Systematic Reviews and Meta-Analysis of Therapeutic Interventions: How to Better Understand Them?

Revisão Sistemática e Metanálise de Intervenções Terapêuticas: Como Melhor Entendê-las?

RAPHAEL CÂMARA MEDEIROS PARENTE[1]; EVANDRO DA SILVA FREIRE COUTINHO[2];
MARCO AURÉLIO PINHO DE OLIVEIRA[3]; ROGER KELLER CELESTE[4]; PAULA DE HOLANDA MENDES[5];
RICARDO BASSIL LASMAR[6]

[1] Doctorate in Gynecology (Human Reproduction) from the Federal University of the State of São Paulo (UNIFESP). Master's in Epidemiology from the Institute of Social Medicine of the University of the State of Rio de Janeiro (UERJ); [2] Senior Researcher of the National School of Public Health (FIOCRUZ). Professor of Epidemiology UERJ. Member of Editorial Board of the Cochrane Library; [3] Doctorate in Epidemiology, by IMS UERJ. Professor and Chief of Gynecology – UERJ; [4] Doctoral in Epidemiology, Social Medicine Institute, UERJ. Ex-Professor of Epidemiology of UFF. Professor of Epidemiology, ULBRA; [5] Master's in Epidemiology, Social Medicine Institute, UERJ; [6] Professor of Gynecology UFF. Doctorate in Gynecology, UNESP

ABSTRACT

In the past decade, the Era of Evidence-Based Medicine, the number of meta-analysis dramatically increased. Meta-analyses statistically combine the results of multiple studies and are considered to be the highest level of evidence when the results of high-quality randomized trials are combined in an appropriate way. Results from a meta-analysis may not correspond to reality because of the large variation in the quality of the studies that have been pooled, and clinical and methodological differences among the included studies. The growing popularity of systematic reviews and meta-analyses has made it important to better understand them. The objective of this article is to help the reader comprehend how a systematic review and meta-analysis is carried out and to be better able to interpret them. We explain some impor tant aspects of conducting a meta-analysis. A better understanding of the basic terminology and the concepts involved in generating a systematic review and meta-analysis may help the clinician better evaluate the quality of a meta-analysis and the real importance of its findings for a specific patient.

Key words: Meta-analysis, Systematic Review, Clinical Trial.

INTRODUCTION

Evidence-Based Medicine (EBM) is the systematic process of searching, quality assessment, and the application of recent research results as a basis for clinical decisions.[1] Systematic reviews seek to present – in a critical and integrated way – the results of existing studies. Using a clear and objective process to search for and evaluate existing research on a given subject, the best available evidence is obtained for clinical decision-making. As a result, it is not surprising that the number of systematic reviews and meta-analysis has been growing in significant ways since the 1990s. A Medline search showed that this technique of reviewing the literature increased 20-fold between 1989 and 1991 (*Marco – check this, a two year interval doesn't seem correct. It seems like the period of comparison should be longer to see a 20 fold increase, unless the initial base was miniscule.*) The change in philosophy brought about by evidence-based medicine, combined with growth in scientific output in the biomedical area, certainly was major factor in this increase. Whereas in 1940 there were about 2,300 biomedical journals, 50 years later this number soared to nearly 25,000.[2] These data give an idea of the problem faced by health professionals to assimilate the knowledge generated and make decisions based on that knowledge.

SYSTEMATIC REVIEW

There are several ways of dealing with this vast bibliography. One of them is using "narrative reviews". Narrative review, however, usually have different goals than systematic reviews. Narrative reviews are broad in terms of content, may express personal opinions and commentaries about the state of the art, selecting studies in a subjective manner, without clear criteria. The style of these reviews is characterized by sequences/series of "who said what?" permeated by a bibliography. The lack of objective criteria and limited integration of findings may lead to erroneous conclusions, if the purpose of such reviews was to provide a summary of all existing literature on the topic.

In contrast, systematic reviews, have as their focus responding to a specific clinical question. Systematic reviews require a search for studies using selective criteria, analysis of the quality of the studies selected, assessment of differences between the results of different studies, and the synthesis of the results of the studies in a qualitative way in the case of the systematic review and in a quantitative way in the case of the meta-analysis, as will be explained later. A systematic review is called a meta-analysis when statistical techniques are used to combine the data of different studies.

Systematic reviews and meta-analyses have their origins in astronomy, proceeding through agriculture, education, whose methods of numerical synthesis of results were developed by statisticians such as R.A. Fisher, L. Tippett, K. Pearson, E.S. Pearson, F. Yates, and W. G. Cochran.[2] Already in the early twentieth century Karl Pearson had published a synthesis of results of studies about the effectiveness of the vaccine against typhoid fever in soldiers, meta-analysis only gained expression in the medical field starting with the study of Thomas Chalmers and Joseph Lau, on the efficacy of streptokinase in reducing the mortality of patients with acute myocardial infarction. This trend got a boost with the creation, in 1992, of the Cochrane Centre at Oxford University, in England, in order to prepare, maintain and disseminate systematic reviews of controlled clinical trials.

Stages of a systematic review

Any systematic review and meta-analysis should be preceded by a protocol in which the strategy to be used must be specified. The steps of a systematic review and meta-analysis are shown in Figure 1. Clearly formulated questions, along with clear criteria for inclusion and exclusion of studies are essential to the process of identifying relevant studies for review and meta-analysis. It is necessary to have clarity about the characteristics of the population for whom the answer the original question is intended, the exposure that you want to investigate, as well as the clinical outcome that one wants to measure. It should also define what types of studies will be included (e.g. controlled clinical trials, case-control studies, cohort studies). Ideally, a systematic review of therapeutic or preventive procedures should include only randomized controlled trials.

Question: Objective and operationalized in order to be tested.

Ex: Does hormone therapy improves osteopenia in postmenopausal women?

Participants: Characterize the population regarding gender, age, clinical characteristics (if applicable). For example, women in the immediate postmenopausal period, regardless of social background, without osteoporosis. Define the degree of osteoporosis permitted in the study.

Intervention: Specify any hormone or one specific type.

Outcome: Specify how the improvement of osteopenia will be is defined and measured.

Type of study: For example, only randomized controlled trials.

This is followed by the phase in which relevant studies are identified. Restricting the search to Medline can lead to the distortions in the results of the systematic review, depending on the topic that you want to investigate. There are several databases of research studies for specific problems such as cancer, non-pharmacological care of the mentally ill, post-traumatic stress disorders, to cite a few examples. On the other hand, it is known that studies with negative results are less likely to be published, especially in major indexed journals; this can lead to an error called publication bias. In the case of therapeutic interventions, publication bias leads to the identification of nonexistent efficacies or exaggerates the magnitude of this efficacy.

One way to minimize the risk of this bias is to expand the search to non-indexed journals and conference proceedings, consulta experts, and search sites that register clinical trials, such as those present at www.york.ac.uk /inst/crd/revs.htm.
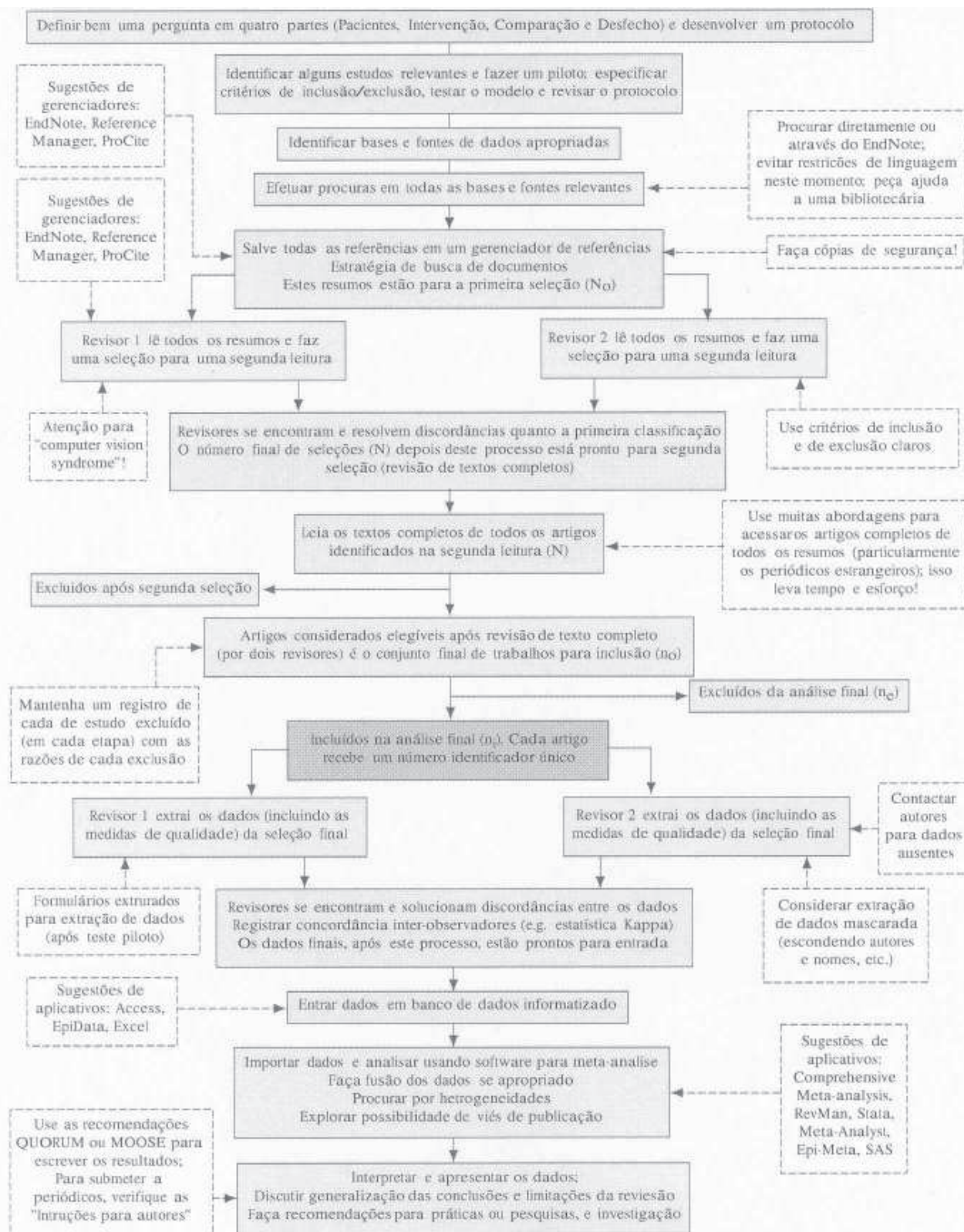
Figure 1 - *Stages of a systematic review.*
Madhukar Pai, Michael McCulloch, and Jack Colford. Systematic Reviews Group, UC Berkeley, 2002 [madhupai@uclink.berkeley.edu]
Translated from Portuguese by: Peter Emmanuel A. A. do Brazil, Master's candidate - IMS / UERJ, 2004 (pemmanuel@ig.com.br)
Available at: http://www.medepi.net/meta/guidelines/Berkeley_Systematic_Reviews_Road_ Map_V22_Versao_Brasileira.pdf

Another important error to avoid is the exclusion of articles written in less common languages. It is known that studies that report favorable results for the tested interventions tend to be published in English. So even if one cannot translate articles published, for example, in German or Japanese, they should be identified in the search so that later one can assess the possible impact of their exclusion on the findings of the systematic review.

Once the search is concluded, the study selection process begins with the evaluation of the titles and abstracts, to see if the articles meet inclusion criteria. In this step it is important, although difficult, that the evaluators are masked ("blind") regarding the origin of the work. This because, there is a chance that an article might be included or excluded solely because the evaluator already knows the group that published it or because the work was published in a particular journal. Having two researchers read each abstract may reduce the chance that an article of interest will be overlooked. Next, complete copies of the articles that meet the criteria or for which there is doubt about the relevance to the review are obtained. Articles can still be excluded at this stage, but the reason for this decision should be noted. The selection process should be documented, preferably in a flowchart. Figure 2 presents the model proposed by the "Quality of Reporting of Meta-Analysis Group – **QUOROM**,"[3] with documentation of how many studies were excluded at each step of the selection and the reason for these exclusions. In the case of observational studies, a proposal for a similar presentation of point was made by "Meta-analysis Of Observational Studies in Epidemiology Group-**MOOSE**". [4]

The selected studies should be evaluated regarding their **methodological quality** according to criteria established in the Protocol. A list of 22 criteria used to describe the quality of randomized clinical trials is described by the Consolidated Standards of Reporting Trials Group - CONSORT".[5] It is suggested that two researchers are involved in this phase, as well as in a later stage - the **extraction of information**.

META-ANALYSIS: QUANTITATIVE SYNTHESIS OF RESULTS

Summary-Measures and Forest Plot

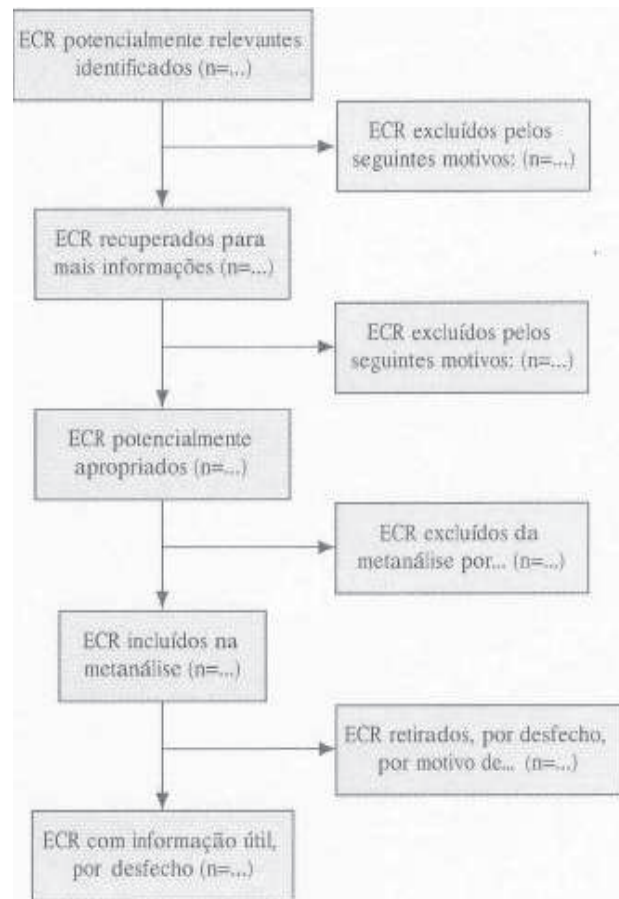If the studies are homogeneous, one can combine their results in a summary-measure. This measure

Figure 2 - Flowchart with the stages of a meta-analysis of clinical trials, proposed by the QUOROM (3).

increases the statistical power[a] and precision of the estimates, by increasing the sample size attained by combining several studies. Statistical techniques, however, are not able to correct biases in the review process. If the raw material is not of good quality, the result is not valid.

The summary-measure is obtained from a weighted average of the results of several studies, in which the weights are the inverse and their variances. In other words, studies with more precision (due to a larger sample size) are given more weight in the combined estimate. One of the statistical methods most commonly used for this purpose is the Mantel-Haenszel.

In Figure 3 we constructed a graph (forest plot) with data from a meta-analysis conducted by

---

[a] *Capacity of the statistical test to detect an effect of the intervention when it differs from the control group.*

Roberts and Dalziel[6] about the effectiveness of corticosteroids for accelerating fetal lung maturation in women at risk of giving birth early/prematurely. With minor variations, these graphs contain the following two elements:

1) Each line represents one study, the estimated relative risk (RR)[b] conveyed by a small square. The horizontal line that bisects the square is the 95% confidence interval.[c] One observes that in 13 of the 18 studies, the confidence interval includes the null value (relative risk = 1); such studies considered inconclusive.

2) The small diamond at the bottom represents the summary-measure. In the example in Figure 3, the combined relative risk (RR) was 0.69, which means a reduction (efficacy) of 31% [d] in the risk of neonatal death in the group in which mothers had used corticosteroids, compared with the control group. The 95% confidence interval of this RR (0.58 to 0.81, p <0.01) does not include the null value. It can be concluded that the prenatal use of corticosteroids during pregnancy reduces the risk of premature birth by 31%, and the probability that this finding is due to chance is less than 5%.

The squares indicating the RR of each study vary in size, and the weight accorded to each study to estimate the pooled RR is proportional to each square's area. The relative weight of each study appears in the right column of the chart.
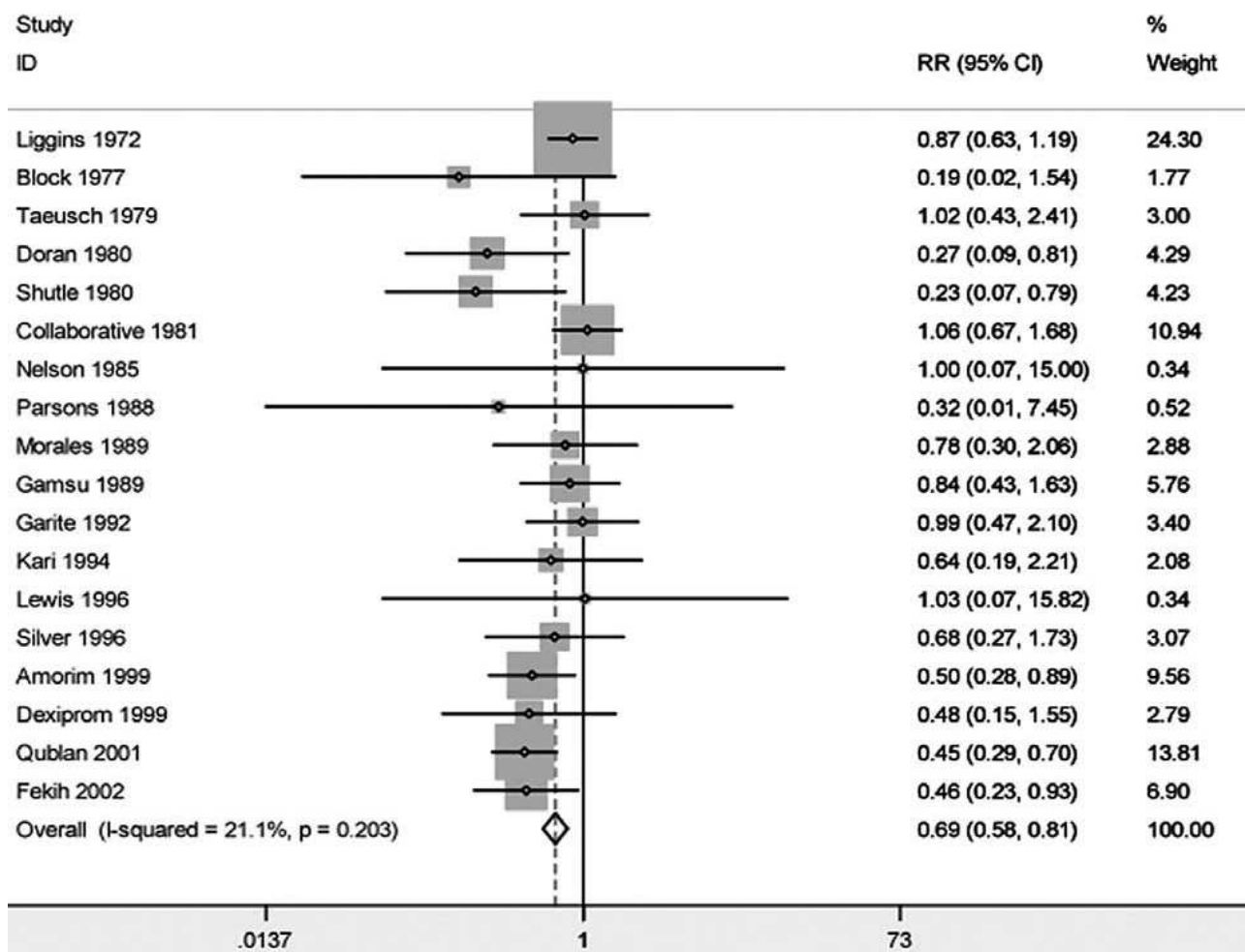
Study ID	RR (95% CI)	% Weight
Liggins 1972	0.87 (0.63, 1.19)	24.30
Block 1977	0.19 (0.02, 1.54)	1.77
Taeusch 1979	1.02 (0.43, 2.41)	3.00
Doran 1980	0.27 (0.09, 0.81)	4.29
Shutle 1980	0.23 (0.07, 0.79)	4.23
Collaborative 1981	1.06 (0.67, 1.68)	10.94
Nelson 1985	1.00 (0.07, 15.00)	0.34
Parsons 1988	0.32 (0.01, 7.45)	0.52
Morales 1989	0.78 (0.30, 2.06)	2.88
Gamsu 1989	0.84 (0.43, 1.63)	5.76
Garite 1992	0.99 (0.47, 2.10)	3.40
Kari 1994	0.64 (0.19, 2.21)	2.08
Lewis 1996	1.03 (0.07, 15.82)	0.34
Silver 1996	0.68 (0.27, 1.73)	3.07
Amorim 1999	0.50 (0.28, 0.89)	9.56
Dexiprom 1999	0.48 (0.15, 1.55)	2.79
Qublan 2001	0.45 (0.29, 0.70)	13.81
Fekih 2002	0.46 (0.23, 0.93)	6.90
Overall (I-squared = 21.1%, p = 0.203)	0.69 (0.58, 0.81)	100.00

.0137   1   73

Figure 3 - *Forest plot of clinical trials comparing the relative risks for neonatal mortality of premature infants in pregnant women who used corticosteroids or received a placebo. Graph produced with the command "metan" (fixed effects) of Stata statistical package, version 9.0, from raw data presented by Roberts and Dalziel.[6]*

---

[b] *Risk of neonatal death in the group of mothers who received corticosteroids divided by the risk of neonatal death in the group of mothers who received placebo. The RR is equal to 1 when there is no difference between the two groups being compared.*

[c] *Range of values that includes, with 95% confidence, the value of RR if all individuals, and not just a sample, had they been studied.*

[d] *Efficiency = (1 − 0.69) x 100.*

Evaluating the heterogeneity

It is common for the selected studies to have findings/results that are inconsistent. The fact that the difference between them exceeds what would be expected by chance is defined as statistical heterogeneity. Such heterogeneity reflects distinctions between studies, with regard to aspects of design, that included differences in the population studied, in the way the intervention or outcome is measured, the methodological quality of studies, among others.[4] In this case, it does not make sense to obtain only a summary measure, but one should explore the reasons for this inconsistency.

Thompson[7] exemplifies this situation with studies about the effect of endoscopic sclerotherapy of esophageal varices on the reduction of the mortality in patients with hepatic cirrhosis, and the efficacy of the reduction of serum cholesterol on the mortality from ischemic heart disease. In the case of the first meta-analysis, the heterogeneity of the results can be attributed to differences between the studies regarding the severity of underlying disease (cirrhosis), the endoscopic technique used (intervention), the management of complications, and length of follow-up of the patients.

Two strategies for investigating the factors related to heterogeneity are: subgroup analysis and meta-regression. In case of the former, the studies are subdivided into levels for the variable that is believed to be causing the heterogeneity. In the case of endoscopic sclerotherapy, the studies could be analyzed separately according to severity of underlying disease, to form more homogeneous groups. This procedure requires a large number of studies.

Meta-regression is a generalization of the subgroup analysis, which examines the relationship between levels of a characteristic of the study (e.g., duration, dose, disease severity, average age of the group) and the variation in the measure of effect (e.g., risk relative, risk difference, difference of means) of the studies.[7] Its implementation requires one makes use of multivariate models, which is beyond the scope of this article.

FINAL THOUGHTS/CONSIDERATIONS

Although the meta-analysis of clinical trials has reached a high degree of acceptance in the clinical and statistical literature, some authors have been critical about its use in general or, more specifically, when applied to non-experimental studies. A careful reading of these articles reveals that much of the criticism is focused on methodological aspects inherent to the designs of the studies upon which the meta-analysis is constructed, including violations of the basic methodological principles or methodological procedures considered unsuitable for meta-analysis. For example, it is not correct to say that the meta-analysis does not consider the quality of studies or the heterogeneity among their findings, mixing "apples and oranges." The quality is often considered both in the process of the inclusion/exclusion of studies and in the evaluation of their possible impact on the conclusion.

As for heterogeneity, several articles on meta-analysis have drawn attention to the need to seek explanations for the inconsistencies among studies and not calculate summary-measures by combining heterogeneous results. For Liberati,[8] this type of criticism stems from a distorted view that considers meta-analysis a simple statistical combination of data.

All the foregoing does not exempt meta-analysis of a series of problems. Because of the fact that it always done after the data have been collected, it is susceptible to hindsight biases of retrospective research. It is common for meta-analysis on the same subject are different results.

Despite the criticism, meta-analysis has been considered by many authors one of the most important innovations in the methodology of clinical research. More recent movements have incorporated the knowledge produced by systematic review and meta-analysis. This is the case of evidence-based medicine and, more recently, evidence-based public health. It is in this context that Liberati [8,9] reminds the critics of this methodology that the only alternative to systematic reviews is to perform non-systematic reviews, whose subjectivity and lack of well-defined criteria are a breeding ground for conclusions of little practical application, or even wrong .

RESUMO

Na última década, Era da Medicina Baseada em Evidências, o número de metanálises cresceu significativamente. A metanálise combina estatisticamente os resultados de vários estudos e estes são considerados o mais alto nível de evidência quando são combinados de forma apropriada os resultados de ensaios clínicos metodologicamente bem conduzidos. Resultados de uma metanálise podem não corresponder à realidade, pelo fato de depender da qualidade dos estudos nela inseridos, além de diferenças clínicas e metodológicas entre os estudos incluídos. A crescente popularidade de metanálises e de revisões sistemáticas faz com que seja necessário melhor compreendê-las. O objetivo deste artigo é fazer com que o leitor entenda como é realizada uma metanálise/revisão sistemática e que tenha melhores condições de interpretá-la. A melhor compreensão da terminologia adotada e dos conceitos envolvidos na sua produção pode ajudar o clínico a avaliar melhor a qualidade de uma metanálise e a real importância de seus resultados para um paciente específico.

Palavras-chave: Metanálise / Revisão Sistemática Ensaio Clínico.

REFERENCES

1. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. BM J. 1995 Apr 29; 310(6987):1122-6.

2. Olkin I. Statistical and theoretical considerations in meta-analysis. J Clin Epidemiol. 1995 Jan; 48(1):133-46; discussion 47.

3. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet. 1999 Nov 7; 354(9193):1896-900.

4. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Jama. 2000 Apr 19; 283(15):2008-12.

5. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001 Apr 17; 134(8):663-94.

6. Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database Syst Rev. 2006; 3:CD004454.

7. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002 Jun 15; 21(11):1559-73.

8. Liberati A. "Meta-analysis: statistical alchemy for the 21st century": discussion. A plea for a more balanced view of meta-analysis and systematic overviews of the effect of health care interventions. J Clin Epidemiol. 1995 Jan; 48(1):81-6.

9. Coutinho, ESF. Metanálise. In: Roberto Medronho; Diana Mahul de Carvalho, Katia Vergetti Bloch, Ronir Raggio Luiz, Guilherme Werneck. (Org.). Epidemiologia. Rio de Janeiro: Atheneu; 2006. p. 447-56.

Correspondence Address:
MARCO AURELIO PINHO DE OLIVEIRA
Rua Coelho Neto, 55 / 201
Tel.: (21) 9987-5843
E-mail: maurelio@infolink.com.br