

ANN M. MØLLER
Herlev University Hospital
Herlev, Denmark

Saturday June 5, 2004

13:00-13:45

Room 3C

Over 2 million scientific articles dealing with topics related to medicine are published every year. It is an overwhelming task to keep updated in medical literature, even in a subspecialty field.

Archie Cochrane, who was an epidemiologist and gynaecologist, wrote in 1979: *“It is surely a great criticism of our profession that we have not organised a critical summary, by speciality or subspecialty, adapted periodically, of all relevant randomised controlled trials”*.

What he meant was that it would be extremely helpful to the clinician to have a systematic, updated review of relevant clinical interventions.

Systematic reviews differ from narrative reviews in that they address a clearly defined question, and use systematic and explicit methods to identify, select and critically appraise all relevant research, to collect and analyse data from the primary studies included in the review, and to present and draw conclusions.

The preparation of a systematic review follows the rules of the scientific process. First of all, the clinically relevant question is defined, and a protocol is prepared. The protocol states in advance what the authors intend to do, what the criteria for trial inclusion are, and what methods will be used in the review. Because of this, the term “secondary research” can be applied to the process of preparing a systematic review.

THE CLINICALLY RELEVANT QUESTION

The choice of the clinically relevant questions generally needs some work. The question needs to be clear and well focused and the following has to be decided:

1. *Who is the question about? A well-defined group of patients has to be described. The group can be wide (thus including a wide range of patients, such as all patients presenting for arthroscopy of the knee joint) or it can be narrow (such as pregnant woman with bleeding disorders presenting for acute Caesarean section). Whatever one chooses depends on the clinical scenario one wants to describe.*
2. *What intervention is the question about? What treatment or manoeuvre are you considering for the patient? (Type of anaesthesia, pain therapy, fluid therapy) and which interventions will be compared?*
3. *What are the outcomes of clinical interest? (mortality, morbidity, pain, costs)*

Working with the question not only gives you the possibility of considering the different elements to construct answerable clinical questions, but also helps you when you are going to do the search for evidence.

LOCATING AND SELECTING STUDIES FOR REVIEWS

Systematic reviews of the effects of health care interventions generally focus on reports from randomised clinical trials (RCTs), when such data are available, because of the general acceptance that this study design will lead to the most reliable estimates of effects(1). A comprehensive search for relevant RCTs, which seeks to minimise bias, is one of the essential steps in doing a systematic review, and one of the factors that distinguishes a systematic review from a traditional review.

A comprehensive search is important, not only for ensuring that as many studies as possible are identified but also to minimise the selection bias for those that are found. Studies that show an intervention to be effective are more likely to be published in English, why limiting language to English may result in an overestimation of effectiveness(2)

When doing a comprehensive search it is logical to start with the electronic bibliographic databases, such as Medline and Embase. These databases usually have an abstract, which allows for judging clinical relevance at an early stage.

MEDLINE AND EMBASE

Medline (and the free electronic version PubMed) contains over 11 million records from 1966 onwards. Embase, which is often considered the European counterpart to Medline, contains approximately 9 million citations. The overlap between the two databases is estimated to about 34%(3). Studies comparing searches of the two databases have generally concluded that a comprehensive search requires that both databases be searched.

The controlled vocabulary search terms for Medline and Embase are not identical. Search strategies needs to be customised for each database. In case of difficulty the local librarian will usually be able to help.

The search strategy must be clear, comprehensive and reproducible. It should be written in the review in its full length permitting readers to redo the search for new studies. The search strategy must contain MeSH search as well as text word search.

CENTRAL (THE COCHRANE CENTRAL REGISTER OF CONTROLLED TRIALS)

Central contains about 350,000 clinically controlled trials (RCTs and CCTs). This database includes citations to reports of controlled trials that might not be indexed in Medline, Embase or other bibliographic databases. (4)

OTHER WAYS OF RETRIEVING TRIALS FOR SYSTEMATIC REVIEWS

Handsearching involves a manual page by page examination of the entire content of a journal issue to identify all eligible reports of trials, whether they appear in articles, abstracts, news columns, editorials, letters or other text.

Checking reference lists sometimes identifies additional relevant trials, but should never be used alone, because of the tendency of investigators to selectively cite positive studies(5).

CRITICALLY APPRAISING THE TRIALS

Quality assessment of individual trials is necessary to limit bias in conducting a systematic review, gain insight into potential comparisons and guide interpretation of findings.

The overall conclusion of the systematic review is reliant on the quality of every trial included. A synergistic effect on the overall result could be seen with the inclusion of several biased trials. Because of this it is important not only to critically appraise all trials thoroughly, but also to describe the quality of each trial.

Sources of bias can be divided into the following four types: Selection bias, Performance bias, Attrition bias and Detection bias.

SELECTION BIAS

Selection bias refers to systematic differences between the groups in RCTs and CCTs. Random allocation with adequate concealment of allocation protects against selection bias. If allocation isn't concealed, different characteristics of the participants may influence the choice of group (sex, weight, patient preferences or what treatment the physician thinks are effective)

Selection bias is also sometimes used to describe a systematic error in reviews due to how the reviews are selected for inclusion. Publication bias is an example of this type of bias.

PERFORMANCE BIAS

This describes any difference in care between the groups apart from the intervention being evaluated. For example, if patients know they are in the control group they may be more likely to use other forms of care, and care providers may treat patients differently according to what group they are in. Blinding of participants (patients and caregivers) is used to protect against performance bias.

ATTRITION BIAS

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

DETECTION BIAS

Detection bias refers to systematic differences between comparison groups in outcome assessment. Trials that have blinded outcome assessors are less likely to be biased than trials that do not. Detection bias is more likely to occur in trials that have subjective outcomes such as complications and pain.

Several scales have been developed in order to systematically assess trials. These are for the most part based on suggested or “generally accepted” criteria for trial quality. Many of the instruments are liable to confuse the quality of reporting with the validity of the design and conduct of a trial. Moreover, scoring is based on whether something was reported rather than whether it was done appropriately in the study.

Because there is no “gold standard” for the true validity of a trial, the possibility of validating any proposed scoring system is limited(6). For these reasons, it is preferable to use simple approaches for assessing trial quality that can be fully reported and explained.

BIAS IN NON-EXPERIMENTAL STUDIES

For non-experimental studies such as cohort studies and case-control studies the general rule is that they are prone to the same biases as the RCTs, but there is less control, and it is more difficult to evaluate bias.

Selection bias is almost always present in this type of studies. Even though the investigators make a great effort in identifying and adjusting for confounders, there will always be many factors that are not known as confounders or that have not been measured.

The bias present in non-randomised studies can alter the results in both directions, causing them to be either larger or smaller than they are. On average, however, selection bias tends to make treatment effects appear larger than they are(7).

Despite these concerns, there is sometimes good reason to rely on observational studies for information about health care interventions. For example, well designed observational studies have provided useful data concerning mandatory use of helmets for motorcyclists, screening for cervical cancer, dissemination of clinical guidelines to change professional practice and rare adverse effects of various interventions.

Generally, non-experimental studies are more prone to be biased; they are more difficult to appraise critically and need a lot of caution when results are interpreted.

COLLECTING DATA

After critical appraisal of the studies the studies are divided into those who meet the inclusion criteria for the systematic review and those who do not.

Now it is time to collect data from the included studies. It is important that the data are collected in the same way from all the studies, using a data collection form. Often, it will be appropriate for the authors of the review to collect the data independently and compare their results afterwards, any disagreement being resolved by rereading and discussion.

During data collection the following issues should be assessed:

Verification of study eligibility (Search and selection process should have eliminated most of the ineligible studies, nevertheless, it is good to verify study eligibility at the time of data collection).

Methods of the review. Adequacy of allocation concealment, trial design, patient, provider and outcome assessor blinding, drop outs, crossovers, and other potential confounders.

Participants. Information about participant characteristics is important to judge heterogeneity between studies and applicability of results.

Interventions. The intervention and how it was delivered should be described. The treatment given to the control group is equally important here. The different intervention types are a common source of heterogeneity and may well be the reason for not combined the study results.

Outcome measures and results. Reports of studies often include more than one outcome, and they may report the same outcome using different measures, may include outcomes for subgroups and may report outcomes measured at different points in time. To avoid hidden mistakes, outcomes should be collected in the reported format and eventually transformed in a subsequent step.

ANALYSING AND PRESENTING RESULTS

The term meta-analysis describes the statistical process of pooling data from several studies, thereby enlarging the size of population and the statistical power. The question whether or not to perform a meta-analysis is often difficult to answer. Before trying to answer this, the following must be taken into consideration:

What comparisons should be made?

What study results should be used in each comparison?

Are the results of studies similar within each comparison?

What is the best summery of effect for each comparison?

How reliable are those summaries?

Decisions about which study results are similar enough that they should be grouped together require an understanding of the problem that the review addresses and judgement by the reviewer and user. The main question is whether the participants, intervention or results are too different to be pooled together statistically (heterogeneity).

The results of almost identical interventions can be presented in many different ways in different studies. When choosing how to present the result it is important to consider whether it should be as dichotomous or continuous data, and whether continuous data should be transformed into dichotomous data by inserting a cut of point. Sometimes it is necessary to contact the investigators to retrieve additional data, since many results may be presented as “odds ratios” or other calculated values, not revealing the real number of outcomes and group sizes.

INVESTIGATING HETEROGENEITY

Whether differences between results from different studies are different because the studies are different or by statistical chance can evaluated by looking at the confidence intervals. If the confidence intervals do not overlap, it is likely that the results are statistically different. There are statistical tests for evaluating heterogeneity more precisely(8).

SENSITIVITY ANALYSIS

Because there are different approaches to conducting systematic reviews, reviewers should ask: How sensitive are the results of the analysis to changes in the way it was done?

This provides reviewers with an approach to testing how robust the result of the review are relative to key decisions and assumptions that were made in the process of conducting the review.

- Changing the inclusion criteria for the types of studies, participants, interventions and outcome measures
- Including or excluding studies where there is some ambiguity as to whether they meet the inclusion criteria
- reanalysing the data imputing a reasonable range of values for missing data

If sensitivity analyses that are done do not materially change the results, it strengthens the confidence that can be placed in these results. If the results change in a way that might lead to different conclusions, this indicates a need for greater caution in interpreting the result and drawing conclusions.

INTERPRETING RESULTS

The systematic review is an important component for clinical decision making. When trying to make conclusion the following should be taken into consideration:

- The strength of evidence
- The applicability of the results
- Other information such as considerations of costs and current practice

STRENGTH OF EVIDENCE

Has the critical appraisal revealed any important methodological limitation of the included trials? How do the included studies fit into the context of other evidence, not included in the review? This could be epidemiological studies of adverse effects, or qualitative studies of patients' perception of the treatment given.

It is logical to consider the quality of the included trials, the magnitude and significance of the observed effects, the consistency of effect across the trials, the existence of supporting evidence from other types of trial and whether there is suspected bias.

APPLICABILITY

Applicability is about whether the results from a systematic review can be user in the clinicians own setting. First the clinician must decide whether the review provides valid information about important potential benefits and harms. If this is the case, the clinician must decide whether the participants and settings of the included studies are reasonably similar to their own. This can be described by another question: Are there any compelling reasons why the evidence should not be applied under the current circumstances? The reasons for this could be biological and cultural variation, variation in compliance (from patients or practitioners) or variations in baseline risk (variation in incidence of a specific disease, or adverse effect)

IN CONCLUSION

A systematic review is a highly reliable document, assuming the authors have followed the overall rules. When using a systematic review for changing practice, it is important to judge the results of the review thoroughly, as well as taking other sources of information into consideration, such as other types of evidence, the local patient characteristics, the costs and current clinical practice.

THE COCHRANE ANAESTHESIA REVIEW GROUP

The Cochrane Anaesthesia Review Group undertakes the preparation, updating and disseminating of systematic reviews dealing with topics related to anaesthesia, intensive care, perioperative medicine and prehospital care(9). The group can be visited at the website www.carg.dk

REFERENCES

1. The Cochrane Handbook. The Cochrane Library issue 4, 2003 . 2003. Wiley.
Ref Type: Electronic Citation
2. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H, Jr. Publication bias and clinical trials. *Control Clin Trials* 1987; 8(4):343-353.
3. Smith BJ, Darzins PJ, Quinn M, Heller RF. Modern methods of searching the medical literature. *Med J Aust* 1992; 157(9):603-611.
4. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof* 2002; 25(1):38-64.
5. Gotzsche PC. Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)* 1987; 295(6599):654-656.
6. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282(11):1054-1060.
7. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. The Cochrane Library, Issue 4, 2003. 2003. John Wiley and Sons, Ltd.
Ref Type: Electronic Citation
8. Walker AM, Martin-Moreno JM, Artalejo FR. Odd man out: a graphical approach to meta-analysis. *Am J Public Health* 1988; 78(8):961-966.
9. Pedersen T, Moller AM, Cracknell J. The mission of the cochrane anesthesia review group: preparing and disseminating systematic reviews of the effect of health care in anesthesiology. *Anesth Analg* 2002; 95(4):1012-8, table.