpublished research reduces the potential for bias, it should be examined explicitly. Several methods exist for examining the representativeness of studies included in the review, all of which are based on the same symmetry assumption.

The most common method for assessing publication bias is the funnel plot, which plots the effect size for each study against some measure of its precision, e.g. sample size or, if the included studies have small sample sizes, 1/standard error of the effect size.

Graph 9.3: Funnel plots with and without publication bias



A plot shaped like a funnel indicates no publication bias, as seen in Plot A above. A funnel shape is expected because trials of decreasing size have increasingly large variation in their effect size estimates due to random variation becoming increasingly influential.

If the chance for publication is greater for larger trials or trials with statistically significant results, some small non-significant studies will not appear in the literature. An absence of such trials will lead to a gap in the bottom right of the plot, and hence a degree of asymmetry in the funnel, as in Plot B above.

Disagreement exists about how best to proceed if publication bias is suspected but, at the very least, the potential for bias should be considered when interpreting review results (see Unit 10).

Synthesis of qualitative data

The synthesis of qualitative data in the context of a systematic review is problematic not only because of difficulties associated with locating qualitative studies, but also because there is no formal method for synthesising qualitative data. The varying theoretical perspectives include

- Cross-case analysis (Miles & Huberman, 1994)

- Nominal group technique (Pope & Mays, 1996)
- Signal-noise technique (Higginson et al., 2002)
- Delphi technique (Jones & Hunter, 2002)
- Meta-ethnography (Noblit & Hare, 1988)
- Integration (Thomas et al., 2004)

The Cochrane Qualitative Methods Groups is conducting research aimed at refining methods for locating and synthesising qualitative research. More information is available on the group's webpage: http://mysite.freeserve.com/Cochrane_Qual_Method/index.htm. Until these methods are more fully developed, the synthesis of qualitative data will remain problematic.

For the time being, although meta-ethnography is the most commonly used method for combining qualitative data, it may be more informative to *integrate* qualitative data into / with the quantitative data used in systematic reviews of health behaviour interventions.

Integrating qualitative and quantitative data

Although systematic reviews may provide an unbiased assessment of the evidence concerning the effectiveness of an intervention, they may be of little use to 'users', such as policy makers and practitioners. Whilst 'users' of reviews want to know about intervention effectiveness, other issues need to be considered when making healthcare decisions. In particular, questions such as

- if the intervention is effective, is it also appropriate, relevant and acceptable to the people / patients who receive it?
- if the intervention is not effective, what are the alternative interventions and to what extent are these appropriate, relevant and acceptable to the people / patients who may receive them?
- Irrespective of effectiveness, what type of intervention is most appropriate, relevant and acceptable to the people / patients who may receive it?

Systematic reviews have mostly neglected these issues, perhaps because providing answers to these questions requires synthesising different types of evidence and methods for integrating different types of evidence are not well-developed. In essence, integrating different types of evidence involves three types of syntheses in the same review (see Thomas et al, 2004):

- a synthesis of quantitative intervention studies tackling a particular problem
- a synthesis of studies examining people's perspectives or experiences of that problem (or the intervention) using qualitative data
- a 'mixed methods' synthesis bringing the quantitative and qualitative together

1 Effectiveness synthesis for trials

Effect sizes from good quality trials are extracted and, if appropriate, pooled using statistical meta-analysis. Heterogeneity is explored either narratively or statistically on a range of categories specified in advance, e.g. study quality, setting and type of intervention.

2 Qualitative synthesis for 'views' studies

The textual data describing the findings from 'views' studies are copied verbatim and entered into a software package to aid qualitative analysis. Two or more reviewers undertake a thematic analysis on this data. Themes are descriptive and stay close to the data, building up a picture of the range and depth of people's perspectives and experiences in relation to the health issue under study.

The content of the descriptive themes are considered in the light of the relevant review question (e.g. what helps and what stops people from quitting smoking?) in order to generate implications for intervention development. The products of this kind of synthesis can be conceptualised as 'theories' about which interventions might work. These theories are grounded in people's own understandings about their lives and health. These methods highlight the theory building potential of synthesis.

3 A 'mixed methods' synthesis

Implications for interventions are juxtaposed against the interventions which have been evaluated by trials included in the 'effectiveness' synthesis. Using the descriptions of the interventions provided in the reports of the trials, matches, miss-matches and gaps can be identified. Gaps may be used for recommending what kinds of interventions need to be developed and evaluated. The effect sizes from interventions which match implications for interventions derived from people's views can be compared to those which do not, using sub-group analysis. This makes it possible to identify the types of interventions that are both effective and appropriate.

Unlike Bayesian methods, which combine qualitative and quantitative studies within systematic reviews by translating textual data into numerical data, these methods integrate 'quantitative' estimates of effect with 'qualitative' understanding from people's lives, whilst preserving the unique contribution of each.

ONE TO READ

Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Brunton G, Kavanagh J. Integrating qualitative research with trials in systematic reviews. BMJ 2004;328:1010-2.

ONE TO REMEMBER

Because health behaviour interventions are complex, being characterised by many known and unknown sources of heterogeneity, the case for conducting a quantitative synthesis needs to be clearly demonstrated – qualitative synthesis should be the default option.

EXERCISE

1. Together we will calculate and interpret effect measures from the data provided in the following worksheet:

Miscarriage and exposure to pesticide

	Miscarriage	**No Miscarriage**	**Total**
Exposed	30 (A)	70 (B)	100 (A+B)
Non-Exposed	10 (C)	90 (D)	100 (C+D)
Total	40 (A+C)	160 (B+D)	200 (A+B+C+D)

1. Calculate the RR of miscarriage for women exposed to pesticide.

Formula: (a/a+b) / (c/c+d) **RR =** _____

Interpretation: A pregnant women exposed to pesticide is _____ times more likely to miscarry than a pregnant women who is not exposed. The risk of miscarriage is _____ times greater among the exposed than those not exposed.

2. Calculate the OR for the association between miscarriage and past exposure to pesticide.

Formula: (axd/bxc) **OR =** _____

Interpretation: The odds of miscarrying are _____ times greater for women exposed to pesticide than for those not exposed. In other words, we are _____ times more likely to find prior exposure to pesticide among women experiencing miscarriage than among women experiencing a normal, full-term pregnancy.

3. Calculate the increased risk (AR) of miscarriage that can be attributed to exposure to pesticide.

Formula: (a/a+b) – (c/c+d) **AR =** _____

The excess or increased risk of miscarriage that can be attributed to pesticide exposure is _____. Thus, if a pregnant woman is exposed to pesticide her risk of miscarriage is increased by _____%.

4. Calculate the NNT.

Formula: 1/ARR **NNT =** _____

Interpretation: We would need to stop _____ pregnant women from being exposed to pesticides in order to prevent one woman from having a miscarriage.

Unit 10: Interpretation of Results

Learning Objectives

- To be able to interpret the results from studies in order to formulate evidence-based conclusions and recommendations
- To understand the factors that impact on the effectiveness of health behaviour interventions

Key considerations

As those who read systematic reviews (e.g. policy makers, practitioners) may not have time to read the whole review, it is important that the conclusions and recommendations are clearly worded and arise directly from the evidence presented in the review. Evidence-based conclusions and recommendations will usefully reflect careful consideration of the following:

- Strength of the evidence
- Integrity of intervention
- Theoretical explanations of effectiveness
- Context as an effect modifier
- Trade-offs between benefits and harms
- Implications for practice and research

Strength of the evidence

Conclusions and recommendations should reflect the strength of the evidence presented in the review. In particular, the strength of the evidence should be assessed in relation to the following:

- Methodological quality of included studies
- Size of intervention effect
- Consistency of intervention effect across studies
- Methodological quality of the review, especially in terms of key review processes, e.g. potential for publication bias

Intervention integrity

The relationship between intervention integrity and effectiveness should be described in relation to key aspects of the intervention:

- dose / intensity, i.e. the amount of intervention provided for participants
- contact, i.e. amount of intervention received by participants
- content, i.e. consistent with theory upon which it is based
- implementation, i.e. monitoring of intervention provision

Theoretical explanation

Reviewers should seek to examine the impact of the theoretical framework on the effectiveness of the intervention. The assessment of theory within systematic reviews:

- provides a framework within which to explore the relationship between findings from different studies, e.g. group interventions by their theoretical basis
- helps to explain success or failure in different interventions, by highlighting the possible impact of differences between what was planned and what actually happened in the implementation of the program
- assists in identifying the key elements or components of an intervention

Context modifiers

Interventions which are effective may be effective due to pre-existing factors of the context into which the intervention was introduced. Where information is available, reviewers should report on the presence of context-related information:

- time and place of intervention
- aspects of the host organisation and staff, e.g. the resources made available to the intervention program, and number, experience / training, morale, expertise of staff
- aspects of the system, e.g. payment and fee structures for services, reward structures, degrees of specialisation in service delivery
- characteristics of the target population, e.g. cultural, socioeconomic, place of residence

The boundary between the particular intervention and its context is not always easy to identify, and seemingly similar interventions can have a different effect depending on the context in which it is implemented.

Benefits and harms

Few health behaviour interventions either consider or report data relating to adverse effects, but the potential for harm should be considered.

- Attrition, e.g. high(er) rates of attrition in intervention groups indicate dissatisfaction / lack of acceptability, perhaps because of adverse effects

- Labelling, e.g. interventions targeting particular populations (e.g. single parent families) may result in stigma and social exclusion

- Differential effectiveness, e.g. interventions may be less effective for certain sub-groups, such as those formed on S-E-S and ethnicity. In fact, interventions that are effective in disadvantaged groups, but to a lesser extent than in non-disadvantaged groups, might be better interpreted as negative or harmful, since they increase heath inequalities.

Implications for practice and research

Reviewers are in an ideal position to identify implications for practice and suggest directions for future research.

- If there are gaps or weaknesses in the evidence base clear and specific recommendations for research should be made, e.g. participants, intervention contexts and settings, study design, sample size, outcome assessment, methods of randomisation, intention-to-treat analysis, etc.

- Current practice and policy should be discussed in the light of the interpretation of review evidence

ONE TO READ

Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health. 2003 Aug;93(8):1261-7.

ONE TO REMEMBER

In many cases the review conclusions will be all that is read, and it is therefore extremely important that conclusions reflect the quality of the evidence, and that the wider health care context has been considered in formulating recommendations.

EXERCISE

1. In small groups, list the types of information required from studies to help you determine the generalisability of results and the transferability of interventions to other settings.

2. In your own time, assess the extent to which key issues have been considered in the interpretation of results presented in the following review:

   Bridle C, Riemsma RP, Pattenden J, Sowden AJ, Mather L, Watt IS, & Walker A. (2005). Systematic review of the effectiveness of health behaviour interventions based on the transtheoretical model. *Psychology and Health*, 20(3), 283-301.

Unit 11: Writing the Systematic Review

Learning Objectives

- To understand the requirements to publish a systematic review
- To be familiar with the criteria that will be used to judged the quality of a systematic review

Publication

Two sets of guidelines are available for reviewers wishing to submit the review to a published journal. Reviewers should read the guidelines relevant to the study designs included in the review:

- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet. 1999 Nov 27;354(9193):1896-900.

  Checklist: http://www.consort-statement.org/QUOROM.pdf

- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000 Apr 19;283(15):2008-12.

  Checklist: http://www.consort-statement.org/Initiatives/MOOSE/Moosecheck.pdf

Critical appraisal

As with other types of research, the quality of a review can be assessed in terms of the systematic manner in which the potential for bias was removed / reduced. Core assessment criteria relate to the key stages of the review process:

- Question: Is the review question clear and specific?
- Search: Have attempts to identify relevant evidence been sufficiently comprehensive?
- Evaluation: Have included studies been critically appraised?
- Synthesis: Is the method of synthesis appropriate? And, have potential sources of heterogeneity been investigated?
- Conclusions: Do conclusions reflect both the quality and quantity of evidence?
- Process: Has the review process limited the potential for bias?

A useful tool to assess the quality of a systematic review is produced by the Critical Appraisal Skills Program (CASP: http://www.phru.nhs.uk/~casp/appraisa.htm). It is useful to keep this tool in mind when writing the final review.

ONE TO READ

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

ONE TO REMEMBER

We have come full circle - the first **'ONE TO REMEMBER'** (p8) highlighted that the key benefit of systematic review is its potential to limit bias when conducted appropriately, it is therefore important to assess the methodological quality of each systematic review before using it to inform decisions concerning healthcare policy, provision and research.

The workshop is finished – don't contact me again.

EXERCISE

1. In groups, critically appraise the following systematic review using the checklist provided:

   DiCenso A, Guyatt G, Willan A, Griffith L. Interventions to reduce unintended pregnancies among adolescents: systematic review of randomised controlled trials. BMJ 2002;324:1426-34.

Appendix A: Glossary of Systematic Review Terminology

Attrition: subject units lost during the experimental/investigational period than cannot be included in the analysis (e.g. units removed due to deleterious side-effects caused by the intervention).

Bias (synonym: systematic error): the distortion of the outcome, as a result of a known or unknown variable other than intervention (i.e. the tendency to produce results that depart from the "true" result).

Confounding variable (synonym: co-variate): a variable associated with the outcome, which distorts the effect of intervention.

Effectiveness: the extent to which an intervention produces a beneficial outcome under ordinary circumstances (i.e. does the intervention work?).

Effect size: the observed association between the intervention and outcome, where the improvement/decrement of the outcome is described in deviations from the mean.

Efficacy: the extent to which an intervention produces a beneficial outcome under ideally controlled circumstances (i.e. can the intervention work?).

Efficiency: the extent to which the effect of the intervention on the outcome represents value for money (i.e. the balance between cost and outcome).

Evidence-based health care: extends the application of the principles of evidence-based medicine to all professions associated with health care, including purchasing and management.

Evidence-based medicine (EBM): is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.

Fixed effects model: a mathematical model that combines the results of studies that assume the effect of the intervention is constant in all subject populations studied. Only within-study variation is included when assessing the uncertainty of results (in contrast to a random effects model).

Forest plot: a plot illustrating individual effect sizes observed in studies included within a systematic review (incorporating the summary effect if meta-analysis is used).

Funnel plot: a graphical method of assessing bias; the effect size of each study is plotted against some measure of study information (e.g. sample size; if the shape of the plot resembles an inverted funnel, it can be stated that there is no evidence of publication bias within the systematic review).

Heterogeneity: the variability between studies in terms of key characteristics (i.e. ecological variables) quality (i.e. methodology) or effect (i.e. results). Statistical tests of heterogeneity may be used to assess whether the observed variability in effect size (i.e. study results) is greater than that expected to occur purely by chance.

Intervention: the policy or management action under scrutiny within the systematic review.

Mean difference: the difference between the means of two groups of measurements.

Meta-analysis: a quantitative method employing statistical techniques, to combine and summarise the results of studies that address the same question.

Meta-regression: A multivariable model investigating effect size from individual studies, generally weighted by sample size, as a function of various study characteristics (i.e. to investigate whether study characteristics are influencing effect size).

Outcome: the effect of the intervention in a form that can be reliably measured.

Power: the ability to demonstrate an association where one exists (i.e. the larger the sample size, the greater the power and the lower the probability of the association remaining undetected).

Precision: the proportion of relevant articles identified by a search strategy as a percent of all articles found (i.e. a measure of the ability of a search strategy to exclude irrelevant articles).

Protocol: the set of steps to be followed in a systematic review. It describes the rationale for the review, the objective(s), 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: the possible result of an unsystematic approach to a review (e.g. research that generates a negative result is less likely to be published than that with a positive result, and this may therefore give a misleading assessment of the impact of an intervention). Publication bias can be examined via a funnel plot.

Random effects model: a mathematical model for combining the results of studies that allow for variation in the effect of the intervention amongst the subject populations studied. Both within-study variation and between-study variation is included when assessing the uncertainty of results (in contrast to a fixed effects model).

Review: an article that summarises a number of primary studies and discusses the effectiveness of a particular intervention. It may or may not be a systematic review.

Search strategy: an *a priori* description of the methodology, to be used to locate and identify research articles pertinent to a systematic review, as specified within the relevant protocol. It includes a list of search terms, based on the subject, intervention and outcome of the review, to be used when searching electronic databases, websites, reference lists and when engaging with personal contacts. If required, the strategy may be modified once the search has commenced.

Sensitivity: the proportion of relevant articles identified by a search strategy as a percentage of all relevant articles on a given topic (i.e. the degree of comprehensiveness of the search strategy and its ability to identify all relevant articles on a subject).

Sensitivity analysis: repetition of the analysis using different sets of assumptions (with regard to the methodology or data) in order to determine the impact of variation arising from these assumptions, or uncertain decisions, on the results of a systematic review.

Standardised mean difference (SMD): an effect size measure used when studies have measured the same outcome using different scales. The mean difference is divided by an estimate of the within-group variance to produce a standardised value without units.

Study quality: the degree to which a study seeks to minimise bias.

Subgroup analysis: used to determine if the effects of an intervention vary between subgroups in the systematic review. Subgroups may be pre-defined according to differences in subject populations, intervention, outcome and study design.

Subject: the unit of study to which the intervention is to be applied.

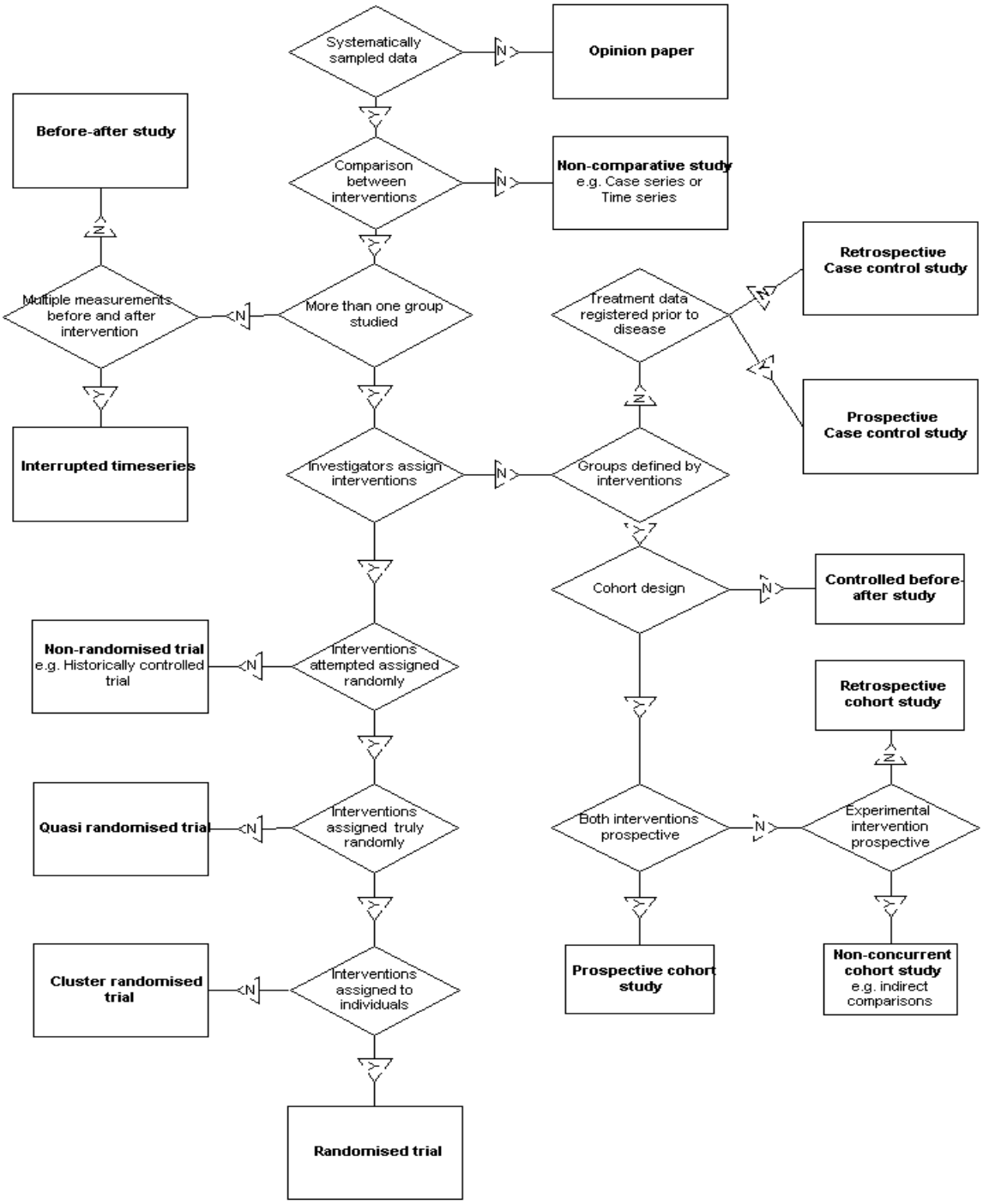
Summary effect size: the pooled effect size, generated by combining individual effect sizes in a meta-analysis.

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 within the review. Statistical methods (meta-analysis) may or may not be used to analyse and summarise the results of the included studies.

Weighted mean difference (WMD): a summary effect size measure for continuous data where studies that have measured the outcome on the same scale have been pooled.

Appendix B:

Figure: Design algorithm for studies of health care interventions



Appendix C: Explanation of key quality criteria for randomised controlled trials

1. Randomisation Method

The process of assigning participants to groups such that each participant has a known and usually an equal chance of being assigned to any given group. The term 'random' is often used inappropriately in the literature to describe non-random, 'deterministic' allocation methods, such as alternation, hospital numbers, or date of birth. Randomisation is intended to prevent performance and ascertainment bias, since group assignment cannot be predicted, and to limit selection bias by increasing the probability that important, but unmeasured, prognostic influences are evenly distributed across groups.

2. Concealment of Randomisation

A technique used to prevent selection bias by concealing the allocation sequence from those assigning participants to intervention groups, until the moment of assignment. Allocation concealment prevents researchers from (unconsciously or otherwise) influencing which participants are assigned to a given intervention group. There is strong empirical evidence that studies with inadequate allocation concealment yield larger estimates of treatment effects (on average, by 30-40%) than trials incorporating adequate concealment (Schulz et al., 1995).

3-6. Blinding

The practice of keeping study participants, health care providers, and sometimes those collecting and analysing clinical data unaware of the assigned intervention, so that they will not be influenced by that knowledge. Blinding is important to prevent performance and ascertainment bias at various stages of a study.

Blinding of patients and health care providers prevents performance bias. This type of bias can occur if additional therapeutic interventions (sometimes called co-interventions) are provided or sought preferentially by participants in one of the comparison groups.

Blinding of patients, health care providers, and other persons involved in evaluating outcomes, minimises the risk for ascertainment bias. This type of bias arises if the knowledge of a patient's assignment influences the process of outcome assessment. For example, in a placebo-controlled multiple sclerosis trial, assessments by unblinded, but not blinded, neurologists showed an apparent benefit of the intervention (Noseworthy et al., 1994). Finally, blinding of the data analyst can also prevent bias. Knowledge of the interventions received may influence the choice of analytical strategies and methods (Gøtzsche, 1996).

7. Blinding Check

Trying to create blind conditions is no guarantee of blindness, and it should be checked in order to assess the potential for performance and ascertainment bias. Questionnaire can be used for patients, care givers, outcome assessors and analysts; the (early) timing of checking the success of blinding is critical because the intervention effect may be the cause of unblinding, in which case it may be used as an outcome measure.

8. Baseline Comparability

The study groups should be compared at baseline for important demographic and clinical characteristics. Although proper random assignment prevents selection bias, it does not guarantee that the groups are equivalent at baseline. Any differences in baseline characteristics are the result of chance rather than bias, but these chance differences can affect the results and weaken the trial's credibility - stratification protects against such imbalances. Despite many warnings of their inappropriateness (e.g. Altman & Doré, 1990) significance tests of baseline differences are still common. Thus, it is inappropriate for authors to state that there were no significant baseline

differences between groups, not least because small, but non-significant, differences at baseline can lead to significant differences post-intervention. Adjustment for variables because they differ significantly at baseline is likely to bias the estimated treatment effect (Bender & Grouven, 1996).

9. Sample Size Calculation

For scientific and ethical reasons, the sample size for a trial needs to be planned in advance. A study should be large enough to have a high probability (power) of detecting, as statistically significant, a clinically important difference of a given size if such a difference exists. The size of effect deemed important is inversely related to the sample size necessary to detect it, i.e. large samples are necessary to detect small differences. Reports of studies with small samples frequently include the erroneous conclusion that the intervention groups do not differ, when too few patients were studied to make such a claim (Altman & Bland, 1995). In reality, small but clinically meaningful differences are likely, but these differences require large trials to be detected (Yusuf, Collins & Peto, 1984).

10. Attrition Rate

Participant attrition during the research process is almost inevitable. Attrition may not be too problematic so long as the level of attrition is not too high (<20%, see 14) and the attrition rate is similar between groups. Systematic differences between groups in the loss of participants from the study is problematic, insofar as non-random differences in attrition after allocation may reflect dissatisfaction, usually with the treatment intervention, e.g. unpleasant, inconvenient, ineffective, etc. Papers should report the attrition rate for each group and, where possible, reasons for attrition.

11. Treatment Comparability

The ability to draw causal inferences is dependent upon study groups receiving identical treatment other than the named intervention. This is much easier in pharmacological studies (e.g. placebo) than in behavioural studies. However, difficulty is no reason for neglect, and in practice many behavioural interventions deal very poorly with this issue. The only difference in participants' contact with the study should be the content of the intervention. Thus, efforts should be made to ensure control participants have the same amount and frequency of contact with the same intervention staff as do intervention group participants. Studies should also assess whether participants sort additional interventions (e.g. smokers in cessation studies often purchase nicotine replacement therapy to use to support their cessation attempt), and the extent to which there was potential for cross-group contamination, i.e. knowledge of the alternative treatment.

12. Intention-To-Treat Analysis

A strategy for analysing data in which all participants are included in the group to which they were assigned, irrespective of whether they completed the study. Excluding participants from the analysis (i.e. failure to use ITT analysis) can lead to erroneous conclusions, e.g. the intervention is effective, when in reality it isn't. Including all participants who started the study in the final analysis provides a conservative estimate of effect. ITT analysis is generally favoured because it avoids bias associated with non-random loss of participants (Lachin, 2000).

13. Outcomes and Estimation

Study results, for each outcome, should be reported as a summary of the outcome in each group (e.g., the proportion of participants with or without the event, or the mean and standard deviation of measurements) together with the effect size (e.g. the contrast between the groups). Confidence intervals should be presented for the contrast between groups, in order to indicate the precision (uncertainty) of the effect size estimate. The use of confidence intervals is especially valuable in relation to non-significant differences, for which they often indicate that the result does not rule out an important clinical difference (Gardner & Altman, 1986).

14. Adequacy of Follow-up

Refers to the number of participants who entered the study and provide data at all follow-ups. Note that, within the same study, loss at follow-up may differ for different outcomes and / or time points. Failure to complete a study usually indicates negative outcomes experienced by the participant. Without this information intervention effects may be interpreted as positive, when in reality many participants may find it unacceptable. A study can be regarded as having inadequate follow-up if outcome data is provided by less than 80% of participants who started the study.

References:

Altman, D.G. and Bland, J.M. (1995). Absence of evidence is not evidence of absence. *BMJ*, 311:485

Altman, D.G. and Doré, C.J. (1990). Randomisation and baseline comparisons in clinical trials. *Lancet*, 335:149-53.

Bender, R. and Grouven, U. (1996). Logistic regression models used in medical research are poorly presented. *BMJ*, 313:628.

Gardner, M.J. and Altman, D.G. (1986). Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ*, 292:746-50.

Gøtzsche, P.C. (1996). Blinding during data analysis and writing of manuscripts. *Control Clin Trials*, 17:285-90.

Lachin, J.L. (2000). Statistical considerations in the intent-to-treat principle. *Control Clin Trials*, 21:526

Noseworthy, J.H., Ebers, G.C., Vandervoort, M.K., Farquhar, R.E., Yetisir, E., and Roberts, R. (1994). The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology*, 44:16-20

Yusuf, S., Collins, R., and Peto, R. (1984). Why do we need some large, simple randomized trials? *Stat Med*, 3:409-22.