

H A N D B O O K

Systematic Reviews of Health Promotion and Public Health Interventions

TRAIN-THE-TRAINER
HANDBOOK

Two-day training
workshop

May 2005



Cochrane Health Promotion
and Public Health Field
www.vichealth.vic.gov.au/cochrane



A Public Health Education and
Research Project funded by the
Australian Government
Department of Health and
Ageing

Author

Nicki Jackson
Cochrane Health Promotion and Public Health Field
Victorian Health Promotion Foundation

Acknowledgements

The production of this handbook was funded by a grant from the Australian Government Public Health Education and Research Program (PHERP).

Thanks to those listed below for providing comments on drafts of the material included in the Handbook:

Professor Elizabeth Waters
Director, Cochrane Health Promotion and Public Health Field, Chair in Public Health, Deakin University

Dr Celia McMichael
School of Public Health, La Trobe University

Dr Lucie Rychetnik
Sydney Health Projects Group, The University of Sydney

John Bennett
Project coordinator, School of Public Health, The University of Sydney

The author has also utilised the work conducted by the Guidelines for Systematic Reviews in Health Promotion and Public Health. This taskforce includes:

Anderson L. (Centers for Disease Control and Prevention, USA); Bailie R. (Menzies School of Health Research and Flinders University NT Clinical School, Australia); Brunton G. (Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre, UK); Hawe P. (University of Calgary, Canada); Jackson N. (Cochrane Health Promotion and Public Health Field, Australia); Kristjansson E. (University of Ottawa, Canada); Naccarella L. (University of Melbourne, Australia); Norris S. (Agency for Healthcare Research and Quality, USA); Oliver S. (EPPI-Centre, UK); Petticrew M. (MRC Social and Public Health Sciences Unit, UK); Pienaar E. (South African Cochrane Centre); Popay J. (Lancaster University, UK); Roberts H. (City University, UK); Rogers W. (Flinders University, Australia); Shepherd J. (University of Southampton, UK); Sowden A. (Centre for Reviews and Dissemination, University of York, UK); Thomas H. (McMaster University and the Effective Public Health Practice Project, Canada); Waters E. (Cochrane Health Promotion and Public Health Field and Deakin University, Australia).

Copyright

The copyright for the course material lies with Deakin University and the Australian Department of Health and Aging. The course materials may be reproduced and used to conduct non-profit systematic review courses for the Australian public health workforce. The materials should not be used for any commercial or profit-making activity unless specific permission is granted by the copyright owners.

Contents

| | |
|---|-----|
| Introduction | 1 |
| Administration of two-day workshop | 3 |
| Overview of two-day workshop | 5 |
| FACILITATOR'S GUIDE – INTRODUCTION | 9 |
| Unit One: Background to Systematic Reviews..... | 11 |
| FACILITATOR'S GUIDE | 15 |
| Unit Two: International Systematic Review Initiatives | 17 |
| FACILITATOR'S GUIDE | 19 |
| Unit Three: Resources Required..... | 21 |
| FACILITATOR'S GUIDE..... | 25 |
| Unit Four: Developing a Protocol | 27 |
| FACILITATOR'S GUIDE | 31 |
| Unit Five: Asking an Answerable Question..... | 33 |
| FACILITATOR'S GUIDE | 41 |
| Unit Six: Finding The Evidence..... | 43 |
| FACILITATOR'S GUIDE | 61 |
| Unit Seven: Data Abstraction | 63 |
| FACILITATOR'S GUIDE | 65 |
| Unit Eight: Principles of Critical Appraisal..... | 67 |
| FACILITATOR'S GUIDE – QUANTITATIVE STUDIES..... | 91 |
| FACILITATOR'S GUIDE – QUALITITATIVE STUDIES | 109 |
| Unit Nine: Synthesising the Evidence | 115 |
| FACILITATOR'S GUIDE | 119 |
| Unit Ten: Interpretation of Results | 121 |
| FACILITATOR'S GUIDE | 129 |
| Unit Eleven: Writing the Systematic Review | 133 |
| FACILITATOR'S GUIDE | 137 |

Introduction

Welcome to the workshop entitled “Systematic reviews of health promotion and public health interventions – “train the trainer””.

This workshop will take you through the complete process of preparing and carrying out a two-day workshop on conducting systematic reviews of public health and health promotion interventions.

This training handbook describes the steps of the systematic review process and provides additional teaching instructions. The handbook, however, is not intended to be used as a single resource for teaching others how to conduct reviews. The additional reading is very important to trainers who are experienced in conducting systematic reviews.

Note: This handbook/workshop is useful for both Cochrane reviewers and reviewers who are completing a systematic review for their workplace, studies, etc. If reviewers wish to complete a Cochrane review, they should visit www.cochrane.org (About us – Contact: Groups and Centres) to find the appropriate Collaborative Review Group to register their interest or contact the Cochrane Health Promotion and Public Health Field for further information cochrane@vichealth.vic.gov.au.

The trainer’s guide provides information on the time taken for each unit. This is only an approximation – the time it takes is dictated by the prior knowledge/experience of the participants and the number of questions. The guide does, however, give you an indication of the time for each unit if the course was to be run in modular format.

Please note: It is likely that every facilitator will have different levels of knowledge and skills. Facilitators should draw on their own experiences when teaching this course. Feel free to adapt any exercises or slides as you see fit.

Overall learning outcomes

This handbook will enable you to learn how to effectively present on the following components of the systematic review process:

- Key challenges of conducting systematic reviews of health promotion and public health interventions
- Formulation of an answerable question about the effectiveness of interventions in health promotion and public health
- Identifying primary studies, including developing evidence-based strategies for searching electronic databases
- Evaluating the quality of both a systematic review and an individual health promotion or public health study
- Synthesising the body of evidence from primary studies
- Formulating conclusions and recommendations from the body of evidence

Additional reading:**Textbooks:**

Oliver S, Peersman P. Using Research for Effective Health Promotion. Open University Press, UK. 2001.

Brownson R, Baker E, Leet T, Gillespie K. Evidence-based Public Health. Oxford University Press, USA. 2003.

Egger M, Smith G, Altman D. Systematic Reviews in Health Care: Meta-analysis in context. BMJ Publishing Group, UK. 2001.

Manuals / Handbooks:

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

Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. <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. March 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

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

Administration of two-day workshop

1) PLANNING AHEAD

Administrative support:

- The course will need to be organised (advertise, receive registrations, find and book venue, receive payments, etc) and course materials will need to be prepared. This may take up to 10 days.

Co-facilitator:

- It is recommended, although not essential, that two facilitators conduct the two-day course (if the days are held consecutively). One facilitator may be adequate if the course is run in a modular format. Up to twenty participants is workable for this workshop (it will also depend on the number of computers available).

Venue:

- You will require a room to hold up to 20 participants, with participants sitting in groups (preferably in groups of 4) around tables.
- A computer training facility, with internet access (you will need a password to access OVID Medline), is required for one half-day.
- Organise payment for venues (if required)
- Familiarise yourself with the facilities of the venue and computer room (air-conditioning, lighting, projector, tea and coffee facilities, toilets, parking, etc)

Costing:

- Determine whether you need to pay for venue hire (especially computer facilities), catering, and printing. Such costs should be recouped by charging participants an administrative fee.

Publicity:

- A draft flyer has been supplied for you on CD to modify as required.

Registration:

- You will need email or postal addresses of all participants in order to send pre-reading materials. Additionally, you may collect such information such as job title, contact details, and prior experience (and food preferences).

Invoicing:

- If participants are required to pay for the course, they will require an invoice.

Catering:

- It is recommended that morning tea, lunch, and afternoon tea are provided, in addition to coffee, tea, and water. You should check food preferences prior to ordering the catering.

Teaching aids:

- If you choose to use the powerpoint slides to deliver the training you will require a computer and projector facilities. The slides may be printed onto overhead transparencies if you do not have projector facilities. Butcher's paper and pens are also required to write down the 4 G's (and refer back to during the course).

2) ONCE REGISTRATIONS HAVE BEEN RECEIVED

Confirmations:

- Email participants to confirm their registration has been received and that they will receive some pre-reading material at least 1 week (preferably 2 weeks) prior to the course.
- Organise name tags.
- You will need to divide the group in half – half appraising the RCT (Sahota), half appraising the controlled before and after study (Gortmaker).
- Send the RCT group the critical appraisal paper (at least one week prior to the course)
- Send the CBA group the critical appraisal paper (at least one week prior to the course)
- Send all participants the Qualitative critical appraisal paper (Cass), at least one week prior to the course.

Printing course materials:

The CD provided contains a pdf file for each unit of the course manual – both outline and powerpoint slides.

- Each unit pdf file should be printed double-sided and then combined to be bound together.
- *** Please note: If you make any changes to the course outline or slides you will need to redo the pdfing for the units you change.
- Simply place a coloured piece of paper/divider at the end of each unit to distinguish between units (and before the first unit).
 - You will also need to print out the evaluation sheet and course certificates (provided).

3) ON THE DAY

You will require:

- Two-day systematic review course slides.ppt
- Who wants to be a systematic reviewer.ppt
- Name tags
- Any resources needed for icebreakers or games
- Course manuals
- Certificates of achievement
- Evaluation forms

IMPORTANT: The copyright for the course materials lies with the authors and the Australian Government Department for Health and Ageing. The course materials may be reproduced and used to conduct non-profit systematic review courses for the Australian public health workforce (including students). The materials should not be used for any commercial or profit-making activity unless specific permission is granted by the copyright owners.

Overview of two-day workshop

DAY ONE:

- Background to systematic reviews
- International systematic review initiatives
- Resources required to complete a systematic review
- Setting the scope of your review
- Asking an answerable question
- Data abstraction
- Principles of critical appraisal

DAY TWO:

- Finding the evidence
- Synthesising the evidence
- Interpretation of results
- Writing the systematic review

Note: If computer facilities are at the same venue as the teaching, finding the evidence should be moved to the afternoon of the first day; the first session of day two would be critical appraisal.

OUTLINE OF TWO-DAY WORKSHOP

| | Day One | Exercise | Day Two | Exercise |
|-----------|---|---|---|-----------------------------------|
| 9:00am | Introductions Overview of workshop | | Unit Six: Introduction to searching (theory) | |
| 9:30am | Unit 1: Background to systematic reviews Unit 2: International systematic review initiatives | Comparing PH/HP interventions | 9:45am Unit 6: Finding the evidence (practice) | Finding the evidence exercise |
| 10:30am | <i>Morning tea</i> | | | |
| 10:45am | Unit 3: Resources required Unit 4: Setting the scope | Advisory group exercise | | |
| 11:30am | Unit 5: Asking an answerable question | Asking an answerable question | | |
| 12:15pm | <i>Lunch</i> | | <i>Lunch</i> | |
| 1:15pm | Unit 7: Data abstraction Unit 8: Principles of critical appraisal (quantitative) | Appraisal exercise (quantitative - CBA and RCT) | Unit 9: Synthesis of evidence | Quiz Discuss narrative reviews |
| | | | Unit 10: Interpretation of results | |
| 3:00pm | <i>Afternoon tea</i> | | <i>Afternoon tea</i> | |
| 3:15pm | Unit 8: Principles of critical appraisal (qualitative) | Appraisal exercise (qualitative) | Unit 11: Writing a systematic review | Appraisal exercise |
| 4:25-4:30 | Review Day One | | Review Day Two and two-day workshop | |

FACILITATOR'S GUIDE - INTRODUCTION

Time required

Approximately 30 minutes

Instructions to facilitator

- 1) Distribute course materials and name tags to participants.
- 2) Trainer introduction: Introduce yourself (and other facilitators if appropriate) and detail your background and experience. Alternatively, you may participate in the group introduction/icebreaker.
- 3) Participant introductions/Icebreakers: There are many choices when it comes to icebreakers. You may have your own preferences.
- 4) Participants' expectations: Ask the group to openly feedback the four 'G's'. You should write these down on butcher's paper or on a whiteboard (or transparency) so you can regularly refer to them during the 2-day course and assess if the workshop is meeting their needs.
 - Gives (what participants can give to the workshop)
 - Gains (what they hope to gain from the workshop)
 - Ghastlies (what they hope does not happen in the workshop (eg. too simple, too advanced, not relevant, etc)
 - Ground rules (what rules can the group agree on (eg. one person talk at a time, no single person to dominate discussion, etc)).
- 5) Discuss course objectives and outline of the two-day workshop.
- 6) Address housekeeping issues – toilets, breaks, coffee/tea/water, any OH&S issues, etc.

Unit One: Background to Systematic Reviews

Learning Objectives

- 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 promotion and public health interventions

Types of reviews

Generally, reviews may be grouped into the following two categories (see Table One):

- 1) Traditional literature reviews/narrative reviews
- 2) Systematic reviews (with or without) meta-analysis

Narrative or traditional literature 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.² Processes such as searching, quality appraisal and data synthesis are not usually described and as such, they are very prone to bias. Although an advantage of these reviews is that they are often conducted by 'experts' who may have a thorough knowledge of the research field, but they are disadvantaged in that the authors may have preconceived notions or biases and may overestimate the value of some studies.³

Note: A narrative review is not to be confused with a narrative systematic review – the latter refers to the type of synthesis of studies (see Unit Nine).

Systematic review

Many of the tools of systematic research synthesis were developed by American social scientists in the 1960s.⁴ However, today's systematic evidence reviews are very much driven by the evidence-based medicine movement, in particular, from the methods developed by the Cochrane Collaboration. 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."¹

What is a meta-analysis?

"A meta-analysis is the statistical combination of at least 2 studies to produce a single estimate of the effect of the health care intervention under consideration."² Note: a meta-analysis is simply the statistical combination of results from studies – the final estimate of effect may not always be the result of a systematic review of the literature. Therefore, it should not be considered as a type of review.

Table One. Comparing different types of reviews

| Review | Characteristics | Uses | Limitations |
|---|--|---|---|
| Traditional literature review / narrative review | Describes and appraises previous work but does not describe specific methods by which the reviewed studies were identified, selected and evaluated | <p>Overviews, discussions, critiques of previous work and the current gaps in knowledge</p> <p>Often used as rationale for new research</p> <p>To scope the types of interventions available to include in a review</p> | <p>The writers assumptions and agenda often unknown</p> <p>Biases that occur in selecting and assessing the literature are unknown</p> <p>Cannot be replicated</p> |
| Systematic review | <p>The scope of the review is identified in advance (eg review question and sub-questions and/or sub-group analyses to be undertaken)</p> <p>Comprehensive search to find all relevant studies</p> <p>Use of explicit criteria to include / exclude studies</p> <p>Application of established standards to critically appraise study quality</p> <p>Explicit methods of extracting and synthesising study findings</p> | <p>Identifies, appraises and synthesises <i>all</i> available research that is relevant to a particular review question</p> <p>Collates all that is known on a given topic and identifies the basis of that knowledge</p> <p>Comprehensive report using explicit processes so that rationale, assumptions and methods are open to scrutiny by external parties</p> <p>Can be replicated / updated</p> | <p>Systematic reviews with narrowly defined review questions provide specific answers to specific questions</p> <p>Alternative questions that have not been answered usually need to be reconstructed by the reader</p> |

Advantages of systematic reviews

- Reduces bias
- Replicable
- Resolves controversy between conflicting findings
- Provides reliable basis for decision making

Reviews of clinical interventions vs. reviews of public health interventions

Some of the key challenges presented by the health promotion and public health field are a focus or emphasis on;

- populations and communities rather than individuals;
- combinations of strategies rather than single interventions;
- processes as well as outcomes;
- involvement of community members in program design and evaluation;
- health promotion theories and beliefs;
- the use of qualitative as well as quantitative approaches to research and evaluation;
- the complexity and long-term nature of health promotion intervention outcomes.⁵

REFERENCES

1. 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. March 2001.
2. Klassen TP, Jadad AR, Moher D. Guides for Reading and Interpreting Systematic Reviews. 1. Getting Started. Arch Pediatr Adolesc Med 1998;152:700-704
3. 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
4. Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. Eval Health Prof 2002;25:12-37.
5. Jackson SF, Edwards RK, Kahan B, Goodstadt M. An Assessment of the Methods and Concepts Used to Synthesize the Evidence of Effectiveness in Health Promotion: A Review of 17 Initiatives. <http://www.utoronto.ca/chp/chp/consort/synthesisfinalreport.pdf>

ADDITIONAL READING

Mulrow CD. Systematic reviews: Rationale for systematic reviews. BMJ 1994;309:597-599.

McQueen D. The evidence debate. J Epidemiol Community Health 2002;56:83-84.

Petticrew M. Why certain systematic reviews reach uncertain conclusions. BMJ 2003;326:756-8.

Petticrew M. Systematic reviews from astronomy to zoology: myths and misconceptions. BMJ 2001;322:98-101.

Grimshaw JM, Freemantle N, Langhorne P, Song F. Complexity and systematic reviews: report to the US Congress Office of Technology Assessment. Washington, DC: Office of Technology Assessment, 1995.

Rychetnik L, Hawe P, Waters E, Barratt A, Frommer M. A glossary for evidence based public health. J Epidemiol Community Health 2004;58:538-45.

EXERCISE

1. In pairs, discuss some of the differences (**using examples**) between reviews of clinical/medical/pharmaceutical interventions *vs.* reviews of health promotion or public health interventions.

Examples

Clinical
E.g. effectiveness of antibiotics for sore throat

Health promotion/public health

E.g. effectiveness of mass media interventions for preventing smoking in young people

| Clinical/medical/pharmaceutical | Health promotion/public health |
|--|--------------------------------|
| 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 (consider external environment (social, political, cultural) and internal factors (training of those implementing intervention, literacy of population, access to services, etc)) | |
| | |
| | |
| | |

FACILITATOR'S GUIDE

Time required:

45 minutes approx. (15minutes slides, 30 minutes exercise)

Learning topics

- Basic systematic review terminology
- Different types of reviews
- Differences between clinical reviews and HP/PH reviews

Summary of activity

Lecture – Powerpoint slides

Group exercise – comparing clinical interventions to health promotion interventions

Description of supporting materials

No supporting materials

Further reading (if required)

1. Klassen TP, Jadad AR, Moher D. Guides for Reading and Interpreting Systematic Reviews. 1. Getting Started. Arch Pediatr Adolesc Med 1998;152:700-704
2. Mulrow CD. Systematic reviews: Rationale for systematic reviews. BMJ 1994;309:597-599.
3. McQueen D. The evidence debate. J Epidemiol Community Health 2002;56:83-84.
4. Petticrew M. Why certain systematic reviews reach uncertain conclusions. BMJ 2003;326:756-8.
5. Petticrew M. Systematic reviews from astronomy to zoology: myths and misconceptions. BMJ 2001;322:98-101.

Instructions for facilitator

Slide presentation

Points to emphasise:

1. Not to confuse narrative reviews with narrative systematic reviews (the latter refers to the type of synthesis within a systematic review).
2. Narrative reviews are useful – they can be used in the scoping stages of a systematic review to determine the different options for interventions that might be included in a review.
3. Systematic reviews may also produce conflicting results – especially when different methodological criteria are used, eg different study designs are included, different quality checklists, different inclusion and exclusion criteria, different databases searched etc.
4. Meta-analysis refers simply to the statistical combination of results from studies. The results produced are not necessarily the result of a systematic review of the evidence.

Exercise

1. Ask participants to break into pairs or small groups (3-4 people), and discuss (using an example determined by the pair/group) the differences between clinical and health promotion interventions.

2. Visit each group to make sure they understand the exercise.
3. Ask a sample of the pairs/groups to report back to the larger group.
4. Look for the following differences (clinical vs. health promotion interventions):
 - a) Patients vs. populations/communities/groups
 - b) Single, specific interventions (surgery, drugs, medical treatment) vs. complex multidisciplinary interventions
 - c) Immediate outcomes vs. changes in attitude, behaviour or long term (eg. BMI)
 - d) Patients not involved vs. community participation in development and evaluation
 - e) Little impact of context vs. potentially large impact of context, population factors and training

Example:

Clinical

Example: Epidural versus non-epidural analgesia for pain relief in labour:

- 1) Participants: women patients giving birth
- 2) Types of interventions: Epidural or non-epidural analgesia
- 3) Types of outcomes: Pain relief in labour
- 4) Participants involved in design on intervention: Generally not involved
- 5) Potential influences on intervention success: Training of person administering epidural

Health promotion or public health

Example: Community interventions for reducing smoking among adults

- 1) Participants: Any adults, male or female
- 2) Types of interventions: Co-ordinated multidimensional program – through schools, workplaces, health professionals, health departments, restaurants, hospitals, retailers. Types of interventions – mass media, policy, counselling, education, etc.

3) Types of outcomes:

Long term: Morbidity and mortality.

Intermediate measures: Biochemical measure of smoking, self-reported smoking status, self-reported cigarette consumption.

Proxy measures and mediating variables: knowledge of harms of smoking, attitudes to smoking, intentions to quit. Process outcomes: Evaluation of the implementation of interventions.

4) Participants involved in design on intervention: Community members usually involved in planning and implementation

5) Potential influences on intervention success: literacy levels of community, training of those implementing the intervention, characteristics of the community – urban/rural, context in which policy was implemented, strength of the organisation implementing the intervention, norms of smoking, etc.

Unit Two: International Systematic Review Initiatives

Learning Objective

- *To be familiar with international groups conducting systematic reviews of the effectiveness of public health and health promotion interventions*

There are a number of groups around the world conducting systematic reviews of public health and health promotion interventions. Reviews are often published on the group's internet website, and follow guidelines/methods developed by the individual organisation. It is useful to visit each of the organisations listed below to view the different styles of systematic reviews. Reviewers seeking to conduct a Cochrane Review should visit the Cochrane website for more information (<http://www.cochrane.org>) or contact the Cochrane Health Promotion and Public Health Field (<http://www.vichealth.vic.gov.au/cochrane/>).

Useful websites of systematic review initiatives:

1. The Cochrane Collaboration – The Cochrane Library:
<http://www.thecochranelibrary.com>

Reviews relevant to health promotion and public health are listed on the Cochrane Health Promotion and Public Health Field website:

<http://www.vichealth.vic.gov.au/cochrane>

2. Guide to Community Preventive Services:
<http://www.thecommunityguide.org>
3. The Evidence for Practice Information and Co-ordinating Centre (EPPI-Centre):
<http://eppi.ioe.ac.uk/>
4. Effective Public Health Practice Project:
<http://www.city.hamilton.on.ca/PHCS/EPHPP/EPHPPResearch.asp>
5. Health Development Agency (HDA):
<http://www.hda-online.org.uk/html/research/effectiveness.html>
Note: These reviews are systematic reviews of systematic reviews (not reviews of individual primary studies).
6. Centre for Reviews and Dissemination:
<http://www.york.ac.uk/inst/crd/>
7. The Campbell Collaboration
<http://www.campbellcollaboration.org/>

ADDITIONAL READING

Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. *Eval Health Prof* 2002;25(1):116-29.

EXERCISE

1. In your own time visit each of the above websites:
 1. Try to read at least one review from each organisation
 2. Compare the different styles of reviews
 3. Try to locate any methodological work or guidelines relating to the conduct of each review (i.e. Methods papers for The Community Guide)

FACILITATOR'S GUIDE

Time required:

15 minutes approx (slides only)

Learning topics

- International sources of reviews of health promotion and public health topics
- The Cochrane Collaboration

Summary of activity

Lecture – Powerpoint slides.

Any participants who raise questions about each organisation can visit the appropriate website for more information (they may choose to do this within the 'Finding the 'Evidence' session if they have time).

Description of supporting materials

No supporting materials

Further reading (if required)

1. Take time to familiarise yourself with each of the websites highlighted in this section. If you have time, read at least one review from each website.
 - Cochrane Collaboration – reviews of healthcare interventions (including health promotion and public health)
 - The Community Guide – excellent source for health promotion and public health reviews
 - Effective Public Health Practice Project – public health reviews
 - Health Development Agency – reviews of reviews (not reviews of primary studies) relating to public health and health promotion
 - EPPI-Centre – reviews relating to health promotion and education
 - Centre for Reviews and Dissemination – note that these databases are included within The Cochrane Library
 - Campbell Collaboration – reviews relating to educational, criminological, social and psychological interventions
2. Read “Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. *Eval Health Prof* 2002;25(1):116-29”.
3. Visit www.cochrane.org to find out more information about The Cochrane Collaboration.

Instructions to facilitator

Databases within The Cochrane Library (emphasise that the Cochrane Library is available to all Australians free of charge (as at 2004)):

- 1) Cochrane Systematic reviews: Cochrane reviews and Cochrane protocols (reviews in progress)
- 2) Database of Reviews of Effectiveness: Systematic reviews from around the world which have been appraised by the Centre for Reviews and Dissemination, UK.

- 3) Cochrane Central Register of Controlled Trials: Bibliography of controlled trials (RCTs and controlled trials) (some not indexed in MEDLINE). Useful source for locating primary studies.
- 4) Cochrane database of Methodology Reviews: Cochrane reviews of methodological studies, i.e. studies examining the methodologies relating to systematic reviews (eg. bias, randomisation, etc).
- 5) The Cochrane Methodology register: Bibliography of primary studies relating to methodological aspects of research synthesis.
- 6) About the Cochrane Collaboration: Information about Review Groups, Fields, Centres, etc. Contact details provided.
- 7) Health Technology Assessment Database: Health Technology Assessment reports from around the world.
- 8) NHS Economic evaluation database: Economic evaluations of health care interventions.

Unit Three: Resources Required

Learning Objective

- To be familiar with the resources required to conduct a systematic review

Conducting a systematic review can be a time-consuming task. Ideally, a minimum of six months is required to complete a review (full-time). However, there will be times which are less busy, for example, when awaiting the retrieval of full-text articles. The following list outlines the requirements to complete a systematic review:

- Topic of relevance or interest
- Team of co-authors (to reduce bias)
- Training and support
- Access to/understanding of the likely users of the review
- Funding
- Time
- Access to electronic searching databases and the internet (for unpublished literature)
- Statistical software (if appropriate)
- Bibliographic software (eg. Endnote)
- Word processing software

The Cochrane Collaboration software, RevMan (abbreviation for Review Manager), can be used for both the text of the review and meta-analysis, and can be downloaded for free from <http://www.ccsims.net/RevMan>.

Time

Although no research has been completed on the overall time it takes to complete a health promotion or public health systematic review, we are given some insight from an analysis of 37 medically-related meta-analyses¹. The analysis by Allen and Olkin¹ found that the average hours for a review were 1139 (~6 months), but ranged from 216 to 2518 hours.

The component mean times were:

| | |
|-----------|---|
| 588 hours | Protocol development, searches, retrieval, abstract management, paper screening and blinding, data extraction and quality scoring, data entry |
| 144 hours | Statistical analysis |
| 206 hours | Report and manuscript writing |
| 201 hours | Other (administrative) |

There was an observed association between the number of initial citations (before exclusion criteria are applied) and the total time it takes to complete a meta-analysis.

Note: The time it takes to complete a health promotion and public health review may be longer due to less standardised definitions (eg. concepts, language, terminology) for public health interventions compared to clinical interventions resulting in a larger number of citations to apply the inclusion and exclusion criteria.

- Searching

The EPPI-Centre² documented the time it took an experienced health promotion researcher in developing and implementing a Medline search strategy to identify sexual health promotion primary studies.

| | |
|----------|---|
| 40 hours | Developing and testing a sensitive search strategy for Medline |
| 8 hours | Implementing the search for the most recent Medline period available at the time (January 1996 to September 1997) and downloading citations |
| 7 hours | Scanning through the 1048 retrieved records |

If such a search strategy was to be implemented over the 30 years covered by Medline, the number of retrieved records would be around 10,000. Consequently, about 70 hours would be needed to identify the relevant citations for the review. Overall, this Medline search strategy would take approximately 120 hours.

A preliminary literature search and contact with relevant experts in the area might help assist in calculating the approximate time required to complete the review.

REFERENCES

1. Allen IE, Olkin I. Estimating Time to Conduct a Meta-analysis From Number of Citations Retrieved. JAMA 1999;282(7):634-5.
2. Evidence for Policy and Practice Information and Co-ordinating Centre. Research Report. Effectiveness Reviews in Health Promotion. 1999.

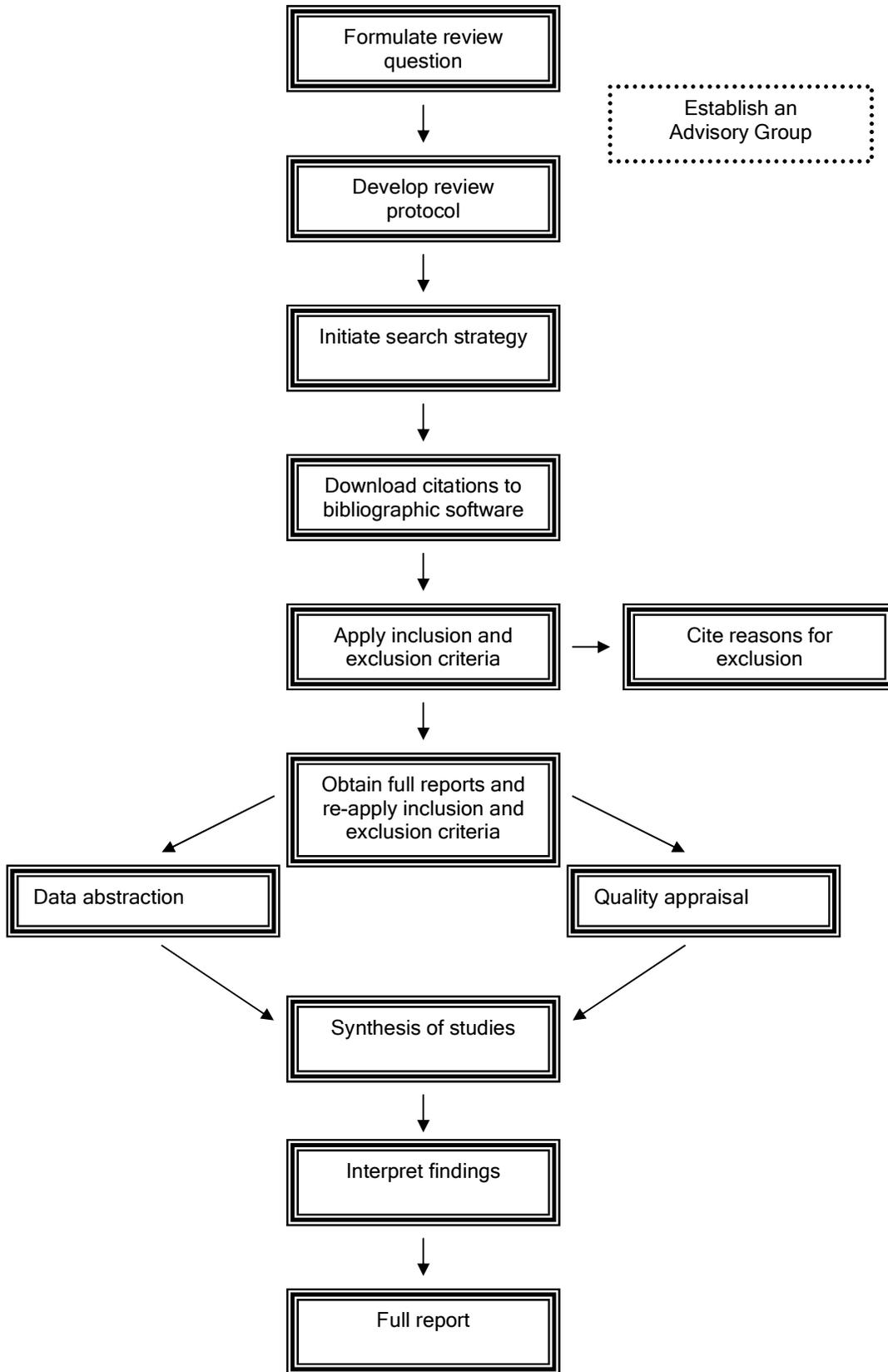


Figure One. Flow chart of a systematic review

FACILTATOR'S GUIDE

Time required:

5-10 minutes approx (slides only)

Learning topics

- Required resources to complete a systematic review

Summary of activity

Lecture – Powerpoint slides

Description of supporting materials

No supporting materials

Further reading (if required)

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

Instructions to facilitator

1. This section is very straightforward; simply emphasise that there are a number of resources available to assist reviewers with the systematic review process. All of the manuals available to reviewers can be obtained over the Internet. The Cochrane Reviewers' Handbook and the CRD Report are perhaps the most useful (in addition to the training handbook provided) as a single information source on the systematic review process.
2. Six months of full-time work is the minimum time required to complete a health promotion or public health review.
3. Bibliographic software is essential; all reviewers need access to such software.
4. Highlight that reviews should ideally be conducted by more than one reviewer, where possible. This reduces bias in many stages of the review.

Unit Four: Developing a Protocol

Learning Objectives

- To understand the rationale for documenting the review plan in the form of a structured protocol
- To understand the importance of setting the appropriate scope for the review

What is a protocol?

A protocol is the plan the reviewers wish to follow to complete the systematic review. It allows thinking to be focused and allocation of tasks to be determined. Methods to be used in the systematic review process must be determined at the outset. The Cochrane Reviewers' Handbook¹ states that "the reviewer's knowledge of the results of the study may influence:

- The definition of the systematic review
- The criteria for study selection
- The comparisons for analyses
- The outcomes to be reported in the review."

Furthermore, spending time at this stage preparing a clear protocol will reduce time spent during the systematic review process.

Information to include in the protocol

Examples of protocols (of Cochrane systematic reviews) can be found in The Cochrane Library (<http://www.thecochranelibrary.com>).

1) 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 and the results of any similar reviews conducted on the topic.

The background should also describe why, theoretically, the interventions under review might have an impact on potential recipients.

Reviewers may refer to a body of:

- empirical evidence such as similar interventions having an impact, or identical interventions having an impact on other populations.
- theoretical literature that justifies the possibility of effectiveness.

If reviewers choose to examine more proximal outcomes (knowledge and attitudes), theory should be used to explain the relationship to more distal outcomes (changes in behaviour).

2) Objectives

Reviewers will need to determine the scope of the review. The scope of a review refers to the type of question being asked and will affect the kind of studies that need to be reviewed, in terms of study topic, population and setting, and, of course, study design.²

The scope of the review should be based on how the results of the review will be used. It is useful to consult with the potential users of the review when determining the review's scope. For example, many health promotion practitioners and policy makers would find it more useful to have systematic

reviews of 'approaches' to health promotion (eg. community development or peer-delivered interventions), rather than topic-focused reviews (eg. healthy eating or accident prevention).

The scope is also likely to depend on how much time is available and the likely volume of research literature.

Lumping the review question, i.e. addressing a wide range of interventions (eg. prevention of injuries in children):

- likely to be time-consuming because of the searching and selecting processes
- will better inform decisions about which interventions to implement when there may be a range of options
- may be ultimately of more use to policy decisions

Splitting the review, i.e. addressing a narrow range of interventions, (eg. prevention of drowning in toddlers)

- may be less time-consuming
- will only inform decisions about whether or not to implement narrowly focused interventions
- may be more useful for practitioners

3) Pre-determined selection criteria

The selection criteria will be determined by the PICO(T) question, which is described in the following unit (Unit Five. Asking an Answerable Question). It is important to take an international perspective – do not restrict the inclusion criteria by nationality or language, if possible.¹

4) Planned search strategy

List the databases that are to be searched and if possible, document the search strategy including subject headings and textwords. Methods to identify unpublished literature should also be described (eg. handsearching, contact with authors, scanning reference lists, internet searching).

5) Planned data extraction

Reviewers should describe whether they are going to extract process, outcome and contextual data and state how many reviewers will be involved in the extraction process. The quality assessment checklists to be used for appraising the individual studies should also be specified at this stage.

6) Proposed method of synthesis of findings

Describe the methods to be used to synthesise the data. For example, reviewers of health promotion and public health interventions often tabulate the included studies and perform a narrative synthesis due to expected heterogeneity. It is worthwhile at this stage to consider the likely reasons for heterogeneity in the systematic review.

Establish an 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, in terms of both the topic and the methodology.² Gaining significant input from the potential users of the review will help bring about a review that is more meaningful, generalisable and potentially more accessible.

Preferably, advisory groups should include persons with methodological and subject/topic area expertise in addition to potential review users.

- Establish an Advisory Group whose members are familiar with the topic and include policy, funders, practitioners and potential recipients/consumers perspectives. Also include methodologists to assist in methodological questions.
- The broader the review, the broader the experience required of Advisory Group members.
- To ensure international relevance consult health professionals in developing countries to identify priority topics/outcomes/interventions on which reviews should be conducted.
- The Effective Public Health Practice Project has found that six members on an Advisory Group can cover all areas and is manageable.
- Develop Terms of Reference for the Advisory Group to ensure there is clarity about the task(s) required. Tasks may include:
 - making and refining decisions about the interventions of interest, the populations to be included, priorities for outcomes and, possibly, 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
- Develop job descriptions and person specifications for consumers and other advisors to clarify expectations. Further information, including how to involve vulnerable and marginalised people in research, is also available at www.invo.org.uk.

An example of the benefits of using an Advisory Group in the planning process

A review of HIV prevention for men who have sex with men (MSM) (<http://eppi.ioe.ac.uk/EPPIWebContent/hp/reports/MSM/MSMprotocol.pdf>) employed explicit consensus methods to shape the review with the help of practitioners, commissioners and researchers.

An Advisory Group was convened of people from research/academic, policy and service organisations and representatives from charities and organisations that have emerged from and speak on behalf of people living with, or affected by, HIV/AIDS. The group met three times over the course of the review.

The group was presented with background information about the proposed review; its scope, conceptual basis, aims, research questions, stages, methods. Discussion focused on the policy relevance and political background/context to the review; the inclusion criteria for literature (interventions, outcomes, sub-groups of MSM); dissemination strategies; and timescales. Two rounds of voting identified and prioritised outcomes for analysis. Open discussion identified sub-groups of vulnerable MSM. A framework for characterising interventions of interest was refined through Advisory Group discussions.

The review followed this guidance by adopting the identified interventions, populations and outcomes to refine the inclusion criteria, performing a meta-analysis as well as sub-group analyses. The subsequent product included synthesised evidence directly related to health inequalities.

REFERENCES

1. Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. <http://www.cochrane.org/resources/handbook/index.htm>
2. 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.

ADDITIONAL READING

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.

Hanley B, Bradburn J, Gorin S, et al. Involving Consumers in Research and Development in the NHS: briefing notes for researchers. Winchester: Help for Health Trust, 2000.

EXERCISE

1. Group exercise: Scenario: You wish to conduct a review on one of the following topics:
 - "Interventions for preventing tobacco sales to minors"
 - "Workplace interventions for smoking cessation"
 - "Post-licence driver education for the prevention of road traffic crashes"
 - "Primary prevention for alcohol misuse in young people"
 - "Support for breastfeeding mothers"
 - "Interventions aimed at improving immunisation rates"

Choose one review. Brainstorm, in small groups, who you might want to include in an Advisory Group for your chosen review. After brainstorming all potential members, try to restrict to 6-7 members. Remember to keep an international focus.

2. In your own time:

1. Search the Cochrane Library for protocols relevant to your area of interest.
2. Familiarise yourself with the essential components of a review protocol.

FACILITATOR'S GUIDE

Time required:

35-40 minutes approx (10 minutes slides, 25-30 minutes exercise)

Learning topics

- Rationale for completing a protocol for a systematic review
- Establishing Advisory Groups
- Determining the scope of the review

Summary of activity

Lecture – Powerpoint slides

Exercise in small groups to list potential Advisory Group members

Description of supporting materials

No supporting materials. However, you may wish to hand out examples of protocols for participants to look at. Protocols downloaded from the Cochrane Library can be used as relevant examples.

Further reading (if required)

1. 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.
2. Hanley B, Bradburn J, Gorin S, et al. Involving Consumers in Research and Development in the NHS: briefing notes for researchers. Winchester: Help for Health Trust, 2000.
3. Oliver S. 1997. Exploring lay perspectives on questions of effectiveness; IN: Maynard A, Chalmers I (eds). *Non-random reflections on health services research*. London BMJ Publishing Group.
4. Website: www.invo.org.uk

Instructions to facilitator

This section highlights the importance of writing a protocol before embarking upon the systematic review process. Emphasise that without thought taken to prepare a protocol one may get lost, as a protocol is essentially a road map for the rest of the review. Time taken at the beginning can really save time throughout the review.

A Cochrane protocol has a standard format – if reviewers choose not to register and complete a Cochrane review they should still write a protocol which includes all of the components of a Cochrane protocol. An advantage of aligning oneself with Cochrane is that the protocol is peer-reviewed before the reviewer starts the review (the final review is also peer-reviewed).

The scope of the review (lumping vs. splitting) is very much determined by the needs of the users and the time available to complete the review. The Advisory Group will be invaluable in guiding this decision.

The differences between what lay people want included in a systematic review and what reviewers include is well-described in the Oliver reference above (3). Examples from this reading can be used to highlight the importance of involving users in the review process.

Exercise: Advisory Group. Participants may come up with the following suggestions for Advisory Group members (add any others you can think of):

- a. Interventions for preventing tobacco sales to minors
 - i. Retailers
 - ii. Young people
 - iii. Environmental health officers
 - iv. School representatives, i.e. teachers
 - v. Parents
 - vi. Health professionals
 - vii. Representative with topic experience

- b. Workplace interventions for smoking cessation
 - i. Counsellors/psychologists
 - ii. Health promotion professionals
 - iii. Workplace representatives – employers
 - iv. Workplace representatives - employees

- c. Post-licence driver education for the prevention of road traffic crashes
 - i. Driving teachers
 - ii. People who have undergone post-driver licence education
 - iii. Parents
 - iv. Traffic/Transport Authority representative
 - v. Health professionals – injuries related
 - vi. Secondary school representatives, i.e. students, teachers

- d. Primary prevention for alcohol misuse in young people
 - i. Young people
 - ii. Parents
 - iii. Police
 - iv. Health promotion professionals/Alcohol related research organisations
 - v. School representatives, i.e. teachers
 - vi. Alcohol industry?

- e. Support for breastfeeding mothers
 - i. Breastfeeding mothers, or women who have breastfed in the past
 - ii. Friends and sources of support for those who breastfeed/breastfed
 - iii. Health professionals – maternal child care nurses, midwives, nutritionists

- f. Interventions aimed at improving immunisation rates
 - i. Parents
 - ii. GPs and health professionals (maternal child care nurses, etc)
 - iii. Schools

Participants may want to openly debate the usefulness of including industry representatives on the Advisory Group. There is no right or wrong answer!

Unit Five: Asking an Answerable Question

Learning Objectives

- To understand the importance of formulating an answerable question
- To be able to formulate an answerable question

Reviewers should seek to answer two questions within their review:

- 1. Does the intervention work (not work)?**
- 2. How does the intervention work?**

Importance of getting the question right

A clearly framed question will guide:

- the reader
 - in their initial assessment of relevance
- the reviewer on how to
 - collect studies
 - check whether studies are eligible
 - conduct the analysis.

Therefore, it is important that the question is formulated before beginning the review. *Post-hoc* questions are also more susceptible to bias than those questions determined *a priori*. Although changes to the review question may be required, the reasons for making the changes should be clearly documented in the completed review.

Components of an answerable question (PICO)

The formula to creating an answerable question is following PICO; Population, Intervention, Comparison, Outcome. It is also worthwhile at this stage to determine the types of study designs to include in the review; PICOT.

Qualitative research can contribute to framing the review question (eg. selecting interventions and outcomes of interest to participants). The Advisory Group can also provide valuable assistance with this task.

Population(s)

In health promotion and public health this may include populations, communities or individuals. Consider whether there is value in limiting the population (eg. street youth, problem drinkers). These groups are often under-studied and may be different in all sorts of important respects from study populations usually included in health promotion and public health reviews.

Reviews may also be limited to the effects of the interventions on disadvantaged populations in order to investigate the effect of the interventions on reducing inequalities. Further information on reviews addressing inequalities is provided below.

Intervention(s)

As described earlier, reviewers may choose to lump similar interventions in a review, or split the review by addressing a specific intervention. Reviewers may also consider ‘approaches’ to health promotion rather than topic-driven interventions, for example, peer-led strategies for changing behaviour. In addition, reviewers may want to limit the review by focusing on the effectiveness of a particular type of theory-based intervention (eg. Transtheoretical model) for achieving certain health outcomes (eg. smoking cessation).

Comparison(s)

It is important to specify the comparison intervention for the review. Comparison interventions may be no intervention, another intervention or standard care/practice. The choice of comparison or control has large implications for the interpretation of results. A question addressing one intervention versus no intervention is a different question than one comparing one intervention versus standard care/practice.

Example: 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.

The majority of the studies included in this review address primary prevention of unintended pregnancy versus standard care/practice. Therefore, this review is not addressing whether primary prevention is effective, it is simply investigating the effect of specific interventions compared to standard practice. This is a much smaller gap to investigate an effect, as it is usually easier to find a difference when comparing one intervention versus no intervention.

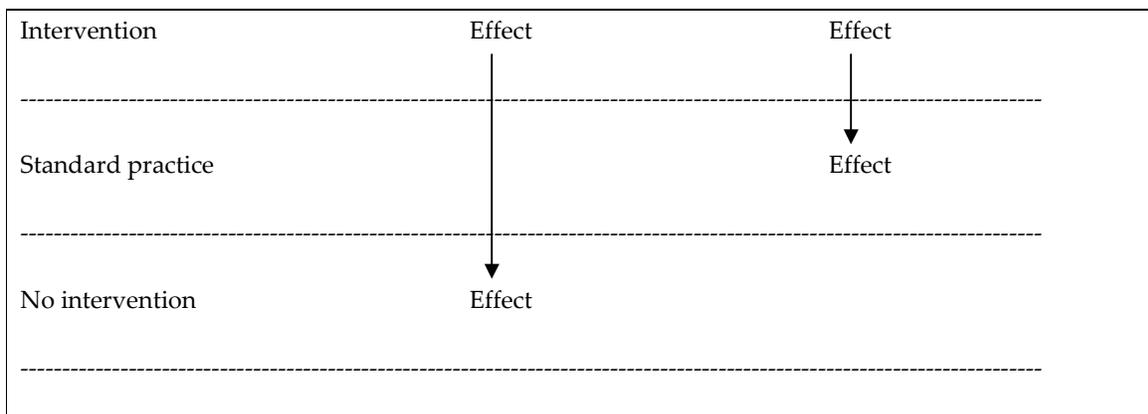


Figure Two. The difference between comparing the effect of one intervention versus no intervention and one intervention versus standard practice.

For example, many of the school-based interventions in the review are compared to normal sexual education in the schools, and are shown to be ineffective for reducing unintended pregnancies. Yet the interpretation of the results read “primary prevention strategies do not delay the initiation of sexual intercourse or improve the use of birth control among young men and women”. This reads that the review question has sought to address primary prevention versus no intervention. Rather, the review addressed whether theory-led interventions are more effective than standard care/practice.

Outcome(s)

The outcome(s) chosen for the review must be meaningful to the users of the review. The discrepancy between the outcomes and interventions that reviewers choose to include in the review and the outcomes and interventions that lay people prefer to be included has been well-described.¹

To investigate both the implementation of the intervention and its effects reviewers will need to include both process indicators as well as outcome measures. Unanticipated (side-effects) as well as anticipated effects should be investigated in addition to cost-effectiveness, where appropriate.

Reviewers will also need to decide if proximal/immediate, intermediate or distal outcomes are to be assessed. If only intermediate outcomes are measured (eg. blood sugar levels in persons with diabetes, change in knowledge and attitudes) reviewers need to determine how strong the linkage is to more distal outcomes (eg. cardiovascular disease, behaviour change). The use of theory can assist with determining this relationship. In addition, reviewers should decide if only objective measures are to be included (eg. one objective measure of smoking status is saliva thiocyanate or alveolar carbon monoxide) or subjective measures (eg. self-reported smoking status), or a combination of both (discussing the implications of this decision).

Examples of review questions

Poorly designed questions:

1. Are condoms effective in preventing HIV?
2. Which interventions reduce health inequalities among people with HIV?

Answerable questions:

1. In men who have sex with men, does condom use reduce the risk of HIV transmission?
2. In women with HIV, do peer-based interventions reduce health inequalities?

Are mass media interventions effective in preventing smoking in young people?

| Problem, population | Intervention | Comparison | Outcome | Types of studies |
|-------------------------------------|---|-----------------|--|--|
| Young people, under 25 years of age | 1. Television 2. Radio 3. Newspapers 4. Billboards 5. Posters 6. Leaflets 7. Booklets | No intervention | 1. objective measures of smoking 2. self-reported smoking behaviour 3. Intermediate measures (intentions, attitudes, knowledge) 4. Process measures (eg. media reach) | 1. RCT (and quasi-RCT) 2. Controlled before and after studies 3. Time series designs |

Types of study designs to include

The decisions about which type(s) of study design to include will influence subsequent phases of the review, particularly the search strategies, choice of quality assessment criteria, and the analysis stage (especially if a statistical meta-analysis is to be performed).

The decision regarding which study designs to include in the review should be dictated by the intervention (the review question) or methodological appropriateness, and not vice versa.^{2,3} If the review question has been clearly formulated then knowledge of the types of study designs needed to

answer it should automatically follow.³ If different types of study designs are to be included in the same review the reasons for this should be made explicit.

Effectiveness studies

Where RCTs are lacking, or for issues relating to feasibility and ethics are not conducted, other study designs such as non-randomised controlled trials, before and after studies, and interrupted time series designs should also be considered for inclusion in the review.

Comparisons with historical controls or national trends may be included when this is the **only** type of evidence that is available, for example, in reviews investigating the effectiveness of policies, and should be accompanied by an acknowledgement that the evidence of evidence is necessarily weaker.

Randomised controlled trial

Subjects are randomly allocated to groups either for the intervention being studied or the control (using a random mechanism, such as coin toss, random number table, or computer-generated random numbers) and the outcomes are compared.¹

Each participant or group has the same chance of receiving each intervention and the investigators cannot predict which intervention is next.

Quasi-randomised controlled trial / pseudo-randomised controlled trial

Subjects are allocated to groups for intervention or control using a non-random method (such as alternate allocation, allocation of days of the week, or odd-even study numbers) and the outcomes are compared.¹

Controlled before and after study / cohort analytic

Outcomes are compared for a group receiving the intervention being studied, concurrently with control subjects receiving the comparison intervention (eg, usual or no care/intervention).¹

Uncontrolled before and after study / cohort study

The same group is pre-tested, given an intervention, and tested immediately after the intervention. The intervention group, by means of the pre-test, act as their own control group.²

Interrupted time series

A time series consists of multiple observations over time. Observations can be on the same units (eg. individuals over time) or on different but similar units (eg. student achievement scores for particular grade and school). Interrupted time series analysis requires knowing the specific point in the series when an intervention occurred.² These designs are commonly used to evaluate mass media campaigns.

Qualitative research

Qualitative research explores the subjective world. It attempts to understand why people behave the way they do and what meaning experiences have for people. Qualitative research relevant to effectiveness reviews may include the following:

Qualitative studies of experience: these studies may use a range of methods, but frequently rely on in-depth tape-recorded interviews and non-participant observational studies to explore the experience of people receiving an intervention.

Process evaluations: these studies can be included within the context of the effectiveness studies. These evaluations use a mixture of methods to identify and describe the factors that promote and/or impede the implementation of innovation in services.³

References:

1. NHMRC (2000). How to review the evidence: systematic identification and review of the scientific literature. Canberra: NHMRC.
2. Thomas H. Quality assessment tool for quantitative studies. Effective Public Health Practice Project. McMaster University, Toronto, Canada.
3. 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. March 2001. <http://www.york.ac.uk/inst/crd/report4.htm>

Cluster-RCTs and cluster non-randomised studies

Allocation of the intervention by group or cluster is being increasingly adopted within the field of public health because of administrative efficiency, lessened risk of experimental contamination and likely enhancement of subject compliance.⁴ Some studies, for example a class-based nutrition intervention, dictate its application at the cluster level.

Interventions allocated at the cluster (eg. school, class, worksite, community, geographical area) level have particular problems with selection bias where groups are formed not at random but rather through some physical, social, geographic, or other connection among their members.^{5,6} Cluster trials also require a larger sample size than would be required in similar, individually allocated trials because the correlation between cluster members reduces the overall power of the study.⁵ Other methodological problems with cluster-based studies include the level of intervention differing from the level of evaluation (analysis) and the often small number of clusters in the study.⁷ Issues surrounding cluster trials have been well described in a Health Technology Assessment report⁷, which should be read for further information if cluster designs are to be included in a systematic review.

The role of qualitative research within effectiveness reviews

- to "provide an in-depth understanding of people's experiences, perspectives and histories in the context of their personal circumstances and settings"⁸

Qualitative studies can contribute to reviews of effectiveness in a number of ways including⁹:

- Helping to frame the review question (eg. selecting interventions and outcomes of interest to participants).
- Identifying factors that enable/impede the implementation of the intervention (eg. human factors, contextual factors)
- Describing the experience of the participants receiving the intervention
- Providing participants' subjective evaluations of outcomes
- Helping to understand the diversity of effects across studies, settings and groups
- 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.

Methods commonly used in qualitative studies may include one or a number of the following; interviews (structured around respondents priorities/interests), focus groups, participant and/or non participant observation, conversation (discourse and narrative analysis), and documentary and video analysis. The unit of analysis within qualitative studies is not necessarily individuals or single cases; communities, populations or organisations may also be investigated. Anthropological research,

which may involve some or all of these methods in the context of wide ranging 'fieldwork' can also be a valuable source of evidence, although may be difficult to subject to many aspects of the systematic review process.

Health inequalities

Health inequalities are defined as "the gap in health status, and in access to health services, between different social classes and ethnic groups and between populations in different geographical areas."¹⁰

There is a need for systematic reviews to consider health inequalities in the assessment of effectiveness of interventions. This is because it is thought that many interventions may not be equally effective for all population subgroups. The effectiveness for the disadvantaged may be substantially lower.

Evans and Brown (2003)¹¹ suggest that there are a number of factors that may be used in classifying health inequalities (captured by the acronym PROGRESS):

- Place of residence
- Race/ethnicity
- Occupation
- Gender
- Religion
- Education
- Socio-economic status
- Social capital

Therefore, it may be useful for a review of public health interventions to measure the effect across different subgroups (as defined by any of the PROGRESS factors).

Example of a review addressing inequalities: Kristjansson E, Robinson VA, MacDonald B, Krasevec J, Greenhalgh T, McGowan J, Francis D, Tugwell P, Petticrew M, Shea B, Wells G. School feeding for improving the physical and psychosocial health of disadvantaged elementary school children (Protocol for a Cochrane Review). In: The Cochrane Library, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd. Disadvantage in this review is defined by income (SES status).

Data required for reviews addressing inequalities:

- A valid measure of health status (or change in health status)
- A measure of disadvantage (i.e., define socio-economic position)
- A statistical measure for summarising the differential effectiveness.

The above review chose to define interventions effective in reducing inequalities as interventions which were more effective for people in lower SES. A potentially effective intervention was one which was:

- equally effective across the socioeconomic spectrum (potentially reducing health inequalities due to the higher prevalence of health problems among the disadvantaged).
- targeted only at disadvantaged groups and was effective.

REFERENCES

1. Oliver S. 1997. Exploring lay perspectives on questions of effectiveness; IN: Maynard A, Chalmers I (eds). Non-random reflections on health services research. London BMJ Publishing Group.
2. Nutbeam D, Harris E. (2004). Theory in a Nutshell. A practical guide to health promotion theories. Sydney, Australia: McGraw-Hill Australia Pty Ltd, vii-9.
3. Petticrew M, Roberts H. (2003). Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health, 57, 527-9.
4. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. Am J Public Health. 2004 Mar;94(3):416-22.
5. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? BMJ. 2001 Feb 10;322(7282):355-7.
6. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. Am J Public Health. 2004 Mar;94(3):423-32.
7. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. Health Technol Assess. 1999;3(5):iii-92.
8. 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.
9. Centre for Reviews and Dissemination (Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD Report Number 4 (2nd Edition) March 2001), at <http://www.york.ac.uk/inst/crd/report4.htm>
10. Public Health Electronic Library. <http://www.phel.gov.uk/glossary/glossaryAZ.asp?getletter=H>. Accessed June 29, 2004
11. Evans T, Brown H. Road traffic crashes: operationalizing equity in the context of health sector reform. Injury Control and Safety Promotion 2003;10(2):11-12.

ADDITIONAL READING

Richardson WS, Wilson MC, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence-based decisions [Editorial]. ACP J Club 1995;123(3):A12-3.

Richardson WS. Ask, and ye shall retrieve [EBM Note]. Evidence Based Medicine 1998;3:100-1.

Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. Health Technol Assess. 1999;3(5):iii-92.

EXERCISE

1. Write an answerable review question (will be used in a later exercise)

P =

I =

C =

O =

Q.....

.....

The effectiveness of (I) versus (C) for (O) in (P)

2. What type(s) of study design(s) should be included to investigate the effectiveness of the intervention?

- Randomised controlled trial / cluster randomised controlled trial
- Quasi-randomised controlled trial/pseudo-randomised trial
- Controlled before and after study/cohort analytic/concurrently controlled comparative study
- Uncontrolled before and after study/cohort study
- Interrupted time series designs
- Qualitative research

FACILATOR'S GUIDE

Time required:

45 minutes approx (15 minutes slides, 30 minutes exercise)

Learning topics

- Formulating an answerable question that is relevant to users
- Types of study designs utilised in health promotion and public health interventions
- The types of questions asked by decision-makers, practitioners and consumers

Summary of activity

Lecture – Powerpoint slides

Individual exercise – to formulate an answerable question

Description of supporting materials

No supporting materials.

Further reading (if required)

1. Richardson WS, Wilson MC, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence-based decisions [Editorial]. ACP J Club 1995;123(3):A12-3.
2. Richardson WS. Ask, and ye shall retrieve [EBM Note]. Evidence Based Medicine 1998;3:100-1.
3. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. Health Technol Assess. 1999;3(5):iii-92. *** Recommended for more information about cluster designed studies. ***

Instructions to facilitator

Emphasise the importance of this component of the systematic review process. A clear question is vital to conducting a comprehensive search and applying the inclusion and exclusion criteria. The type of question asked in health promotion and public health is expanded to include investigating how the interventions worked (or did not work) and why they may have worked (not worked).

Participants will see the acronym PICO used everywhere in evidence-based medicine. Sometimes it will be seen as PICOT which includes types of study designs (T).

Emphasise that reviewers consider conducting reviews which examine inequalities. Although the methodology is in its infancy, more reviews are needed in this area. Conducting reviews will contribute to the methodological base.

Explain that qualitative research should be integrated into reviews to address the issues that are of concern to users – the implementation and appropriateness of interventions. ** Reviewers may want to only include qualitative research relating to the outcome evaluations (effectiveness studies). For example, a process evaluation may be included if it is part of an effectiveness (outcome) study.

Exercise: Make sure you go round the room to make sure everyone is clear about the exercise. As there will be insufficient time to ask all participants to report back to the class, choose a few

participants to describe their PICO question, and the study designs they would choose that would help them answer the question. You will have time in the searching session to ensure that all PICO questions are answerable.

Common mistakes: The PICO question does not relate to whom the reviewer wishes to apply the results, eg. wanting to prevent disease in Indigenous people in the Northern Territory of Australia. The PICO question relates to the kind of research required to answer the question, which in this example may come from a number of studies of different Indigenous populations. Therefore, the P of PICO would be any indigenous population. Again, the question must reflect the research being carried out – do not let reviewers restrict age limits, for example, if age is not standardised throughout the world. For example, school aged children and adolescents could range from 4 years to 19 years throughout the world, or interventions may have different names or comprise slightly different components between countries.

Unit Six: Finding The Evidence

Learning Objectives

- To understand the complexities of searching for health promotion and public health studies
- To gain knowledge of how to locate primary studies of health promotion and public health interventions
- To gain basic skills to carry out a search for primary studies

Identifying health promotion and public health primary studies

The inclusion of an unbiased sample of relevant studies is central to the validity of systematic reviews. Time-consuming and costly literature searches, which cover the grey literature and all relevant languages and databases, are normally recommended to prevent reporting biases.¹

Searching for primary studies on health promotion and public health topics can be a very time-intensive task, as search strategies will need to be adapted for a number of databases, and broad searches using a wide range of terms may result in a large number of citations requiring application of the inclusion and exclusion criteria. This is partly due to health promotion and public health terminology being very non-specific or non-standardised; day to day words are often used to describe interventions and populations. In addition, it may not be appropriate to add a randomised controlled trial (RCT) filter to limit the search because the question may be best answered using other types of study designs.

Components of the searching process

The key components of the search strategy comprise of subject headings and textwords that describe each element of the PICO(T) question.

However, it is usually recommended not to include the O (outcome) of the PICO question in the search strategy because outcomes are described in many different ways and may not be described in the abstract of the article. Search terms to describe outcomes should *only* be used if the number of citations is too large to apply the inclusion and exclusion criteria.

Pilot the search strategy first – complete a scoping search on a database most likely to yield studies using a sample of keywords to locate a few relevant studies. Check the subject headings that are used to index the studies and the relevant textwords in the abstract of the citation. Also, it may be useful to find the citations of key articles in PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) and click on Related Articles to find other relevant studies in order to determine additional relevant subject headings and textwords.

The search strategy developed to identify studies will not search the entire full-text of the article. The following complete reference for the citation demonstrates the information that is available for each citation (example provided using the OVID interface): therefore searching the subject headings **and** textwords in the abstract will help us to find this study. **Always** use a combination of subject headings and textwords for each PICO element.

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.

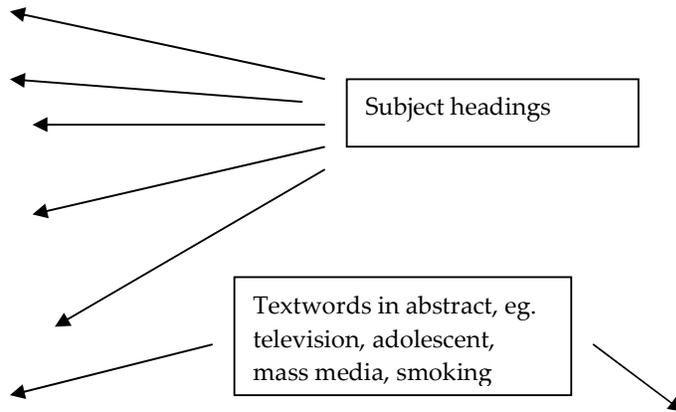
ISSN

0090-0036

Publication Type

Journal Article.

Grant Number



CA38392 (NCI)
Language
English
Entry Date
19910516
Revision Date
20021101
Update Date
20031209

Subject headings /descriptors (eg. MESH headings in Medline)

Subject headings are used in different databases to describe the subject of each journal article indexed in the database. For example, MeSH (Medical Subject Headings) are used within the Medline database; there are more than 22,000 terms used to describe studies and the headings are updated annually to reflect changes in medicine and medical terminology.

Examples of subject headings relevant to health promotion and public health:

Mass media, smoking, adolescent, health promotion, health education, students, sports

Remember, each database will have different controlled vocabulary (subject headings). Also, subject headings are assigned by human beings, so mistakes can be made. For example, the mass media article was not assigned with the mass media subject heading in the PyscINFO database. Therefore, search strategies should always include textwords in addition to subject headings.

For many health promotion topics there may be few subject headings available (eg. community-based interventions). Therefore, the search strategy may comprise mainly of textwords.

Textwords

These are words that are used in the abstract of articles (and title) to assist with finding the relevant literature. Textwords in a search strategy always end in .tw, eg. 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 textwords, and add any other words such may also describe the component of PICO.

- Truncation \$ - this picks up various forms of a textword.

Eg. teen\$ will pick up teenage, teenagers, teens, teen

Eg. Smok\$ will pick up smoke, smoking, smokes, smoker, smokers

- Wildcards ? and #

These syntax commands will pick up different spellings.

? will substitute for one or no characters, so is useful for locating US and English spellings

Eg. 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

Eg. 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.

Eg. sport ADJ1 policy will pick up sport policy and policy for sport

Eg. mental ADJ2 health will pick up mental health and mental and physical health

Note: Databases may use different syntax to retrieve records (eg. \$ or * may be used in different databases or interfaces). Therefore, reviewers will need to become well-acquainted with the idiosyncrasies of each database. Due to the different subject headings used between databases, reviewers will also need to adapt their search strategy for each database (only adapt the subject headings, not textwords).

Combining each element of the PICO questions

| Element of question | | |
|--|----|-----------|
| <i>P - Population</i> | | |
| Subject headings | OR | Textwords |
| <i>I - Intervention</i> | | |
| Subject headings | OR | Textwords |
| <i>C - Comparison (if necessary)</i> | | |
| Subject headings | OR | Textwords |
| <i>O - Outcome</i> | | |
| Subject headings | OR | Textwords |
| <i>T - Type of study (if necessary)</i> | | |
| Subject headings | OR | Textwords |
| Use a validated filter | | |
| To find studies using all of the PICO elements | | |
| P AND I AND C AND O (AND T) | | |

A lumped review (review of a number of different interventions) is simply a review comprising a number of different PICO(T) questions. This is exemplified in the following pages outlining the search strategy to locate “Interventions for preventing obesity in children”.

Using study design to limit search

RCTs: If the review is limited to evidence from RCTs a study design filter can be added to the search strategy. The Cochrane Reviewer’s Handbook² details the appropriate filter to add.

Non-RCTs: Limiting the search strategy by using non-randomised study terms can be very problematic, and is generally not recommended. This is because:

- Few studies may be indexed by study design
- The vocabulary required to identify different study designs can vary extensively between electronic databases. Terms vary from ‘control groups’ to ‘follow-up studies’, to ‘longitudinal studies’ or even ‘program effectiveness’ or ‘program evaluation’, to index the same studies
- Some databases, eg. PsycINFO, are poorly indexed with respect to methodology.

Therefore, after a PICO search is completed all citations will require application of the inclusion and exclusion criteria.

Qualitative research: A filter for the CINAHL database is available from the Edward Miner Library http://www.urmc.rochester.edu/hslt/miner/digital_library/tip_sheets/Cinahl_eb_filters.pdf

Where to locate studies

a) Electronic databases of relevance to health promotion and public health

Reviewers should ensure that the search strategy (subject headings and textwords) is developed for a number of databases that cover the variety of domains where the literature may be located.

A full list of free public health databases and subscription-only databases is available at <http://library.umassmed.edu/ebpph/dblist.cfm>. This website contains a number of databases that have not been included in the following list.

Some examples of electronic databases that may be useful to identify public health or health promotion studies include (websites listed for databases available freely via the internet):

| | |
|--------------------|--|
| Psychology: | PsycINFO/PscyLIT |
| Biomedical: | CINAHL, LILACS (Latin American Caribbean Health Sciences Literature) http://www.bireme.br/bvs/l/ibd.htm , Web of Science, Medline, EMBASE, CENTRAL (http://www.update-software.com/clibng/cliblogon.htm), Combined Health Information Database (CHID) http://chid.nih.gov/ , Chronic Disease Prevention Database (CDP) http://www.cdc.gov/cdp/ |
| 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 |
| Transport: | NTIS (National Technical Information Service), TRIS (Transport Research Information Service) http://ntl.bts.gov/tris , IRRD (International Road Research Documentation), TRANSDOC (from ECMT (European Conference of Ministers of Transport)) |
| Physical activity: | SportsDiscus |
| HP/PH: | BiblioMap (EPPI-Centre) http://eppi.ioe.ac.uk , HealthPromis (HDA, UK) http://www.hda-online.org.uk/evidence/ , Global Health |
| Other: | Popline (population health, family planning) http://db.jhuccp.org/popinform/basic.html , Enviroline (environmental health) – available on Dialog, Toxfile (toxicology) – available on Dialog, Econlit (economics) |
| 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). |

b) Handsearching health promotion and public health journals

It may be useful to handsearch specialist journals relevant to the review topic area to identify further primary research studies. Also consider non-health promotion and public health journals which may cover the topic of interest, i.e., marketing journals, etc.

Two lists of health promotion and public health journals have been produced which may help to determine which journals to search.

- 1) The Lamar Soutter Library list of public health journals, <http://library.umassmed.edu/ebpph/> (a list of freely available journals is also included)
- 2) The Core Public Health Journals List compiled by Yale University, <http://www.med.yale.edu/eph/library/phjournals/>

The Effective Public Health Practice Project (Canada) has found that the most productive journals to handsearch to locate public health and health promotion articles are: American Journal of Health Promotion, American Journal of Preventive Medicine, American Journal of Public Health, Canadian Journal of Public Health, BMJ. Other useful journals include Annual Review of Public Health, Health Education and Behavior (formerly Health Education Quarterly), Health Education Research, JAMA, Preventive Medicine, Public Health Reports, Social Science and Medicine.

c) Grey literature

Methods to locate unpublished, difficult-to-find literature include:

- Scanning reference lists of relevant studies
- Contacting authors/academic institutions of key studies
- Searching for theses, dissertations, conference proceedings (one source of dissertations and theses is the Networked Digital Library of Theses and Dissertations (NDLTD) which can be accessed from <http://www.theses.org/>)
- Searching the internet for national public health reports, local public health reports, reviews serving as background documentation for legislation, quality assurance reports, etc. A useful internet search engine for locating academic work is Google Scholar (<http://scholar.google.com>).

Save, document and export the search

Always save and print out the search strategy for safe record-keeping. It is essential to have bibliographic software (Endnote, Reference Manager, GetARef) to export the retrieved citations to apply the inclusion/exclusion criteria. Citations from unpublished literature cannot usually be exported, so will require individual entry by hand into the reference managing system. Bibliographic software will also assist with the referencing when writing the final review.

REFERENCES

1. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;7(1).
2. Clarke M, Oxman AD, editors. *Cochrane Reviewers' Handbook* 4.2.0 [updated March 2003]. <http://www.cochrane.dk/cochrane/handbook/handbook.htm>

ADDITIONAL READING

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.

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 textwords that would help describe each of the PICO components of the review question.

| | MeSH/descriptors | Textwords |
|-----|------------------------------------|------------------------------|
| | eg. Adolescent (Medline) | student, highschool, teenage |
| | eg High School Students (PsycINFO) | |
| P = | | |
| | | |
| | | |
| | | |
| | | |
| I = | | |
| | | |
| | | |
| | | |
| | | |
| C = | <i>May not be required</i> | |
| | | |
| | | |
| | | |
| | | |
| O = | | |
| | | |
| | | |
| | | |
| | | |

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

Examples of searching strategies

Campbell K, Waters E, O'Meara S, Kelly S, Summerbell C. Interventions for preventing obesity in children (Cochrane Review). In: The Cochrane Library, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd.

MEDLINE, 1997

1. explode "Obesity"/ all subheadings
2. "Weight-Gain"/ all subheadings
3. "Weight-Loss"/ all subheadings
4. obesity or obese
5. weight gain or weight loss
6. overweight or over weight or overeat* or over eat*
7. weight change*
8. (bmi or body mass index) near2 (gain or loss or change)
9. #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8

10. "Child-" in MIME,MJME
11. "Adolescence"/ all subheadings
12. "Child-Preschool"/ all subheadings
13. "Infant-" in MIME,MJME
14. child* or adolescen* or infant*
15. teenage* or young people or young person or young adult*
16. schoolchildren or school children
17. p?ediatr* in ti,ab
18. boys or girls or youth or youths
19. #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18

20. explode "Behavior-Therapy"/ all subheadings
21. "Social-Support" in MIME,MJME
22. "Family-Therapy"/ all subheadings
23. explode "Psychotherapy-Group"/ all subheadings
24. (psychological or behavio?r*) adj (therapy or modif* or strateg* or intervention*)
25. group therapy or family therapy or cognitive therapy
26. (lifestyle or life style) adj (chang* or intervention*)
27. counsel?ing
28. social support
29. peer near2 support
30. (children near3 parent?) near therapy
31. #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30

32. explode "Obesity"/ drug-therapy
33. explode "Anti-Obesity-Agents"/ all subheadings
34. lipase inhibitor*
35. orlistat or xenical or tetrahydrolipstatin
36. appetite adj (suppressant* or depressant*)
37. sibutramine or (meridia in ti,ab)
38. dexfenfluramine or fenfluramine or phentermine
39. bulking agent*
40. methylcellulose or celevac
41. (antiobesity or anti obesity) adj (drug* or agent*)
42. guar gum
43. #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42

44. explode "Obesity"/ diet-therapy

45. "Diet-Fat-Restricted"/ all subheadings
46. "Diet-Reducing"/ all subheadings
47. "Diet-Therapy"/ all subheadings
48. "Fasting"/ all subheadings
49. diet or diets or dieting
50. diet* adj (modif* or therapy or intervention* or strateg*)
51. low calorie or calorie control* or healthy eating
52. fasting or modified fast*
53. explode "Dietary-Fats"/ all subheadings
54. fruit or vegetable*
55. high fat* or low fat* or fatty food*
56. formula diet*
57. #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56

58. "Exercise"/ all subheadings
59. "Exercise-Therapy"/ all subheadings
60. exercis*
61. aerobics or physical therapy or physical activity or physical inactivity
62. fitness adj (class* or regime* or program*)
63. aerobics or physical therapy or physical training or physical education
64. dance therapy
65. sedentary behavior reduction
66. #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65

67. explode "Obesity"/ surgery
68. "Surgical-Staplers"/ all subheadings
69. "Surgical-Stapling"/ all subheadings
70. "Lipectomy"/ all subheadings
71. "Gastric-Bypass"/ all subheadings
72. "Gastroplasty"/ all subheadings
73. dental splinting or jaw wiring
74. gastroplasty or gastric band* or gastric bypass
75. intragastric balloon* or vertical band*
76. stomach adj (stapl* or band* or bypass)
77. liposuction
78. #67 or #68 or #69 or #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77

79. explode "Alternative-Medicine"/ all subheadings
80. alternative medicine or complementary therap* or complementary medicine
81. hypnotism or hypnosis or hypnotherapy
82. acupuncture or homeopathy or homoeopathy
83. chinese medicine or indian medicine or herbal medicine or ayurvedic
84. #79 or #80 or #81 or #82 or #83

85. (diet or dieting or slim*) adj (club* or organi?ation*)
86. weightwatcher* or weight watcher*
87. correspondence adj (course* or program*)
88. fat camp* or diet* camp*
89. #85 or #86 or #87 or #88

90. "Health-Promotion"/ all subheadings
91. "Health-Education"/ all subheadings
92. health promotion or health education
93. media intervention* or community intervention*
94. health promoting school*
95. (school* near2 program*) or (community near2 program*)

- 96. family intervention* or parent* intervention*
- 97. parent* near2 (behavio?r or involve* or control* or attitude* or educat*)
- 98. #90 or #91 or #92 or #93 or #94 or #95 or #96 or #97

- 99. "Health-Policy"/ all subheadings
- 100. "Nutrition-Policy"/ all subheadings
- 101. health polic* or school polic* or food polic* or nutrition polic*
- 102. #99 or #100 or #101

- 103. explode "Obesity"/ prevention-and-control
- 104. "Primary-Prevention"/ all subheadings
- 105. primary prevention or secondary prevention
- 106. preventive measure* or preventative measure*
- 107. preventive care or preventative care
- 108. obesity near2 (prevent* or treat*)
- 109. #103 or #104 or #105 or #106 or #107 or #108

- 110. explode "Controlled-Clinical-Trials"/ all subheadings
- 111. "Random-Allocation" in MIME,MJME
- 112. "Double-Blind-Method" in MIME,MJME
- 113. "Single-Blind-Method" in MIME,MJME
- 114. "Placebos"/ all subheadings
- 115. explode "Research-Design"/ all subheadings
- 116. (singl* or doubl* or trebl* or tripl*) near5 (blind* or mask*)
- 117. exact{CONTROLLED-CLINICAL-TRIAL} in PT
- 118. placebo*
- 119. matched communities or matched schools or matched populations
- 120. control* near (trial* or stud* or evaluation* or experiment*)
- 121. comparison group* or control group*
- 122. matched pairs
- 123. outcome study or outcome studies
- 124. quasiexperimental or quasi experimental or pseudo experimental
- 125. nonrandomi?ed or non randomi?ed or pseudo randomi?ed
- 126. #110 or #111 or #112 or #113 or #114 or #115 or #116 or #117 or #118 or #119 or #120 or #121 or #122 or #123 or #124 or #125

- 127. #9 and #19
- 128. #31 or #43 or #57 or #66 or #78 or #84 or #89 or #98 or #102 or #109
- 129. #126 and #127 and #128
- 130. animal in tg
- 131. human in tg
- 132. #130 not (#130 and #131)
- 133. #129 not #132
- 134. #133 and (PY >= "1997")

Brunton G, Harden A, Rees R, Kavanagh J, Oliver S, Oakley A (2003). Children and Physical Activity: A systematic Review of Barriers and Facilitators. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

1. Exp child/
2. Exp adolescence/ or exp child/ hospitalized/ or exp child institutionalized/ or exp disabled children/ or infant
3. 1 not 2
4. exp child preschool/
5. exp students/
6. ((university or college or medical or graduate or post graduate) adj2 student\$.ti.ab.
7. 5 not 6
8. (school adj3 (child\$ or pupil\$ or student\$ or kid of kids of primary or nursery or infant\$)).ti.ab.
9. or/3-4,7-8
10. exp health promotion/
11. exp health education/
12. exp preventive medicine/
13. (prevent\$ or reduc\$ or promot\$ or increase\$ or program\$ or curricul\$ or educat\$ or project\$ or campaign\$ or impact\$ or risk\$ or vulnerab\$ or resilien\$ or factor\$ or correlate\$ or predict\$ or determine\$ or behavior\$).ti.ab.
14. (health\$ or ill or illness or ills or well or wellbeing or wellness or poorly or unwell or sick\$ or disease\$).ti.ab.
15. ((prevent\$ or reduc\$ or promot\$ or increase\$ or program\$ or curricul\$ or educat\$ or project\$ or campaign\$ or impact\$ or risk\$ or vulnerab\$ or resilien\$ or factor\$ or correlate\$ or predict\$ or determine\$ or behavior\$) adj3 (health\$ or ill or illness or ills or well or wellbeing or wellness or poorly or unwell or sick\$ or disease\$).ti.ab.
16. or/10-12,15
17. (determine\$ or facilitate\$ or barrier\$).ti
18. Risk factors/
19. Culture/
20. Family/ or Internal-external control/ or Life style/ or Prejudice/ or Psychology, social/ or Psychosocial deprivation/
21. Child behavior/
22. Habits/
23. Poverty/
24. Social class/
25. Social conditions/
26. Socioeconomic factors/
27. Family characteristics/
28. Ethnicity.ti,ab.
29. Attitude to health/
30. Or/17-29
31. Exp sports/
32. Exp physical fitness/
33. Exp exertion/
34. "Physical education and training"/
35. exp leisure activities/
36. Recreation/
37. ((sedentary or inactive\$) adj3 child\$).ti,ab.
38. ((physical\$ or sport\$ or exercise\$ or game\$1) adj3 (activit\$ or exercise\$ or exert\$ or fit or fitness or game\$1 or endurance or endure\$ or educat\$ or train\$1 or training)).ti,ab.
39. Or/31-38
40. Or/16,30
41. And/9,39-40

WORKED EXAMPLE

We will work through the process of finding primary studies for a systematic review, using the review below as an example: ** This search has been modified from the original version **

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: John 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 relating to P - population

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 relating to O - outcome

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 relating to I - intervention

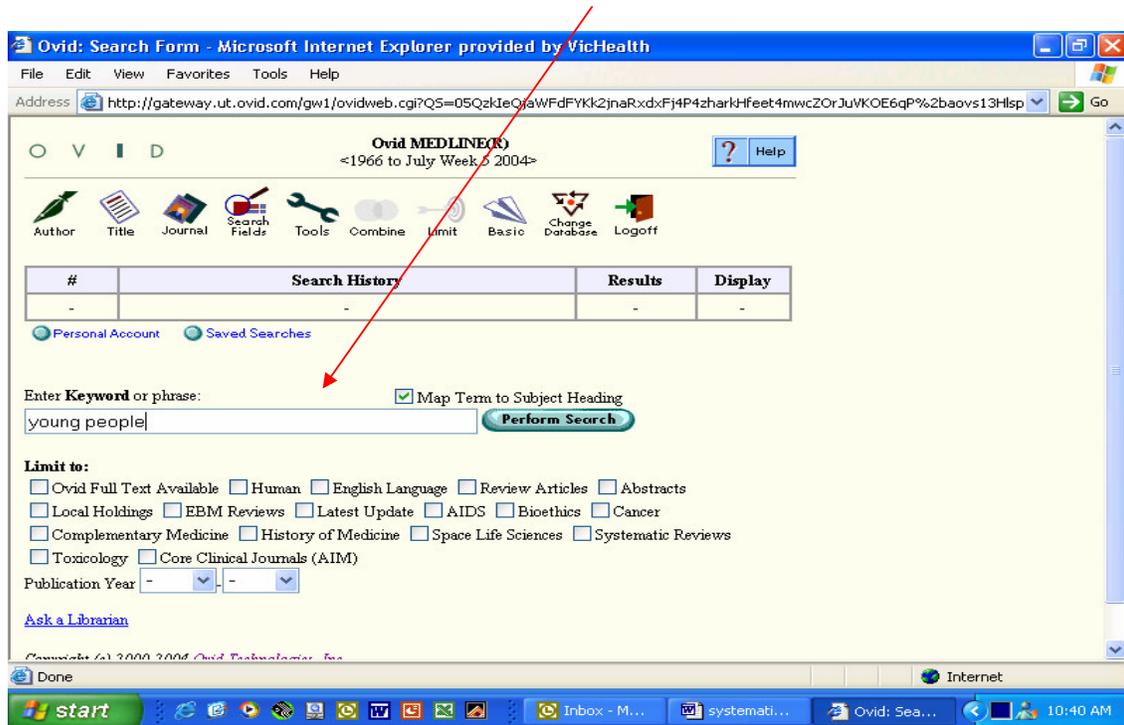
26 mass media/
27 audiovisual aids/
28 exp television/
29 motion pictures/
30 radio/
31 exp telecommunications/
32 videotape recording/
33 newspapers/
34 advertising/
35 (tv or televis\$).tw.
36 (advertis\$ adj4 (prevent or prevention)).tw.
37 (mass adj media).tw.
38 (radio or motion pictures or newspaper\$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

This review wants to exclude mass media interventions as a community based intervention (a review has already been completed on this topic)
- see search line 42

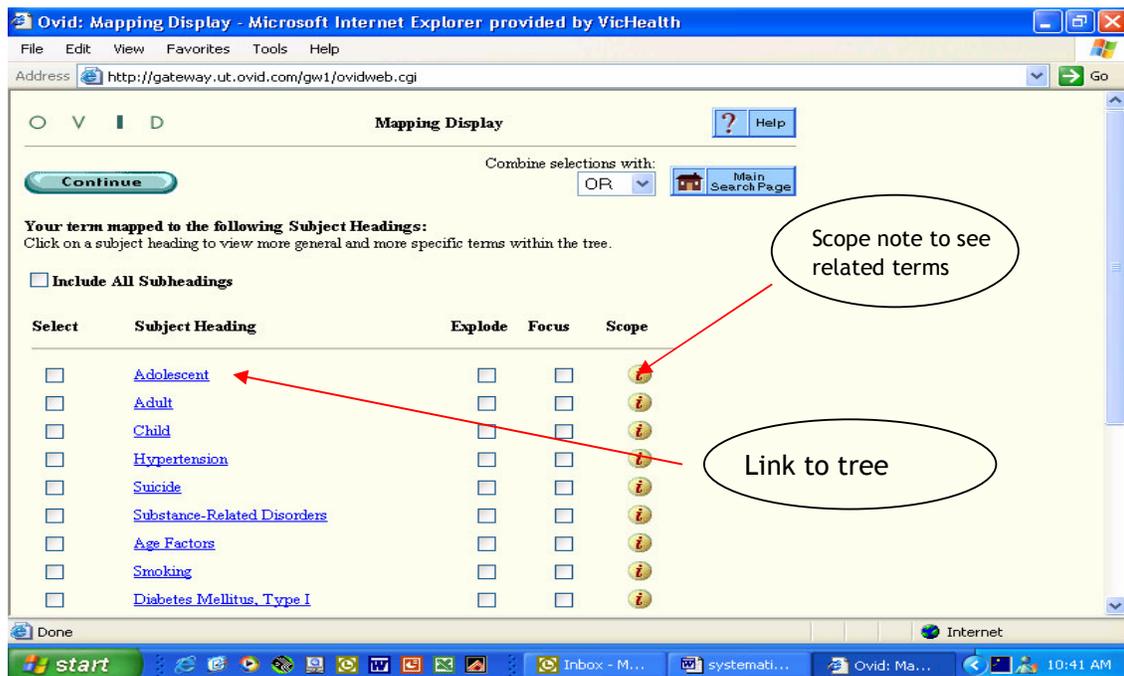
40 7 and 12 and 25
41 7 and 12 and 39
42 40 not 41

40 – young people and smoking and community-based intervention
41 – young people and smoking and mass media interventions
42 - community interventions not including mass media interventions

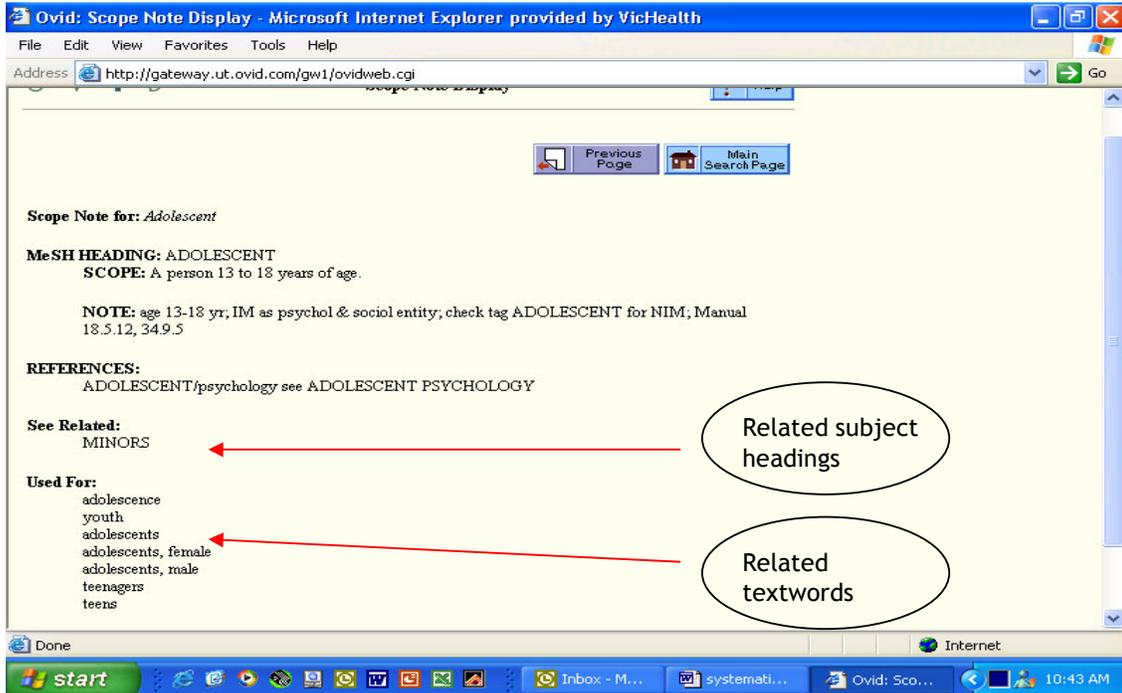
1. Start with the primary concept, i.e. young people.
2. The Ovid search interface allows plain language to be 'mapped' to related subject headings, terms from a controlled indexing list (called controlled vocabulary) or thesaurus (eg. MeSH in MEDLINE). **Map the term 'young people'**



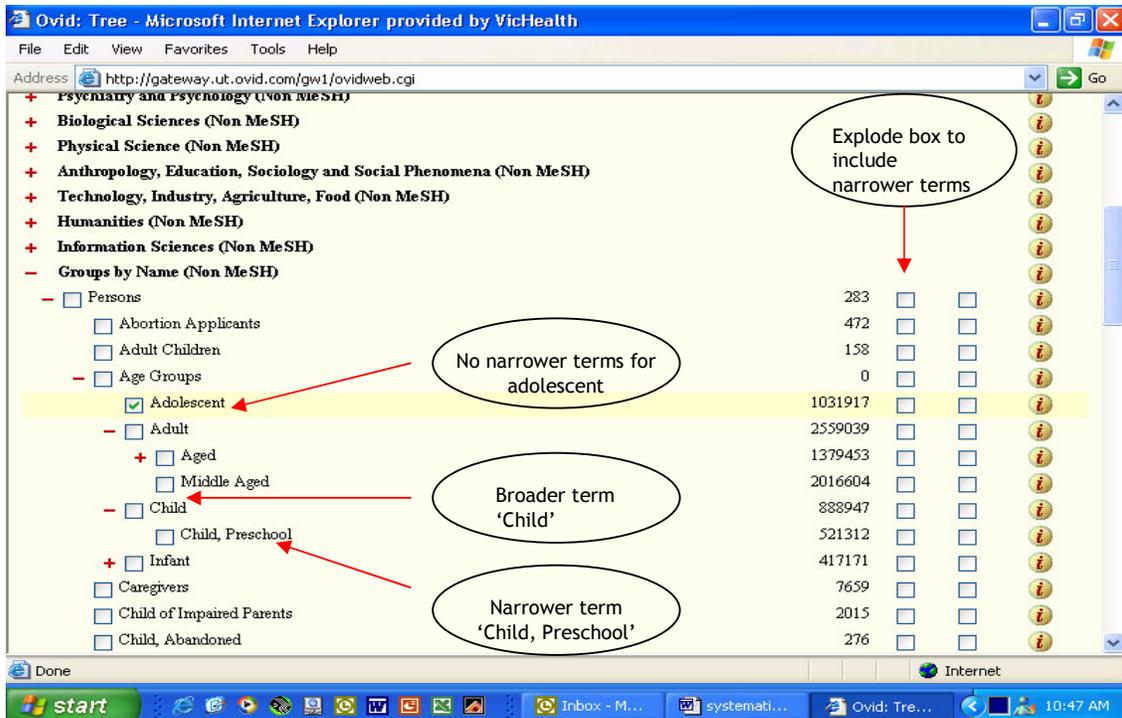
3. The result should look like this:



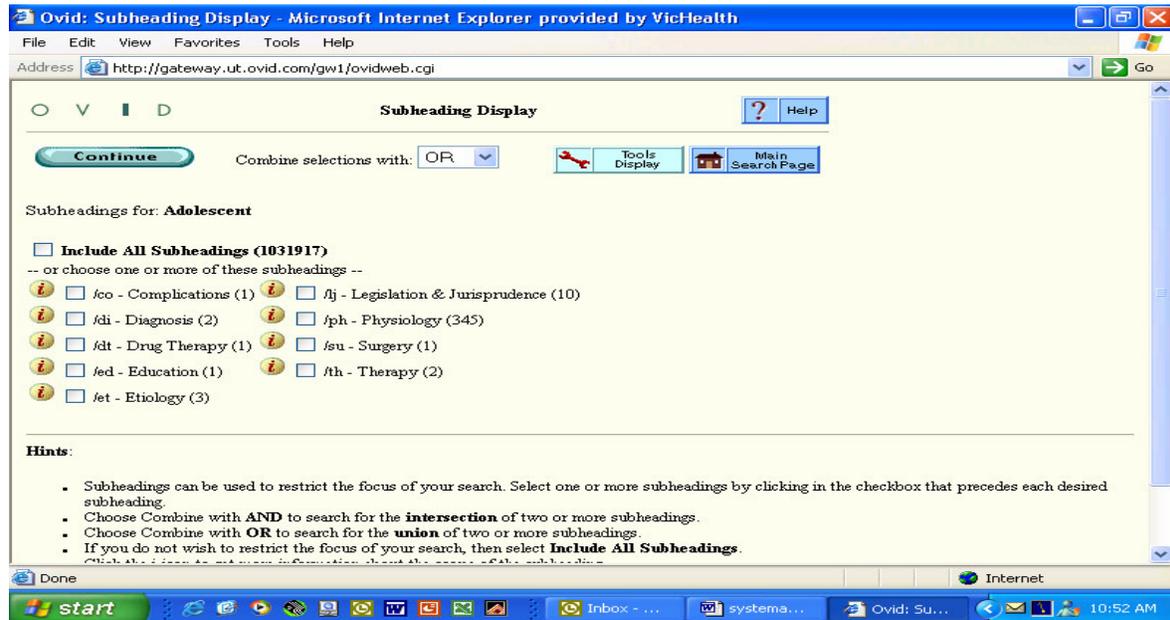
4. Click on the scope note for the **Adolescent** term (i symbol) to find the definition of adolescent and terms related to adolescent that can also be used in the search strategy. Note that Minors can also be used for the term adolescent.



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

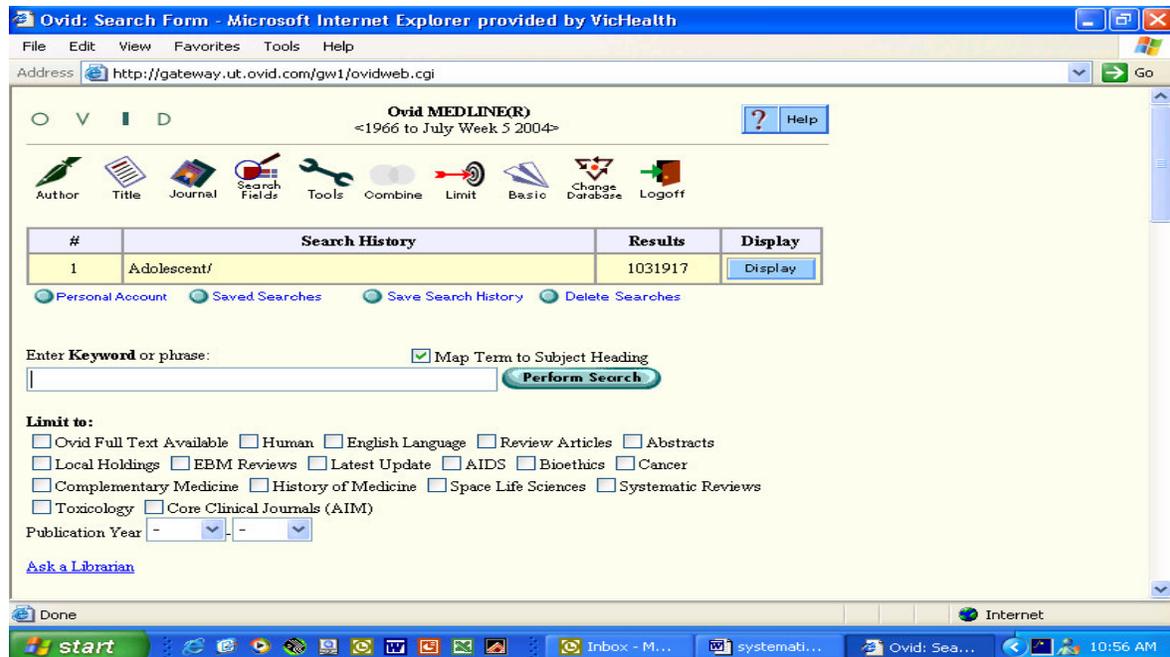


5. Because adolescent has no narrower terms click 'continue' at the top of the screen. This will produce a list of all subheadings.
 (If adolescent had narrower terms that are important to include the explode box would be checked).

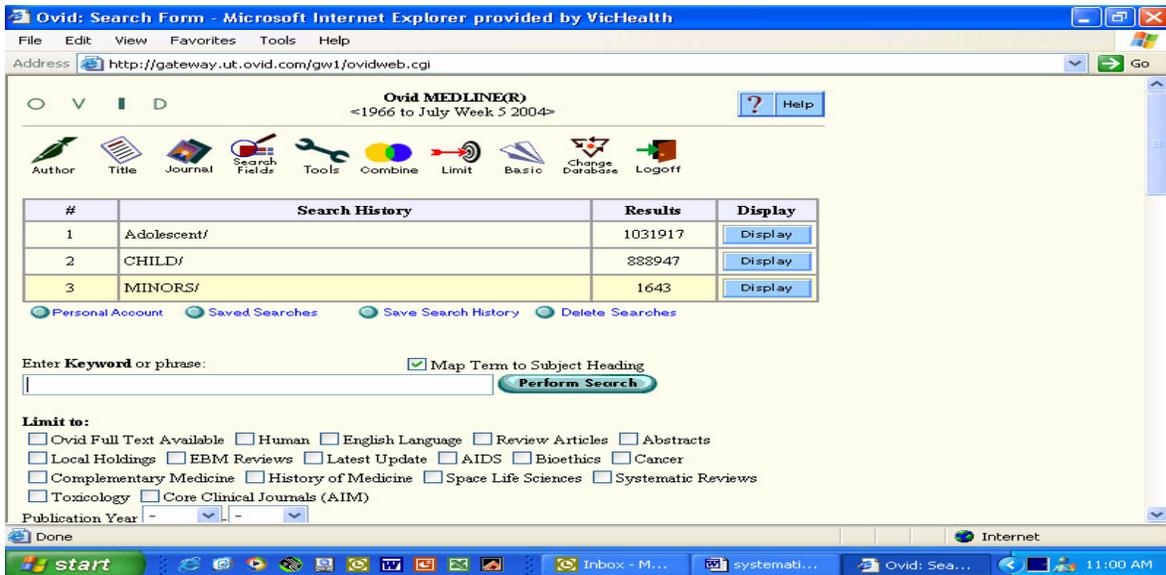


6. Press continue (it is not recommended to select any of the subheadings for public health reviews).

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



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



9. Using freetext or text-words to identify articles.

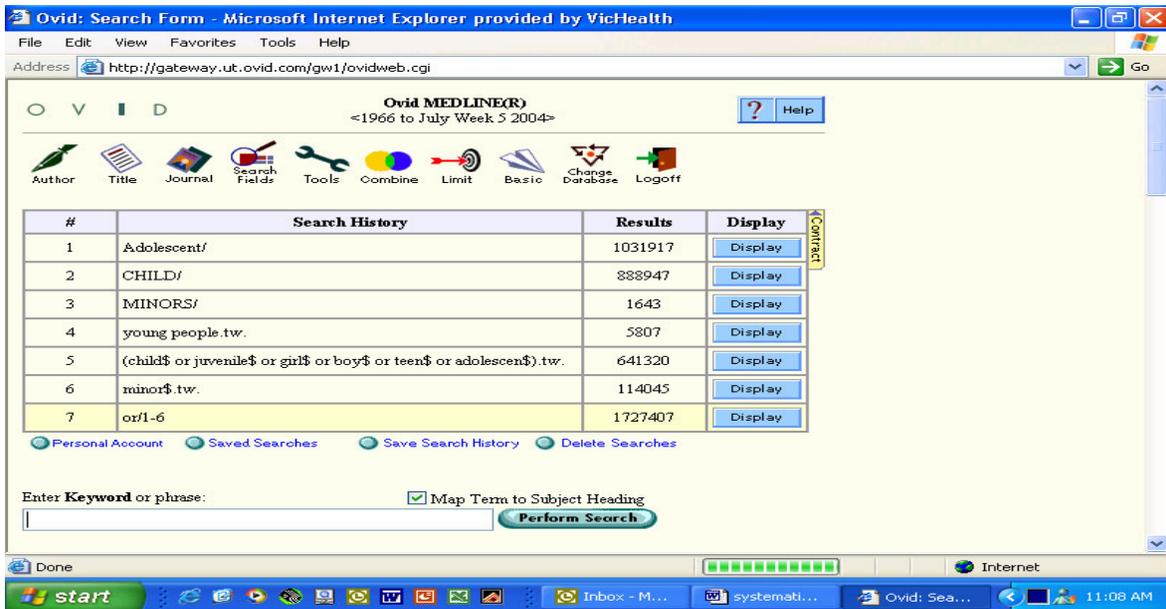
Truncation - \$ - Unlimited truncation is used to retrieve all possible suffix variations of a root word. Type the desired root word or phrase followed by either of the truncation characters "\$" (dollar sign). Another wild card character is "?" (question mark). It can be used within or at the end of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

10. Freetext words for searching – **type in young people.tw**.

You can also combine all textwords 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 adolescens\$.tw**.



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



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

13. Well done! Now try a search using the PICO question developed earlier 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 textwords.

Good luck!

FACILITATOR'S GUIDE

Time required:

3 hours and 15 minutes approx (slides 45 minutes, 2 hrs and 30 minutes practical session)

Note: this time includes 15-20 minutes for morning tea

Learning topics

- Developing a search strategy
- Using electronic databases
- Grey literature

Summary of activity

Lecture – Powerpoint slides

Practical exercise on the computer

Description of supporting materials

Examples of searching strategies and worked example (provided in handbook)

Instructions to facilitator

*** Note: You will need to obtain access to Medline for your searching. Contact your nearest library to seek access/passwords or try contacting OVID (www.ovid.com or 1800 226 474 (in Australia).

This can be quite difficult to teach, especially to participants who are new to electronic searching. Go through the worked example and slides slowly to make sure participants understand all of the terminology used.

Reiterate that MeSH headings only relate to Medline – they are called subject headings or descriptors in other databases. In addition, note that other databases may have different syntax (NEXT, ADJ, NEAR, \$, *, ?, etc) so it is important to be familiar with the database before starting database searching. Finally, emphasise that search strategies have to be adapted for each database, as subject headings are likely to be different. Textwords can remain the same in each database (although some databases may not have truncation, etc).

When in the computer room, go through the worked example with participants, rather than having them go through it at their own pace. When participants move on to their own question, encourage them to brainstorm as many terms as possible to describe each element of their PICO question.

If participants complete both exercises (worked example and their own PICO question) they can log on to the Cochrane Library or visit the websites of organisations that have databases of primary studies (eg. EPPI-Centre, Health Development Agency) or systematic reviews.

Unit Seven: Data Abstraction

Learning Objectives

- To understand the importance of a well-designed, unambiguous data abstraction form
- To identify the necessary data to abstract/extract from the primary studies

Once data has been abstracted from primary studies the synthesis of findings becomes much easier. The data abstraction form becomes a record to refer back to during the latter stages of the review process. In addition, the forms may be of use to future reviewers who wish to update the review.

Different study designs will require different data abstraction forms, to match the quality criteria and reporting of the study. The data abstraction form should mirror the format for which the results will be presented.

Details to collect:

Sometimes, the data required for synthesis is not reported in the primary studies, or is reported in a way that isn't useful for synthesis. Studies vary in the statistics they use to summarise the results (medians rather than means) and variation (standard errors, confidence intervals, ranges instead of standard deviations).¹ It is therefore important that authors are contacted for any additional details of the study.

** It is possible that one study is reported in more than one journal (duplication of publication). In addition, different aspects of the study (process outcomes, intervention details, outcome evaluations) may be reported in different publications. All of the papers from the study can be used to assist with data abstraction. However each paper should have a unique identifier in the data abstraction form to record where the information was located.

The data abstraction form should be piloted on a small group of studies to ensure the form captures all of the information required. In addition, if there is more than one reviewer a selection of studies should be tested to see if the reviewers differ in the interpretation of the details of the study and data abstraction form. If reviewers do not reach a consensus they should try to determine why their accounts differ.

The data abstraction form should contain the criteria used for quality appraisal. If the study does not meet the pre-determined criteria for quality there is no point continuing with the data abstraction process.

Useful data to collect:

- Publication details
- Study details (date, follow-up)
- Study design
- Population details (n, characteristics)
- Intervention details
- Theoretical framework
- Provider
- Setting
- Target group
- Consumer involvement

- Process measures – adherence, exposure, training, etc
- Context details
- Outcomes and findings

Examples of data abstraction forms:

A number of data abstraction forms are available in the following publication: 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

Other data abstraction forms can be found at:

- The Effective Public Health Practice Project reviews – (appendices in reviews) <http://www.city.hamilton.on.ca/phcs/EPHPP/default.asp>
- The Community Guide <http://www.thecommunityguide.org/methods/abstractionform.pdf>
- Effective Practice and Organisation of Care Review Group <http://www.epoc.uottawa.ca/tools.htm>
- NHS CRD Report Number 4. http://www.york.ac.uk/inst/crd/crd4_app3.pdf

Please note: No single data abstraction form is absolutely suitable for every review. Forms will need to be adapted to make them relevant to the information required for the review.

REFERENCES

1. Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. <http://www.cochrane.org/resources/handbook/index.htm>

FACILITATOR'S GUIDE

Time required:

10 minutes approx (slides only)

Learning topics

- Rationale for well-designed and ambiguous data collection forms
- Types of data to collect
- Grey literature

Summary of activity

Lecture – Powerpoint slides

Description of supporting materials

None

Instructions to facilitator

1. Direct participants to online examples of data abstraction forms. Emphasise that the data abstraction form should be designed to match the needs of the review – it is difficult to create a standard data abstraction form that would meet the needs of all reviews.
2. The data abstraction form should contain the quality criteria for critical appraisal. There is no point extracting data from a study if it has not met the quality criteria/standard determined at the beginning of the review.
3. In situations where there is more than one reviewer, the form should be tested to ensure that the form is not interpreted differently by reviewers, i.e. is reliable.
4. Data in primary studies is often not reported in a way that is useful for completing systematic reviews. Therefore, it is usually necessary to contact the authors for the data required (eg. means and SDs).

Unit Eight: Principles of Critical Appraisal

Learning Objectives

- To understand the components that relate to quality of a quantitative and qualitative primary study
- To understand the term 'bias' and types of bias
- To gain experience in the assessment of the quality of a health promotion or public health primary study (qualitative and quantitative)

1) QUANTITATIVE STUDIES

Validity

Validity is the degree to which a result from a study is likely to be true and free from bias.¹ Interpretation of findings from a study depends on both internal and external validity.

Internal validity

The extent to which the observed effects are true for people in a study.¹ Common types of bias that affect internal validity include; allocation bias, confounding, blinding, data collection methods, withdrawals and dropouts, statistical analysis, and intervention integrity (including contamination). Unbiased results are internally valid.

External validity (generalisability or applicability)

The extent to which the effects in a study truly reflect what can be expected in a target population beyond the people included in the study.¹ Note: Only results from internally valid studies should be considered for generalisability.

Critical appraisal tools

1) RCTs, non-randomised controlled studies, uncontrolled studies

- The Quality Assessment Tool for Quantitative Studies (<http://www.city.hamilton.on.ca/phcs/EPHPP/>).

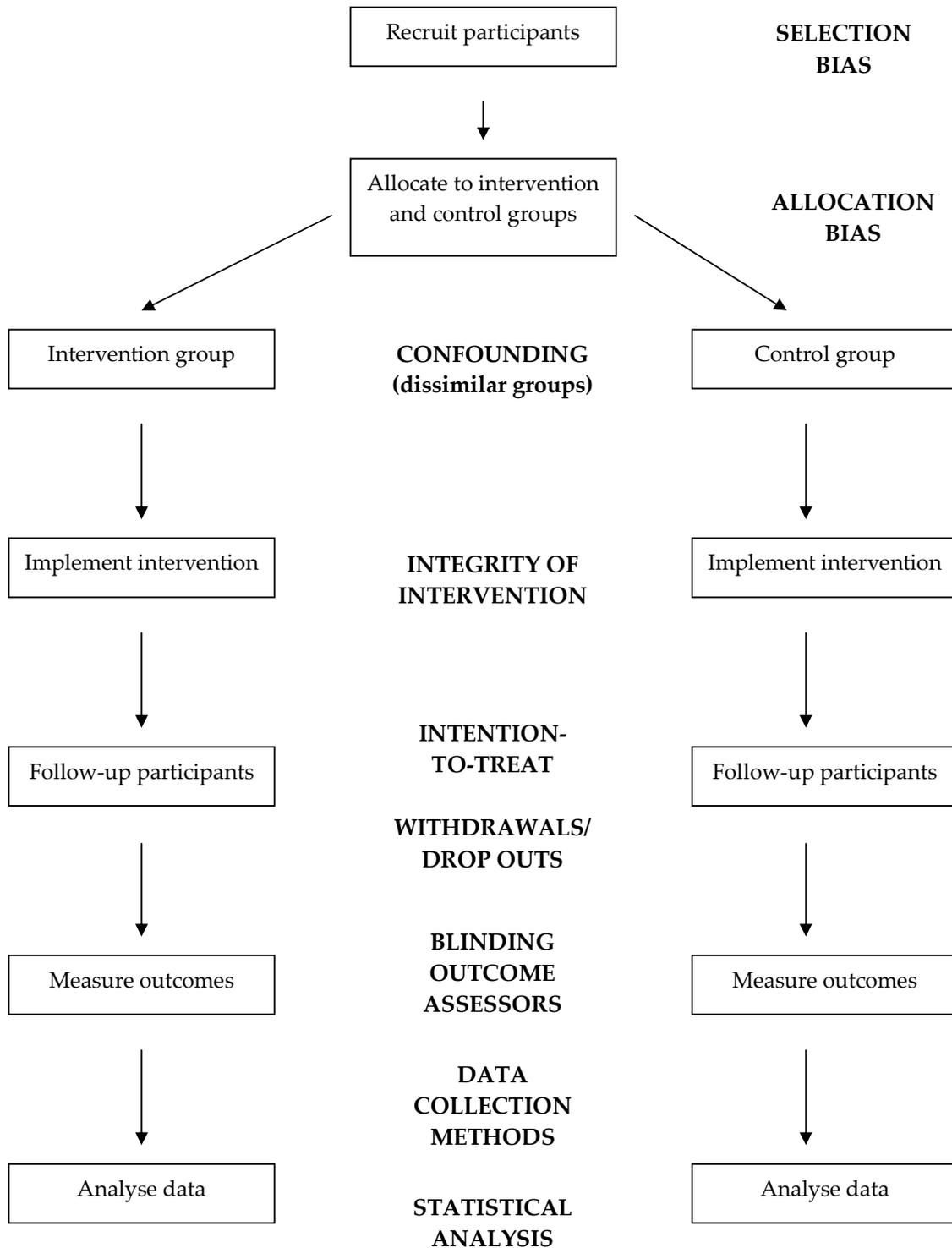
Developed by the Effective Public Health Practice Project, Canada. This tool assesses both internal and external validity. Content and construct validity have been established.² Rates the following criteria relevant to public health studies:

- | | |
|---------------------------------------|--|
| 1) selection bias (external validity) | 6) withdrawals and dropouts (attrition bias) |
| 2) allocation bias | 7) statistical analysis |
| 3) confounding | 8) intervention integrity |
| 4) blinding (detection bias) | |
| 5) data collection methods | |

2) Interrupted time series designs

- Methods for the appraisal and synthesis of ITS designs are included on the Effective Practice and Organisation of Care website (www.epoc.uottawa.ca).

Introduction of bias into the conduct of a primary study



Types of bias in health promotion and public health studies

Bias

A systematic error or deviation in results. Common types of bias in health promotion and public health studies arise from systematic differences in the groups that are compared (allocation bias), the exposure to other factors apart from the intervention of interest (eg. contamination), withdrawals from the study (attrition bias), assessment of outcomes (detection bias), including data collection methods, and inadequate implementation of the intervention.

The following sections of this unit describe the types of bias to be assessed using The Quality Assessment Tool for Quantitative Studies (<http://www.city.hamilton.on.ca/phcs/EPHPP/>) developed by the Effective Public Health Practice Project, Canada. Further information is also provided in the Quality Assessment Dictionary provided in the following pages.

1) Selection bias

Selection bias is used to describe a systematic difference in characteristics between those who are selected for study and those who are not. As noted in the Quality Assessment Dictionary, it occurs when the study sample (communities, schools, organisations, etc) does not represent the target population for whom the intervention was intended. Examples:

- Results from a teaching hospital may not be generalisable to those in non-teaching hospitals
- Results which recruited volunteers may not be generalisable to the general population
- Results from low SES schools or inner city schools may not be generalisable to all schools

Examples from www.re-aim.org

Example: Eakin and her associates (1998) illustrate selection bias in a smoking cessation study offered to participants in a planned-parenthood program. They begin by explicitly reporting their inclusion criteria --female smokers between 15 and 35 years of age who are patients at a planned-parenthood clinic. During a routine visit to the clinic the patient services staff described the study and solicited participants. Those women who declined (n=185) were asked to complete a short questionnaire that included questions to assess demographics, smoking rate, and reasons for non-participation. Participants (n=518) also completed baseline demographic and smoking rate assessments. They tracked recruitment efforts and reported that 74% percent of the women approached agreed to participate in the study. To determine the representativeness of the sample two procedures were completed. First, based on information from patient medical charts, those who were contacted were compared on personal demographics to those who were not contacted. Second, participants were compared to non-participants on personal demographics and smoking rate. The study found that those contacted did not differ from those not contacted on any of the test variables. Also, the results suggested that participants were slightly younger than non-participants, but there were no other differences between these groups. This suggests that Eakin and her associates were fairly successful in contacting and recruiting a fairly representative sample of their target population.

Example: The Language for Health (Elder et al., 2000) nutrition education intervention provides a good example of determining the representativeness of study participants to a given target population. The behaviour change intervention was developed to target Latino participants in English as a second language (ESL) classes at seven schools. To examine representativeness, the 710 participants in the study were compared to the overall Latino ESL student population in the city. This comparison revealed that the intervention participants did not differ from the general ESL student

population on gender, age, or education level. As such, the authors concluded that the study had strong generalisability to the greater target population (Elder et al., 2000).

Example: All the participating schools were state primary schools sited outside the inner city area. Socio-demographic measures suggested that the schools' populations generally reflected the Leeds school aged population, although there was a slight bias towards more advantaged children. The schools had 1-42% children from ethnic minorities and 7-29% entitled to free school meals compared with 11% and 25% respectively for Leeds children as a whole.

2) Allocation bias

Bias can result from the way that the intervention and control groups are assembled.³ Unless groups are equivalent or balanced at baseline, differences in outcomes cannot confidently be attributed to the effects of the intervention.⁴ Studies which show that comparison groups are not equivalent at baseline have high allocation bias.

Random allocation is the best method to produce comparison groups that are balanced at baseline for known and unknown confounding factors, and therefore reduce allocation bias. This is usually achieved by coin-tossing or developing computer-generated random number tables. This ensures that every participant in the study has an equal chance (50%/50%) of being in the intervention or control group.

Ideally, the coin-tossing or computer-generated randomisation should be carried out by individuals external to the study. Once the allocation scheme is developed, the allocation of participants to intervention and control group should be carried out by someone who is not responsible for the study to prevent manipulation by researchers and participants. Therefore, once the allocation scheme has been developed it is important that allocation to intervention and control group is concealed. Concealment of allocation is the process to prevent foreknowledge of group assignment.¹ Methods to conceal allocation include allocation by persons external to the study and sequentially numbered, sealed opaque envelopes. Unfortunately, information on concealment of allocation is very rarely reported in primary studies.

Example: Worksites were randomised within blocks: unionised versus non-unionised; single versus multiple buildings; and three worksites that were part of a single large company. Worksites were randomly assigned by the study biostatistician using a process conducted independently from the intervention team.

Example: Subjects were randomised to one of three arms: (1) Direct Advice, (2) Brief Negotiation or (3) Control by household with each monthly batch forming a single permuted block. Randomisation of intervention arms were sent to CF (the investigator) in sealed opaque envelopes. At the health check participants were asked to consent to a randomised trial of the effect of health professionals' communication style on patient's health behaviour, namely physical activity. If consent was given, the envelope was opened and the appropriate intervention carried out.

There are also quasi-randomised methods of allocating participants into intervention and control groups. These include alternation (eg. first person intervention, second person control), allocation by date of birth, day of week, etc. These methods are not able to conceal allocation, do not guarantee that every participant has an equal chance of being in either comparison group, and consequentially do not guarantee that groups will be similar at baseline.

Example: Families then were randomly assigned to an intervention (n = 65) or control group (n = 70). An alternate-day randomisation system was used to simplify intervention procedures and more importantly to avoid waiting-room contamination of control families by intervention families exiting the rooms with books and handouts.

Non-randomised studies often involve the investigators choosing which individuals or groups are allocated to intervention and control groups. Therefore, these study designs have high allocation bias and are likely to produce uneven groups at baseline. Even if every attempt has been made to match the intervention and control groups it is impossible to match for unknown confounding factors. Furthermore, there are inherent problems in assessing known confounding factors, as measurement tools for collecting the information may not be valid.

3) Confounding

Confounding is a situation where there are factors (other than the intervention) present which influence the outcome under investigation. A confounding factor has to be related to both the intervention and the outcome. For example, Body Mass Index at baseline would be a confounding factor when investigating the effect of school based nutrition intervention on preventing obesity. A factor can only confound an association if it differs between the intervention and control groups.

The assessment of confounding is the next stage in the critical appraisal process after determining the method of allocation. Remember, randomisation of participants or groups to intervention/control group is the best way to distribute known and unknown confounding factors evenly. Differences between groups in baseline characteristics that relate to the outcome may distort the effect of the intervention under investigation.

Before beginning to answer this critical appraisal question it is important to determine the potential confounding factors relating to the particular intervention under question. Good knowledge of the subject area is essential when determining potential confounders.

Example:

Presence of confounders: Intervention and control subjects were similar on baseline variables.

Adjustment for confounders: We assessed the effect of the intervention after adjusting for sex, age, baseline BMI and type of school.

4) Blinding of outcome assessors (detection bias)

Outcome assessors who are blind to the intervention or control status of participants should logically be less biased than outcome assessors who are aware of the status of the participants.

Detection bias is important in health promotion studies where outcomes are generally subjective. For example, if outcome assessors were required to interview children regarding their food consumption in the past 24 hours, they may be more likely to prompt the intervention group to respond favourably.

Example: Questionnaires were developed based on a review of other STD/HIV risk questionnaires and our findings from focus groups and in-depth interviews. When administering the 3- and 9-month follow-up questionnaires, interviewers were blind to the study group assignment of adolescents.

5) Data collection methods

As highlighted, a number of outcomes measured in health promotion are subjectively reported. Although a number of outcomes can be measured objectively, such as Body Mass Index or pregnancy, generally health promotion interventions are trying to change behaviour, which usually requires subjective self-reporting (unless behaviour is directly observed). Subjective outcome data must be collected with valid and reliable instruments.

Critical appraisal therefore requires the reader to assess whether the outcomes have been measured with valid and reliable instruments.

Example: We used three validated tools to evaluate the effect of the intervention on psychological well-being; the self-perception profile for children; a measure of dietary restraint; and the adapted body shape perception scale.

6) Withdrawals (attrition bias)

Attrition bias relates to the differences between the intervention and control groups in the number of withdrawals from the study. It arises because of inadequacies in accounting for losses of participants due to dropouts, leading to missing data on follow-up.⁴

If there are systematic differences in losses to follow-up the characteristics of the participants in the intervention and control groups may not be as similar as they were at the beginning of the study. For randomised controlled trials, the effect of randomisation is lost if participants are lost to follow-up. An intention-to-treat analysis, where participants are analysed according to the group they were initially allocated, protects against attrition bias.

For cluster-level interventions all members of the cluster should be included in the evaluation, regardless of their exposure to the intervention.⁵ Thus, a sample of eligible members of the cluster is generally assessed, not only those who were sufficiently motivated to participate in the intervention.⁵ Therefore, it is said that studies tracking change in entire communities are likely to observe smaller effect sizes than other studies tracking change in intervention participants alone.⁵

Example: Twenty one (14%) of the 148 patients who entered the trial dropped out, a rate comparable to that in similar trials. Of these, 19 were in the intervention group and dropped out during treatment (eight for medical reasons, seven for psychiatric reasons, four gave no reason, one emigrated, and one was dissatisfied with treatment).

Example: Completed follow-up responses were obtained from 87% of surviving intervention patients and 79% of surviving control patients. There were no significant differences between respondents and non-respondents in age, sex, educational achievement, marital status, or baseline health status.

7) Statistical analysis

A trial/study must have a sufficient sample size to have the ability (or power) to detect significant differences between comparison groups. A lack of a significant effect could be due to the study having insufficient numbers, rather than the intervention being ineffective.

The publication of the study should report whether a sample size calculation was carried out. For group/cluster studies the study should report that it took the clustering into account when calculating sample size. These types of study designs should also analyse the data appropriately; if schools/classrooms were allocated to intervention and control groups then they must be analysed at

this level. Often this is not the case, as the intervention is allocated to schools (for practical reasons) and individual outcomes (eg. behaviour change) are analysed. In these instances, a cluster analysis (taking into account the different levels of allocation and analysis) should be reported.

For further information about cluster studies read "Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess.* 1999;3(5):iii-92."

Example: A power calculation indicated that with five schools in each arm, the study would have 80% power to detect and underlying difference in means of a normally distributed outcome measure of ≥ 1.8 standard deviations at the 5% significance level and 65% to detect a difference of 1.5 SD. This took into account the cluster randomisation design.

Example: The statistical model took into account the lack of independence among subjects within the school, known as the clustering effect.

8) Integrity of intervention

Critical appraisal should determine if results of ineffectiveness within primary studies is simply due to incomplete delivery of the intervention (failure of implementation) or a poorly conceptualised intervention (failure of intervention concept or theory)^{6,7}. Evaluating a program that has not been adequately implemented is also called a Type III error⁸. Assessing the degree to which interventions are implemented as planned is important in preventive interventions which are often implemented in conditions that present numerous obstacles to complete delivery.⁶ A review of smoking cessation in pregnancy⁹ found that in studies which measured the implementation of the intervention the implementation was less than ideal.

In order to provide a comprehensive picture of intervention integrity five dimensions of the intervention should be measured. These five factors are adherence, exposure, quality of delivery, participant responsiveness, and program differentiation (to prevent contamination).⁶

Adherence: the extent to which specified program components were delivered as prescribed in program manuals.

Exposure: an index that may include any of the following: (a) the number of sessions implemented; (b) the length of each session; or (c) the frequency with which program techniques were implemented.

Quality of delivery: a measure of qualitative aspects of program delivery that are not directly related to the implementation of prescribed content, such as implementer enthusiasm, leader preparedness and training, global estimates of session effectiveness, and leader attitude towards the program.

Participant responsiveness: a measure of participant response to program sessions, which may include indicators such as levels of participation and enthusiasm.

Program differentiation: a manipulation check that is performed to safeguard against the diffusion of treatments, that is, to ensure that the subjects in each experimental condition received only the planned interventions. Contamination may be a problem within many public health and health promotion studies where intervention and control groups come into contact with each other. This bias is minimised through the use of cluster RCTs.

These data provide important information that enhances the ability to interpret outcome assessments, identify competing explanations for observed effects and measure exposure to the intervention.⁵ However, very few studies disentangle the factors that ensure successful outcomes, characterise the failure to achieve success, or attempt to document the steps involved in achieving successful implementation of complex interventions.^{10,11}

In relation to the appraisal of process evaluations the EPPI-Centre has developed a 12-question checklist, available at:

http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/hp/reports/phase/phase_process.htm.

Does the study focus on the **delivery** of a health promotion intervention?

Screening questions

1. Does the study focus on a health promotion intervention?
2. Does the intervention have clearly stated aims?
3. Does the study describe the key processes involved in delivering the intervention?

Detailed questions

4. Does the study tell you enough about planning and consultation?
 5. Does the study tell you enough about the collaborative effort required for the intervention?
 6. Does the study tell you enough about how the target population was identified and recruited?
 7. Does the study tell you enough about education and training?
- B) What are the results?
8. Were all the processes described and adequately monitored?
 9. Was the intervention acceptable?
- C) Will the results help me?
10. Can the results be applied to the local population?
 11. Were all important processes considered?
 12. If you wanted to know whether this intervention promotes health what outcomes would you want to measure?

Examples of assessment of the intervention implementation

Example: This study evaluated a 19-lesson, comprehensive school-based AIDS education program lasting one year in rural southwestern Uganda. Quantitative data collection (via questionnaire) found that the program had very little effect on overall knowledge, overall attitude, intended condom use, and intended assertive behaviour. Data from the focus group discussions suggested that the program was incompletely implemented, and that key activities such as condoms and the role-play exercises were only completed superficially. The main reasons for this were a shortage of classroom time, as well as teachers' fear of controversy (condoms are an unwelcome intrusion into African tradition and may be associated with promiscuity). Teachers tended to teach only the activities that they preferred, leaving out the activities they were reluctant to teach. One problem with the intervention was that the program was additional to the standard curriculum, so teaching time was restricted. It was also found that neither teachers nor students were familiar with roleplay. Furthermore, a number of teachers also left the intervention schools (or died).

Therefore, it is suggested that AIDS education programs in sub-Saharan Africa may be more fully implemented if they are fully incorporated into national curricula (see interpretation or results unit) and examined as part of school education.

References:

Kinsman J, Nakiyingi J, Kamali A, Carpenter L, Quigley M, Pool R, Whitworth J. Evaluation of a comprehensive school-based AIDS education programme in rural Masaka, Uganda.

Health Educ Res. 2001 Feb;16(1):85-100.

Kinsman J, Harrison S, Kengeya-Kayondo J, Kanyesigye E, Musoke S, Whitworth J. Implementation of a comprehensive AIDS education programme for schools in Masaka District, Uganda. *AIDS Care*. 1999 Oct;11(5):591-601.

Example: Gimme 5 Fruit, Juice and Vegetable intervention. This school-based intervention included components to be delivered at the school and newsletters with family activities and instructions for intervention at home. Overall, there were small changes in fruit, juice, and vegetable consumption. Teacher self-reported delivery of the intervention was 90%. However, all teachers were observed at least once during the 6-week intervention and it was found that only 51% and 46% of the curriculum activities were completed in the 4th and 5th grade years. Reference: Davis M, Baranowski T, Resnicow K, Baranowski J, Doyle C, Smith M, Wang DT, Yaroch A, Hebert D. Gimme 5 fruit and vegetables for fun and health: process evaluation. *Health Educ Behav*. 2000 Apr;27(2):167-76.

REFERENCES

1. Cochrane Reviewers' Handbook Glossary, Version 4.1.5. www.cochrane.org/resources/handbook/glossary.pdf, 6 December 2004 [date last accessed]
2. Thomas H, Micucci S, Thompson O'Brien MA, Briss P. Towards a reliable and valid instrument for quality assessment of primary studies in public health. Unpublished work. 2001.
3. Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. <http://www.cochrane.dk/cochrane/handbook/handbook.htm>
4. 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. March 2001. <http://www.york.ac.uk/inst/crd/report4.htm>
5. Sorensen G, Emmons K, Hunt MK, Johnston D. Implications of the results of community intervention trials. *Annu Rev Public Health*. 1998;19:379-416.
6. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin Psychol Rev* 1998;18:23-45.
7. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;56:119-27.
8. Basch CE, Sliepecevic EM, Gold RS, Duncan DF, Kolbe LJ. Avoiding type III errors in health education program evaluations: a case study. *Health Educ Q*. 1985 Winter;12(4):315-31.
9. Lumley J, Oliver S, Waters E. Interventions for promoting smoking cessation during pregnancy. In: *The Cochrane Library*, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd.
10. Steckler A, Linnan L (eds). *Process Evaluation for Public Health Interventions and Research*. Jossey-Bass, USA, 2002.

11. Green J, Tones K. Towards a secure evidence base for health promotion. *Journal of Public Health Medicine* 1999;21(2):133-9.

ADDITIONAL READING

Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2000;56:119-27.

Kahan B, Goodstadt M. The IDM Manual for Using the Interactive Domain Modal Approach to Best Practices in Health Promotion.

www.utoronto.ca/chp/bestp.html#Outputs/Products

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.

Notes on terms/statistics used in primary studies: Adapted from the Cochrane Reviewers' Handbook Glossary, Version 4.1.5. Available at www.cochrane.org/resources/handbook/glossary.pdf

Bias

A systematic error or deviation in results. Common types of bias in health promotion and public health studies arise from systematic differences in the groups that are compared (allocation bias), the exposure to other factors apart from the intervention of interest (eg. contamination), withdrawals from the study (attrition bias), assessment of outcomes (detection bias), including data collection methods, and inadequate implementation of the intervention.

Blinding

Keeping secret group assignment (intervention or control) from the study participants or investigators. Blinding is used to protect against the possibility that knowledge of assignment may affect subject response to the intervention, provider behaviours, or outcome assessment. The importance of blinding depends on how objective the outcome measure is; blinding is more important for less objective measures.

Confidence Interval (CI)

The range within which the 'true' value (eg. size of effect of the intervention) is expected to lie within a given degree of certainty (eg. 95%). It is about the precision of the effect. CI's therefore indicate the spread or range of values which can be considered probable. The narrower the CI the more precise we can take the findings to be.

Confounding

A situation in which the measure of the effect of an intervention or exposure is distorted because of the association of exposure with other factors that influence the outcome under investigation.

Intention to treat

An intention-to-treat analysis is one in which all the participants in the trial are analysed according to the intervention to which they are allocated, whether they received it or not.

Odds ratios

The ratio of the odds of an event (eg. prevention of smoking, unintended pregnancy) in the intervention group to the odds of an event in the control group.

p-value

The probability (from 0 to 1) that the results observed in a study could have occurred by chance. They are used as a benchmark of how confident we can be in a particular result. You will often see statements like 'this result was significant at $p < 0.05$ '. This means that we could expect this result to occur by chance no more than 5 times per 100 (one in twenty). The level of $p < 0.05$ is conventionally regarded as the lowest level at which we can claim statistical significance.

Relative risk

The ratio of the risks of an event (eg. prevention of smoking, unintended pregnancy) in the intervention group to the odds of an event in the control group. eg. $RR = 0.80$ for unintended pregnancy – the intervention group had a 20% reduced risk of unintended pregnancy compared to those in the control group. Note: a $RR < 1$ is good if you want less of something (pregnancy, death, obesity), a $RR > 1$ is good if you want more of something (people stopping smoking, using birth control).

EXERCISE

1. Join the group who are appraising the same paper that you received prior to the workshop.

(a) Randomised controlled study

Use the quality checklist provided (and considering the Public Health Schema) to appraise: "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;323:1029-1032". This exercise also includes the process evaluation: "Sahota P, Rudolf MC, Dixey R, Hill AJ, Barth JH, Cade J. Evaluation of implementation and effect of primary school based intervention to reduce risk factors for obesity. *BMJ* 2001;323:1027-9."

(b) Controlled before and after study

Use the quality checklist provided to appraise: "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 Adolesc Med* 1999;153:975-983".

(Q2) If there were differences between groups for important confounders, were they adequately managed in the analysis?

Yes No Not Applicable

(Q3) Were there important confounders not reported in the paper?

Yes No

Relevant Confounders NOT reported in the study:

| | | | |
|---|---------------|-----------------|-------------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|---|---------------|-----------------|-------------|

D) BLINDING

(Q1) Was (were) the outcome assessor(s) blinded to the intervention or exposure status of participants?

Yes No Not reported Not applicable

| | | | |
|---|---------------|-----------------|-------------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|---|---------------|-----------------|-------------|

E) DATA COLLECTION METHODS

(Q1) Were data collection tools shown or are they known to be valid?

Yes No

(Q2) Were data collection tools shown or are they known to be reliable?

Yes No

| | | | |
|---|---------------|-----------------|-------------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|---|---------------|-----------------|-------------|

F) WITHDRAWALS AND DROP-OUTS

(Q1) Indicate the percentage of participants completing the study. (If the percentage differs by groups, record the lowest).

80 -100% 60 - 79% Less than 60% Not Reported Not Applicable

| | | | |
|---|---------------|-----------------|-------------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|---|---------------|-----------------|-------------|

G) ANALYSIS

(Q1) Is there a sample size calculation or power calculation?

Yes Partially No

(Q2) Is there a statistically significant difference between groups?

Yes No Not Reported

(Q3) Are the statistical methods appropriate?

Yes No Not Reported

(Q4a) Indicate the unit of allocation (circle one)

Community Organization/ Group Provider Client Institution

(Q4b) Indicate the unit of analysis (circle one)

Community Organization/ Group Provider Client Institution

(Q4c) If 4a and 4b are different, was the cluster analysis done?

Yes No Not Applicable

(Q5) Is the analysis performed by intervention allocation status (i.e. intention to treat) rather than the actual intervention received?

Yes No Can't Tell

H) INTERVENTION INTEGRITY

(Q1) What percentage of participants received the allocated intervention or exposure of interest?

80 -100% 60 - 79% Less than 60% Not Reported Not Applicable

(Q2) Was the consistency of the intervention measured?

Yes No Not reported Not Applicable

Q3) Is it likely that subjects received an unintended intervention (contamination or cointervention) that may influence the results?

Yes No Can't tell

SUMMARY OF COMPONENT RATINGS

Please transcribe the information from the grey boxes on pages 1-3 onto this page.

A SELECTION BIAS

| | | | |
|------------------------------------|--------|----------|------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|------------------------------------|--------|----------|------|

B STUDY DESIGN

| | | | |
|------------------------------------|--------|----------|------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|------------------------------------|--------|----------|------|

C CONFOUNDER

| | | | |
|------------------------------------|--------|----------|------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|------------------------------------|--------|----------|------|

D BLINDING

| | | | |
|------------------------------------|--------|----------|------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|------------------------------------|--------|----------|------|

E DATA COLLECTION METHODS

| | | | |
|------------------------------------|--------|----------|------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|------------------------------------|--------|----------|------|

F WITHDRAWALS AND DROPOUTS

| | | | |
|------------------------------------|--------|----------|------|
| Rate this section (see dictionary) | Strong | Moderate | Weak |
|------------------------------------|--------|----------|------|

G ANALYSIS

Comments _____

H INTERVENTION INTEGRITY

Comments _____

WITH BOTH REVIEWERS DISCUSSING THE RATINGS:

Is there a discrepancy between the two reviewers with respect to the component ratings?

No Yes

If yes, indicate the reason for the discrepancy

Oversight

Differences in
Interpretation of Criteria

Differences in
Interpretation of Study

DICTIONARY
for the
Effective Public Health Practice Project
Quality Assessment Tool for Quantitative Studies

INTRODUCTION

The purpose of this tool is to assess the methodological quality of relevant studies since lesser quality studies may be biased and could over-estimate or under-estimate the effect of an intervention. Each of two raters will independently assess the quality of each study and complete this form. When each rater is finished, the individual ratings will be compared. A consensus must be reached on each item. In cases of disagreement even after discussion, a third person will be asked to assess the study.

When appraising a study, it is helpful to first look at the design then assess other study methods. It is important to read the methods section since the abstract (if present) may not be accurate. Descriptions of items and the scoring process are located in the dictionary that accompanies this tool.

The scoring process for each component is located on the last page of the dictionary.

INSTRUCTIONS FOR COMPLETION

Circle the appropriate response in each component section (A-H). Component sections (A-F) are each rated using the roadmap on the last page of the dictionary. After each individual rater has completed the form, both reviewers must compare their ratings and arrive at a consensus.

The dictionary is intended to be a guide and includes explanations of terms.

The purpose of this dictionary is to describe items in the tool thereby assisting raters to score study quality. Due to under-reporting or lack of clarity in the primary study, raters will need to make judgements about the extent that bias may be present. When making judgements about each component, raters should form their opinion based upon information contained in the study rather than making inferences about what the authors intended.

A) SELECTION BIAS

Selection bias occurs when the study sample does not represent the target population for whom the intervention is intended. Two important types of biases related to sample selection are referral filter bias and volunteer bias. For example, the results of a study of participants suffering from asthma from a teaching hospital are not likely to be generalisable to participants suffering from asthma from a general practice. In volunteer bias, people who volunteer to be participants may have outcomes that are different from those of non-volunteers. Volunteers are usually healthier than non-volunteers.

Q1 Are the individuals selected to participate in the study likely to be representative of the target population?

| | |
|---|-----------------|
| The authors have done everything reasonably possible to ensure that the target population is represented. | Very likely |
| Participants may not be representative if they are referred from a source within a target population even if it is in a systematic manner (eg. patients from a teaching hospital for adults with asthma, only inner-city schools for adolescent risk). | Somewhat likely |
| Participants are probably not representative if they are self-referred or are volunteers (eg. volunteer patients from a teaching hospital for adults with asthma, inner-city school children with parental consent for adolescent risk) or if you can not tell. | Not likely |

Q2 What percentage of selected individuals agreed to participate?

| | |
|---|----------------|
| The % of subjects in the control and intervention groups that agreed to participate in the study before they were assigned to intervention or control groups. | % |
| There is no mention of how many individuals were approached to participate. | Not Reported |
| The study was directed at a group of people in a specific geographical area, city, province, broadcast audience, where the denominator is not known, eg. mass media intervention. | Not Applicable |

B) ALLOCATION BIAS

In this section, raters assess the likelihood of bias due to the allocation process in an experimental study. For observational studies, raters assess the extent that assessments of exposure and outcome are likely to be independent. Generally, the type of design is a good indicator of the extent of bias. In stronger designs, an equivalent control group is present and the allocation process is such that the investigators are unable to predict the sequence.

Q1 Indicate the study design

| | |
|---|--|
| Investigators randomly allocate eligible people to an intervention or control group. | RCT |
| <i>Cohort (two group pre and post)</i> Groups are assembled according to whether or not exposure to the intervention has occurred. Exposure to the intervention may or may not be under the control of the investigators. Study groups may not be equivalent or comparable on some feature that affects the outcome. | Two-group Quasi Experimental |
| <i>Before/After Study (one group pre + post)</i> The same group is pretested, given an intervention, and tested immediately after the intervention. The intervention group, by means of the pretest, act as their own control group. | Case-control, Before/ After Study or No Control Group |
| <i>Case control study</i> A retrospective study design where the investigators gather 'cases' of people who already have the outcome of interest and 'controls' that do not. Both groups are then questioned or their records examined about whether they received the intervention exposure of interest. | |
| <i>No Control Group</i> | |

Note: The following questions are not for rating but for additional statistics that can be incorporated in the writing of the review.

(i) **If the study was reported as an RCT was the method of random allocation stated?**

| | |
|--|-----|
| The method of allocation was stated. | YES |
| The method of allocation was not stated. | NO |

(ii) **Is the method of random allocation appropriate?**

| | |
|---|-----|
| The method of random allocation is appropriate if the randomization sequence allows each study participant to have the same chance of receiving each intervention and the investigators could not predict which intervention was next. eg. an open list of random numbers of assignments or coin toss | YES |
| The method of random allocation is not entirely transparent, eg. the method of randomization is described as alternation, case record numbers, dates of birth, day of the week. | NO |

(iii) **Was the method of random allocation concealed?**

| | |
|---|-----|
| The randomization allocation was concealed so that each study participant had the same chance of receiving each intervention and the investigators could not predict which group assignment was next. Examples of appropriate approaches include assignment of subjects by a central office unaware of subject characteristics, or sequentially numbered, and sealed in opaque envelopes. | YES |
| The method of random allocation was not concealed or not reported as concealed. | NO |

C) CONFOUNDERS

A confounder is a characteristic of study subjects that:

- is a risk factor (determinant) for the outcome to the putative cause, or
- is associated (in a statistical sense) with exposure to the putative cause

Note: Potential confounders should be discussed within the Review Group and decided a priori.

Q1 Prior to the intervention were there differences for important confounders reported in the paper

| | |
|--|-----|
| The authors reported that the groups were balanced at baseline with respect to confounders (either in the text or a table) | NO |
| The authors reported that the groups were not balanced at baseline with respect to confounders. | YES |

Q2 Were the confounders adequately managed in the analysis?

| | |
|--|-----|
| Differences between groups for important confounders were controlled in the design (by stratification or matching) or in the | YES |
| No attempt was made to control for confounders. | NO |

Q3 Were there important confounders not reported?

| | |
|--|-----|
| describe | YES |
| All confounders discussed within the Review Group were reported. | NO |

D) BLINDING

The purpose of blinding the outcome assessors (who might also be the care providers) is to protect against detection bias.

Q1 Was (were) the outcome assessor(s) blinded to the intervention or exposure status of participants?

| | |
|--|----------------|
| Assessors were described as blinded to which participants were in the control and intervention groups. | YES |
| Assessors were able to determine what group the participants were in. | NO |
| The data was self-reported and was collected by way of a survey, questionnaire or interview. | Not Applicable |
| It is not possible to determine if the assessors were blinded or not. | Not Reported |

E) DATA COLLECTION METHODS

Some sources from which data may be collected are:

Self reported data includes data that is collected from participants in the study (eg. completing a questionnaire, survey, answering questions during an interview, etc.).

Assessment/Screening includes objective data that is retrieved by the researchers. (eg. observations by investigators).

Medical Records / Vital Statistics refers to the types of formal records used for the extraction of the data.

Reliability and validity can be reported in the study or in a separate study. For example, some standard assessment tools have known reliability and validity.

Q1 Were data collection tools shown or known to be valid for the outcome of interest?

| | |
|--|-----|
| The tools are known or were shown to measure what they were intended to measure. | YES |
| There was no attempt to show that the tools measured what they were intended to measure. | NO |

Q2 Were data collection tools shown or known to be reliable for the outcome of interest?

| | |
|--|-----|
| The tools are known or were shown to be consistent and accurate in measuring the outcome of interest (eg., test-retest, Cronback's alpha, interrater reliability). | YES |
| There was no attempt to show that the tools were consistent and accurate in measuring the outcome of interest. | NO |

F) WITHDRAWALS AND DROP-OUTS

Q1 Indicate the percentage of participants completing the study.

| | |
|---|----------------|
| The percentage of participants that completed the study. | % |
| The study was directed at a group of people in a specific geographical area, city, province, broadcast audience, where the percentage of participants completing, withdrawing or dropping-out of the study is not known, eg. mass media intervention. | Not Applicable |
| The authors did not report on how many participants completed, withdrew or dropped-out of the study. | Not Reported |

G) ANALYSIS

If you have questions about analysis, contact your review group leader.

- Q1. The components of a recognized formula are present. There's a citation for the formula used.
- Q2. The appropriate statistically significant difference between groups needs to be determined by the review group before the review begins.
- Q3. The review group leader needs to think about how much the study has violated the underlying assumptions of parametric analysis?
- Q5. Whether intention to treat or reasonably high response rate (may need to clarify within the review group).

H) INTERVENTION INTEGRITY

Q1 What percentage of participants received the allocated intervention or exposure of interest?

| | |
|---|----------------|
| The number of participants receiving the intended intervention is noted. For example, the authors may have reported that at least 80% of the participants received the complete intervention. | Not Applicable |
| describe | Not Reported |
| describe | Not applicable |

Q2 Was the consistency of the intervention measured?

The authors should describe a method of measuring if the intervention was provided to all participants the same way.

| | |
|----------|--------------|
| describe | Yes |
| describe | No |
| describe | Not reported |

Q3 Is it likely that subjects received an unintended intervention (contamination or cointervention) that may influence the results?

The authors should indicate if subjects received an unintended intervention that may have influenced the outcomes. For example, co-intervention occurs when the study group receives an additional intervention (other than that intended). In this case, it is possible that the effect of the intervention may be over-estimated. Contamination refers to situations where the control group accidentally receives the study intervention. This could result in an under-estimation of the impact of the intervention.

| | |
|----------|------------|
| describe | Yes |
| describe | No |
| describe | Can't tell |

Component Ratings for Study

A) SELECTION BIAS

Strong

Q1 = Very Likely AND Q2 = 80-100% Agreement
OR
Q1 = Very Likely AND Q2 = Not Applicable

Moderate

Q1 = Very Likely AND Q2 = 60 - 79% Agreement
OR
Q1 = Very Likely AND Q2 = Not Reported
OR
Q1 = Somewhat Likely AND Q2 = 80-100%
OR
Q1 = Somewhat Likely AND Q2 = 60 - 79% Agreement
OR
Q1 = Somewhat Likely AND Q2 = Not Applicable

Weak

Q1 = Not Likely
OR
Q2 = Less than 60% agreement
OR
Q1 = Somewhat Likely AND Q2 = Not Reported

B) ALLOCATION BIAS

Strong

Study Design = RCT

Moderate

Study Design = Two-Group Quasi-Experimental

Weak

Study Design = Case Control, Before/After Study, No Control Group

C) CONFOUNDERS

Strong

| | | |
|----------|--------------|-------------|
| Q1 = No | AND Q2 = N/A | AND Q3 = No |
| Q1 = Yes | AND Q2 = YES | AND Q3 = No |

Moderate

| | | |
|----------|--------------|--------------|
| Q1 = Yes | AND Q2 = YES | AND Q3 = Yes |
|----------|--------------|--------------|

Weak

| | | |
|-----------------|--------------|--------------|
| Q1 = Can't tell | | |
| Q1 = Yes | AND Q2 = No | AND Q3 = Yes |
| Q1 = Yes | AND Q2 = No | AND Q3 = No |
| Q1 = No | AND Q2 = N/A | AND Q3 = Yes |

D) BLINDING

Strong

Q1=Yes

Weak

Q1=No

Q1= Not reported

Not applicable

E) DATA COLLECTION METHODS

Strong

Q1 = Yes AND Q2 = Yes

Moderate

Q1 = Yes AND Q2 = No

Weak

Q1 = No AND Q2 = Yes

OR

Q1 = No AND Q2 = No

F) WITHDRAWALS AND DROP-OUTS

Strong

Q1 = 80-100%

Moderate

Q1 = 60-79%

Weak

Q1 = Less than 60%

OR

Q1 = Not Reported Not Applicable

Not applicable

FACILITATOR'S GUIDE - QUANTITATIVE STUDIES

Time required:

1 hour 35 minutes approx (slides 35 minutes, quantitative appraisal exercise 1 hour minutes)

Learning topics

- Sources of bias within health promotion and public health studies
- Critical appraisal tools/checklists
- Integrity of interventions

Summary of activity

Lecture – Powerpoint slides

Group exercise 1 – class divided into half depending on the type of study design they received as part of the pre-reading material (1) randomised controlled trial, (2) controlled before and after study

Description of supporting materials

- 1) Quality Assessment Tool for Quantitative Studies – provided in handbook
- 2) Dictionary for the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies – provided in handbook
- 3) The two RCT papers by Sahota – print out and give to participants
- 4) The CBA study by Gortmaker – print out and give to participants

Further reading (if required)

1. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2000;56:119-27.
2. 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.
3. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess.* 1999;3(5):iii-92.

Instructions to facilitator

The aims of this section are to 1) describe the types of biases associated with different study designs, 2) describe the impact the implementation of the intervention has on effectiveness, and 3) provide practical experience to participants in working through a critical appraisal checklist.

Due to copyright restrictions, you will need to locate and print the required readings for this session (they should also be emailed to participants one week prior to the course). Both study designs do not have to be appraised; it may be more effective for the whole group to appraise one of the studies. The study by Gortmaker is included for participants with more advanced knowledge. The RCT is a good study design for beginners.

Emphasise that the criteria to appraise studies may not be suitable for every health promotion or public health intervention. For example, blinding in educational interventions is difficult and the potential for contamination is high. Therefore, for such interventions to score highly on a quality checklist, aspects such as blinding need to be given less weight than aspects such as data collection methods and comparability of groups at baseline.

Integrity of intervention: Many studies may lack adequate reporting of factors describing integrity.

Information on cluster designs and cluster analysis issues: It is highly recommended to read the executive summary of "Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess.* 1999;3(5):iii-92.". This can be found at the following website http://www.nchta.org/ProjectData/3_publication_listings_ALL.asp.

Crib sheets for the appraisal exercises are attached. The checklist for the randomised controlled trial and the controlled before and after study is the Effective Public Health Practice Project (EPHPP) quality appraisal tool for quantitative studies. The EPHPP checklist was chosen as it was one of the six checklists recommended by an HTA report by Deeks et al (2003) for assessing quality when conducting a systematic review. This checklist is the only one of the six to ask reviewers to determine the integrity of the intervention. It also has the advantage of being able to be used for a variety of study designs.

The aim of the exercise is for participants to become familiar with the checklists provided. Do not let participants get swayed into conversations that do not relate directly to the task at hand. It is more important that participants complete the checklist than get caught up on one particular question.

** Feel free to choose critical appraisal studies which are more relevant to your audience.

CRIB SHEETS FOR CRITICAL APPRAISAL EXERCISES

1. Randomised controlled trial

Using the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies

Sahota P, Rudolf CJ, Dixey R, Hill AJ, Barth JH, Cade J. Randomised controlled trial of primary school based intervention to reduce risk factors for obesity. *BMJ* 2001;323:1029-32.

2. Controlled before and after study

Using the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies

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 Adolesc Med* 1999;153:975-983.

Randomised controlled trial – critical appraisal exercise

Sahota P, Rudolf CJ, Dixey R, Hill AJ, Barth JH, Cade J. Randomised controlled trial of primary school based intervention to reduce risk factors for obesity. *BMJ* 2001;323:1029-32.

| Selection bias | |
|--|---|
| (Q1) Are the individuals selected to participate in the study likely to be representative of the target population? <i>Not likely</i> <i>Somewhat likely</i> <i>Not likely</i> | Somewhat likely All of the schools were state primary schools sited outside the inner city area. Slight bias towards more advantaged children – 1-42% of children from ethnic minorities, 7-29% entitled to free school meals. |
| (Q2) What percentage of selected individuals agreed to participate? <i>80-100% agreement</i> <i>60-79% agreement</i> <i>less than 60% agreement</i> <i>Not reported</i> <i>Not Applicable</i> | Not reported We are only told that 10 schools took part in the intervention; we do not know how many schools were approached. |
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Weak |
| Study design | |
| Indicate the study design | Randomised controlled trial |
| Is the method of random allocation stated? <i>No</i> <i>Yes</i> | Yes Group randomised controlled trial |
| If the method of random allocation is stated is it appropriate? <i>No</i> <i>Yes</i> | Yes Schools were first paired (size, ethnicity, level of social disadvantaged) and then randomised using the toss of a coin. |
| Was the method of random allocation reported as concealed? <i>No</i> <i>Yes</i> | No (no information reported) |
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Strong |
| Confounders | |
| (Q1) Prior to the intervention were there between group differences for important confounders reported in the paper? <i>Yes</i> <i>No</i> <i>Can't tell</i> | No The groups were similar in terms of age, school year, BMI, etc. There were no significant differences found between any of the intervention and comparison pupils for any of the measures. |
| (Q2) If there were differences between groups for important confounders, were they adequately managed in the analysis? <i>Yes</i> <i>No</i> <i>Not applicable</i> | No differences, but still adjusted results. Calculations adjusted for sex, age, initial BMI SD score, and type of school. |
| (Q3) Were there important confounders not reported in the paper? <i>Yes</i> | No |

| | |
|--|---|
| No | |
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Strong |
| Blinding | |
| (Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants? Yes No Not reported Not applicable | Yes The measures could not be obtained blind to the schools' intervention status. Measurement of BMI by the same auxologist. (Would you expect BMI to change in such a short time?) Diet – 24-hour recall, 3-day diary Physical activity – measured by questionnaire, activity during the past week, frequency of sedentary behaviour in last 24 hours. Psychological measures- 3 validated tools – self-perception profile, measure of dietary restraint, adapted body shape perception scale. Knowledge and attitudes towards healthy eating – focus groups. |
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Weak |
| Data collection methods | |
| (Q1) Were the data collection tools shown to be valid? Yes No | No Standard measurements for height and weight. No information on validity of 24-hour recall and 3-day food diaries. No information provided on validity of physical activity questionnaire. Three validated tools were used for measuring psychological status. |
| (Q2) Were data collection tools shown to be reliable? Yes No | No No information provided. |
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Weak |
| Withdrawals and drop-outs | |
| (Q1) Indicate the percentage of participants completing the study. 80-100% 60-79% Less than 60% Not reported Not applicable | 80-100% The figure presented shows the progress of the schools throughout the trial. 21 children declined to participate in the data collection, 613 children were measured at baseline and 595 at one-year follow-up. All persons were accounted for. However, 42 children left the study and 40 new children joined. We are not told which groups these children dropped-out of. |
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Strong |
| Analyses | |
| (Q1) Is there a sample size calculation or power | Yes |

| | |
|---|--|
| <p>calculation?</p> <p><i>Yes</i></p> <p><i>Partially</i></p> <p><i>No</i></p> | <p>A power calculation indicated that with five schools in each arm the study would have 80% power to detect an underlying difference in means of normally distributed outcome measure of $\geq 1.8SD$ at the 5% significance level and 65% power to detect a difference at 1.5SD. This took into account the cluster randomisation design.</p> |
| <p>(Q2) Is there a statistically significant difference between groups?</p> <p><i>Yes</i></p> <p><i>No</i></p> <p><i>Not reported</i></p> | <p>BMI – no significant difference</p> <p>Vegetable intake – 24-hour recall: weighted mean difference of 0.3 (95% CI 0.2 to 0.4). No p values reported. Higher intake of foods and drinks high in sugar in overweight children in the intervention group, and lower fruit intake in obese children in the intervention group. Same difference (0.3) for the overweight children.</p> <p>Three-day diary: No differences.</p> <p>Physical activity: No differences.</p> |
| <p>(Q3) Are the statistical methods appropriate?</p> <p><i>Yes</i></p> <p><i>No</i></p> <p><i>Not reported</i></p> | <p>Yes</p> |
| <p>(Q4a) Indicate the unit of allocation</p> <p><i>Community/Organisation/group/provider/client</i></p> | <p>School (institution)</p> |
| <p>(Q4b) Indicate the unit of analysis</p> <p><i>Community/Organisation/group/provider/client</i></p> | <p>Individual</p> |
| <p>(Q4c) If 4a and 4b are different, was the cluster analysis done?</p> <p><i>Yes</i></p> <p><i>No</i></p> <p><i>Not applicable</i></p> | <p>Yes</p> <p>Cluster analysis performed</p> |
| <p>(Q4) Is the analysis performed by intervention allocation status (i.e. ITT) rather than the actual intervention received?</p> <p><i>Yes</i></p> <p><i>No</i></p> <p><i>Can't tell</i></p> | <p>Can't tell – not reported.</p> |
| <p>Intervention integrity</p> | |
| <p>(Q1) What percentage of participants received the allocated intervention or exposure of interest?</p> <p><i>80-100%</i></p> <p><i>60-79%</i></p> <p><i>Less than 60%</i></p> <p><i>Not reported</i></p> <p><i>Not applicable</i></p> | <p>80-100%</p> <p>- Use process evaluation paper</p> <p>Schools had a different number of activities that they implemented (6 to 14 action plans per school) – 89% of these action plans were successfully achieved.</p> |
| <p>(Q2) Was the consistency of the intervention measured?</p> <p><i>Yes</i></p> <p><i>No</i></p> | <p>Yes – use process evaluation paper.</p> |

| | |
|----------------------------------|---------------|
| <i>Not reported</i> | |
| <i>Not applicable</i> | |
| COMPONENT RATINGS | |
| Selection Bias | Weak |
| Study design | Strong |
| Confounders | Strong |
| Blinding | Weak |
| Data collection methods | Weak |
| Withdrawals and drop-outs | Strong |

Controlled before and after study – critical appraisal exercise

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 Adolesc Med 1999;153:975-983.

| Selection bias | |
|--|--|
| (Q1) Are the individuals selected to participate in the study likely to be representative of the target population? <i>Not likely</i> <i>Somewhat likely</i> <i>Not likely</i> | Somewhat likely or Not likely Median household income of both groups was lower than that for all households in the US (\$33952 vs. \$22708 intervention, \$22651 control). Public school students only. 91% were African-American. Intervention schools were chosen as they had initial interest in rapidly receiving the intervention. May not be generalisable to schools that are less motivated. Only grades 4 and 5. |
| (Q2) What percentage of selected individuals agreed to participate? <i>80-100% agreement</i> <i>60-79% agreement</i> <i>less than 60% agreement</i> <i>Not reported</i> <i>Not applicable</i> | Not reported. Article did not state how many schools were approached to participate. |
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Weak |
| Study design | |
| Indicate the study design | Quasi-experimental - Controlled before and after design/ Cohort analytic. |
| Is the method of random allocation stated? <i>1 No</i> <i>2 Yes</i> | NA |
| If the method of random allocation is stated is it appropriate? <i>1 No</i> <i>2 Yes</i> | NA |
| Was the method of random allocation reported as concealed? <i>1 No</i> <i>2 Yes</i> | NA |
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Moderate |
| Confounders | |
| (Q1) Prior to the intervention were there between group differences for important confounders reported in the paper? <i>Yes</i> <i>No</i> <i>Can't tell</i> | No Groups were well matched in average school enrolment, percentage of students receiving free/reduced cost lunches, race, and reading and math achievement scores. Similar in median household income. Nutrition content of school lunch offerings in groups was similar. |

| | |
|--|--|
| <p>(Q2) If there were differences between groups for important confounders, were they adequately managed in the analysis?</p> <p>Yes No Not applicable</p> | <p>No differences, but results were still adjusted by possible confounders. Covariates included in regression: Baseline scores, sex, ethnicity, baseline total energy intake, baseline knowledge, having been held back in school, mobility, number of adults in the household, frequency of sit-down dinners, whether both parents live with the child, number of children living at home, and how often the mother and father exercised.</p> |
| <p>(Q3) Were there important confounders not reported in the paper?</p> <p>Yes No</p> | <p>No</p> |
| <p>RATE THIS SECTION</p> <p>Strong Moderate Weak</p> | <p>Moderate</p> |
| <p>Blinding</p> | |
| <p>(Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants?</p> <p>Yes No Not reported Not applicable</p> | <p>24-hour recall – no information provided on whether the interviewers were aware which group participants were allocated. FAS – teachers, who would have been aware of which group they were allocated to, supervised the completion of the FAS.</p> |
| <p>RATE THIS SECTION</p> <p>Strong Moderate Weak</p> | <p>Weak</p> |
| <p>Data collection methods</p> | |
| <p>(Q1) Were the data collection tools shown to be valid?</p> <p>Yes No</p> | <p>No – FAS, Yes – 24-hr recall Dietary intake and physical activity: Student food and activity survey (FAS). Authors say that FAS has indicated lower validity in this age group. Therefore they added a 24-hour recall to enhance validity (only for the post-intervention data!). The 24-hour recall method has been shown to be reliable and valid in children as young as grade 3.</p> |
| <p>(Q2) Were data collection tools shown to be reliable?</p> <p>Yes No</p> | <p>Yes – only 24-hr recall Dietary recalls were performed by trained interviewers. The 24-hour recall methods have been shown to be reliable and valid in children as young as grade 3.</p> |
| <p>RATE THIS SECTION</p> <p>Strong Moderate Weak</p> | <p>Weak (because 24-hr recall results were adjusted using the FAS)</p> |
| <p>Withdrawals and drop-outs</p> | |
| <p>(Q1) Indicate the percentage of participants completing the study.</p> <p>80-100% 60-79% Less than 60% Not reported Not applicable</p> | <p>60-79% Baseline data – 90% of students completed Follow-up – 88% in control and 89% in intervention groups. Overall response rate with complete predata and postdata was 61%.</p> |

| | |
|--|---|
| RATE THIS SECTION <i>Strong Moderate Weak</i> | Moderate |
| Analyses | |
| (Q1) Is there a sample size calculation or power calculation? <i>Yes</i> <i>Partially</i> <i>No</i> | No – not reported. |
| (Q2) Is there a statistically significant difference between groups? <i>Yes</i> <i>No</i> <i>Not reported</i> | 24-hour recall adjusted using FAS: Significant differences in total energy from fat and saturated fat in the intervention group compared to the control group. (-1.4%, 95% CI -2.8 to -0.04, p=0.04, -0.60%, 95% CI -1.2 to -0.01, p=0.05). Increase in F & V intake (p=0.01), in vitamin C intake (p=0.01), and in fibre consumption (p=0.05). FAS and cross sectional studies: Significant differences in total energy from fat (p=0.02, p=0.02). Cross-sectional study also found differences in total energy from saturated fat (p=0.04). |
| (Q3) Are the statistical methods appropriate? <i>Yes</i> <i>No</i> <i>Not reported</i> | Yes Cluster analysis performed |
| (Q4a) Indicate the unit of allocation <i>Community/Organisation/group/provider/client</i> | Organisation. i.e. school |
| (Q4b) Indicate the unit of analysis <i>Community/Organisation/group/provider/client</i> | Individual |
| (Q4c) If 4a and 4b are different, was the cluster analysis done? | Yes |
| (Q5) Is the analysis performed by intervention allocation status (i.e. ITT) rather than the actual intervention received? <i>Yes</i> <i>No</i> <i>Can't tell</i> | Yes ITT analysis was used. |
| Intervention integrity | |
| (Q1) What percentage of participants received the allocated intervention or exposure of interest? <i>80-100%</i> <i>60-79%</i> <i>Less than 60%</i> <i>Not reported</i> <i>Not applicable</i> | 60-79% In 1995, 71% of teachers in the intervention group returned evaluation forms, 81% in 1996. Survey data indicate that on average, 22 (71%) of the possible 31 nutrition and physical activity lessons were completed during the intervention. |
| (Q2) Was the consistency of the intervention measured? <i>1 Yes</i> <i>2 No</i> | Yes Teachers completed evaluation forms. |

| | |
|----------------------------------|-----------------|
| 3 <i>Can't tell</i> | |
| COMPONENT RATINGS | |
| Selection Bias | Weak |
| Study design | Moderate |
| Confounders | Moderate |
| Blinding | Weak |
| Data collection methods | Strong |
| Withdrawals and drop-outs | Moderate |

2) QUALITATIVE STUDIES

Qualitative research explores the subjective world. It attempts to understand why people behave the way they do and what meaning experiences have for people.¹

Qualitative research may be included in a review to shed light on whether the intervention is suitable for a specific target group, whether special circumstances have influenced the intervention, what factors might have contributed if an intervention did not have the expected effects, what difficulties must be overcome if the study is to be generalised to other populations.² These are all important questions often asked by the users of systematic reviews.

Reviewers may choose from a number of checklists available to assess the quality of qualitative research. Sources of information on quality appraisal include:

- CASP appraisal tool for Qualitative Research – included in this manual, <http://www.phru.nhs.uk/casp/qualitat.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
- Health Care Practice Research and Development Unit (HCPDRU), University of Salford, UK. Evaluation Tool for Qualitative Studies, <http://www.fhsc.salford.ac.uk/hcprdu/tools/qualitative.htm>
- Greenhalgh T, Taylor R. Papers that go beyond numbers (qualitative research). BMJ 1997;315:740-3.
- Popay J, Rogers A, Williams G. Rationale and standards for the systematic review of qualitative literature in health services research. Qual Health Res 1998;8:341-51.
- Mays N, Pope C. Rigour and qualitative research. BMJ 1995;311:109-12.

In relation to the appraisal of process evaluations the EPPI-Centre has developed a 12-question checklist, available at:

http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=hp/reports/phase/phase_process.htm.

REFERENCES

1. 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. March 2001. <http://www.york.ac.uk/inst/crd/report4.htm>
2. 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

ADDITIONAL READING

Jones R. Why do qualitative research? BMJ 1995;311:2.

Pope C, Mays N. Qualitative Research: Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ* 1995;311:42-45.

EXERCISE

1. Appraise the qualitative study using the Critical Appraisal Skills Programme (CASP) qualitative worksheet in small groups and report back to the group.
"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."

Critical Appraisal Skills Programme (CASP)
making sense of evidence

10 questions to help you make sense of qualitative research

This assessment tool has been developed for those unfamiliar with qualitative research and its theoretical perspectives. This tool presents a number of questions that deal very broadly with some of the principles or assumptions that characterise qualitative research. It is not a definitive guide and extensive further reading is recommended.

How to use this appraisal tool

Three broad issues need to be considered when appraising the report of qualitative research:

- **Rigour: has a thorough and appropriate approach been applied to key research methods in the study?**
- **Credibility: are the findings well presented and meaningful?**
- **Relevance: how useful are the findings to you and your organisation?**

The 10 questions on the following pages are designed to help you think about these issues systematically.

The first two questions are screening questions and can be answered quickly. If the answer to both is "yes", it is worth proceeding with the remaining questions.

A number of italicised prompts are given after each question. These are designed to remind you why the question is important. Record your reasons for your answers in the spaces provided.

The 10 questions have been developed by the national CASP collaboration for qualitative methodologies.

Screening Questions

1 Was there a clear statement of the aims of the research?

Yes

No

Consider:

- what the goal of the research was*
 - why it is important*
 - its relevance*
-

2 Is a qualitative methodology appropriate?

Yes

No

Consider:

- if the research seeks to interpret or illuminate the actions and/or subjective experiences of research participants*
-

Is it worth continuing?

Detailed questions

.....
Appropriate research design

3 Was the research design appropriate to the aims of the research?

Write comments here

Consider:

- if the researcher has justified the research design (eg. have they discussed how they decided which methods to use?)*
-

.....
Sampling

4 Was the recruitment strategy appropriate to the aims of the research?

Write comments here

Consider:

- if the researcher has explained how the participants were selected*
 - if they explained why the participants they selected were the most appropriate to provide access to the type of knowledge sought by the study*
 - if there are any discussions around recruitment (eg. why some people chose not to take part)*
-

.....
Data collection

5 Were the data collected in a way that addressed the research issue?

Write comments here

Consider:

- if the setting for data collection was justified
- if it is clear how data were collected
(eg. focus group, semi-structured interview etc)
- if the researcher has justified the methods chosen
- if the researcher has made the methods explicit
(eg. for interview method, is there an indication of how interviews were conducted, did they use a topic guide?)
- if methods were modified during the study.
If so, has the researcher explained how and why?
- if the form of data is clear (eg. tape recordings, video material, notes etc)
- if the researcher has discussed saturation of data

.....
Reflexivity (research partnership relations/recognition of researcher bias)

6 Has the relationship between researcher and participants been adequately considered?

Write comments here

Consider whether it is clear:

- if the researcher critically examined their own role, potential bias and influence during:
 - formulation of research questions
 - data collection, including sample recruitment and choice of location
 - how the researcher responded to events during the study and whether they considered the implications of any changes in the research design

.....
Ethical Issues

7 Have ethical issues been taken into consideration?

Write comments here

Consider:

- if there are sufficient details of how the research was explained to participants for the reader to assess whether ethical standards were maintained
 - if the researcher has discussed issues raised by the study (e. g. issues around informed consent or confidentiality or how they have handled the effects of the study on the participants during and after the study)
 - if approval has been sought from the ethics committee
-

.....
Data analysis

8 Was the data analysis sufficiently rigorous?

Write comments here

Consider:

- *if there is an in-depth description of the analysis process*
- *if thematic analysis is used. If so, is it clear how the categories/themes were derived from the data?*
- *whether the researcher explains how the data presented were selected from the original sample to demonstrate the analysis process*
- *if sufficient data are presented to support the findings*
- *to what extent contradictory data are taken into account*
- *whether the researcher critically examined their own role, potential bias and influence during analysis and selection of data for presentation*

.....
Findings

9 Is there a clear statement of findings?

Write comments here

Consider:

- *if the findings are explicit*
- *if there is adequate discussion of the evidence both for and against the researcher's arguments*
- *if the researcher has discussed the credibility of their findings*
- *if the findings are discussed in relation to the original research questions*

.....
Value of the research

10 How valuable is the research?

Write comments here

Consider:

- *if the researcher discusses the contribution the study makes to existing knowledge or understanding eg. do they consider the findings in relation to current practice or policy, or relevant research-based literature?*
 - *if they identify new areas where research is necessary*
 - *if the researchers have discussed whether or how the findings can be transferred to other populations or considered other ways the research may be used*
-

FACILITATOR'S GUIDE - QUALITATIVE STUDIES

Time required:

One hour and 10 minutes (slides 20 minutes, qualitative appraisal 50 minutes)

Learning topics

- Sources of bias within health promotion and public health studies (qualitative and quantitative studies)
- Critical appraisal tools/checklists
- Integrity of interventions

Summary of activity

Lecture – Powerpoint slides

Group exercise – groups can sit around their tables completing the quality appraisal checklist

Description of supporting materials

- 1) CASP tool for Qualitative Studies
- 2) The qualitative study by Cass

Further reading (if required)

Jones R. Why do qualitative research? BMJ 1995;311:2.

Pope C, Mays N. Qualitative Research: Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995;311:42-45.

Instructions to facilitator

The aim of this section is simply to introduce participants to qualitative research and provide them with the opportunity to practice the appraisal of a qualitative study.

Due to copyright restrictions, you will need to locate and print the required reading for this session (the reading should also be emailed to participants one week prior to the course).

The qualitative study uses the Critical Appraisal Skills Programme checklist for Qualitative research. The aim of the exercise is for participants to become familiar with the checklist provided. Do not let participants get swayed into conversations that do not relate directly to the task at hand. It is better for participants to complete the checklist than to get caught up on one particular question.

*** Note: If this workshop is to be run in its current format participants will need to be advised that they must read the critical appraisal paper “DiCenso A, Guyatt G, Griffith WL. Interventions to reduce unintended pregnancies among adolescents: systematic review of randomised controlled trials. BMJ 2002;324:1426-34” before the afternoon of the second day. Remember, you will need to obtain copies of this article, as copyright restrictions prevent the reproduction of this article.

** Feel free to choose critical appraisal studies which are more relevant to your audience.

The Schema for Evaluating Evidence on Public Health Interventions

The Schema includes questions that encourage reviewers of evidence to consider whether the evidence demonstrates that an intervention was adequately implemented in the evaluation setting(s), whether information is provided about the implementation context, and whether interactions that occur between public health interventions and their context were assessed and reported. It is used to appraise individual papers and to formulate a summary statement about those articles and reports. The Schema can be downloaded from:

<http://www.nphp.gov.au/publications/phpractice/schemaV4.pdf>.

A Checklist for Evaluating Evidence on Public Health Interventions

SECTION 1: THE SCOPE OF YOUR REVIEW

Items to record about the scope of your review

1. What is the question you want to answer in the review?
2. How are you (and possibly others) going to use the findings of the review?
3. Who asked for the review to be done?
4. How has the review been funded?
5. Who is actually carrying out the review?

SECTION 2: THE PAPERS IN THE REVIEW

2A Publication details

- Identify the publication details for each paper or report to be appraised (eg title, authors, date, publication information, type of article or report). Also note what related papers or reports have been published (eg process evaluations or interim reports).

2B Specifying the intervention

1. Exactly what intervention was evaluated in the study?
2. What was the origin of the intervention?
3. If the origin of the intervention involved a degree of formal planning, what was the rationale for the strategies selected?
4. What organisations or individuals sponsored the intervention (with funding or in-kind contributions)? Where relevant, give details of the type of sponsorship provided.

2C Identifying the intervention context

5. What aspects of the context in which the intervention took place were identified in the article?
6. Was enough information provided in the article to enable you to describe the intervention and its context as requested above? (Identify major deficiencies)
7. How relevant to the scope of your review (as recorded in Section 1) are the intervention and the context described in this article?

Decision Point

If you conclude that the article is relevant (or partly relevant) to the scope of your review, go to sub-section 2D.

If the article is not relevant record why not, and then move on the next paper or report to be appraised.

2D The evaluation context – background, purpose and questions asked

8. Who requested or commissioned the evaluation and why?
9. What research questions were asked in the evaluation reported in the study?
10. What measures of effect or intervention outcomes were examined?
11. What was the anticipated sequence of events between the intervention strategies and the measures of effect or intended intervention outcomes?
12. Were the measures of effect or intervention outcomes achievable and compatible with the sequence of events outlined above?
13. What was the timing of the evaluation in relation to the implementation of the intervention?
14. Was the intervention adequately implemented in the setting in which it was evaluated?
15. Was the intervention ready for the type of evaluation that was conducted?
16. Were the measures of effect or intervention outcomes validated or pilot tested? If so, how?
17. Did the observations or measures include the important individual and group-level effects?
18. Was there a capacity to identify unplanned benefits and unanticipated adverse effects?
19. If the research was not primarily an economic evaluation, were economic factors considered?
20. Was there a significant potential for conflict of interest (in the way the intervention and/or its evaluation were funded and implemented) that might affect interpretation of the findings?

2E The methods used to evaluate the intervention

21. What types of research methods were used to evaluate the intervention?
22. What study designs were used in the evaluation?
23. How appropriate were the research methods and study designs in relation to the questions asked in the study?
24. Was the evaluation conducted from a single perspective or multiple perspectives? Give details.
25. Appraise the rigour of the research methods used in the study using the relevant critical appraisal checklist(s) (see Table 1)
26. What are your conclusions about the adequacy of the design and conduct of the research methods used to evaluate the intervention?
27. Are the reported findings of the evaluation likely to be credible?

Decision Point

If you conclude from Section 2 that the reported findings are likely to be credible go to Section 3. If the findings are unlikely to be credible go to Section 4 to answer question 2 only, and then move to the next paper to be appraised.

SECTION 3: DESCRIBING THE RESULTS FROM THE PAPERS SELECTED

The study findings

1. What findings were reported in the study?
2. If the study specified measurable or quantifiable targets, did the intervention achieve these objectives?
3. Were reported intervention effects examined among sub-groups of the target population?
4. Should any other important sub-group effects have been considered that were not considered?
5. Was the influence of the intervention context on the effectiveness of the intervention investigated in the study?
6. How dependent on the context is the intervention described in the article?
7. Were the intervention outcomes sustainable?
8. Did the study examine and report on the value of the measured effects to parties interested in or affected by them?

SECTION 4: INTERPRETING EACH ARTICLE

Your interpretations

1. How well did the study answer your review question(s)? Give details.
2. Are there other lessons to be learned from this study (eg lessons for future evaluations)

Decision Point

If you are conducting the review for the purpose of making recommendations for a particular policy or practice setting, continue in Section 4 to answer questions 3 – 8. Otherwise move on to Section 5.

3. Are the essential components of the intervention and its implementation described with sufficient detail and precision to be reproducible?
4. Is the intervention context, as described in the article being examined, comparable to the intervention context that is being considered for future implementation of the intervention?
5. Are the characteristics of the target group studied in the article comparable to the target group for whom the intervention is being considered?
6. If an economic evaluation was conducted, did the paper or report include and address the details required in order to make an informed assessment about the applicability and transferability of the findings to other settings?
7. If enough information was provided, are the findings of the economic evaluation relevant and transferable to your setting?
8. Are the effects of the intervention likely to be considered important in your setting?

SECTION 5: SUMMARISING THE BODY OF EVIDENCE

5A Grouping, rating and weighing up the papers and reports (see Table 2 for example of presenting findings)

1. Group articles with similar research questions and similar intervention strategies. With each group, complete the following:
2. Rate the quality of each study, from 1 (weak) to 3 (strong).
3. Assess the consistency of the findings among the stronger studies, from 1 (inconsistent) to 3 (consistent).
4. Determine the degree to which the stronger studies with consistent findings are applicable to your review context.

5A Formulating a summary statement

5. Did studies that examined similar intervention strategies, with similar research questions, produce consistent results?
6. Did studies with different research questions produce compatible results?
7. Overall, what does the body of evidence tell you about the intervention?
8. Are there important gaps in the evidence? If so, what are they?
9. To what degree are the review findings useful for your purposes, as identified in Section 1?
10. What are your recommendations based on this review?

Unit Nine: 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

Generally, there are two approaches to synthesising the findings from a range of studies:

Narrative synthesis – findings are summarised and explained in words

Quantitative/statistical synthesis – data from individual studies are combined statistically and then summarised
(*meta-analysis*)

The Cochrane Reviewers' Handbook¹ suggests the following framework for synthesis of primary studies (regardless of the method (narrative/meta-analysis) used to synthesise data):

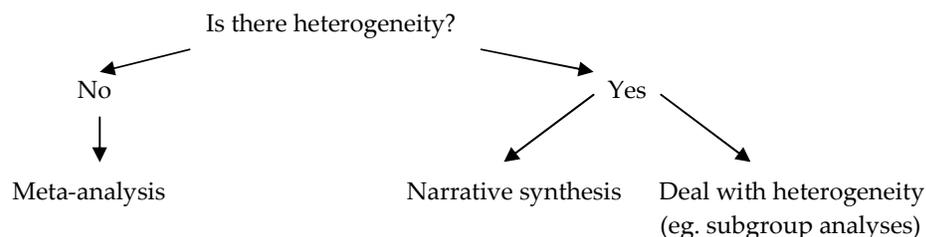
- What is the direction of the effect?
- What is the size of the effect?
- Is the effect consistent across studies?
- What is the strength of evidence for the effect?

Before deciding which synthesis approach to use it is important to tabulate the findings from the studies. This aids the reviewer in assessing whether studies are likely to be homogenous or heterogenous, and tables greatly assist the reader in eyeballing the types of studies that were included in the review. Reviewers should determine which information should be tabulated; some examples are provided below:

| | | |
|----------------------------|-------------------|---------------------------------|
| Authors | Year | Intervention details |
| Comparison details | Theoretical basis | Study design |
| Quality assessment | Outcomes | Setting/context (incl. country) |
| Population characteristics | | |

Example: An example of tabulating studies can be found in the following systematic review: DiCenso A, Guyatt G, Willan A, Griffith L. Interventions to reduce unintended pregnancies among adolescents: systematic review of randomised controlled trials. *BMJ*. 2002 Jun 15;324(7351):1426.

The choice of analysis usually depends on the diversity of studies included in the review. Diversity of studies is often referred to as 'heterogeneity'. Because some reviews may include studies that differ in such characteristics as design, methods, or outcome measures, a quantitative synthesis of studies is not always appropriate or meaningful.



Where studies are more homogenous, i.e., we can compare like with like, it may be appropriate to combine the individual results using a meta-analysis. If the results are similar from study to study we can feel more comfortable that a meta-analysis is warranted. Heterogeneity can be determined by presenting the results graphically and examining the overlap of confidence intervals (CI) (if CI overlap studies are more likely to be homogenous) and by calculating a statistical measure of heterogeneity. Both of these methods are further outlined in Chapter Eight of the Cochrane Reviewers' Handbook (Analysing and presenting results).

Meta-analysis produces a weighted summary result (more weight given to larger studies). By combining results from more than one study it has the advantage of increasing statistical power (which is often inadequate in studies with a small sample size). The final estimate is usually in the form of an odds ratio: the ratio of the probability of an event happening to that of it not happening. The odds ratio is often expressed together with a confidence interval (CI). A confidence interval is a statement of the range within which the true odds ratio lies - within a given degree of assurance (eg. usually estimates of effect like odds ratios are presented with a 95% confidence interval).

Guidelines for narrative synthesis are not yet available, although research is currently underway to develop guidelines for systematic reviews. Ideally, the reviewer should²:

- Describe studies
- Assess whether quality is adequate in primary studies to trust the results
- Demonstrate absence of data for planned comparisons
- Demonstrate degree of heterogeneity
- Stratify results by populations, interventions, settings, context, outcomes, validity (*if appropriate*)

Example: A number of Cochrane systematic reviews of health promotion and public health topics synthesise the results narratively. Visit The Cochrane Library to read examples. Another example can be found in the following article: Riemsma RB, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, Walker A. Systematic review of the effectiveness of stage based interventions to promote smoking cessation. *BMJ* 2003;326:1175-77.

Integrating qualitative and quantitative data

The Evidence for Policy and Practice Information and Co-ordinating Centre has developed methods for synthesising the findings from diverse types of studies within one review³. These methods involve conducting three types of syntheses in the same review: 1) a statistical meta-analysis to pool trials of interventions tackling particular problems (or a narrative synthesis when meta-analysis is not appropriate or possible); 2) a synthesis of studies examining people's perspectives or experiences of that problem using qualitative analysis ('views' studies); and 3) a 'mixed methods' synthesis bringing the products of 1) and 2) together. These developments have been driven by particular review questions rather than methodology; 'users' of the reviews want to know about the effects of interventions, but also want to know which interventions will be most appropriate and relevant to people. However, they do illustrate how qualitative studies can be integrated into a systematic review as 'views' studies are often, but not always, qualitative in nature. The methods for each of the three syntheses are described in brief below:

Synthesis 1) Effectiveness synthesis for trials

Effect sizes from good quality trials are extracted and, if appropriate, pooled using statistical meta-analysis. Heterogeneity is explored statistically by carrying out sub-group analyses on a range of categories specified in advance (eg. study quality, study design, setting and type of intervention).

Synthesis 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 then considered in the light of the relevant review question (eg. what helps and what stops children eating fruit and vegetables?) 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 synthesis methods have much in common with the work of others who have emphasised the theory building potential of synthesis.⁴

Synthesis 3) A 'mixed methods' synthesis

Implications for interventions are juxtaposed against the interventions which have been evaluated by trials included in Synthesis 1. Using the descriptions of the interventions provided in the reports of the trials, matches, miss-matches and gaps are identified. Gaps are used for recommending what kinds of interventions need to be newly developed and tested. The effect sizes from interventions which matched implications for interventions derived from people's views can be compared to those which do not, using sub-group analysis. This provides a way to highlight which types of interventions are both effective and appropriate. Unlike Bayesian methods, another approach to combining 'qualitative' and 'quantitative' studies within systematic reviews which translates textual data into numerical data, these methods integrate 'quantitative' estimates of benefit and harm with 'qualitative' understanding from people's lives, whilst preserving the unique contribution of each.³

REFERENCES

1. Clarke M, Oxman AD, editors. Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. <http://www.cochrane.dk/cochrane/handbook/handbook.htm>
2. 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. March 2001. <http://www.york.ac.uk/inst/crd/report4.htm>
3. 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.
4. Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, Oakley A. Applying systematic review methods to studies of people's views: an example from public health research. *J Epidemiol Community Health*. 2004 Sep;58(9):794-800.

EXERCISE

1. Read the methods and results section of the article:
"Riemsma RB, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, Walker A. Systematic review of the effectiveness of stage based interventions to promote smoking cessation. *BMJ* 2003;326:1175-77". Look at the type of narrative synthesis of results and contrast with the meta-analysis approach.

FACILITATOR'S GUIDE

Time required:

45 minutes approx. (30 minutes slides, 15 minutes exercise)

Learning topics

- Techniques to synthesise evidence (meta-analysis and narrative synthesis)
- Terminology used in synthesis (heterogeneity, odds ratio, relative risk, p value, confidence interval, funnel plots, sensitivity analysis)

Summary of activity

Lecture – Powerpoint slides

Exercise – examining narrative synthesis

Description of supporting materials

Possible examples of narrative synthesis:

1. Riemsma RB, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, Walker A. Systematic review of the effectiveness of stage based interventions to promote smoking cessation. *BMJ* 2003;326:1175-77.
2. Cochrane reviews of health promotion or public health interventions (most have been narratively summarised).

Instructions to facilitator

The aim of this section is to provide an overview of evidence synthesis; participants should not expect to be able to understand the details of meta-analysis (there should be other classes/workshops available for participants to enrol in). Further information on data synthesis can be obtained from the Cochrane Reviewers' Handbook.

For participants who are new to relative risks and odds ratios - concepts need to be described very slowly. Any person whose role is to read and interpret articles of interventions (individual studies or systematic reviews) needs to be able to understand what the results mean and how they are calculated. You may need to spend extra time on this unit if the basics of relative risks, odds ratios, confidence intervals and p-values are to be fully explained first.

This section should enable participants to clarify the difference between a narrative review (literature review) and a narrative systematic review (a systematic review which uses a narrative synthesis).

Exercise: participants only need to look at the results section to examine how Riemsma et al carried out narrative synthesis. Further examples of narrative syntheses can be found in many health promotion and public health Cochrane reviews, or reviews from other organisations. (You may want take along other examples to demonstrate to participants or use in the exercise).

Unit Ten: Interpretation of Results

Learning Objectives

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

The following issues should be included in the discussion and recommendations section of a systematic review of a health promotion or public health intervention:

- 1) Strength of the evidence
- 2) Integrity of intervention on health-related outcomes
- 3) Theoretical explanations of effectiveness
- 4) Context as an effect modifier
- 5) Sustainability of interventions and outcomes
- 6) Applicability
- 7) Trade-offs between benefits and harms
- 8) Implications for practice and future health promotion and public health research

As those who read systematic reviews (eg. policy makers) 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 findings of the review.¹

1) Strength of the evidence

The discussion should describe the overall strength of the evidence, including the quality of the evidence and the size and consistency of the results. The size of the results is particularly important in population-based studies, where a small effect at the community level may have a much more practical significance than the effect of comparable size at the individual level.² Using statistical significance alone as the standard for interpretation of the results of community intervention trials is inappropriate for research at the population level.³

This section of the review should also describe the biases or limitations of the review process. Difficulties in locating health promotion/public health literature may have resulted in the inability to carry out a comprehensive search. For many reviewers, a further limitation of the review process is the inability to translate non-English articles, or search non-English electronic databases. Furthermore, interpretations may be limited due to studies missing important information relating to such factors as the implementation of the intervention, context, and methodological features (eg. blinding, data collection tools, etc) required in order to determine study quality.

2) Intervention integrity

Reviewers should discuss whether the studies included in the review illuminated the key process factors that led to effective interventions. In addition, the relationship between intervention integrity and effectiveness should be described, i.e., did studies that address integrity thoroughly show a greater impact?

An important outcome of process evaluation is the assessment of intervention 'dose', or the amount of intervention delivered and received by participants or the target group.³ Intervention dose varies markedly between community level interventions, and may be one of the factors that explain differences in effectiveness between studies. Investigators have postulated that the small effect sizes

resulting from some community interventions is a result of an insufficient intervention dose or intensity, or because participation rates were too low.³ Or alternatively, the dose of the intervention may have been inadequate relative to other forces in the environment, such as an information environment already saturated with sophisticated advertisements and product promotions.³ Mittlemark and colleagues⁴ have suggested that intervention effectiveness has been limited by the length of the intervention, recommending that for community-based interventions the intervention period be at least five years, given the time it typically takes for the community to be mobilised into action. This is because it may not be realistic to expect large individual changes in lifetime habits to occur with complex behaviours, such as eating patterns, within the timeframe of most community studies.⁴ Mittlemark et al⁴ further suggest that at the organisational or community level, additional time must be built in for “institutionalisation”; that is, the continuing process of building local, regional, and national capacity to mount permanent health promotion programs.

Information is also needed in reviews on whether it is more effective to spread a given dose out over an extended period of time, rather than to compress it into a shorter time frame to maximise the population’s focus on the intervention messages.

3) Theoretical explanations of effectiveness

Although many public health interventions are planned and implemented without explicit reference to theory, there is substantial evidence from the literature to suggest that the use of theory will significantly improve the chances of effectiveness.⁵

Types of theories:

- Theories that explain health behaviour and health behaviour change at the individual level (eg. Health belief model, Stages of Change)
- Theories that explain change in communities and communal action for health (eg. Diffusion of Innovation)
- Theories that guide the use of communication strategies for change to promote health (eg. social marketing, communication-behaviour change model)
- Models that explain changes in organisations and the creation of health-supportive organisational practices (eg. theories of organisational change)
- Models that explain the development and implementation of health public policy (eg. evidence-based policy making to promote health)

Depending on the level of intervention (individual, group, or organisation) or the type of change (simple, one-off behaviour, complex behaviour, organisational or policy change), different theories will have greater relevance.⁵

Reviewers should seek to examine the impact of the theoretical framework on the effectiveness of the intervention. The assessment of theory within systematic reviews⁵:

- 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, aiding the dissemination of successful interventions.

Theory may also provide a valuable framework within which to explore the relationship between findings from different studies. For example, when combining the findings from different studies, reviewers can group interventions by their theoretical basis. Alternatively, reviewers may consider grouping interventions depending of whether they seek to influence individual behaviour, interpersonal relationships, or community or structural factors or whether they used a Program Logic or Program Theory approach.

Systematic reviews would also be greatly enhanced if in the discussion attention was paid to the gaps in theoretical coverage of interventions. For example, many interventions seek to focus on single level changes rather than seeking to change the environment within which people make their choices.

4) Context as an effect modifier

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 information⁶:

- social and political factors surrounding the intervention, eg. local/national policy environment, concurrent social changes
- time and place of intervention
- structural, organisational, physical environment
- aspects of the host organisation and staff, eg. number, experience/training, morale, expertise of staff, competing priorities to the staff's attention, the organisation's history of innovation, size of the organisation, the status of the program in the organisation, the resources made available to the program;
- aspects of the system, eg. payment and fee structures for services, reward structures, degrees of specialisation in service delivery; and
- characteristics of the target population (eg. 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.

5) Sustainability of interventions and outcomes

The extent to which the intended outcomes or interventions are sustained should be an important consideration in systematic reviews, as decision-makers and funders become increasingly concerned with allocating scarce resources effectively and efficiently.⁷

It is believed that interventions which isolate individual action from its social context would be unlikely to produce sustainable health gain in the absence of change to the organisational, community and institutional conditions that make up the social context.⁷

Reviewers may choose from a number of frameworks which describe the factors that determine sustainability⁸⁻¹⁰

- Bossert⁸ suggests that both contextual (eg. political, social, economic and organisational) factors and project characteristics (eg. institution management, content, community participation) are related to sustainability.
- Swerissen and Crisp⁹ propose that the relationship between the intervention level (individual, organisational, community, institutional) and strategies (eg. education, policies, social planning, social advocacy) indicates the likely sustainability of programmes and effects.
- A framework outlining the four integrated components of sustainability has also been produced.¹⁰

6) Applicability

Applicability is a key part of the process of summarising evidence, since the goal of systematic reviews is to recommend interventions that are likely to be effective in different settings.

Reviewers should use the RE-AIM model¹¹ (Reach, Efficacy, Adoption, Implementation, and Maintenance) for conceptualising the potential for translation and the public health impact of an intervention. The user can then compare their situation to the RE-AIM profile of the included studies or the body of evidence.

RE-AIM:

Reach – the absolute number, proportion, and representativeness of individuals (characteristics that reflect the target population’s characteristics) who are willing to participate in a given initiative, intervention, or program. Individual levels of impact.

Efficacy/Effectiveness – the impact of the intervention on important outcomes, including potential negative effects, quality of life, and economic outcomes. Individual levels of impact.

Adoption - the absolute number, proportion, and representativeness of settings and intervention agents (people who deliver the program) who are willing to initiate a program. Comparisons should be made on basic information such as resource availability, setting size and location, and interventionist expertise. Organisational levels of impact.

Implementation – at the setting level, implementation refers to the intervention agents’ integrity to the various elements of an intervention’s protocol, including consistency of delivery as intended and the time and cost of the intervention. At the individual level, implementation refers to clients’ use of the intervention strategies. Organisational levels of impact.

Maintenance – The extent to which a program or policy becomes institutionalised or part of the routine organisational practices and policies. At the individual level, it refers to the long-term effects of a program on outcomes after 6 or more months after the most recent intervention contact. Both individual and organisational levels of impact.

Example – taken from www.re-aim.org

A school-based intervention that has a large impact in terms of reach and efficacy at the individual-level but is only adopted, implemented and maintained at a small number of organisations (with specific resources that are not available in typical ‘real-world’ schools) could potentially be described as an intervention that has a large potential for impact (if the RE-AIM model was not used). In reality, when considering organisational-level impact, in addition to individual –level impact, this intervention would have little hope of resulting in a large public health impact because it could not be adopted, implemented and maintained in real-world settings.

This is also true of the converse situation where an intervention has systemic organisational adoption, implementation, and maintenance, but little reach, efficacy or maintenance at the individual level. So if only one level was assessed (i.e. the organisational level) the impact of the intervention would be considered large even though there is no individual-level reach, efficacy or maintenance.

Case study - The Victoria Council on Fitness and General Health Inc. (VICFIT)

VICFIT was established through the Ministers for Sport and Recreation and Health to provide advice to government and to coordinate the promotion of fitness in Victoria. One of VICFIT’s initiatives, the Active Script Program (ASP), was designed to enable all general practitioners in Victoria to give consistent, effective and appropriate physical activity advice in their particular communities. The evaluation of the initiative utilised the RE-AIM framework, which is available at <http://www.vicfit.com.au/activescript/DocLib/Pub/DocLibAll.asp>.

Reviewers should describe the body of evidence with respect to the main domains relevant to the applicability of public health and health promotion interventions to the users' needs – see Table Two.

Table Two. Evaluation of the applicability of an individual study or a body of evidence

| RE-AIM evaluation factor | Domain | Characteristic | Data to be collected from the study* | Applicability to the user's needs* |
|--------------------------|-------------------|--|--|--|
| Reach | Sample | Sampling frame | How well the study population resembles the target population the authors indicate they would like to examine Inclusion and exclusion criteria | Does the study population resemble that of the user's with respect to relevant characteristics, eg., disease risk factors? |
| | | Sampling method | Participation rate The representativeness of the study population to the target population, eg., volunteers, provider/researcher selected, random sample Characteristics of the non-participants | If the study population was selected (i.e. not a random sample with a high participation rate), how might the user's population differ? Might they be less receptive to the intervention? |
| | Population | Age | Age of the population | What age of population do the data likely apply to, and how does this relate to the user's needs? |
| | | Sex | Percentage of each sex in the population | What sex do the data likely apply to, and how does this relate to the user's needs? |
| | | Race/ethnicity | Race/ethnicities are represented in the study population | Are the data likely specific to a specific racial/ethnic group, or are they applicable to other groups? |
| | | Health status and baseline risk | Percentage of the population affected at baseline by diseases or risk factors | How does the baseline health status of the user's population compare to that of the study population? |
| | | Other | Other population characteristics that are relevant to outcomes of this intervention | Are there other population characteristics that are relevant to outcomes of this intervention? |
| Efficacy | Internal validity | Internal validity | Assess internal validity for the study | Can the study results be attributed to the intervention or are there important potential confounders? |
| | Outcomes | Process and intermediate outcomes | Process (eg., number of telephone calls to clients) and intermediate outcomes (eg., dietary change) examined in the study | Are the outcomes examined in the study relevant to your population? Are the linkages between more proximal (intermediate and process) outcomes based on sufficient evidence to be useful in the current situation? |
| | | Distal health and quality of life outcomes | Health and quality of life outcomes examined in the study | Are the outcomes examined in the study relevant to user's population? |

| | | | | |
|-----------------------|------------------|--|--|--|
| | | Economic efficiency | Economic outcomes: cost, cost effectiveness, cost-benefit, or cost-utility | Is economic efficiency part of the decision-making process? If so, are the data on cost or economic efficiency relevant to the user's situation? |
| | | Harms | Any harms from the intervention that are presented in the data | Are these harms relevant to the user's population? Are there other potential harms? How is the user balancing potential benefits with potential harms? |
| Adoption | Intervention | Provider | Who delivered the intervention Training and experience of the interventionists If the intervention is delivered by a team, indicate its members and their specific tasks | Are the described interventions reproducible in the situation under consideration? Is the provider expertise and training available? |
| | | Contacts | Number of contacts made between the providers and each participant Duration of each contact | Is the frequency of contacts in the study feasible in the current situation? |
| | | Medium | Medium by which the intervention was delivered: in-person, telephone, electronic, mail | Is this medium feasible in the user's situation? |
| | | Presentation format | To individuals or groups With family or friends present | Is this format feasible in the current situation? |
| | | Content | Based on existing tools and materials or developed <i>de-novo</i> Tailoring of the intervention to individuals or subgroups | Is this feasible in the current situation? |
| | Setting | Infrastructure of the health care delivery system or the community | Organisational or local infrastructure for implementing the intervention | Is the needed infrastructure present in the current situation? |
| | | Access to the intervention | Access to the intervention among the target population | Does the current situation provide the resources to ensure access to the intervention? |
| Implementation | Individual level | Adherence | Individual rate of adherence to the intervention Attrition rate from the program | Are there barriers to adherence in the current situation? Are their local factors that might influence the attrition rate? |
| | Program level | Integrity | The extent to which the intervention delivered as planned | Are there barriers to implementation in the current situation? |
| Maintenance | Individual level | Sustainability of outcomes | Change in behaviour or other important outcomes in the long term | What is the relative importance of short- versus long-term outcomes to the user? |
| | Program level | Sustainability of the intervention | Facets of the intervention that were sustainable in the long term Infrastructure that supported a sustained | Is the intervention feasible in the long term in the user's setting? Does the necessary infrastructure exist? Are there available resources? What |

| | | | | |
|--|--|--|--|---|
| | | | intervention Barriers to long-term use of the intervention | barriers to sustainability might be anticipated? |
|--|--|--|--|---|

* “Data to be collected” and “applicability” can be applied to the individual study or to the body of evidence

7) Trade-offs between benefits and harms

Reviewers should discuss whether there were any adverse effects of the interventions, or describe if there were certain groups that received more/less benefit from the interventions (differential effectiveness). If cost data is provided for the interventions studies this should also be reported.

8) Implications for practice and future health promotion and public health research

Public health and health promotion reviewers are in an ideal position to determine the implications for practice and future research to be conducted to address any gaps in the evidence base. For example, where evidence is shown to be lacking, reviewers should clearly describe the type of research required, including the study design, participants, intervention details and contexts and settings. If the reviewed evidence base is flawed due to particular methodological issues (eg. outcome assessment tools, allocation bias, etc) these quality issues can be addressed in future studies.

REFERENCES

1. 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. March 2001. <http://www.york.ac.uk/inst/crd/report4.htm>
2. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health*. 2004 Mar;94(3):416-22.
3. Sorensen G, Emmons K, Hunt MK, Johnston D. Implications of the results of community intervention trials. *Annu Rev Public Health*. 1998;19:379-416.
4. Mittelmark MB, Hunt MK, Heath GW, Schmid TL. Realistic outcomes: lessons from community-based research and demonstration programs for the prevention of cardiovascular diseases. *J Public Health Policy*. 1993 Winter;14(4):437-62.
5. Nutbeam D, Harris E. *Theory in a Nutshell. A practical guide to health promotion theories*. McGraw-Hill Australia Pty Ltd, 2004.
6. Hawe P, Shiell A, Riley T, Gold L. Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *J Epidemiol Community Health*. 2004 Sep;58(9):788-93.
7. Shediach-Rizkallah MC, Bone LR. Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy. *Health Educ Res* 1998;13:87-108.
8. Bossert TJ. Can they get along without us? Sustainability of donor-supported health projects in Central America and Africa. *Soc Sci Med* 1990;30:1015-23.

9. Swerrissen H, Crisp BR. The sustainability of health promotion interventions for different levels of social organization. *Health Promot Int* 2004;19:123-30.
10. The Health Communication Unit. Overview of Sustainability, University of Toronto, Centre for Health Promotion, 2001. Available from: http://www.thcu.ca/infoandresources/publications/SUS%20Master%20Wkbk%20and%20Wkshts%20v8.2%2004.31.01_formatAug03.pdf
11. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999;89:1322-7.

ADDITIONAL READING

Rychetnik L, Frommer MS. Schema for Evaluating Evidence on Public Health Interventions; Version 4. National Public Health Partnership, Melbourne 2002.

Visit <http://www.re-aim.org> for information relating to generalising the results from primary studies.

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.

Dzewaltowski DA, Estabrooks PA, Klesges LM, Bull S, Glasgow RE. Behavior change intervention research in community settings: how generalizable are the results? *Health Promot Int*. 2004 Jun;19(2):235-45.

EXERCISE

1. In small groups, list the types of information required from studies to help you determine the applicability of the results to other settings and the transferability of interventions to other settings. Do **not** use the table provided.

FACILITATOR'S GUIDE

Time required:

One hour approx. (20 minutes slides, 40 minutes exercise) – may not take this long, depends on size of group

Learning topics

- Components of the discussion of the review
- Theoretical frameworks
- Influence on context
- Applicability of the results from the review
- Sustainability of interventions and/or outcomes

Summary of activity

Lecture – Powerpoint slides

Group activity to determine factors affecting applicability of results and transferability of interventions to other settings

Description of supporting materials

None.

Further reading (if required)

1. Nutbeam D, Harris E. Theory in a Nutshell. A practical guide to health promotion theories. McGraw-Hill Australia Pty Ltd, 2004. *** highly recommended
2. Visit <http://www.re-aim.org> for information relating to applicability of recommendations and the table included in the training manual.
3. 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.
4. Dzewaltowski DA, Estabrooks PA, Klesges LM, Bull S, Glasgow RE. Behavior change intervention research in community settings: how generalizable are the results? Health Promot Int. 2004 Jun;19(2):235-45.
5. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health. 1999 Sep;89(9):1322-7.

Instructions to facilitator

The assessment of sustainability and applicability is also determined by the quality of reporting in the primary studies. However, reviewers should make some attempt to guide the user on the applicability of the results.

Mention to participants that information in primary studies is often lacking in terms of reporting the theoretical frameworks of interventions and the context in which they are implemented.

Exercise: Allow participants 20 minutes for the exercise and then 10-20 minutes reporting back – put all ideas onto a whiteboard and determine if they match all the factors listed in the table provided in the handbook.

Additional information:

Reach – require information on the representativeness of the study sample. This is defined as the similarity or differences between those who participate and those who are eligible but do not. The intervention may have a differential impact based on the variables that differ between participants and non-participants. Less differences, better generalisability.

Adoption - Comparisons should be made on basic information such as resource availability, setting size and location, and interventionist expertise.

Examples from www.re-aim.org

Reach:

Eakin and her associates (1998) illustrate how Reach issues can be incorporated in a smoking cessation study offered to participants in a planned-parenthood program. They begin by explicitly reporting their inclusion criteria --female smokers between 15 and 35 years of age who are patients at a planned-parenthood clinic. During a routine visit to the clinic the patient services staff described the study and solicited participants. Those women who declined (n=185) were asked to complete a short questionnaire that included questions to assess demographics, smoking rate, and reasons for non-participation. Participants (n=518) also completed baseline demographic and smoking rate assessments. They tracked recruitment efforts and reported that 74% percent of the women approached agreed to participate in the study. To determine the representativeness of the sample two procedures were completed. First, based on information from patient medical charts, those who were contacted were compared on personal demographics to those who were not contacted. Second, participants were compared to non-participants on personal demographics and smoking rate. The study found that those contacted did not differ from those not contacted on any of the test variables. Also, the results suggested that participants were slightly younger than non-participants, but there were no other differences between these groups. This suggests that Eakin and her associates were fairly successful in contacting and recruiting a fairly representative sample of their target population.

The Language for Health (Elder et al., 2000) nutrition education intervention provides a good example of determining the representativeness of study participants to a given target population. The behaviour change intervention was developed to target Latino participants in English as a second language (ESL) classes at seven schools. To examine representativeness, the 710 participants in the study were compared to the overall Latino ESL student population in the city. This comparison revealed that the intervention participants did not differ from the general ESL student population on gender, age, or education level. As such, the authors concluded that the study had strong generalisability to the greater target population (Elder et al., 2000).

Efficacy:

Project GRAD (Graduate Ready for Activity Daily; Zabinski, Calfas, Gehrman, Wilfley, & Sallis, 2001). Project GRAD was an efficacy trial of a physical activity promotion course offered to university students. The results of the trial revealed that the intervention was successful in increasing total energy expenditure, strength exercises, and flexibility exercises of female students. Unfortunately, there were also some unintended negative consequences for women who participated. Female participants who participated in the intervention showed a significant increase in drive for thinness as measured by the Eating Disorder Inventory. This increase was not observed in females in the control condition. These findings suggest that there is a need to assess both

positive and potential negative outcomes associated with physical activity trials. By doing so, determining whether an intervention does more good than harm can be assessed in a comprehensive manner. Incidentally, in project GRAD, the mean score for drive for thinness of females in the experimental condition, although statistically significant, was still in the normal range when compared to normative data. As a result, the trial may be considered to have done more good than harm.

Adoption:

The SPARK physical activity and fitness trial that targeted elementary students provides a good example of the proportion component of adoption (Sallis et al., 1997). In this study, the principals of 16 elementary schools were approached for participation in the trial. Twelve of the 16 schools were willing to participate; however, because of the level of research funding only seven of the 12 were selected for participation. Although there were no tests of representativeness of school resources, location, staff-to-student ratio, or other school-level variables, they did document that the seven smallest schools were selected for participation. Based upon this information one could conclude that the effects of the intervention could generalize to other small schools with similar resources, but effects of the intervention may not generalize to larger schools.

Implementation:

Baranowski and colleagues provided a good example of rigorously documenting implementation rates of the Gimme 5 Fruit, Juice, and Vegetables for Fun and Health Trial. In that study of fourth and fifth grade students, the intervention curriculum included components to be delivered at the school and newsletters with family activities and instructions for intervention at home. Researchers documented the delivery of the curriculum as intended through classroom observations and teacher self-report of the completion of the curriculum activities. All teachers were observed at least once during the 6-week intervention. The observations revealed that only 51% and 46% of the curriculum activities were completed in the fourth and fifth grade years of intervention, respectively. In contrast, teacher self-reported delivery was 90%.

Resnicow, et al. (1997) also demonstrated the need to track implementation of treatment delivery. One component of their self-help smoking cessation program was a telephone booster call. Of the 650 participants in the intervention arm of their study only 31% were reached for the intervention telephone call. They found that those who received the call had a significantly higher abstinence rate than those in the control and those in the intervention who had not received the booster call. Had the authors not documented the delivery of the intervention as intended, the null finding could have been attributed to an ineffective intervention rather than to an ineffective delivery of the intervention.

Maintenance:

Project ACTIVE (Dunn et al., 1999), a 24-month randomised clinical trial comparing the effects of two treatment arms on physical activity and cardiorespiratory fitness in adults, provides an example of maintenance data at the participant level. Both treatment groups received six months of intensive intervention. Measures, including physical activity, were obtained at the beginning of the clinical trial, at six months, and at 24 months. Findings indicated that both groups increased activity from the beginning to six months (i.e., during the intensive intervention) but decreased activity from six to 24 months. These findings support the need for multiple assessments of behaviour in order to determine the pattern of behaviour and, thus, whether participants maintain activity.

Although few studies have documented setting level maintenance or institutionalization, Richmond and colleagues (1998) provide an example of including this information in addition to reporting patient level long-term follow-up results and attrition rates. They followed up on family physicians

who had participated in their smoking cessation training program. They found that 6 months after the 2-hour training, 93% of intervention condition physicians reported still using the program.

Unit Eleven: 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 judge the quality of a systematic review

When others read your review they will be assessing it for the systematic manner in which bias was reduced. A useful tool to assess the quality of a systematic review is produced by the Critical Appraisal Skills Programme (CASP) and can be found at <http://www.phru.nhs.uk/~casp/appraisa.htm> (provided overleaf). It is useful to keep this tool in mind when writing the final review.

Reviewers may consider submitting their review to:

- 1) The Cochrane Collaboration – must go through the Cochrane editorial process
- 2) The Database of Abstracts of Reviews of Effects (DARE) – this database is held by the University of York - <http://www.york.ac.uk/inst/crd/crddatabases.htm>
- 3) The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) to be included in The Database of Promoting Health Effectiveness Reviews (DoPHER) - <http://eppi.ioe.ac.uk>
- 4) A published journal relevant to the topic of the review.

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:

1) *Systematic reviews of RCTs:*

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.

2) *Systematic reviews of observational studies:*

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.

ADDITIONAL READING

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.

EXERCISE

1. Critically appraise in small groups "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", using the 10-question CASP checklist.

Critical Appraisal Skills Programme (CASP)
making sense of evidence

10 questions to help you make sense of reviews

How to use this appraisal tool

Three broad issues need to be considered when appraising the report of a systematic review:

Is the study valid?

What are the results?

Will the results help locally?

The 10 questions on the following pages are designed to help you think about these issues systematically.

The first two questions are screening questions and can be answered quickly. If the answer to both is "yes", it is worth proceeding with the remaining questions.

You are asked to record a "yes", "no" or "can't tell" to most of the questions. A number of italicised prompts are given after each question. These are designed to remind you why the question is important. Record your reasons for your answers in the spaces provided.

The 10 questions are adapted from Oxman AD, Cook DJ, Guyatt GH, Users' guides to the medical literature. VI. How to use an overview. JAMA 1994; 272 (17): 1367-1371

© Milton Keynes Primary Care Trust 2002. All rights reserved.

Screening Questions

1 Did the review ask a clearly-focused question? Yes Can't tell No

Consider if the question is 'focused' in terms of:

- the population studied
 - the intervention given or exposure
 - the outcomes considered
-

2 Did the review include the right type of study? Yes Can't tell No

Consider if the included studies:

- address the review's question
 - have an appropriate study design
-

Is it worth continuing?

Detailed questions

.....

3 Did the reviewers try to identify all the relevant studies? Yes Can't tell No

Consider:

- which bibliographic databases were used
 - if there was follow-up from reference lists
 - if there was personal contact with experts
 - if the reviewers searched for unpublished studies
 - if the reviewers searched for non-English language studies
-

4 Did the reviewers assess the quality of the included studies? Yes Can't tell No

Consider:

- if a clear, pre-determined strategy was used to determine which studies were included. Look for:
 - a scoring system
 - more than one assessor
-

5 If the results of the studies have been combined, was it reasonable to do so? Yes Can't tell No

Consider whether:

- the results of each study are clearly displayed
 - the results were similar from study to study (look for tests of heterogeneity)
 - the reasons for any variations in results are discussed
-

.....

6 How are the results presented and what is the main result?

Yes Can't tell No

Consider:

- how the results are expressed (eg. odds ratio, relative risk, etc.)
 - how large this size of result is and how meaningful it is
 - how you would sum up the bottom-line result of the review in one sentence
-

7 How precise are these results?

Yes Can't tell No

Consider:

- if a confidence interval were reported. Would your decision about whether or not to use this intervention be the same at the upper confidence limit as at the lower confidence limit?
 - if a p-value is reported where confidence intervals are unavailable
-

8 Can the results be applied to the local population?

Yes Can't tell No

Consider whether:

- the population sample covered by the review could be different from your population in ways that would produce different results
 - your local setting differs much from that of the review
 - you can provide the same intervention in your setting
-

9 Were all important outcomes considered?

Yes Can't tell No

Consider outcomes from the point of view of the:

- individual
 - policy makers and professionals
 - family/carers
 - wider community
-

10 Should policy or practice change as a result of the evidence contained in this review?

Yes Can't tell No

Consider:

- whether any benefit reported outweighs any harm and/or cost. If this information is not reported can it be filled in from elsewhere?
-

FACILITATOR'S GUIDE

Time required

1 hour (10 minute discussion, 50 minutes exercise)

Learning topics

- Guidelines for publishing systematic reviews
- Appraising the quality of a systematic review

Summary of activity

Lecture – Powerpoint slides

Exercise in groups – critical appraisal of a systematic review

Description of supporting materials

- 1) Critical appraisal paper to be included in course materials. “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”.
- 2) CASP critical appraisal checklist for systematic reviews.
- 3) Rapid responses to the article in *BMJ* (optional)

Further reading (if required)

1. 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.

Instructions to facilitator

There are very few slides for this unit – use your knowledge or the critical appraisal checklist to discuss or review the components of a good systematic review. Review the past two days to provide examples to participants of the steps of the systematic review process. Highlight that these are the same steps that are appraised.

The appraisal exercise utilises the quality appraisal checklist developed by the Critical Appraisal Skills Programme. The crib sheet for the appraisal exercise is attached.

The exercise should highlight to participants everything that they have been covered throughout the two days, i.e. the need for a clear question, comprehensive search, critical appraisal, etc.

Issues for further discussion:

- 1) Are single factor interventions limited in their capacities to address the complex, multiple, and ongoing influences of parents, peers, health service providers, schools, socioeconomic conditions, religion, and the media which shape the values, beliefs and attitudes determining sexual risk-taking? For example, the multi-factor intervention included in the review was the most effective intervention.
- 2) Was it the right decision to only include RCTs in this review? What would be the value of including other study designs? i.e. compare to the other systematic review which showed positive results, and included non-RCTs. The RCTs in this review had flaws (only 8 of 22 studies scored over

2 points out of the possible 4 points for quality); including high quality experimental evidence (from non-RCTs) may be advisable.

3) The PICO question is not P- adolescents, I – primary prevention C- no intervention O- delay in initiation of sexual intercourse, consistent use of birth control, avoidance of unintended pregnancy
The PICO questions is P- adolescents, I – primary prevention using theory based interventions C- standard practice or care (only a few interventions are compared to no intervention in the review) O- delay in initiation of sexual intercourse, consistent use of birth control, avoidance of unintended pregnancy. Therefore, the first line of the discussion of the article is too strong and does not reflect what the review was actually answering “..show that primary prevention strategies do not delay the initiation of sexual intercourse or improve use of birth control among young men and women”.

Note: There are a number of rapid responses to this article in the BMJ. They can be found on-line when printing out the article of the systematic review. Participants have found it useful to look at these rapid responses (especially as many think the study is very high quality and useful for decision-making).

Systematic review – critical appraisal exercise

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

| Screening questions | |
|---|---|
| <p>1. Did the review address a clearly focused research question?</p> <p><i>HINT: A research question can be 'focused' in terms of</i></p> <ul style="list-style-type: none"> - the population studied - the intervention given - the outcomes considered | <p>Yes.</p> <p>Population – children 11-18 years</p> <p>Intervention – Primary pregnancy prevention programmes – sex education classes, school based clinics, family planning clinics, community based programmes.</p> <p>Control: NOTE: Comparison is conventional sex education, comparison is not no intervention. This review is only measuring the difference between the two, and is in fact comparing theory-led approaches to conventional approaches. It is not answering the question of sex education versus no education.</p> <p>Outcomes – delay in sexual intercourse, consistent use of birth control, avoidance of unintended pregnancy</p> |
| <p>2. Did the review include the right type of studies?</p> <p><i>HINT: These would</i></p> <ul style="list-style-type: none"> - address the review's research question - have an appropriate study design | <p>Probably not</p> <p>Although the effectiveness of interventions is best measured using RCTs, this review could have benefited from including high quality non-randomised controlled trials. This type of intervention would normally be measured using non-RCTs (in 'real-life') in most countries, and high quality non-RCTs in this review could contribute more than poorly-designed RCTs.</p> <p>Therefore, this review should have included both kinds of studies. The authors could have separated the results for non-RCTs and RCTs to see if the inclusion criteria made any difference to the final results.</p> |
| Detailed questions | |
| <p>3. Did the reviewers try to identify all relevant studies?</p> <p><i>HINT: Look for</i></p> <ul style="list-style-type: none"> - which bibliographic databases were used - follow up from reference lists - personal contact with experts - search for unpublished as well as published studies | <p>Yes.</p> <p>Searched databases from 1970 to December 2000.</p> <p>Used:</p> <ul style="list-style-type: none"> McMaster Teen Project database Electronic databases, conference papers Dissertation abstracts Reviewed content lists of journals |

| | |
|---|---|
| <p>- search for non-English language studies</p> | <p>Reviewed reference lists Sent the lists to experts to review for completeness</p> <p>26 RCTS found in 22 reports that met inclusion criteria.</p> |
| <p>4. Did the reviewers assess the quality of the included studies?</p> <p><i>HINT: A clear pre-determined strategy should be used to determine which studies are included. Look for:</i></p> <ul style="list-style-type: none"> - a scoring system - more than one assessor | <p>Yes.</p> <p>Assessed quality of RCTs using modified version of Jadad.</p> <p>Assessed according to:</p> <ul style="list-style-type: none"> Appropriateness of randomisation Extent of bias in data collection Proportion of study participants followed to the last point of follow-up (adequate $\geq 80\%$ of study participants) Similarity of attrition rates in the comparison groups (within 2%) <p>Scoring system used: 1 point for each, poor studies ≤ 2.</p> <p>Data extraction section: two people independently extracted data on setting, participants, unit of randomisation and analysis, etc.</p> |
| <p>5. If the results of the review have been combined, was it reasonable to do so?</p> <p><i>HINT: Consider whether</i></p> <ul style="list-style-type: none"> - the results of all the included studies are clearly displayed - the results of the studies are similar from study to study (look for tests of heterogeneity) - the reasons for any variations in results are discussed | <p>Maybe, but given the likely heterogeneity of the studies (settings, populations, interventions) a narrative synthesis should have been employed.</p> <p>10 of 22 studies had cluster randomisation. Assessed correlation within clusters. Could only get correlation information from one study, so used that as the estimates for other studies.</p> <p>Tested for heterogeneity using X^2, considered $p < 0.1$ as indication of heterogeneity.</p> <p>Although heterogeneity was present for birth control in young women and birth control at last intercourse (young women) the authors tested 10 hypotheses a priori for heterogeneity. These hypotheses did not explain the heterogeneity. Therefore, a narrative synthesis may have been more meaningful.</p> |
| <p>6. What are the main results of the review?</p> <p><i>HINT: Consider</i></p> <ul style="list-style-type: none"> - how were the results expressed (odds ratio, relative risk etc.) | <p>1) Initiation of sexual intercourse (Fig 1)</p> <p>Young women: 13 studies of 9642 young women showed no delay in sexual intercourse. Pooled odds ratio 1.12; (95% CI 0.96 to 1.30).</p> <p>Young men: 11 studies of 7418 young men showed no delay in initiation of sexual</p> |

| | |
|--|---|
| | <p>intercourse. Pooled odds ratio 0.99 (95% CI 0.84 to 1.16).</p> <p>2) Use of birth control</p> <p>Every time they had intercourse (Fig 2) Young women: 8 studies of 1967 women showed no improvement in use of birth control at every intercourse (0.95; 0.69 to 1.30). There was significant heterogeneity among studies (p=0.08). Unexplained heterogeneity. Young men: 3 studies of 1505 young men indicated that the programmes did not improve the use of birth control at every intercourse (0.90;0.70 to 1.16). Heterogeneous studies.</p> <p>Use of birth control last time they had intercourse (Fig 3) Young women: 5 studies of 799 young women showed no improvement (1.05;0.5 to 2.19) with significant heterogeneity, p=0.007). Heterogeneity not explained by hypotheses. Young men: 4 studies of 1262 young men showed no improvement (1.25;0.99 to 1.59), no heterogeneity.</p> <p>3) Pregnancy Young women: 12 studies of 8019 young women showed that the interventions did not reduce pregnancy rates (1.04;0.78 to 1.4), no heterogeneity. Young men – pregnancies with partners – 5 studies of 3759 young men showed no effect of interventions on reducing pregnancies among partners of young men, (1.54;1.03 to 2.29), no heterogeneity.</p> |
| <p>7. Could these results be due to chance? <i>HINT: Look for tests of statistical significance eg. p values and confidence intervals (CIs)</i></p> | <p>Most results are not significant. The only significant difference was for pregnancies in partners of young men, CI (1.54;1.03 to 2.29). Confidence interval does not include one, the null hypothesis.</p> |
| <p>8. Can the results be applied to the local population? <i>HINT: Consider whether</i> - the population sample covered by the review could be sufficiently different to your population to cause concern. - your local setting is likely to differ much from that of the review</p> | <p>Most of the participants in over half of the studies were African-American or Hispanic, thus over-representing low socioeconomic groups. Interventions may be more successful in other populations.</p> <p>In all but 5 studies, participants received conventional sex education in the control group.</p> <p>It is possible that these interventions had some effect, and tested interventions were not potent</p> |

| | |
|---|--|
| | enough to exceed this effect. What is conventional sex education? |
| <p>9. Were all important outcomes considered?</p> <p><i>HINT: Consider outcomes from the point of view of the:</i></p> <ul style="list-style-type: none"> - individual - policy makers and practitioners - family/carers - wider community | <p>Costs not included.</p> <p>Knowledge and attitudes not included.</p> |
| <p>10. Should policy or practice be changed as result of the evidence contained in this review?</p> <p><i>HINT: Consider whether the benefits are worth the harms and costs</i></p> | <p>No.</p> <p>Recommendations:</p> <ol style="list-style-type: none"> 1) Review RCTs and non-RCTs. 2) Conduct new studies with young people designing the intervention. 3) Only 12 studies conducted lasted >1 yr duration – need longer term studies. |