CHAPTER VIII

META-ANALYSIS

Francois Delahaye

Meta-analysis is a particular type of study which contrasts with the previous ones: the researcher did not collect the data useful for the study himself, he was not in direct contact with the subjects studied (or their file).

Meta in Greek means "through". Meta-analysis is therefore literally an analysis of analysis. The researcher collects the studies whose data are likely to be combined.

Meta-analysis first appeared in the medical literature in the 1970s. This method, which is increasingly used, is a way of circumventing insurmountable logistical difficulties when a very large number of subjects is needed to demonstrate an effect.

Meta-analysis is therefore a type of study in its own right which, more than any other, requires exhaustive bibliographic research.

The best possible critical review and synthesis of the available information is essential for all those who have to make decisions, whether in front of a patient, to establish a common strategy for similar groups of patients, or to formulate research hypotheses in medicine, epidemiology, or health policy and administration.

Invented by researchers in the educational sciences and psychology in the early 1970s, meta-analysis, a term coined by Glass in 1976, is a qualitative and quantitative evaluation of medical information, and its synthesis and structured integration . “The analysis of analyses” (meta-means “that which exceeds, encompasses”), is “the statistical analysis of a large number of data coming from several analyses, in order to integrate the results”.

The words “meta-analysis” and “systematic review” are often used synonymously, although they do not have quite the same meaning. The systematic review uses a structured procedure (for example for the search of the literature). Meta-analysis is a statistical technique of combining results.

In medicine, the first meta-analyses were published in the mid-1970s. The technique was skeptical until the mid-1980s, then gained momentum with Peto's team in Oxford. In 1993, Iain Chalmers, an epidemiologist from Oxford, founded the Cochrane Collaboration (named after Archie Cochrane, a researcher who contributed greatly to the development of epidemiology), an international non-profit organization whose aim is to produce , disseminate and update meta-analyses in the medical field (www.cochrane.org). There are now standards for conducting and reporting a meta-analysis (Cochrane, PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses).

Meta-analysis is any systematic method that uses statistical techniques to combine data from independent studies to obtain an estimate of the overall effect of a variable on a defined event. We can thus carry out a meta-analysis of descriptive studies, intervention studies, or studies validating clinical tools, for example diagnostic methods, but most often the meta-analyses relate to therapeutic trials.

Meta-analysis allows:

- decide on conflicting conclusions;
- increase power for major events and subgroup analyses;
- to narrow the limits of the size of the effect (increase the precision);
- answer new questions.

Six reasons can lead to a meta-analysis:

- obtain more stable estimates of the effect of a treatment;
- help to interpret the "generalizability" of the results;
- conduct analyzes on sub-groups;
- help with marketing authorization requests;
- assist in the planning of clinical trials;
- counteract the over-enthusiasm that often accompanies the introduction of new drugs.

I - QUALITATIVE OR QUANTITATIVE META-ANALYSIS

Two main approaches are possible:

A - The qualitative approach

It consists of giving different importance to the various studies according to their methodological quality. The studies are reviewed according to a set of criteria to judge the scientific validity and the possibilities of clinical application.

The purpose of such a meta-analysis is to draw conclusions from studies deemed methodologically superior.

The qualitative approach includes fundamental steps:

- wording of the question;
- search for studies;
- definition of criteria for judging the scientific credibility of studies;
- application of these criteria to each study;
- analysis of the relationship between the scientific credibility of a study and its conclusions.

Let's take the example of BCG and the prevention of tuberculosis. BCG has been widely used to prevent tuberculosis for over 70 years, but its effectiveness is controversial. This, at least in part, is due to discordant results from different trials.

Clemens et al. first describe how the ideal clinical trial should be conducted, then analyze the available literature and compare the methods of these trials with their results *(Table 1).* According to Clemens, "Adequate demonstration of unbiased detection of tuberculosis was only available for the three trials reporting efficacy of 75% and above; in most of the trials reporting low efficacy, the confidence intervals were wide, not being able to exclude high

efficacy, but in all the trials reporting high efficacy, the confidence intervals were narrow, excluding low efficacy". The authors conclude that BCG may provide protection, and that bias or insufficient statistical power may have contributed to the discordant results.

Table 1 *- Protection against bias and accuracy of statistical precision in 8 major BCG trials*

Trial	Adequate protection against bias				Adequate statistical precision	Observed protective efficacy
	susceptibility	monitoring	diagnostic method	interpretation		
American Indians	Yes	Yes	Yes	Yes	Yes	80%
England	Yes	Yes	Yes	Yes	Yes	76%
Chicago	Probable	Yes	Yes	Yes	Yes	75%
Puerto Rico	Yes	No	No	No	No	29%
Madanapalle	Equivocal	No	No	Probable	No	20%
Georgia-Alabama	Yes	No	No	Equivocal	No	6%
Chingleput	Probable	No	No	Yes	Yes	-32%
Georgia	Yes	No	No	No	No	-56%

B - The quantitative approach

It consists of a quantitative summary of the results of the different studies, so as to create a single, large study with greater statistical power.

The quantitative approach must also follow several steps, but most often, the two approaches, qualitative and quantitative, are combined, and the steps are:

- wording of the question;
- bibliographic search ;
- development of criteria specifying the attributes (clinical conditions, treatments, events) that will be grouped and compared;
- classification and coding of the studies selected;
- definition of criteria for judging the scientific credibility of studies;
- study of the quality of the studies;
- statistical analysis of data;
- formulation of results;
- sensitivity analyses;
- analysis of the relationship between the scientific credibility of a study and its conclusions;
- interpretation of results and conclusions.

II - THE SEVEN STEPS OF META-ANALYSIS *(figure 1)*

A- Objectives

Of course, the objective of the meta-analysis must always be clearly specified. A protocol should always be written, and this before the execution of the meta-analysis. Precise and rigorous, it must in particular specify the hypotheses, and all the procedures used, in particular those of the literature search, the selection criteria for the trials, the definition of the events, the technique for analyzing heterogeneity, and statistical methods.

Figure 1 *- Stages of meta-analysis*

OBJECTIVES

SEARCHING THE LITERATURE

EXTRACTING DATA FROM EACH STUDY

QUALITY ASSESSMENT OF EACH STUDY

DATA POOLING

SENSITIVITY ANALYZES

CONCLUSIONS

B - Literature search

This colossal task is fundamental! The exhaustiveness of the research indeed depends on the quality of the meta-analysis.

Research must use several methods simultaneously. The collection is made:

- of course by tracking through bibliographic databases, manual (Index Medicus, Excerpta Medica, etc.) or automated (Medline, Cancerlit, Pascaline, etc.), and by consulting conference proceedings and bibliographies of articles and books on the subject;
- but also thanks to discussions with colleagues and experts, consultation with pharmaceutical companies and government bodies funding trials.

While you might think they are infallible, automated searches are not perfect. All the research methods mentioned above should be used, not just bibliographic databases.

A big pitfall of meta-analysis, but common to all kinds of literature reviews, is publication bias, which is the preferential submission and preferential acceptance of studies showing positive results. There is no perfect solution yet to this publication bias. Some have proposed calculating the number of studies it would take to change the conclusions of the meta-analysis. Another response is to keep up-to-date records of all ongoing trials. Thus, we know the fate of all trials, including those that have been interrupted and those whose results have been negative. A representation in the form of an inverted funnel plot is often used (figure 2).

The choice of inclusion and exclusion criteria for studies in the meta-analysis can be based on various variables: the study design (a meta-analysis is often limited to randomized trials), the size of the study (we can require a minimum number of subjects per group), the study population, the type of groups treated and control groups (a certain dose, etc.), the event studied, etc. The criteria, which depend on the objectives of the meta-analysis, must be listed in the protocol, with the reason for their choice.

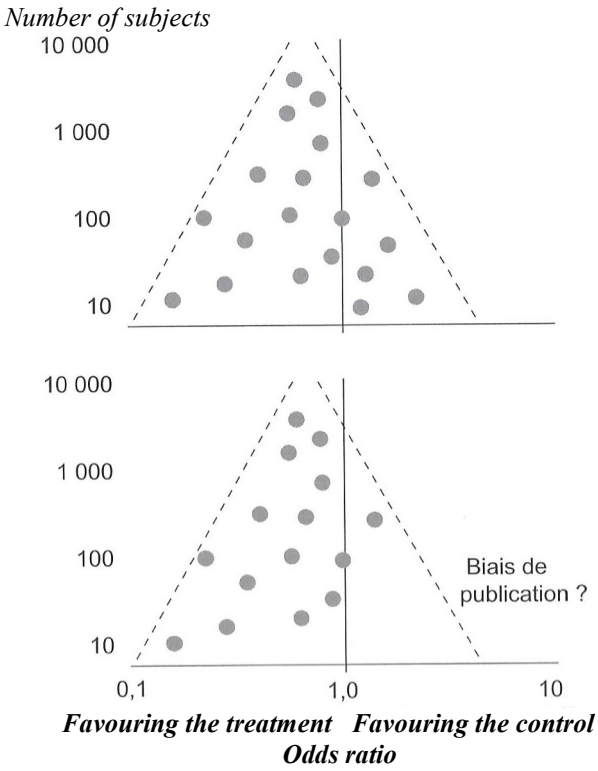
Should we put all the studies?

This increases the representativeness of the conclusions, but decreases the statistical validity of the synthesis by including less rigorous studies. This decision depends on the objective of the meta-analysis, so for example we are much stricter in the selection of trials for a meta-analysis that is part of a marketing authorization dossier than for a exploratory meta-analysis.

The decision that studies are similar enough that their results can be aggregated is subjective, and it is difficult to develop universal criteria to ensure appropriate selection of studies.

The report should contain the list of studies included and the list of studies excluded, so that the reader can know what the meta-analysis is based on, and which studies have been refuted, as well as the reason for the exclusion.

Figure 2 *- Inverted funnel plot looking for publication bias*
Each trial is presented as a point at the intersection of the effect size and the number of subjects in the study 'test. A symmetrical look suggests that he is not missing an attempt, an asymmetrical look suggests publication bias.



C - Extraction of data from each study

There are three main types of data used in meta-analyses.

1- Individual data

One of the first examples is given by Canner.

Canner analyzes the six most important trials of the efficacy of aspirin in the secondary prevention of mortality after myocardial infarction. None of these trials show a statistically significant effect of aspirin. The meta-analysis of five of these trials makes it possible to objectify a beneficial effect of aspirin ($p = 0.014$).

The addition of AMIS (Aspirin Myocardial Infarction Study) changes everything: in this trial, by far the largest, aspirin has an adverse effect. When the six trials are pooled, the odds

ratio favoring aspirin increases from 0.76 (five trials) to 0.90 (six trials), and statistical significance disappears. The heterogeneity test is borderline significant.

Since then, there have been many meta-analyses on individual data. Let us cite, for example, in cardiology, the work of the Prospective studies collaboration, the Antiplatelet Trialists' Collaboration, the Antithrombotic Trialists' Collaboration or the INDANA group.

How can this heterogeneity be explained?

There are no obvious differences in the study design, the doses used, or the time elapsed between infarction and therapy. On the other hand, there is a poor distribution of the basic characteristics. Having access to individual data from three of the trials, Canner can fit (multiple linear regression) according to these baseline characteristics. The heterogeneity test becomes much less significant (0.22), and the association test (six trials) becomes significant (0.04), in favor of a favorable effect of aspirin.

This technique, much more laborious, and which requires the acceptance of the investigators to lend their data, but which makes it possible to answer many more questions, has experienced significant growth in recent years.

2 - Summaries from publications

It is currently from these publications that the vast majority of meta-analyses are made. But often the necessary data are missing, and the biases that may have arisen from the inadequate inclusion of certain randomized subjects are retained in the analysis, hence the interest in requesting the data from the experimenter.

3 - Summaries by sub-groups (in particular by gender or age, in the same publication or in subsequent publications)

This data extraction process is long, boring, subject to error, and therefore biased. To guard against this as much as possible, data extraction slips must have been designed, and they must be completed by several people if possible, with consensus meetings to settle disagreements.

There are different kinds of interesting events:

- continuous variable (blood pressure, quality of life score);
- binary variable (mortality, complications);
- ordinal variable (tumor stage);
- time-related variable (disease-free survival).

The analyzes mainly focus on binary variables. Four measures of treatment effect are often used: if Pc is the proportion of events in the control group and Pt the proportion of events in the treated group, the effect can be measured as:

- absolute difference: $Pc - Pt$
- relative risk: Pt/Pc and odds ratio
- relative risk reduction: $\frac{Pc - Pt}{Pc}$
- number of subjects to be treated: $\frac{1}{Pc - Pt}$

Two clues are particularly telling:

- the relative risk reduction is the difference in risk between the two groups, relative to the risk in the control group; if mortality is 10% in the control group and 5% in the treated group, the relative risk reduction is 50%; however, the isolated presentation of the relative risk reduction is misleading: to hear that the relative reduction in the risk of an event thanks to an intervention is 50% is often, unconsciously, to think that the intervention avoids one event out of two ; there must also be an absolute risk: the relative risk reduction is 50% when the absolute risk drops from 80% to 40%, a clinically very significant reduction; the relative reduction in risk is also 50% when the absolute risk goes from 2 / 1 billion to 1 / 1 billion, a clinically… infinitesimal reduction! ;
- the number of subjects to be treated for a certain period of time to avoid an event is the inverse of the absolute difference in risk; if mortality is 10% in the control group and 5% in the treated group, the number needed to treat is 20.

Another index, widely used in psychology, is the effect size, the difference between the mean in the treatment group and that in the control group, divided by the standard deviation in the control group.

D - Quality assessment of each study

This process being particularly subjective, study quality assessment reports are mandatory, with reading by at least two people and disagreement resolution meetings, after the articles have been prepared, removing any identification.

At the end of the review, a quality score is given to each study. This score can be used:

- either as a threshold, for inclusion or exclusion from a study;
- either to give a respective weight to each study;
- or to compare the result of the study and its quality score.

E - Data grouping

This step is the one that most distinguishes meta-analysis from other information synthesis techniques. But before carrying out the grouping itself, it is first necessary to check the absence of heterogeneity.

1 - Homogeneity

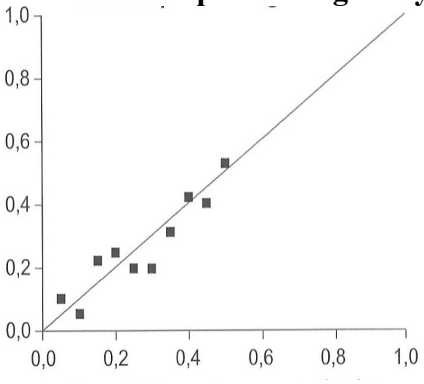
An underlying assumption when combining multiple studies is that differences between study results are due to chance alone, and therefore all results are consistent. But this hypothesis must be discussed. If the variations are not due to chance alone, the data clustering techniques are more complicated, and possibly unreasonable.

A first step in the analysis of heterogeneity consists in a graphical study *(Figure 3)*. There are, of course, more formal statistical techniques for testing homogeneity, in particular a Mantel-Haenszel $\chi 2$ or regression techniques. But their power is limited, and the associated graphical analysis is particularly useful.

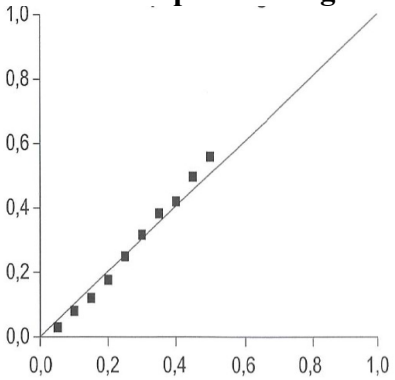
Figure 3 *- Graphical analysis of heterogeneity*

The squares represent for each trial the rate of events in the control group and in the treated group. In Figure 3A, the squares are fairly evenly distributed; in Figure 3B, the distribution of the squares is heterogeneous: the squares are below the diagonal when the event rates are low and above the diagonal when the event rates are high.

A- Example: homogeneity



B- Example: heterogeneity

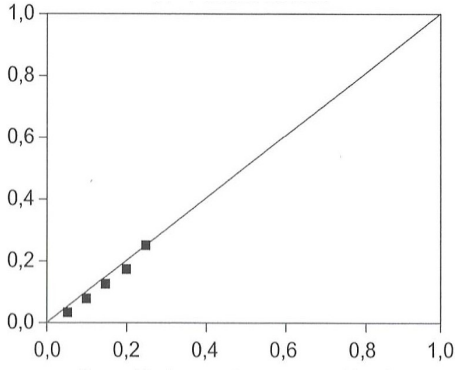


The non-uniformity may be due to a certain characteristic, for example the dose used. A new graph separating the trials into different groups according to the dose, makes it possible to find homogeneity within each group *(figure 4)*.

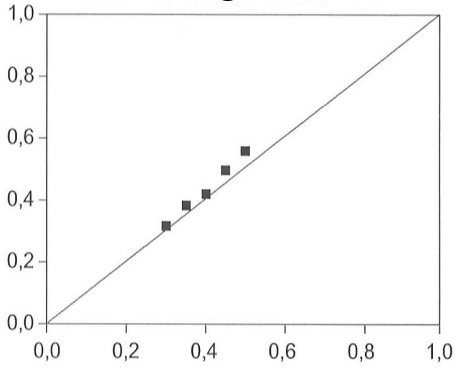
Figure 4 *- Graphical analysis of heterogeneity*

The squares represent for each trial the rate of events in the control group and in the treated group. In Figure 3B, heterogeneity is due to dose. Homogeneity appears when the trials are distinguished according to the dose, low or high.

A- Low doses



B- High doses



2 - Statistical methods

They can be grouped into four categories:

a- methods combining the p;

b- methods combining the values of statistical tests (z, t): the oldest and simplest, they are like the previous ones very limited;

c- model-based methods:

• for a binary event, binomial model: logarithm of the odds ratio, difference in rates, Mantel-Haenszel, Peto or Cochran method; these methods have several advantages: the events are compared within each trial, which increases the accuracy of the overall result;

the difference for each event rate is weighted by its variance, the trials with the most stable events (usually the larger trials) have the most influence; the aggregation of "Observed – Expected" provides an overall estimate in addition to the statistical test;

• for a quantitative variable: analysis of variance;

d- modeling methods (multiple linear regression, logistic regression). It is advisable to use several techniques (the results are "robust"), to choose techniques that give weight to each trial, allow a reasonable definition of the underlying model and allow testing for heterogeneity.

F - Sensitivity analyzes

We have to ask ourselves how sensitive the results of the meta-analysis are to the way of doing this meta-analysis. In other words, sensitivity analyzes can be done. For example, are the results different if one includes randomized and non-randomised trials instead of only including randomized trials?

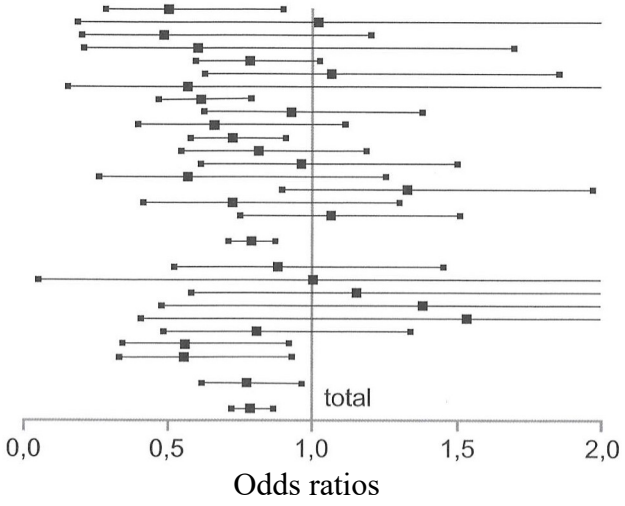
The studies can be grouped according to the characteristics of the different groups of patients or of the study design (random or non-random allocation, dose of the active drug, etc.), in order to determine the influence of these characteristics on the results of the meta-analyse.

Subgroup analyzes within a clinical trial pose several problems: multiple comparisons, misinterpretation of differences, interactions. An additional problem for meta-analyses is represented by the heterogeneity of the characteristics of the various studies subject to the meta-analysis. If the power of the tests is increased by the increase in the number of subjects thanks to the grouping, it is however necessary to remain very careful in the realization, and the interpretation, of the sensitivity analyses.

G - Presentation of results

After the objectives and the methods, in particular the statistical techniques and the quality control procedures, have been specified, the results are presented in the form of tables, but also very often thanks to figures representing for each test the estimation of the effect and its confidence interval, then the same data for the total ("forest plot") *(Figure 5)*.

Figure 5 *- Graphic representation of the results of the meta-analysis*



The index used is the odds ratio, represented by the large square, the horizontal line representing the 95% confidence interval, with its limits (small squares). Each line represents the results of one trial, the last (below) showing the results of the meta-analysis.

Finally, we conclude by discussing the results of the meta-analysis according to the choice of studies, their quality and homogeneity. The quality and limitations of the meta-analysis should also be discussed, and the significance of the results assessed.

III - EVALUATION OF THE QUALITY OF THE META-ANALYSIS

The reader of a meta-analysis must ask several questions before adopting the conclusions of the work:

- is the objective clearly specified?
- is there evidence of a working protocol?
- are the literature search techniques specified? Is the issue of publication bias considered?
- Are the inclusion and exclusion criteria specified, are the included and excluded items listed, are the reasons for exclusion given?
- Are the treatments similar enough to allow the results to be pooled? The same for the control groups?
- are the homogeneity tests, graphs and statistics presented?
- Is the statistical technique for grouping the data correct?
- have sensitivity analyzes been carried out?
- Are conclusions about the effectiveness of the treatment and for future research drawn?

IV - ADVANTAGES AND LIMITS OF META-ANALYSIS

We are currently witnessing a flowering of meta-analyses, and several teams have clarified their methodology.

This technique has several advantages:

- it makes it possible to estimate the importance of an effect;
- it increases the statistical power;
- it increases “generalizability”;
- it requires rigor in the methods, the reading, the collection of data;
- it reduces the part of the subjective.

However, many authors criticize it:

- it ignores the quality of the studies: we have seen that tools exist, but it is true that improvements are possible;
- it is illogical to combine the results of studies using different patients, different measurement techniques, and carried out at different times: the meta-analyst must present a summary of the characteristics for each study, highlight the differences, test

heterogeneity, and discuss its results and "generalizability" in terms of these various elements;
• there is a potential publication bias: but this is the fate of any synthesis of information;
• the validity of the meta-analysis depends on the degree of completeness and precision of the information reported in the various studies: again, this is not specific to it;
• the validity of the statistical techniques of the meta-analysis must be established.

Despite the criticisms, the limitations, this technique of meta-analysis, when it is prudent and well carried out, brings additional information, allowing to improve the quality of our answer to a particular question.

References

Boissel JP, Blanchard J, Panak E, Peyrieux JC, Sacks H. Considerations for the meta-analysis of randomized clinical trials. Summary of a panel discussion. Controlled Clin Trials 1989;10:254-81

Canner PL. An overview of six clinical trials of aspirin in coronary heart disease. Stat Med 1987;6:255-63

Clemens JD, Chuong JJH, Feinstein AR. The BCG controversy. A methodological and statistical reappraisal. JAMA 1983;249:2362-9

Colton T, Freedman LS, Johnson AL, eds. Proceedings of the workshop on methodologic issues in overviews of randomized clinical trials. Stat Med 1987;6:217-410

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

Leucht S, Kissling W, Davis JM. How to read and understand and use systematic reviews and meta-analyses. Acta Psychiatr Scand 2009;119:443-50

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009;151:W65-94

Moher D, Tricco AC. Issues related to the conduct of systematic reviews: a focus on the nutrition field. Am J Clin Nutr 2008;88:1191-9

Tseng TY, Dahm P, Poolman RW, Preminger GM, Canales BJ, Montori VM. How to use a systematic literature review and meta-analysis. J Urol 2008;180:1249-56

Some articles on the subject of meta-analysis:

Mulrow CD, Cook DJ, Davidoff F. Systematic reviews: Critical links in the great chain of evidence. Ann Intern Med 1997;126:389-91

Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: Synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126:376-80

Hunt DL, MacKibbon KA. Locating and appraising systematic reviews. Ann Intern Med 1997;126:532-8

MacQuay HJ, Moore. Using numerical results from systematic reviews in clinical practice. Ann Intern Med 1997;126:712-20

Badgett RG, Okeefe M, Henderson MC. Using systematic reviews in clinical education. Ann Intern Med 1997;126:886-91

Bero LA, Jadad AR. How consumers and policymakers can use systematic reviews for decision making. Ann Intern Med 1997;127:37-42

Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR. The relation between systematic reviews and practice guidelines. Ann Intern Med 1997;127:210-6

Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med 1997;127:380-7

Books:

Cucherat M. Méta-analyse des essais thérapeutiques. Éditions Masson, 1997

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Éditions Wiley, 2009