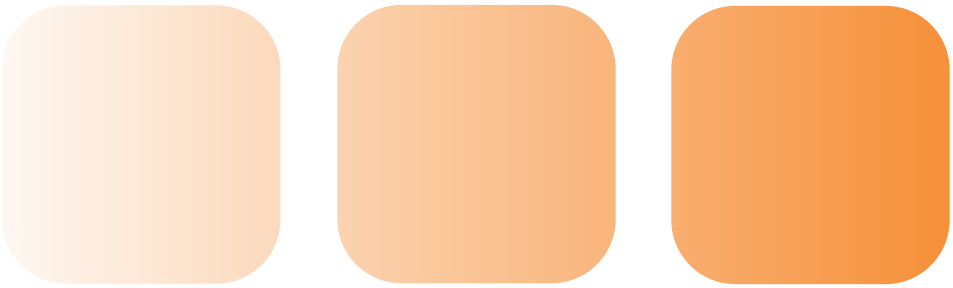
Joanna Briggs Institute
Comprehensive Systematic
Review - Training Programme

CSRTP **Study Guide**

The Systematic Review of Evidence Generated through Quantative Research



THE JOANNA BRIGGS INSTITUTE

THE UNIVERSITY of ADELAIDE

Joanna Briggs Institute
Comprehensive Systematic
Review - Training Programme

CSRTP **Study Guide**

The Systematic Review of Evidence Generated through Quantative Research

JBI/CSRTP/2012-13/0002

Published by The Joanna Briggs Institute, Adelaide, South Australia, Australia.

First published 2009

THE JOANNA BRIGGS INSTITUTE | THE UNIVERSITY of ADELAIDE

Contents

Introduction

Welcome to the quantitative module of the Joanna Briggs Institute Comprehensive Systematic Review Training. This Module is designed to provide participants skills in the systematic review of evidence generated by experimental studies, observational studies and descriptive studies. The program has a hands-on orientation and includes computer laboratory sessions. Underpinning this approach is clearly defined content contained in extensive printed course material.

Aim and Objectives

The aim of this module is to enable participants to develop a comprehensive understanding of the purposes and principles of appraising, extracting and pooling the results of quantitative studies.

The objectives of this module are to prepare participants to:

- critically appraise experimental, observational and descriptive studies;
- extract numerical data from studies;
- combine the results of two or more studies in statistical meta-analysis;
- develop a systematic review protocol; and
- fully utilize the JBI- CReMS and JBI-MAStARI software programs.

Recommended Readings:

Pearson A, Field J. and Jordan Z. (2007) 'Evidence-Based Clinical Practice in Nursing and Health Care.' Oxford, Blackwell Publishing.

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press.

Readings (Attached in accompanying Reader)

Joanna Briggs Institute (2011) Reviewers Handbook. Adelaide, Joanna Briggs Institute.

Joanna Briggs Institute (2011) SUMARI User Guide. Adelaide, Joanna Briggs Institute

Lassen M.,R., Borris L.,C., Nakov R.,L.(2002) Use of the low-molecular-weight heparin reviparin to prevent deep-vein thrombosis after leg injury requiring immobilisation. N Engl J Med 2002;347(10):726-30.

Lapidus, L.J., Ponzer, S., Elvin, A., Levander, C., Larfars, G., Rosfors, F., De Bri, E. (2007). Prolonged thromboprophalaxsis with Dalteparin during immbolization after ankle fracture surgery: A randomized, placebo- controlled, double blind study. Acta Orthopaedica, 78:4, 528-535.

Mehta, A., Sequeira, P.S., Sahoo, R.C., Kaur, G. (2009) Is Bronchial Asthma a Risk Factor for Gingival Diseases? A Control Study. NYSDA, January, 44-47

Troisi, R.J., Speizer, F.E., Willet, W.C., Trichopolous, D. and Rosner, B. (1995) Menopause, Postmenopausal Estrogen Preparations, and the risk of Adult-onset asthma. A prospective cohort study. Am J Respir Crit Care 152: 1183-1188.

Assessment

Multiple choice question assessment of 30 minutes.

Presentation of a draft, preliminary review protocol.

Program of Study

Day 1		
Time	**Session**	**Group Work**
09.00	Introductions and overview of Module 3	
09.30	Session 1: The Critical Appraisal of Studies	
10.00	**Morning Tea**	
10.30	Session 2: Appraising RCTs and experimental studies	Group Work 1: Critically appraising RCTs and experimental studies. Report back
11.45	Session 3: Appraising observational Studies	
12.30	**Lunch**	
13.30		Group Work 2: Critically appraising observational studies. Report back
14.15	Session 4: Study data and data extraction	
15.15	**Afternoon tea**	
15.30		Group Work 3: Data extraction. Report back
16.00	Session 5: Protocol development	Protocol development
17.00	End	

Day 2		
09.00	Overview of Day 1	
09.15	Session 6: Data analysis and meta-analysis	
10.30	**Morning Tea**	
11.00	Session 7: Appraisal extraction and synthesis using JBI MAStARI	Group Work 4: MAStARI trial. Report back
12.30	**Lunch**	
13.30	Session 8: Protocol Development	Protocol development
14.15	Session 9: Assessment	MCQ Assessment
14.45	**Afternoon tea**	
15.00	Session 10: Protocol Presentations	Protocol Presentations
17.00	End	

Session 1:
The Critical Appraisal of Studies

The Introductory Module to the Comprehensive Systematic Review Training program introduced the following four steps in the systematic review process common to reviews of any evidence type on Day 1:

- Developing a question (PICO);
- Inclusion Criteria;
- Search Strategy;
- Selecting Studies for Retrieval.

This module considers the three steps in the systematic review of studies that are retrieved from the initial search results – appraising the studies to determine whether they are of sufficient quality to include in the review; extracting data from studies that are included in the review; and combining the results of two or more studies in statistical meta-analysis.

Critical Appraisal

All papers selected for inclusion in the systematic review (that is – those that meet the inclusion criteria described in the protocol) need to be subjected to rigorous appraisal by two critical appraisers. Critical appraisal is probably the most difficult component of the systematic review and a good understanding of research design is required. Many checklists have been developed that can be used by appraisers and different checklists are used for different research designs.

The purpose of appraisal is to include only those studies that are of high quality and to exclude those of poor quality. Given that the systematic review aims to summarize the best available evidence through, where possible, pooling the results of sufficiently similar studies, it is important to note that the pooling of poor quality evidence may lead to outcomes than are less than desirable for patients.

The major aim of critical appraisal of quantitative evidence is to establish the validity of the evidence. Validity refers to the soundness of the evidence; in other words it is about the degree to which we can accept the evidence as trustworthy and believable.

The validity of quantitative studies refers to the degree to which possible bias has been limited. Bias refers to any influence that may distort the results of a study. Relying on the results of studies with variable validity and pooling these results may lead to a conclusion that is incorrect and thus mislead us to conclude that an activity or intervention is effective when it is not, or vice versa. Sources of bias in randomized controlled trial (RCT) designs include selection bias, performance bias, attrition bias and detection bias.

Sources of Bias

Selection bias, or allocation bias, usually arises out of an inadequate randomization of subjects. There are strategies that can be used to minimize the chance of bias (e.g. allocation concealment). Randomization should ensure that every participant has an equal chance of being in any of the study groups and appraisers therefore need to establish how well this was achieved. Randomization of participants is of vital importance in a clinical trial. This prevents both the participants and the researchers from influencing the outcomes of the research. It is best achieved by an individual who is not directly linked to the study or by using a system whereby randomization of participants cannot be in any way manipulated, such as blinding or allocation concealment. When critically appraising a randomized controlled trial there are a number of indicators as to whether or not true randomization has occurred.

Performance bias arises when there are differences in care received other than the designated treatment under investigation. Performance bias is avoided by blinding of all involved in the study, both participants and

investigators, to what intervention any given participant has been randomly allocated to receive. It is essential that performance bias is limited in order to ensure that the study participants are protected against unintended differences in care and placebo effects. Again, this may be achieved through blinding both participants and carers to prevent 'contamination' or 'co-intervention' that may affect study results.

Detection bias, also referred to a s measurement bias, relates to differences that may occur when assessing the outcomes of study groups. Again, blinding is an important factor to consider when examining detection bias. Detection bias may arise from differences in the subjective interpretation of the level of some outcomes such as pain for example.

Attrition bias relates to differences in terms of losses of subjects, or loss to follow up, between groups. It is important that all differences between study groups are reported. Losses to a study include things such as participants who withdraw or drop out or deviate from the original protocol in some way. Losses to follow-up should be reported but this is sometimes difficult to assess.

The same sources of bias are relevant in appraising both observational and descriptive studies but, because these types of studies are less able to address questions of cause and effect than the RCT, the sources of bias are interpreted differently in the appraisal process. In the case of case control studies selection bias is minimized by identifying and dealing with potential confounding factors. In cohort studies, which are described later in this chapter, the degree to which the study sample is representative of the population is central to assessing validity.

Assessing the methodological Quality of a Study

There is disagreement internationally on the most appropriate way of carrying out the appraisal of RCT's, case control and cohort studies. These disagreements largely concern the number of appraisers required and whether or not (if there are two or more appraisers) the process should be carried out independently by each of the reviewers; what should occur if reviewers disagree; and what degree of expertise in research methods is necessary in reviewers. Notwithstanding the lack of consensus in this area, there is growing acceptance (as evidenced by the approaches adopted by many systematic review groups such as the Joanna Briggs Institute, the NHS Centre for Reviews and Dissemination and the Scottish Intercollegiate Guidelines Network) of the need for papers to be appraised independently by two reviewers, and for these reviewers to confer and reach agreement. There is also increasing allegiance to the view that reviewers should have previous training in the critical appraisal process and experience and training in research design.

Criteria for the appraisal of studies vary from centre to centre and it is generally accepted that criteria may be developed specifically for each review. Reviewers often design an appraisal sheet and pilot the criteria on a small number of papers to establish that the criteria are appropriate and that they discriminate between papers with obvious low or high-risk bias. Other reviewers suggest, however, that as the aim is to appraise the degree to which bias has been limited, the criteria should be common to all studies of the same design.

Session 2:
Appraising RCTs and Experimental Studies

RCTs and quasi (pseudo) RCTs/CCTs provide the most robust form of evidence for effects because they provide evidence related to whether or not a causal relationship exists between a stated intervention, and a specific, measurable outcome, and the direction and strength of the relationship

Properly performed RCTs reduce bias, confounding factors, and results by chance. They have three essential elements

- Randomization (where possible);
- Researcher-controlled manipulation of the independent variable; and
- Researcher control of the experimental situation.

RCTs are often used to evaluate how effective a new treatment/therapy/intervention is for patients with a certain condition. Individuals (or other units) are randomly allocated to a treatment group. Randomization is essential as this ensures that all treatment groups are comparable at the beginning. Confounding factors (variables), which may somehow impact upon the results of the study such as age, gender, etc will be spread evenly across groups to ensure treatment arms are as comparable as possible prior to receiving the intervention. Properly designed and performed randomized controlled trials reduce the risk of bias, confounding factors, and results by chance. However, poorly conducted randomized controlled trials are susceptible to bias and may produce misleading information or exaggerated treatment effects. (Kao et al., 2008; Moher et al 2001; Altman et al 2001)

Randomization

An effective method of randomization may be computer-generated sets of random allocations (true randomization method) to be put in opaque envelopes (allocation concealment) and assigned to participants (Kendall 2003; Togerson and Roberts, 1999).

Simple methods of randomization, such as tossing a coin or rolling a dice, may result in unequal group sizes, which can have an effect on the results of the trial, especially in smaller scale studies with smaller samples. Block randomization has been developed and is commonly used to address this issue. To ensure comparable sample sizes, participants are put in blocks (groups). Randomization then occurs in the block to ensure equal numbers are assigned to each treatment arm. It is important to blind the investigator from block size, and where this is not concealed, block sizes should vary randomly (Kendall 2003; Schultz et al., 1995).

Another issue that may occur with small sample sizes is chance imbalances between the treatment groups in terms of certain confounding factors. This can be reduced through stratification. This technique ensures that important baseline characteristics are even in both groups (Kendall 2003; Schultz et al., 1995). Stratification occurs prior to randomization by placing subjects in strata on the basis of differing characteristics of the participants, this may include differences in age, sex, co morbidities and prognosis for example, with randomization occurring within the strata.

Sampling

Sampling is the process of selecting individuals/groups from the target population and including them in the trial. The target population is the population, which the results of the trial will be relevant/applicable. Inclusion/exclusion criteria are set to define a specific study group for the trial. The most important issue to consider when selecting a sample is that the sample is representative of the target population. There are different methods of sampling from the population (Kendall 2003).

Probabilistic (Random) sampling is the random sampling of individuals from the target population. This ensures different individuals in the population all have an equal chance of being involved with the trial. Where and if possible probabilistic sampling should occur. Cluster or multi-stage sampling can be used when a large sample is required, where a random sample of hospitals is drawn and random patients from within each hospital.

Consecutive sampling is the consecutive sampling of every patient who meets the inclusion criteria from the population over a period. (Kendall 2003)

Systematic sampling occurs where samples are decided on a system, such as every third patient is to be enrolled in the trial. This may be hazardous if the investigator can affect the order in which patients are seen.

Convenience sampling is sampling by convenience. This sample technique will save time and money and is quite simple, but may not be representative of the target population. (Kendall 2003)

Blinding

Blinding is a method used to eliminate bias resulting from human behavior (Day and Altman, 2000). Therefore blinding is relevant for individuals included in the trial and also the investigators and assessors. When adequate blinding has not occurred, studies report a larger treatment effect on average (Day and Altman, 2000). Blinding ensures that those involved in the trial (including study participants, data collectors, those managing the participants) do not know which treatment group has been assigned (Day and Altman, 2000). Blinding is most important when subjective measures are used to assess outcomes, as these are more likely to be affected by knowledge of the treatment. Blinding is not always possible in every circumstance. Single-blind and double blind are terms used to state whether blinding has occurred at either both the participant and investigator or only at one (Day and Altman 2000). However, there are different interpretations of the terms single, double and triple blinded. It is more important for the report to state what specific means of blinding took place (Devereaux et al 2001, Moher et al 2001).

Critical appraisal/assessment of risk of bias in included studies

The key to high quality systematic reviews is a clearly formulated research question. Assessing the quality of studies refers to the extent to which the study design, conduct, and analysis minimize the potential for bias. The three features that have been empirically shown to influence or bias the results of study of intervention are randomization, blinding and allocation concealment (Schulz et al., 1995).

Assessment of the quality of the trials includes (Jadad et al., 1996):

- clinical relevance of the review question
- internal validity of the trial
- external validity of the trial
- appropriateness of data analysis and presentation

An important issue in designing studies like this is the generation of the comparison groups. Ideally, the two groups should be identical in every respect, except for the different interventions they receive. The process of randomization should be clearly documented in the published report of the study to enable the reviewers to assess this important aspect of study quality. If each person enrolled in the study has an equal chance of going into either group, over the long run all factors, known and unknown, will be equally distributed if there is a large enough number of participants or techniques such as block randomization and stratification are used. Systematic bias is avoided by selecting a sample of participants from a particular population and allocating them randomly to the different groups (Higgins and Green 2006).

Quality assessment of RCTs is essential in conducting high quality systematic reviews. The JBI checklist is used as a guide assessing the quality of RCTs in all JBI reviews and addresses the following questions:

- Were patients truly randomized?
- Was allocation to treatment groups concealed?
- Were patients analyzed in the groups to which they were randomized? (intention-to-treat analysis)
- Were patients in the treatment and control groups similar at entry?
- Were patients aware of group allocation?
- Were those assessing outcomes aware of group allocation?
- Were groups treated identically other than the intervention/s being investigated?
- Were outcomes measured in the same way for all the groups?

Intention to treat analysis

Definition

Intention to treat, abbreviated by ITT is a strategy for the analysis of randomized controlled trials (RCTs) that compares patients in the groups to which they were originally randomly assigned. This is generally interpreted as including all patients, regardless of whether they actually satisfied the entry criteria, the treatment actually received, and subsequent withdrawal or deviation from the protocol.

Purposes

The ITT approach has two main purposes (Hollis, 1999). Firstly, the approach maintains treatment groups that are similar apart from random variation. This is the reason for randomization, and the feature may be lost if analysis is not performed on the groups produced by the randomization process (Hollis, 1999). For example, in a trial comparing medical and surgical treatment for stable angina pectoris, some patients allocated to surgical intervention died before being operated on. If these deaths are not attributed to surgical intervention using an ITT analysis, surgery seems to have a falsely low mortality. Secondly, ITT analysis allows for noncompliance and deviations from policy by clinicians (Hollis, 1999). Some types of deviations from randomized allocation may occur only within the trial setting and would not be expected in routine practice. For example, in a trial comparing active and placebo vaccination there is the potential for placebo vaccine to be incorrectly administered in place of active, but this could not occur outside the trial and so need not be accounted for in estimates of potential efficacy. However, most types of deviations from protocol would continue to occur in routine practice and so should be included in the estimated benefit of a change in treatment policy. Intention to treat analysis is therefore most suitable for pragmatic trials of effectiveness rather than for explanatory investigations of efficacy.

The basic ITT principle is that participants in trials should be analyzed in the groups to which they were randomized, regardless of whether they received or adhered to the allocated intervention. Two issues are involved here (Alderson and Green, 2002). The first issue is that participants who strayed from the protocol (for example by not adhering to the prescribed intervention, or by being withdrawn from active treatment) should still be kept in the analysis. An extreme variation of this is participants who receive the treatment from the group they were not allocated to, who should be kept in their original group for the analysis. This issue causes no problems provided that, as a systematic reviewer, you can extract the appropriate data from trial reports. The rationale for this approach is that, in the first instance, we want to estimate the effects of allocating an intervention in practice, not the effects in the subgroup of participants who adhere to it.

The second issue in ITT analyses is the problem of loss to follow-up. People are lost from clinical trials for many reasons. They may die, or move away; they may withdraw themselves or be withdrawn by their clinician, perhaps due to adverse effects of the intervention being studied.

If participants are lost to follow-up then the outcome may not be measured on them. But the strict ITT principle suggests that they should still be included in the analysis. There is an obvious problem - we often do not have the data that we need for these participants. In order to include such participants in an analysis, we must either find out whether outcome data are available for them by contacting the coordinators or researchers of RCTs, or we must impute (make up) their outcomes. This involves making assumptions about outcomes in the lost participants.

Recommendations for intention to treat analysis "… is better regarded as a complete trial strategy for design, conduct and analysis rather than as an approach to analysis alone" (Lewis and Machin, 1993). For example:

Design

- Decide whether the aim is pragmatic or explanatory. For pragmatic trials, intention to treat should be considered essential;
- Justify in advance any inclusion criteria that when violated would merit exclusion from ITT analysis.

Conduct

- Minimize missing response on the primary outcome;
- Follow up subjects who withdraw from treatment.

Analysis

- Include all randomized subjects in the groups to which they were allocated;
- Investigate the potential effect of missing response.

Reporting

- Specify whether or not ITT analysis has been carried out, explicitly describing the handling of deviations from randomized allocation and missing response;
- Report deviations from randomized allocation and missing response;
- Discuss the potential effect of missing response;
- Base conclusions on the results of ITT analysis.

Group Work 1

Working in pairs, critically appraise the two RCT papers in your workbook

Reporting Back

Session 3:
Appraising observational studies

Observational studies represent an effective way of collecting information and data that is different from an experimental study or trial. Experimental studies are often not feasible because of ethical issues, costs, and difficulties in recruiting participants (Black, 1996). Therefore, a great deal of health research is conducted using observational studies. In observational studies, the investigator takes a less active role compared to trials. Observational studies take advantage of the fact that people are exposed to any number of possibilities throughout their day-to-day life and are not limited to preventions and treatments.

The two most commonly used designs for observational studies are cohort studies and case-control studies. In a cohort study the comparison groups are identified according to an exposure, or potentially treatment, of interest. In case-control studies, the study groups are chosen on the basis of their disease or outcome of interest. Hoppe et al., (2009) presents an excellent synopsis of common design features of observational studies and the indications for their use in scientific research.

Cohort Studies

A cohort study is a form of longitudinal study that is commonly used to study exposure-disease associations. Outcomes are tracked forward in time. A cohort is a group of people who share a common characteristic within a defined period and hence are useful to determine the incidence and natural history of a disorder or exposure (Hoppe et al., 2009). Examples could be a group of people born in the same year (birth cohort), or who may have been exposed to a drug or a vaccine. The comparison group may be the general population from which the cohort is drawn, or it may be another cohort of persons who have had little or no exposure to the substance under investigation, but otherwise similar. Groups like doctors, civil servants and surgery patients are often chosen as the source of the groups because they are relatively easy to define and/or monitor. Subgroups within a cohort may be compared with each other. As individuals are not randomly assigned in cohort studies they are more prone to biases than RCTs (Tay and Tinmouth, 2007).

Cohort studies can come in two types, prospective and retrospective. In a prospective cohort study, the cohort is identified before the appearance of the exposure/disease under investigation. This is an important point! Prospective studies begin with a sample whose members are free of the disease or disorder under study for example, free of lung cancer. The cohort cannot therefore be defined as a group of people who already have the disease. All the individuals in the sample are followed over time. The individual differences that exist in a population will therefore be represented in any sample of that population, for example amongst the sample, some people smoke whilst others do not. The incidence rates for the disease under study are determined in key subgroups. The study groups, so defined, are observed over a period of time to determine the frequency of new incidence of the studied disease among them. For example, in a sample that was free of lung cancer at the outset of the study, after 25 years, you may anticipate that non-smokers have the lowest incidence rate of the disease after 25 years, followed by moderate smokers, and that lung cancer will be most common in people classified as heavy smokers. The prospective study is important for research on the etiology of diseases and disorders in humans because, for ethical reasons, people cannot be deliberately exposed to suspected risk factors in controlled experiments (Hoppe et al., 2009).

Prospective cohort studies investigating the relationship between exposure and disease strongly aide in studying causal associations, although distinguishing true causality may well require corroboration from further experimental trials. The advantage of prospective cohort study data is the longitudinal observation of the individual through time, and the collection of data at regular intervals, so recall error is reduced (this is a recognized potential source of bias in retrospective studies) (Hoppe et al., 2009). However, cohort studies are expensive to conduct, are sensitive to attrition and take a long follow-up time to generate useful data. Nevertheless, the results that are obtained from long-term cohort studies are of substantially superior quality to retrospective/cross-sectional studies, and prospective cohort studies are generally considered the gold standard in observational epidemiology (Hoppe et al., 2009).

A retrospective cohort study is quite different from a prospective cohort study in the manner in which it is conducted (Tay and Tinmouth, 2007). A retrospective cohort study, also called a historic cohort study, is a health research study in which the records of groups of individuals who are alike in many ways but differ by a certain characteristic - again, as an example, patients who smoke and those who do not smoke - are compared for a particular outcome, such as the development of lung cancer. Effectively, in a retrospective cohort study, all the events - exposure, latent period, and subsequent development of disease have already occurred in the past. The data is simply collected now, and the risk of developing a disease established if exposed to a particular risk factor. There is no follow up of patients, as is the case with a prospective study.

A particular variation of a cohort study is a before and after study (or time series) (Tay and Tinmouth, 2007). In a before and after study, a measurement of interest is taken before and after a population sample is exposed to an intervention.

Clearly, a retrospective study has the benefits of being cheaper and less time consuming with resources mainly directed at data collection. Statistically, the two forms of the cohort study differ also. Whereas prospective cohorts should be summarized with the relative risk, retrospective cohorts should be summarized with the odds ratio.

Case-Control Studies

Case-control studies use patients who already have a disease or other condition ('cases') and look back to see if there are characteristics of these patients that differ from those who don't have the disease ('controls') (Hoppe et al., 2009). The main purpose of matching is to control for confounding (see below). Case-control studies are a relatively inexpensive and frequently used type of epidemiological study that can be carried out by small teams or individual researchers in single facilities without the rigid structure of an experimental study. Case control studies are often used as a rapid means of studying risk factors (Hoppe et al., 2009). Despite the retrospective, non-randomized nature of a case-control study limiting the conclusions that can be drawn, this study design has led to a number of important discoveries and advances including the demonstration of the link between tobacco smoking and lung cancer by Sir Richard Doll. It is worth noting also that the term retrospective study is sometimes used as another term for a case-control study. One recognized disadvantage of case-control studies is that they do not give any indication of the absolute risk of the factor in question. For instance, a case-control study may tell you that a certain behavior may be associated with a tenfold increased risk of death as compared with the control group. Although this sounds distressing, it would not tell you that the actual risk of death would change from one in ten million to one in one million, which is less alarming. Further data from outside the case-control study would have to be considered to generate such a finding (Hoppe et al., 2009). In the case-control study, the association is determined for each individual case-control pair, then aggregated. This provides a more specific analysis of the possible associations, and potentially determines more accurately which possible causes are directly related to the effect being studied, and which are merely related by a common cause.

Case series/Case reports

A case series (also known as a clinical series) is a health research study that tracks patients with a known exposure given similar treatment or examines their records for exposure and outcome. A case series can be retrospective or prospective and usually involves a smaller number of patients than more powerful case-control studies or RCTs. Case series may be described as consecutive or non-consecutive, depending on whether all cases presenting to the reporting authors over a period of time were included, or only a selection. Case series may be confounded by selection bias, which limits statements on the causality of correlations observed.

In health care, a case report is a detailed report of the symptoms, signs, diagnosis, treatment, and follow-up of an individual patient. Case reports may contain a demographic profile of the patient, but usually describe an unusual or novel occurrence. A case report is a type of anecdotal evidence. As such, it is less scientifically rigorous than controlled clinical data involving a larger sample size. Proponents argue that case reports have value within scientific method in that they permit discovery of new diseases and unexpected effects (adverse or beneficial) as well as the study of mechanisms, and they play an important role in medical education. Case reports and series have a high sensitivity for detecting novelty and therefore remain one of the cornerstones of medical progress; they provide many new ideas in health care (Vandenbroucke, 2001).

Cross-sectional Studies

Cross-sectional studies can be thought of as providing a "snapshot" of the frequency and characteristics of a disease in a population at a particular point in time. They are also referred to as prevalence studies. Cross-sectional research takes a 'slice' of its target group and bases its overall finding on the views or behaviors of those targeted, assuming them to be typical of the whole group. These studies often use survey research methods, which aim to find the same kind of relationships that might be shown over time in a cohort study, but at far less cost. Cross-sectional analysis studies the relationship between different variables at a point in time. This type of data can be used to assess the prevalence of acute or chronic conditions in a population. It is worth noting, since exposure and disease status are measured at the same point in time, it may not always be possible to distinguish whether the exposure preceded or followed the disease.

In a cross-sectional survey, a specific group is looked at to see if a substance or activity, say smoking, is related to the health effect being investigated - for example, lung cancer. If a significantly greater number of smokers already have lung cancer than those who don't smoke, this would support the hypothesis that lung cancer is correlated with smoking. In epidemiology, cross-sectional studies often involve secondary analysis of data collected for another purpose. Data from these studies are useful in providing information about the health status and needs of a population. Major sources of such data are often large institutions like a national Census Bureau. Such studies can cover study groups as large as the entire population of the United States or groups from different countries around the world, but others are small and geographically limited.

Cross-sectional studies that reveal clues to exposure/disease relationships are often used as precursors to subsequent studies using more robust experimental design to study a relationship, such as case-control, cohort studies or sometimes even RCTs.

Systematic reviews and meta-analyses of observational studies are as common as reviews of RCTs. As mentioned, observational studies are necessary, particularly when investigating causation of harms, for example, as no one will gain ethical approval to conduct an RCT to determine if smoking causes lung cancer or a particular drug is responsible for producing physical defects at birth, this is the realm of the observational study (or laboratory research) (Egger et al., 1998). Furthermore, most RCTs are conducted to establish efficacy and safety of a single agent in a specific clinical situation. These studies are limited in their sample size, therefore the only way to detect, for example less common but often significant adverse effects of drugs is in a case-control study. This allows follow-up of adverse events often not possible with an RCT. The subject pool available to an RCT is often limited also - often the elderly and women are not included. Where intervention is unsavory or invasive, patient preference may preclude RCT (Egger et al., 1998).

When the size of underlying risk is small, or there is disagreement in results from individual studies it is tempting to do meta-analysis; particularly so with observational studies which investigate either causation or effectiveness. The overall estimate of association or treatment effect can often be misleading or incorrect in meta-analyses of observational studies due to factors such as heterogeneity, confounding and bias.

Confounding and bias

Meta analysis of observational studies is fundamentally different from that of RCTs. In meta-analyses of RCTs, variability of the results between studies is attributable to random variation as each trial is assumed to provide an unbiased estimate of the effect of the experimental treatment. In general in scientific research, we draw conclusions based on statistical methods constrained by the assumptions of the statistical test we decide to use. In principle, what distinguishes observational studies from randomized controlled trials is the validity of some of these assumptions (Hoppe et al., 2009). In observational studies estimates of association may deviate beyond the effects of chance due to confounding factors or biases even where statistical significance is apparent. Confounding occurs when the effects of two processes are not separated so that the apparent effect is not the true effect and any interpretation of the results is potentially faulty (Clancy, 2002). In other words, subjects exposed to the factor under investigation may in fact differ in a number of aspects that are relevant to the risk of developing the disease in question. Even if adjustments for confounding factors have been made in the analysis, residual confounding remains a potentially serious problem in observational research and arises when a confounding factor cannot be, or has not been, measured with sufficient precision – quite common in epidemiological studies (see below). Egger et al (1998) give some examples of studies where statistical associations may appear causal, but are in fact, implausible. Confounding is the most important threat to validity of results from cohort studies whereas many more difficulties, including selection biases arise in case-control studies. Selection bias occurs when study subjects differ systematically from the population with the

same condition (Clancy, 2002). For example, subjects who present to hospital may not be representative of all patients with the condition and this affects the ability to generalize the result outside the study sample, reducing external validity (Clancy, 2002). Furthermore, for example, in a prospective cohort study where subjects are being followed up in the future, the researchers who assess any outcome measures should, if possible, be blinded to exposure status to reduce biases.

Implausible results are no protection from misleading conclusions. Researchers look for plausible explanations for the findings. Epidemiological studies produce a large number of seemingly plausible associations. Some will be wrong but reported in the media nonetheless. Collecting more detailed and accurate data on a smaller number of participants is, in many cases, more likely to produce a correct result than a poorly conducted study on an enormous number of participants (Egger et al., 1998). The statistical power of the study, although important, is not the best indicator of which study is likely to have the least biased or confounded results. It is worth noting however, in a meta-analysis a higher power study, will still contribute more 'weight' to the overall analysis.

Meta-analysis of observational studies is contentious. Some authors suggest rather than avoid it completely, statistical combination of observational studies should not be a prominent component of systematic reviews of observational studies (Egger et al., 1998) These same authors again highlight potential and confounding sources of heterogeneity, using examples, when dealing with the methods of data collection inherent in observational study design. Many examples exist where extraction of data from published articles which present data in different, complex formats is prone to error. Many of these errors can be reduced or avoided completely in analyses where there is access to, or investigators make their primary data available. Overall measures of effect will often be biased and more relevant information will be provided by thorough consideration of possible sources of heterogeneity – access to primary data allows this. Regardless of consideration of possible explanation of results and reasons they may be incorrect which are inherent to observational methodologies, in a comparison of the same treatments for particular conditions, little difference was found between estimates of treatment effects in observational studies versus RCTs (Concato et al., 2000; Benson and Hartz, 2000).

Group Work 2

Working in pairs:

- critically appraise the cohort study in your workbook

- critically appraise the case control study in your workbook

Reporting Back

Session 4:
Study Data and Data Extraction

Once the studies have been included in a review, the relevant results have to be abstracted from the reports. This requires definition of the comparison and outcome to be assessed, and is often a quite complex process. It is also open to subjective influences, and so the extraction of the key data from each study should involve two or more investigators working independently using a data extraction instrument. The data extracted includes:

- Source - citation and contact details
- Eligibility - confirm eligibility for review
- Methods - study design, concerns about bias
- Participants - total number, setting, diagnostic criteria
- Interventions - total number of intervention groups
- Outcomes - outcomes and time points
- Results - for each outcome of interest

Difficulties related to the extraction of data include different populations outcome measures, interventions administered differently and the reliability of data extraction (i.e. between reviewers). Errors in data extraction can be minimized by using a data extraction form; pilot testing the extraction form prior to commencement of the review; training and assessing data extractors; having two people extract data from each study; and blinding extraction before conferring

Data most frequently extracted

Dichotomous data

Risk

The risk of an event is the probability that an event will occur within a stated time period (P). This is sometimes referred to as the absolute risk. For example:

– The risk of developing anaemia during pregnancy for a particular group of pregnant women would be the number of women who develop anaemia during pregnancy divided by the total number of pregnant women in the group.

– The risk of a further stroke occurring in the year following an initial stroke would be the number who have another stroke within a year divided by the total number of stroke patients being followed up.

Relative risk

When we use the term ‘relative risk’, we are referring to the ratio of risk in the intervention group to the risk in the control group. A risk ratio of one indicates no difference between comparison groups. For undesirable outcomes a risk ratio of less than one indicates that the intervention was effective in reducing the risk of that outcome.

Risk difference

Risk differences are the absolute difference in the event rate between two comparison groups. A risk difference of zero indicates no difference between comparison groups. For undesirable outcomes a risk difference that is less than zero indicates that the intervention was effective in reducing the risk of that outcome.

Number Needed to Treat

This is the additional number of people you would need to give a new treatment to in order to cure one extra person compared to the old treatment. Alternatively for a harmful exposure, the number needed to treat becomes the number needed to harm and it is the additional number of individuals who need to be exposed to the risk in order to have one extra person develop the disease, compared to the unexposed group.

Odds ratio

An odds ratio is the ratio of the odds of an event in the intervention group to the odds of an event in the control group. An odds ratio of one indicates no difference between comparison groups. For undesirable outcomes an odds ratio of less than one indicates that the intervention was effective in reducing the risk of that outcome.

Relative Risk and Odds Ratio

The odds ratio can be interpreted as a relative risk when an event is rare and the two are often quoted interchangeably. For case-control studies it is not possible to calculate the RR and thus the odds ratio is used. For cross-sectional and cohort studies both can be derived and if it is not clear which is the causal variable and which is the outcome should use the odds ratio as it is symmetrical, in that it gives the same answer if the causal and outcome variables are swapped. Odds ratios have mathematical properties that make them more often quoted for formal statistical analyses

Continuous data

Mean difference

Mean difference, as the name implies, is the difference between the means (i.e. the average values) of two groups.

Weighted mean difference

Weighted mean difference refers to the situation where studies have measured an outcome on the same scale and the weight given to the mean difference in each study is usually equal to the inverse of the variance.

Standardized differences in mean

Standardized differences in mean refers to the situation where studies have measured an outcome using different scales or units of measurement (for example inches and centimeters) and the mean difference may be divided by an estimate of the within group standard deviation to produce a standardized value without any units.

Group Work 3

Working in pairs, extract the data from the two papers referred to in your Workbook.

Reporting Back

Session 5:
Protocol Development

Using CReMS, develop a draft Protocol of a question of effectiveness of an intervention or therapy

Session 6:
Data Synthesis/Meta-analysis

When data has been extracted a general analysis is conducted and, where possible, results are synthesised. General analysis focuses on:

- What interventions/activities have been evaluated
- The effectiveness/appropriateness/feasibility of the intervention/activity
- What interventions may be effective
- Contradictory findings and conflicts
- Limitations of study methods
- Issues related to study quality
- The use of inappropriate definitions
- Specific populations excluded from studies
- Future research needs

Meta-analysis

A meta-analysis is performed to calculate a more precise estimate of the outcome of interest. Meta-analysis is a quantitative method of combining the results of independent studies. This is achieved by pooling the results of various studies, in effect increasing the total sample size of the analysis, and improving the precision of the outcome estimate. Meta-analysis of RCTs aims to derive an overall estimate of effect. The principle purpose of the meta-analysis of observational studies is to investigate the reasons for differences in risk estimates between studies and to discover patterns of risk among study results (Kheifets et al., 1997). During the process of combining studies, you will undoubtedly encounter the problem related to the appropriateness of pooling or combining the results of different studies. This problem arises due to heterogeneity between studies. Heterogeneity may be problematic due to methodological issues or despite similar methodology, different outcomes being measured! Meta-analyses of observational studies, more so than for clinical trials often have the added challenge of incorporating various designs and levels of quality. This issue is a problem when there is more variation between studies than would be expected based on sampling alone. Heterogeneity between studies is often more common and extreme in observational studies than clinical studies (Sutton and Abrams, 2000).

Tests of heterogeneity are based on the assumption that all studies in the systematic review are essentially the same, therefore, these tests effectively measure the extent to which the observed study outcomes deviate from the calculated summary outcome. Visual inspection of the meta-analysis Forest plot can be the first stage of assessing heterogeneity. Longer confidence intervals (CI) indicate less certain estimates. Statistically, a χ^2 Test for Homogeneity can be used. This test calculates a P value using an individual studies weight, effect size and overall effect size. The Q-test is also efficient for determining heterogeneity (DerSimonin and Laird, 1986). A funnel plot can be used to visually detect sources of heterogeneity such as publication and selection bias. If the plot appears asymmetrical, it may suggest a heterogeneous sample.

For a meta-analysis to be feasible all outcomes of interest must be similar and measured in the same way, that is, they must be homogeneous. The relative risk (risk ratio), risk difference, and odds ratio are common numerical measures for dichotomous (binary) or "yes-no" outcomes. The hazard ratio is similarly used to present dichotomous survival data. Continuous outcomes, like blood pressure for example will most commonly be presented with the mean difference (effect size) (Deeks et al., 2001). The effect size is commonly standardized by the pooled estimate of the within-group variance. Where continuous outcomes are skewed, you may encounter transformed data (e.g. logarithmic) or use of the median rather than the mean.

The focus of the remainder of this discussion will be on the more commonly used dichotomous outcomes. A single summary measure of study outcomes is a weighted average of all study outcomes. The weight indicates the 'influence' of the study and in a meta-analysis a study with a large number of subjects is thus more influential than a study with a small number of subjects. The estimate of the precision of this summary measure is the CI. As a meta-analysis aims to improve the precision of the outcome measure, the CI around your summary measure should ideally shrink when compared to the individual studies included in the meta-analysis - the smaller the better!

There are various statistical methods for the combination of study outcomes, including fixed effects analysis, and random effects analysis. These can be distinguished by their methods used for estimating the CI, or precision of the overall summary outcome.

When using fixed effects analysis for dichotomous outcomes there are various methods available including Woolf's Method, Mantel-Haenszel Method and Peto's Method, which will in most instances yield similar results (Deeks et al., 2001). Each of these methods are referred to as fixed effects, as all studies are measuring the same parameter and any difference in outcome observed across studies is only due to chance – that is, it is assumed there is no variation inherent in the source population. In essence, each of the studies in the meta-analysis these methods take into account, considers within study variation rather than the between study variation, and hence these methods are not used if there is significant heterogeneity apparent. The CI of the summary measure therefore, will reflect variability between patients within the sample.

Where there is evidence of statistical heterogeneity between studies the fixed effects model will not fit the observed data well and therefore it is more appropriate to use the random effects model (DerSimonian and Laird, 1986). Random effects are often applied to compensate for the heterogeneity apparent in observational studies. In this model, variability in data arises from variability between the patients (or within the sample) and also from the differences between the studies also. It is assumed that all studies are different, and that the outcome of a study will fluctuate around its own true value. It is assumed that each of these true values is drawn randomly from the same normal distribution within the population. The resultant summary outcome is the estimate of the mean of the normal probability distribution of sample outcomes from which our sample of outcomes was randomly drawn. The summary value from a random effects model will often have a wider CI than seen for the fixed effects model. Where there is no heterogeneity present, the results of fixed and random effects models will be similar. Sometimes, when heterogeneity is indicated, it may be an indication that it is not appropriate to proceed with meta-analysis and the results of included studies should be summarized solely as a narrative review.

Conducting a meta-analysis

Meta-analysis is useful if studies report different treatment effects or if studies are too small (insufficient power) to detect meaningful effect. It can be used if studies:

- have the same population;
- use the same intervention administered in the same way;
- measure the same outcomes; and
- studies are homogeneous (i.e. sufficiently similar to estimate an average effect).

Calculating an Average

Most commonly this involves calculating:

The Odds Ratio

– for dichotomous data eg. the outcome is either present or not

– 51/49 = 1.04

– (no difference between groups = 1)

OR

The Weighted mean difference

– Continuous data, such as weight

– (no difference between groups = 0)

AND

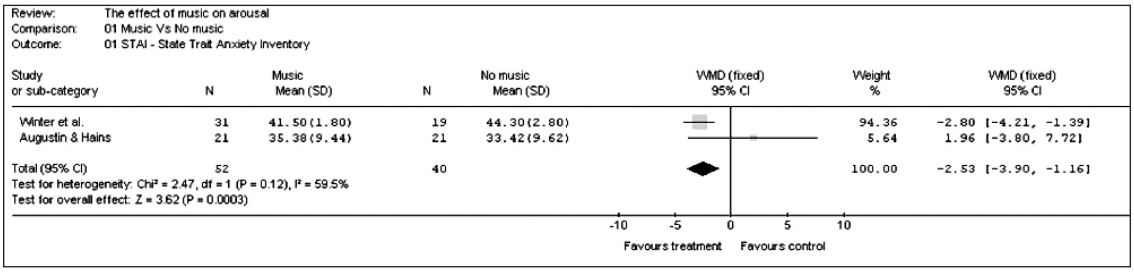
The Confidence Interval

– The range in which the real result lies, with the given degree of certainty

Confidence intervals are an indication of how precise the findings are. Sample size greatly impacts the CI-i.e., the larger the sample size the smaller the CI, the greater the power and confidence of the estimate. When calculated for Odds Ratio, the CI provides the upper and lower limit of the odds that a treatment may or may not work. If the odds ratio is 1, odds are even and therefore, not significantly different (recall the odds of having a boy).

The Meta-view Graph

The results of a meta- analysis are conventionally presented in a forest plot – often referred to as a meta-view graph:



Odds ratio (for categorical data) and standard or weighted mean differences (for continuous data) and their 95% confidence intervals are calculated in the meta-view graph.

Assessing Heterogeneity

Heterogeneity is assessed using the standard Chi-square. When used in relation to meta-analysis, the term 'heterogeneity' refers to the amount of variation in the characteristics of included studies. For example, if three studies are to be included in a meta-analysis, do each of the included studies have similar sample demographics, and assess the same intervention? (Note that the method by which the intervention is measured does not need to be identical.) While some variation between studies will always occur due to chance alone, heterogeneity is said to occur if there are significant differences between studies, and under these circumstances meta-analysis is not valid and should not be undertaken. But how does one tell whether or not differences are significant? Visual inspection of the meta-analysis is the first stage of assessing heterogeneity. JBI-MAStARI plots the results of individual studies and thus indicates the magnitude of any effect between the treatment and control groups. Do the individual studies show a similar direction and magnitude of effect – i.e. are the rectangular symbols at similar positions on the X-axis? A formal statistical test of the similarity of studies is provided by the test of homogeneity. This test calculates a probability (P value) from a Chi-square statistic calculated using estimates of the individual study's weight, effect size and the overall effect size. However, note that this test suffers from a lack of power – and will often fail to detect a significant difference when a difference actually exists – especially when there are relatively few studies included in the meta-analysis. Because of this low power, some review authors use a significance level of $P < 0.01$, rather than the conventional 0.05 value, in order to protect against the possibility of falsely stating that there is no heterogeneity present.

In meta-analysis, the results of similar, individual studies are combined to determine the overall effect of a particular form of health care intervention (the treatment) compared to another standard or control intervention for a specified patient population and outcome. In meta-analysis, the effect size and weight of each study are calculated. The effect size indicates the direction and magnitude of the results of a particular study (i.e. do the results favor the treatment or control, and if so, by how much), while the weight is indicative of how much information a study provides to the overall analysis when all studies are combined together.

Subgroup analyses

If there were some types of participant, intervention or outcome you thought were likely to be quite different to the others, you might plan a subgroup analysis. The number of planned subgroup analyses should be kept to a minimum to avoid spurious findings. Where there is significant heterogeneity in the results and no subgroup analysis has been stated a priori, subgroup analysis may be used, but the results interpreted with caution.

Two types of subgroup analyses:

- Between trial (trials classified into subgroups)
- Within trial (each trial contributes to all subgroups)

Problems with subgroup analyses: potentially misleading, more number of analyses leads to spurious findings, and most importantly, they are observational not randomized.

When to do subgroup analyses – pre-specified in the protocol, if there are enough studies, and if there are good clinical reasons.

Sensitivity analysis

The process of undertaking a systematic review and meta-analysis involves many decisions. Ideally, most of these are made while designing the protocol. The role of a sensitivity analysis is to determine whether the assumptions or decisions we have made do in fact have a major effect on the results of the review. A sensitivity analysis addresses the question 'Are the findings robust to the method used to obtain them?' Sensitivity analyses involve comparing the results of two or more meta-analyses calculated using different assumptions. If a study is of doubtful eligibility for the systematic review, then comparing meta-analyses excluding and including that study might be undertaken as a sensitivity analysis (Higgins and Green, 2006).

Results may be calculated using all studies and then excluding poorer quality studies. Both fixed and random effects meta-analyses might be undertaken to assess the robustness of the results to the method used. If a study appears to be an outlier (has results very different from the rest of the studies), then its influence on the meta-analysis might be assessed by excluding it.

Fixed Effects and Random Effects

Meta-analysis can be based on either of two assumptions. In a fixed effects model, it is assumed that any differences between treatment and control are the same (or fixed) in each study. Thus any observed differences among the studies' results are due solely to chance and there is no heterogeneity between the studies. However, when there is heterogeneity apparent (for example, the test of homogeneity is significant), the validity of the assumption of a fixed effect is questionable, and thus another approach is to consider that the treatment effects for the individual studies are not identical and, in fact, follow a distribution related to an overall average treatment effect. That is, the effect size is random, and is assumed to follow a Normal distribution and consequently has a mean and variance.

Essentially, the test for homogeneity is asking the statistical question "is the variance around the estimate of the effect size zero or non zero?" If the variance around the estimate of the effect size is zero, then there is no heterogeneity present, and the results of the fixed and random effects models will be similar.

There is no consensus about whether fixed or random effects models should be used in meta-analysis. In many cases when heterogeneity is absent, the two methods will give similar overall results. When heterogeneity is present, the random effects estimate provides a more conservative estimate of the overall effect size, and is less likely to detect significant differences. For this reason, random effects models are sometimes employed when heterogeneity is not severe; however, the random effects model does not actually analyze the heterogeneity away and should not be considered as a substitute for a thorough investigation into the reasons for the heterogeneity. Additionally, random effects models give relatively more weight to the results of smaller studies – this may not be desirable because smaller studies are typically more prone to bias and of lower quality than larger studies.

Meta-analytical techniques available in MAStARI

There are a number of meta-analytical techniques available. The selection of a particular technique is governed by three things: the study type, nature of the data extracted and assumptions underlying the meta-analysis. Here, we introduce the tests that are available in JBI-MAStARI and discuss when it is appropriate to use each of the tests.

When the outcomes of included studies are dichotomous, JBI-MAStARI can be used to generate two overall effect sizes: odds ratios (OR) and relative risks (also known as risk ratios, RR). The choice of whether OR or RR are calculated is important and should be carefully considered with due reference to three criteria.

Dichotomous data – methods of meta-analysis

There are several different methods available to pool results of dichotomous data, depending on the data type and whether a random or fixed effects model is required: Mantel-Haenszel, Peto's; and DerSimonian and Laird.

Mantel-Haenszel

Mantel-Haenszel is the default meta-analytical method for dichotomous data using a fixed effects model. Both OR and RR can be pooled using Mantel-Haenszel methods; the calculation of study weights and effect sizes, and overall effect sizes differs slightly between OR and RR. The Mantel- Haenszel method is generally preferred in meta-analysis to another method (inverse variance) because it has been shown to be more robust when data are sparse (in terms of event rates being low and/or the number of trials being small).

Peto's odds ratio

Peto's odds ratio is an alternative method for meta-analysis of OR using a fixed effects method. It employs an approximation that can be inaccurate if treatment affects are very large, and when the sample sizes between treatment and control groups are unbalanced. However, the method is appropriate when event rates are very low and effect sizes are not overly large.

DerSimonian and Laird

DerSimonian and Laird methods are used in the meta-analysis of OR and RR using a random effects model. Although the study effect sizes and heterogeneity statistics are calculated as for the fixed effects model, the study weights and overall effect sizes in DerSimonian and Laird random effects models are calculated slightly differently to fixed models.

Meta-analysis of continuous data

When the outcomes of included studies are continuous, JBI-MAStARI can be used to generate two overall effect size calculations using weighted mean differences (WMD) or standardized mean differences (SMD).

Weighted mean difference

The WMD measures the difference in means of each study when all outcome measurements are made using the same scale. It then calculates an overall difference in mean for all studies (this is equivalent to the effect size) based on a weighted average of all studies, which is, in turn related to the SD. JBI-MAStARI uses the inverse variance method of calculating WMD for fixed effects models and the DerSimonian and Laird method for random effects models.

Alternatively, different studies may measure the same outcome using different scales. For example, pain can be measured on a range of different scales including non-verbal scales (e.g.. visual analogue scale) and verbal scales (e.g. 5 point categorical scale). These studies can be combined in a meta-analysis that incorporates SMD. If the measurement scales operate in the same direction (e.g. An increase in pain is measured as an increase in on both scales), then using SMD is straightforward.

However, if two measurement scales operate in a different direction – for example a score of 10 is the worst pain imaginable on one scale but a score of 1 is the worst pain imaginable on another scale – then data from one scale need to be reversed. This is relatively simply achieved by multiplying the mean data from one scale (for both treatment and control groups) by -1. Standard deviations do not need to be modified.

Standard mean difference

JBI-MAStARI provides two options for calculation of the SMD using fixed effects: Cohen's SMD and Hedges' SMD. Both options produce a similar result, although Hedges' SMD is generally preferred as it includes an adjustment to correct for small sample size bias. [6] As per WMD, the DerSimonian and Laird method is used for random effects models calculations for SMD.

Summarizing the Findings in Narrative Form

Where meta-analysis is not possible, a narrative summary should draw upon the data extraction, with an emphasis on a textual summary of study characteristics as well as data relevant to the specified outcomes.

Reporting findings

There is no standardized international approach to structuring how the findings of reviews will be reported. Guidance has been developed for authors reporting on systematic reviews and meta-analyses. The Preferred Reporting Items for Systematic reviews and Meta-Analyses or PRISMA statement provides a 27-item checklist and a four phase flow diagram to help improve reporting of systematic reviews and meta-analyses with particular focus on randomized trials and intervention studies and is also relevant to reviews of harms (Liberati et al., 2009, Moher et al., 2009).

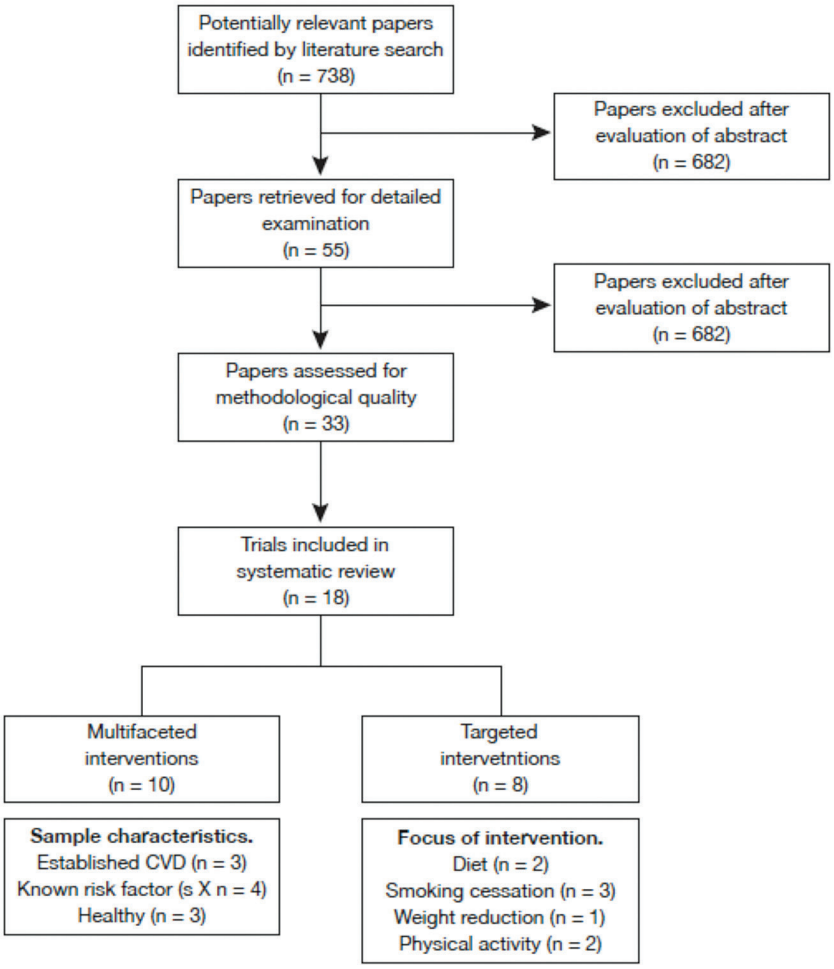
The audience for the review should be considered when structuring and writing the findings up. Meta-view graphs represent a specific item of analysis that can be incorporated in to the results section of a review. However, the results are more than the meta-view graphs, and whether it is structured based on the intervention of interest, or some other structure, the content of this section needs to present the results with clarity using the available tools (meta-view graphs, tables, figures) supported by textual descriptions.

The results section should be framed in such a way that as a minimum, the following fields are described in the protocol as either planned for reporting, or given consideration by the reviewers in preparing their systematic review report:

Studies:

- Numbers of studies identified,
- Numbers of retrieved studies,
- Numbers of studies matching preferred study design (i.e. RCTs),
- Numbers and designs of other types of studies,
- Numbers of appraised studies,
- Numbers of excluded studies and overview of reasons for exclusion,
- Numbers of included studies.

These results are commonly written in narrative style, and illustrated with a flow diagram as per the example below:



Drawing Conclusions in a Review

The conclusions provide a detailed discussion of issues arising from the findings of the review and demonstrate the significance of the review findings to practice and research. Areas that may be addressed include:

- A summary of the major findings of the review
- Issues related to the quality of the research within the area of interest
- Other issues of relevance
- Implications for practice and research, including recommendations for the future
- Potential limitations of the systematic review

The discussion does not bring in new literature or findings that have not been reported in the results section. The discussion does seek to establish a line of argument based on the findings regarding the effectiveness of an intervention, or its impact on the outcomes identified in the protocol.

Session 7:
Appraisal, extraction and synthesis using JBI-MAStARI

Information pertinent to this session of the program can be located and referred to in the SUMARI user manual.

Group Work 4

MAStARI Trial and Meta Analysis

Session 8:
Protocol Development

Using CReMS, develop a draft Protocol of a question of effectiveness of an intervention or therapy

Session 9:
Assessment

Assessment

Take 30 minutes to complete the multiple choice questions. Your facilitator will go through the answers with the group after this. Raise any points you have for discussion or clarity.

Session 10:
Protocol Presentation

All participants are required to give a 5 minute presentation covering the initial aspects of their protocols. This includes their title, question, brief background and PICO criteria using 3-4 power point slides.

Take this opportunity to receive constructive feedback on your protocol from your facilitator and other CSR participants.

References

Alderson P, Green S. 2002, Cochrane Collaboration open learning material for reviewers, viewed 7 June 2009 http://www.cochrane net.org/openlearning/PDF/Openlearning-full.pdf

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials) 2001, The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration, Annals of internal medicine, vol 134, no 8, pp 663-694.

Benson, K. and Hartz, A.J. (2000) A comparison of observational studies and randomized controlled trials. N Engl J Med 342:1878-1886.

Black, N. (1996) Why we need observational studies to evaluate effectiveness of healthcare. BMJ 312:1215-1218.

Clancy, M.J. (2002) Overview of research designs. Emerg Med J 19:546-549.

Concato, J., Shah, N. and Horwitz, R.I. (2000) Randomized, controlled trials, observational studies, and the hierarchy or research designs. N Engl J Med 342:1887-1892.

Day S, Altman DG 2000, Statistics Notes: blinding in clinical trials and other studies, British Medical Journal, vol 321, pp 504.

Deeks, J.J., Altman, D.G. and Bradburn, M.J. (2001) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger, M., Smith, G.D., Altman, D. (eds). Systematic review in health care: meta-analysis in context. 2nd ed. London: BMJ Publishing group. P. 157-175.

DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. Control Clin Trials 7:177-188.

Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, Bhandari M, Guyatt GH 2001, Physician Interpretations and Textbook Definitions of Blinding Terminology in Randomized Controlled Trials, Journal of the American Medical Association, vol 285, no 15, pp 2000 – 2003.

Egger, M., Schneider, M. and Smith, G.D. (1998) Meta-analysis spurious precision? Meta-analysis of observational studies BMJ 316:140-144.

Harris, M. and G. Taylor (2003) Medical Statistics made easy. London and New York, Taylor and Francis Group.

Higgins JPT, Green S (eds). Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 (updated September 2006). In The Cochrane Library, Issue 4, 2006, John Wiley & Sons, Ltd. Chichester, UK.

Hollis, S (1999) What is meant by intention to treat analysis? Survey of published randomised controlled trials BMJ 319:670.

Hoppe, D.J., Schemitsch, E.H., Morshed, S., Tornetta, P. and Bhandari, M. (2009) Hierarchy of Evidence: Where observational studies fit in and why we need them. J Bone Joint Surg Am 91 Suppl 3:2-9.

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds JM, Gavaghan DJ, McQuay DM. 1996, Assessing the quality of reports on randomized clinical trials: Is blinding necessary? Controlled Clinical Trials, vol. 17, pp.1-12.

Kao LS, Tyson JE, Blakely ML, Lally KP 2008, Clinical research methodology I: Introduction to randomized trials, Journal of the American College of Surgeons, vol 2, pp 361 -369.

Kendall JM 2003, Emergency medicine journal, vol 20, pp 164-168.

Kheifets, L.I., Afifi, A.A., Buffler, P.A., Zhang, Z. and Matkin., C.C. (1997) Occupational and electric and magnetic field exposure in leukaemia: A meta-analysis. J Occ Environ Med 39:1074-1091.

Laupacis, A., Wells G, et al. (1994). “Users guides to the medical literature. V. How to use an article about prognosis. .” JAMA 272: 234–237.

Levine, M., Walter SS, et al. (1994). “Users’ guides to the medical literature: IV. How to use an article about harm. Evidence-Based Medicine Working Group.” JAMA 271: 1615-9.

Lewis JA, Machin D. 1993, Intention to treat-who should use ITT? British Journal of Cancer, vol. 68, no. 647-650.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzche, P.C., Ioannidis, J.P.A., Clarke, M., Devereaux, P.J., Kleijnen, J. and Moher, D. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of Clinical Epidemiology, 62, e1-e34.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med, 151, 264-9, W64.

Moher D, Schulz KF, Altman D, for the CONSORT Group. 2001, The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. Journal of American Medical Association, vol. 285, pp.1987-1991.

Schulz K, Chalmers I, Hayes RJ, Altman DG. 1995, Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled clinical trials. Journal of American Medical Association, vol. 273, pp. 408-12.

Steuber, T., P. Helo, et al. (2007). “Circulating biomarkers for prostate cancer.” World Journal of Urology 25: 111-119.

Tay, J. and Tinmouth, A. (2007) Observational studies: what is a cohort study? Transfusion 47:1115-1117.

Togerson DJ, Roberts C 1999, Understanding controlled trials: randomisation methods: concealment, British Medical Journal, vol 319, pp 37-376.

Vandenbroucke, J.P. 2001. In Defense of Case Reports and Case Series. Annals of Internal Medicine, 134, 330-334.

Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gotzsche, P.C., Mulrow, C.D., Pocock, S.J., Poole, C., Schlesselman, J.J. and Egger, M. (2007) Strengthening the reporting of Observational studies in Epidemiology (STROBE): Explanation and elaboration. PLoS Med 4:e297