

Franco M. Impellizzeri, PhD<sup>1</sup>Mario Bizzini, PT, PhD<sup>1</sup>**ABSTRACT**

The use of an evidence-based approach to practice requires “the integration of best research evidence with clinical expertise and patient values”, where the best evidence can be gathered from randomized controlled trials (RCTs), systematic reviews and meta-analyses. Furthermore, informed decisions in healthcare and the prompt incorporation of new research findings in routine practice necessitate regular reading, evaluation, and integration of the current knowledge from the primary literature on a given topic. However, given the dramatic increase in published studies, such an approach may become too time consuming and therefore impractical, if not impossible. Therefore, systematic reviews and meta-analyses can provide the “best evidence” and an unbiased overview of the body of knowledge on a specific topic. In the present article the authors aim to provide a gentle introduction to readers not familiar with systematic reviews and meta-analyses in order to understand the basic principles and methods behind this type of literature. This article will help practitioners to critically read and interpret systematic reviews and meta-analyses to appropriately apply the available evidence to their clinical practice.

**Key words:** evidence-based practice, meta-analysis, systematic review

**CORRESPONDING AUTHOR**

Mario Bizzini, PT, PhD

FIFA Medical Assessment and Research  
Center, Schulthess Clinic

Lengghalde 2, 8008 Zurich, Switzerland

Phone: + 41 44 385 75 85

Fax: + 41 44 385 75 90

E-mail: [mario.bizzini@f-marc.com](mailto:mario.bizzini@f-marc.com)

<sup>1</sup> Department of Research and Development and FIFA Medical Assessment and Research Centre, Schulthess Clinic, Zurich, Switzerland

**Acknowledgment**

We would like to thank Kirsten Clift for the English revision of the manuscript.

---

## INTRODUCTION

Sacket et al<sup>1,2</sup> defined evidence-based practice as “the integration of best research evidence with clinical expertise and patient values”. The “best evidence” can be gathered by reading randomized controlled trials (RCTs), systematic reviews, and meta-analyses.<sup>2</sup> It should be noted that the “best evidence” (e.g. concerning clinical prognosis, or patient experience) may also come from other types of research designs particularly when dealing with topics that are not possible to investigate with RCTs.<sup>3,4</sup> From the available evidence, it is possible to provide clinical recommendations using different levels of evidence.<sup>5</sup> Although sometimes a matter of debate,<sup>6-8</sup> when properly applied, the evidence-based approach and therefore meta-analyses and systematic reviews (highest level of evidence) can help the decision-making process in different ways:<sup>9</sup>

1. Identifying treatments that are not effective;
2. Summarizing the likely magnitude of benefits of effective treatments;
3. Identifying unanticipated risks of apparently effective treatments;
4. Identifying gaps of knowledge;
5. Auditing the quality of existing randomized controlled trials.

The number of scientific articles published in biomedical areas has dramatically increased in the last several decades. Due to the quest for timely and informed decisions in healthcare and medicine, good clinical practice and prompt integration of new research findings into routine practice, clinicians and practitioners should regularly read new literature and compare it with the existing evidence.<sup>10</sup> However, this is time consuming and therefore is impractical if not impossible for practitioners to continuously read, evaluate, and incorporate the current knowledge from the primary literature sources on a given topic.<sup>11</sup> Furthermore, the reader also needs to be able to interpret both the new and the past body of knowledge in relation to the methodological quality of the studies. This makes it even more difficult to use the scientific literature as reference knowledge for clinical decision-making. For this reason, review articles are important tools available for practitioners to summarize and synthesize the available evidence on a

particular topic,<sup>10</sup> in addition to being an integral part of the evidence-based approach.

International institutions have been created in recent years in an attempt to standardize and update scientific knowledge. The probably best known example is the Cochrane Collaboration, founded in 1993 as an independent, non-profit organisation, now regrouping more than 28,000 contributors worldwide and producing systematic reviews and meta-analyses of healthcare interventions. There are currently over 5000 Cochrane Reviews available (<http://www.cochrane.org>). The methodology used to perform systematic reviews and meta-analyses is crucial. Furthermore, systematic reviews and meta-analyses have limitations that should be acknowledged and considered. Like any other scientific research, a systematic review with or without meta-analysis can be performed in a good or bad way. As a consequence, guidelines have been developed and proposed to reduce the risk of drawing misleading conclusions from poorly conducted literature searches and meta-analyses.<sup>11-18</sup>

In the present article the authors aim to provide an introduction to readers not familiar with systematic reviews and meta-analysis in order to help them understand the basic principles and methods behind this kind of literature. A meta-analysis is not just a statistical tool but qualifies as an actual observational study and hence it must be approached following established research methods involving well-defined steps. This review should also help practitioners to critically and appropriately read and interpret systematic reviews and meta-analyses.

## NARRATIVE VERSUS SYSTEMATIC REVIEWS

Literature reviews can be classified as “narrative” and “systematic” (Table 1). Narrative reviews were the first form of literature overview allowing practitioners to have a quick synopsis on the current state of science in the topic of interest. When written by experts (usually by invitation) narrative reviews are also called “expert reviews”. However, both narrative or expert reviews are based on a subjective selection of publications through which the reviewer qualitatively addresses a question summarizing the findings of previous studies and drawing a conclusion.<sup>15</sup> As such, albeit offering

**Table 1.** Characteristics of narrative and systematic reviews, modified from *Physiotherapy Evidence Database*.<sup>37</sup>

	<b>Systematic review</b>	<b>Narrative review</b>
Research question	Strictly formulated	Broadly formulated
Methodology	Clearly defined	Not or insufficiently described
Search strategy	Clearly defined	Not described
Selection of the studies	Clearly defined	Not described
Ranking of the studies	By levels of evidence	Not performed
Analysis of the studies	Clearly described	Not described
Interpretation of results	Objective	Subjective

interesting information for clinicians, they have an obvious author's bias since not performed by following a clear methodology (i.e. the identification of the literature is not transparent). Indeed, narrative and expert reviews typically use literature to support authors' statements but it is not clear whether these statements are evidence-based or just a personal opinion/experience of the authors. Furthermore, the lack of a specific search strategy increases the risk of failing to identify relevant or key studies on a given topic thus allowing for questions to arise regarding the conclusions made by the authors.<sup>19</sup> Narrative reviews should be considered as opinion pieces or invited commentaries, and therefore they are unreliable sources of information and have a low evidence level.<sup>10,11,19</sup>

By conducting a "systematic review", the flaws of narrative reviews can be limited or overcome. The term "systematic" refers to the strict approach (clear set of rules) used for identifying relevant studies;<sup>11,15</sup> which includes the use of an accurate search strategy in order to identify all studies addressing a specific topic, the establishment of clear inclusion/exclusion criteria and a well-defined methodological analysis of the selected studies. By conducting a properly performed systematic review, the potential bias in identifying the studies is reduced, thus limiting the possibility of the authors to select the studies arbitrarily considered the most "relevant" for supporting their own opinion or research hypotheses. Systematic reviews are considered to provide the highest level of evidence.

### **META-ANALYSIS**

A systematic review can be concluded in a qualitative way by discussing, comparing and tabulating the

results of the various studies, or by statistically analysing the results from independent studies: therefore conducting a meta-analysis. Meta-analysis has been defined by Glass<sup>20</sup> as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings". By combining individual studies it is possible to provide a single and more precise estimate of the treatment effects.<sup>11,21</sup> However, the quantitative synthesis of results from a series of studies is meaningful only if these studies have been identified and collected in a proper and systematic way. Thus, the reason why the systematic review always precedes the meta-analysis and the two methodologies are commonly used together. Ideally, the combination of individual study results to get a single summary estimate is appropriate when the selected studies are targeted to a common goal, have similar clinical populations, and share the same study design. When the studies are thought to be too different (statistically or clinically), some researchers prefer not to calculate summary estimates. Reasons for not presenting the summary estimates are usually related to study heterogeneity aspects such as clinical diversity (e.g. different metrics or outcomes, participant characteristics, different settings, etc.), methodological diversity (different study designs) and statistical heterogeneity.<sup>22</sup> Some methods, however, are available for dealing with these problems in order to combine the study results.<sup>22</sup> Nevertheless, the source of heterogeneity should be always explored using, for example, sensitivity analyses. In this analysis the primary studies are classified in different groups based on methodological and/or clinical characteristics and subsequently compared. Even

---

after this subgroup analysis the studies included in the groups may still be statistically heterogeneous and therefore the calculation of a single estimate may be questionable.<sup>11,19</sup> Statistically heterogeneity can be calculated with different tests but the most popular are the Cochran's  $Q^{23}$  and  $I^{23}$ . Although the latter is thought to be more powerful, it has been shown that their performance is similar<sup>24</sup> and these tests are generally weak (low power). Therefore, their confidence intervals should always be presented in meta-analyses and taken into consideration when interpreting heterogeneity. Although heterogeneity can be seen as a "statistical" problem, it is also an opportunity for obtaining important clinical information about the influences of specific clinical differences.<sup>11</sup> Sometimes, the goal of a meta-analysis is to explore the source of diversity among studies.<sup>15</sup> In this situation the inclusion criteria are purposely allowed to be broader.

### **Meta-analyses of observational studies**

Although meta-analyses usually combine results from RCTs, meta-analyses of epidemiological studies (case-control, cross-sectional or cohort studies) are increasing in the literature, and therefore, guidelines for conducting this type of meta-analysis have been proposed (e.g. Meta-analysis Of Observational Studies in Epidemiology, MOOSE<sup>25</sup>). Although the highest level of evidence study design is the RCT, observational studies are used in situations where RCTs are not possible such as when investigating the potential causes of a rare disease or the prevalence of a condition and other etiological hypotheses.<sup>3,4,11</sup> The two designs, however, usually address different research questions (e.g. efficacy versus effectiveness) and therefore the inclusion of both RCTs and observational studies in meta-analyses would not be appropriate.<sup>11,15</sup> Major problems of observational studies are the lack of a control group, the difficulty controlling for confounding variables, and the high risk of bias.<sup>26</sup> Nevertheless, observational studies and therefore the meta-analyses of observational studies can be useful and are an important step in examining the effectiveness of treatments in healthcare.<sup>3,4,11</sup> For the meta-analyses of observational studies, sensitivity analyses for exploring the source of heterogeneity is often the main aim. To note, meta-analyses themselves can be considered "observational studies of the evidence"<sup>11</sup> and, as a consequence, they may be

influenced by known and unknown confounders similarly to primary type observational studies.

### **Meta-analyses based on individual patient data**

While "traditional" meta-analyses combine aggregate data (average of the study participants such as mean treatment effects, mean age, etc.) for calculating a summary estimate, it is possible (if data are available) to perform meta-analyses using the individual participant data on which the aggregate data are derived.<sup>27-29</sup> Meta-analyses based on individual participant data are increasing.<sup>28</sup> This kind of meta-analysis is considered the most comprehensive and has been regarded as the gold standard for systematic reviews.<sup>29,30</sup> Of course, it is not possible to simply pool together the participants of various studies as if they come from a large, single trial. The analysis must be stratified by study so that the clustering of patients within the studies is retained for preserving the effects of the randomization used in the primary investigations and avoiding artifacts such as the Simpson's paradox, which is a change of direction of the associations.<sup>11,15,28,29</sup> There are several potential advantages of this kind of meta-analysis such as consistent data checking, consistent use of inclusion and exclusion criteria, better methods for dealing with missing data, the possibility of performing the same statistical analyses across studies, and a better examination of the effects of participant-level covariates.<sup>15,31,32</sup> Unfortunately, meta-analyses on individual patient data are often difficult to conduct, time consuming, and it is often not easy to obtain the original data needed for performance of a such an analysis.

### **Cumulative and Bayesian meta-analyses**

Another form of meta-analysis is the so-called "cumulative meta-analysis". Cumulative meta-analyses recognize the cumulative nature of scientific evidence and knowledge.<sup>11</sup> In cumulative meta-analysis a new relevant study on a given topic is added whenever it becomes available. Therefore, a cumulative meta-analysis shows the pattern of evidence over time and can identify the point when a treatment becomes clinically significant.<sup>11,15,33</sup> Cumulative meta-analyses are not updated meta-analyses since there is not a single pooling but the results are summarized as each new study is added.<sup>33</sup> As a consequence, in the

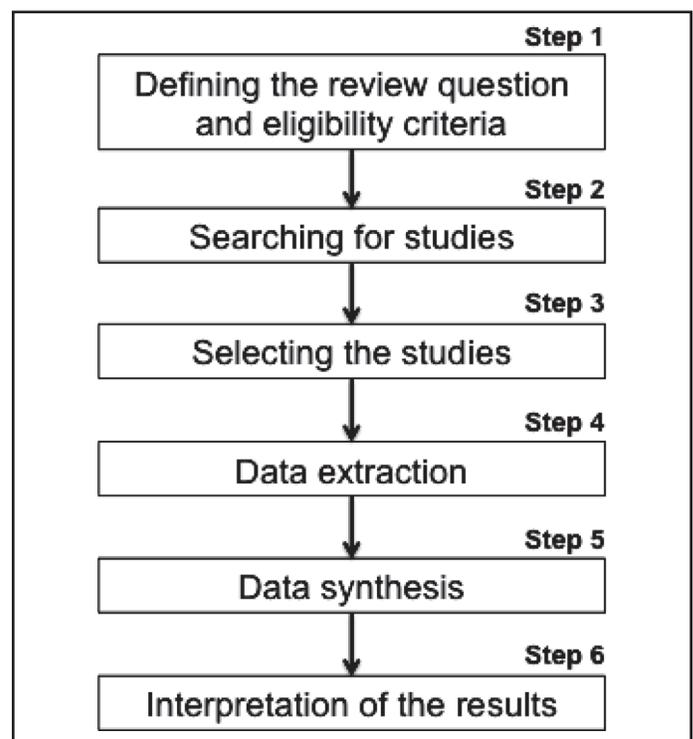
forest plot, commonly used for displaying the effect estimates, the horizontal lines represent the treatment effect estimates as each study is added and not the results of the single studies. The cumulative meta-analysis should be interpreted within the Bayesian framework even if they differ from the “pure” Bayesian approach for meta-analysis.

The Bayesian approach differs from the classical, or frequentist methods to meta-analysis in that data and model parameters are considered to be random quantities and probability is interpreted as an uncertainty rather than a frequency.<sup>11,15,34</sup> Compared to the frequentist methods, the Bayesian approach incorporates prior distributions, that can be specified based on *a priori beliefs* (being unknown random quantities), and the evidence coming from the study is described as a likelihood function.<sup>11,15,34</sup> The combination of prior distribution and likelihood function gives the posterior probability density function.<sup>34</sup> The uncertainty around the posterior effect estimate is defined as a credibility interval, which is the equivalent of the confidence interval in the frequentist approach.<sup>11,15,34</sup> Although Bayesian meta-analyses are increasing, they are still less common than traditional (frequentist) meta-analyses.

### Conducting a systematic review and meta-analysis

As aforementioned, a systematic review must follow well-defined and established methods. One reference source of practical guidelines for properly apply methodological principles when conducting systematic reviews and meta-analyses is the *Cochrane Handbook for Systematic Reviews of Interventions* that is available for free online.<sup>12</sup> However other guidelines and textbooks on systematic reviews and meta-analysis are available.<sup>11,13,14,15</sup> Similarly, authors of reviews should report the results in a transparent and complete way and for this reason an international group of experts developed and published the QUOROM (Quality Of Reporting Of Meta-analyses),<sup>16</sup> and recently the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)<sup>17</sup> guidelines addressing the reporting of systematic reviews and meta-analyses of studies which evaluate healthcare interventions.<sup>17,18</sup>

In this section the authors briefly present the principal steps necessary for conducting a systematic



**Figure 1.** Steps in conducting a systematic review. Modified from<sup>11,14</sup>

review and meta-analysis, derived from available reference guidelines and textbooks in which all the contents (and much more) of the following section can be found.<sup>11,12,14</sup> A summary of the steps is presented in Figure 1. As with any research, the methods are similar to any other study and start with a careful development of the review protocol, which includes the definition of the research question, the collection and analysis of data, and the interpretation of the results. The protocol defines the methods that will be used in the review and should be set out before starting the review in order to avoid bias, and in case of deviation this should be reported and justified in the manuscript.

#### Step 1. Defining the review question and eligibility criteria

The authors should start by formulating a precise research question, which means they should clearly report the objectives of the review and what question they would like to address. If necessary, a broad research question may be divided into more specific questions. According to the PICOS framework,<sup>35,36</sup> the question should define the **P**opulation(s), **I**ntervention(s), **C**omparator(s), **O**utcome(s) and

---

Study design(s). This information will also provide the rationale for the inclusion and exclusion criteria for which a background section explaining the context and the key conceptual issues may be also needed. When using terms that may have different interpretations, operational definitions should be provided. An example may be the term “neuromuscular control” which can be interpreted in different ways by different researchers and practitioners. Furthermore, the inclusion criteria should be precise enough to allow the selection of all the studies relevant for answering the research question. In theory, only the best evidence available should be used for the systematic reviews. Unfortunately, the use of an appropriate design (e.g. RCT) does not ensure the study was well-conducted. However, the use of cut-offs in quality scores as inclusion criteria is not appropriate given their subjective nature, and a sensitivity analysis comparing all available studies based on some methodological key characteristics is preferable.

### **Step 2. Searching for studies**

The search strategy must be clearly stated and should allow the identification of all the relevant studies. The search strategy is usually based on the PICOS elements and can be conducted using electronic databases, reading the reference lists of relevant studies, hand-searching journals and conference proceedings, contacting authors, experts in the field and manufacturers, for example.

Currently, it is possible to easily search the literature using electronic databases. However, the use of only one database does not ensure that all the relevant studies will be found and therefore various databases should be searched. The Physiotherapy Evidence Database (PEDro: <http://www.pedro.org.au>) provides free access to RCTs (about 18,000) and systematic reviews (almost 4000) on musculoskeletal and orthopaedic physiotherapy (sports being represented by more than 60%). Other available electronic databases are MEDLINE (through PubMed), EMBASE, SCOPUS, CINAHL, Web of Science of the Thomson Reuters and The Cochrane Controlled Trials Register. The necessity of using different databases is justified by the fact that, for example, 1800 journals indexed in MEDLINE are not indexed in EMBASE, and vice versa.

The creation and selection of appropriate keywords and search term lists is important to find the relevant literature, ensuring that the search will be highly sensitive without compromising precision. Therefore, the development of the search strategy is not easy and should be developed carefully taking into consideration the differences between databases and search interfaces. Although Boolean searching (e.g. AND, OR, NOT) and proximity operators (e.g. NEAR, NEXT) are usually available, every database interface has its own search syntax (e.g. different truncation and wildcards) and a different thesaurus for indexing (e.g. MeSH for MEDLINE and Emtree for EMBASE). Filters already developed for specific topics are also available. For example, PEDro has filters included in search strategies (called SDIs) that are used regularly and automatically in some of the above mentioned databases for retrieving guidelines, RCTs, and systematic reviews.<sup>37</sup>

After performing the literature search using electronic databases, however, other search strategies should be adopted such as browsing the reference lists of primary and secondary literature and hand searching journals not indexed. Internet sources such as specialized websites can be also used for retrieving grey literature (e.g. unpublished papers, reports, conference proceedings, thesis or any other publications produced by governments, institutions, associations, universities, etc.). Attempts may be also performed for finding, if any, unpublished studies in order to reduce the risk of publication bias (trend to publish positive results or results going in the same direction). Similarly, the selection of only English-language studies may exacerbate the bias, since authors may tend to publish more positive findings in international journals and more negative results in local journals. Unpublished and non-English studies generally have lower quality and their inclusion may also introduce a bias. There is no rule for deciding whether to include or not include unpublished or exclusively English-language studies. The authors are usually invited to think about the influence of these decisions on the findings and/or explore the effects of their inclusion with a sensitivity analysis.

### **Step 3. Selecting the studies**

The selection of the studies should be conducted by more than one reviewer as this process is quite

---

subjective (the agreement, using kappa statistic, between reviewers should be reported together with the reasons for disagreements). Before selecting the studies, the results of the different searches are merged using reference management software and duplicates deleted. After an initial screening of titles and abstracts where the obviously irrelevant studies are removed, the full papers of potentially relevant studies should be retrieved and are selected based on the previously defined inclusion and exclusion criteria. In case of disagreements, a consensus should be reached by discussion or with the help of a third reviewer. Direct contact with the author(s) of the study may also help in clarifying a decision.

An important phase at this step is the assessment of quality. The use of quality scores for weighting the study entered in the meta-analysis is not recommended, as it is not recommended to include in a meta-analysis only studies above a cut-off quality score. However, the quality criteria of the studies must be considered when interpreting the results of a meta-analysis. This can be done qualitatively or quantitatively through subgroup and sensitivity analyses based on important methodological aspects, which can be assessed using checklists that are preferable over quality scores. If quality scores would like to be used for weighting, alternative statistical techniques have been proposed. e.g.<sup>38</sup> The assessment of quality should be performed by two independent observers. The Cochrane handbook, however, makes a distinction between study quality and risk of bias (related for example to the method used to generate random allocation, concealment, blindness, etc.), focusing more on the latter. As for quality assessment, the risk of bias should be taken into consideration when interpreting the findings of the meta-analysis. The quality of a study is generally assessed based on the information reported in the studies thus linking the quality of reporting to the quality of the research itself, which is not necessarily true. Furthermore, a study conducted at the highest possible standard may still have high risk of bias. In both cases, however, it is important that the authors of primary studies appropriately report the results and for this reason guidelines have been created for improving the quality of reporting such as the CONSORT (Consolidated Standards of Reporting Trials<sup>39</sup>) and the STROBE (Strengthening the Report-

ing of Observational Studies in Epidemiology<sup>40</sup>) statements.

#### **Step 4. Data extraction**

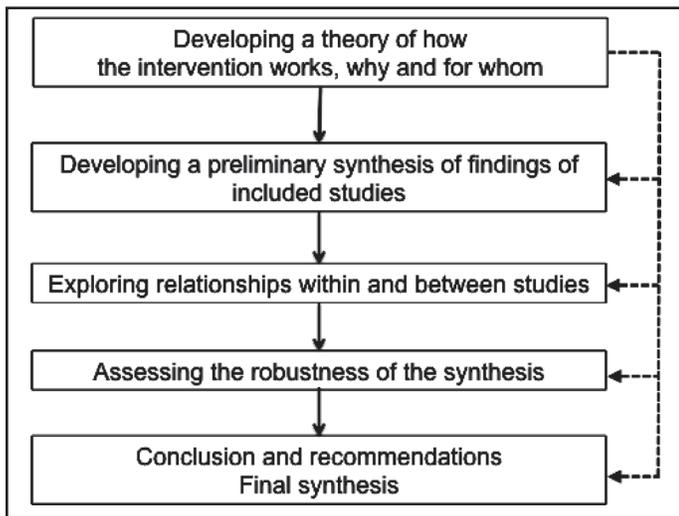
Data extraction must be accurate and unbiased and therefore, to reduce possible errors, it should be performed by at least two researchers. Standardized data extraction forms should be created, tested, and if necessary modified before implementation. The extraction forms should be designed taking into consideration the research question and the planned analyses. Information extracted can include general information (author, title, type of publication, country of origin, etc.), study characteristics (e.g. aims of the study, design, randomization techniques, etc.), participant characteristics (e.g. age, gender, etc.), intervention and setting, outcome data and results (e.g. statistical techniques, measurement tool, number of follow up, number of participants enrolled, allocated, and included in the analysis, results of the study such as odds ratio, risk ratio, mean difference and confidence intervals, etc.). Disagreements should be noted and resolved by discussing and reaching a consensus. If needed, a third researcher can be involved to resolve the disagreement.

#### **Step 5. Analysis and presentation of the results (data synthesis)**

Once the data are extracted, they are combined, analyzed, and presented. This data synthesis can be done quantitatively using statistical techniques (meta-analysis), or qualitatively using a narrative approach when pooling is not believed to be appropriate. Irrespective of the approach (quantitative or qualitative), the synthesis should start with a descriptive summary (in tabular form) of the included studies. This table usually includes details on study type, interventions, sample sizes, participant characteristics, outcomes, for example. The quality assessment or the risk of bias should also be reported. For narrative reviews a comprehensive synthesis framework (Figure 2) has been proposed.<sup>14,41</sup>

#### **Standardization of outcomes**

To allow comparison between studies the results of the studies should be expressed in a standardized format such as effect sizes. The appropriate effect size for standardizing the outcomes should be



**Figure 2.** Narrative synthesis framework. Modified from<sup>14,41</sup>

similar between studies so that they can be compared and it can be calculated from the data available in the original articles. Furthermore, it should be interpretable. When the outcomes of the primary studies are reported as means and standard deviations, the effect size can be the raw (unstandardized) difference in means (D), the standardized difference in means (d or g) or the response ratio (R). If the results are reported in the studies as binary outcomes the effect sizes can be the risk ratio (RR), the odds ratio (OR) or the risk difference (RD).<sup>15</sup>

### Statistical analysis

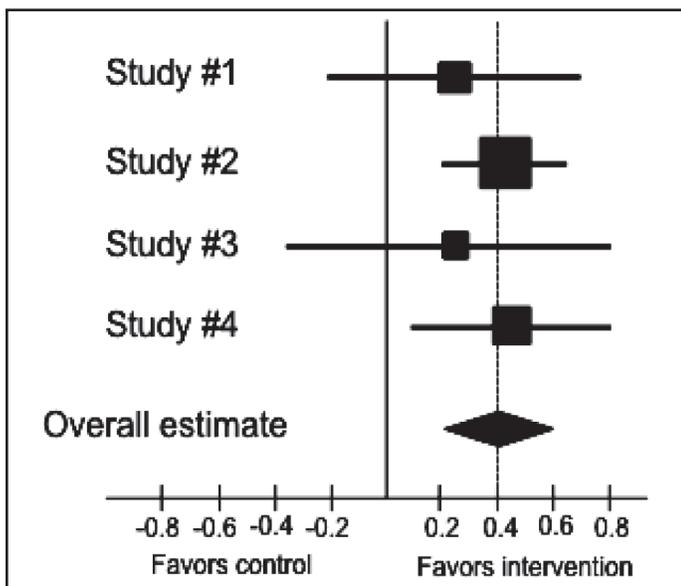
When a quantitative approach is chosen, meta-analytical techniques are used. Textbooks and courses are available for learning statistical meta-analytical techniques. Once a summary statistic is calculated for each study, a “pooled” effect estimate of the interventions is determined as the weighting average of individual study estimates, so that the larger studies have more “weight” than the small studies. This is necessary because small studies are more affected by the role of chance.<sup>11,15</sup> The two main statistical models used for combining the results are the “fixed-effect” and the “random-effects” model. Under the fixed effect model, it is assumed that the variability between studies is only due to random variation because there is only one true (common) effect. In other words, it is assumed that the group of studies give an estimate of the same treatment effect and therefore the effects are part of the same distribution. A common method for weighting each study is

the inverse-variance method, where the weight is given by the inverse of variance of each estimate. Therefore, the two essential data required for this calculation are the estimate of the effect with its standard error. On the other hand, the “random-effects” model assumes a different underlying effect for each study (the true effect varies from study to study). Therefore the study weight will take into account two sources of error: the between- and within-studies variance. As in the fixed-effect model, the weight is calculated using the inverse-variance method, but in random-effects model the study specific standard errors are adjusted incorporating both within and between-studies variance. For this reason, the confidence intervals obtained with random-effect models are usually wider. In theory, the fixed-effect model can be applied when the studies are heterogeneous while the random-effects model can be applied when the results are not heterogeneous. However, the statistical tests for examining heterogeneity lack power and, as aforementioned, the heterogeneity should be carefully scrutinized (e.g. interpreting the confidence intervals) before taking a decision. Sometimes, both fixed- and random-effects models are used for examining the robustness of the analysis. Once the analyses are completed, results should be presented as point estimates with the corresponding confidence intervals and exact *p*-values.

Other than the calculations of the individual studies and summary estimates, other analyses are necessary. As mentioned various time, the exploration of possible source of heterogeneity is important and can be performed using sensitivity, subgroup, or regression analyses. Using meta-regressions is also possible to examine the effects of differences in study characteristics on the treatment effect estimate. When using meta-regression, the larger studies have more influence than smaller studies; and regarding other analyses, recall that the limitations should be taken into account before deciding to use it and when interpreting the results.

### Graphic display

The results of each trial are commonly displayed with their corresponding confidence intervals in the so-called “forest plot” (Figure 3). In the forest plot the study is represented by a square and a horizontal line indicating the confidence interval, where the



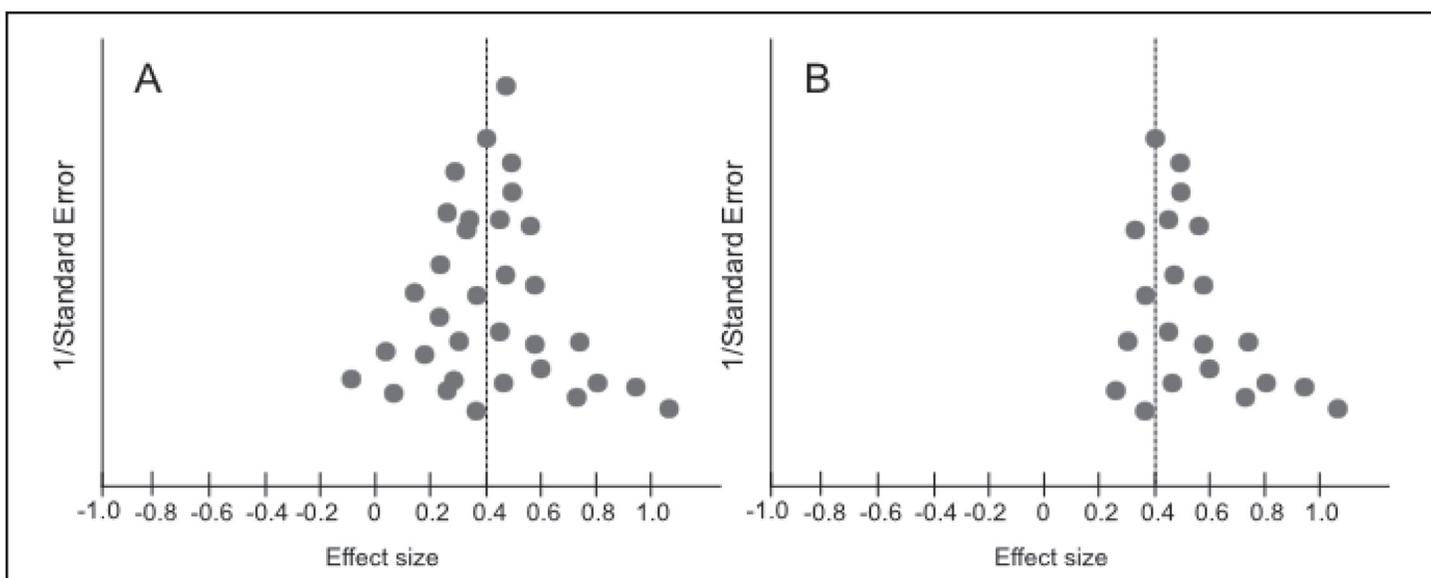
**Figure 3.** Example of a forest plot: the squares represent the effect estimate of the individual studies and the horizontal lines indicate the confidence interval; the dimension of the square reflects the weight of each study. The diamond represent the summary point estimate is usually represented with a diamond at the bottom of the graph with the horizontal extremities indicating the confidence interval. In the example as standardized outcome measure the authors used *d*.

dimension of the square reflects the weight of each study. A solid vertical line usually corresponds to no effect of treatment. The summary point estimate is usually represented with a diamond at the bottom of the graph with the horizontal extremities indicating the confidence interval. This graphic solution gives an immediate overview of the results.

An alternated graphic solution called a funnel plot can be used for investigating the effects of small studies and for identifying publication bias (Figure 4). The funnel plot is a scatter-plot of the effect estimates of individual studies against measures of study size and precision (commonly, the standard error, but the use of sample size is still common). If there is no publication bias the funnel plot will be symmetrical (Figure 4B). However, the funnel plot examination is subjective, based upon visual inspection, and therefore can be unreliable. In addition, other causes may influence the symmetry of the funnel plot such as the measures used for estimating the effects and precision, and differences between small and large studies.<sup>14</sup> Therefore, its use and interpretation should be done with caution.

### Step 6. Interpretation of the results

The final part of the process pertains to the interpretation of the results. When interpreting or commenting on the findings, the limitations should be discussed and taken into account, such as the overall risk of bias and the specific biases of the studies included in the systematic review, and the strength of the evidence. Furthermore, the interpretation should be performed based not solely using *P*-values, but rather on the uncertainty and the clinical/practical importance. Ideally, the interpretation should help the clinician in understanding how to apply the findings in practice, provide recommendations or implications for policies, and offer directions for further research.



**Figure 4.** Example of symmetric (A) and asymmetric (B) funnel plots.

---

## CONCLUSIONS

Systematic reviews have to meet high methodological standards, and their results should be translated into clinically relevant information. These studies offer a valuable and useful summary of the current scientific evidence on a specific topic and can be used for developing evidence-based guidelines. However, it is important that practitioners are able to understand the basic principles behind the reviews and are hence able to appreciate their methodological quality before using them as a source of knowledge. Furthermore, there are no RCTs, systematic reviews, or meta-analyses that address all aspects of the wide variety of clinical situations. A typical example in sports physiotherapy is that most available studies deal with recreational athletes, while an individual clinician may work with high-profile or elite athletes in the clinic. Therefore, when applying the results of a systematic review to clinical situations and individual patients, there are various aspects one should consider such as the applicability of the findings to the individual patient, the feasibility in a particular setting, the benefit-risk ratio, and the patient's values and preferences.<sup>1</sup> As reported in the definition, evidence-based medicine is the integration of both research evidence **and** clinical expertise. As such, the experience of the sports PT should help in contextualizing and applying the findings of a systematic review or meta-analysis, and adjusting the effects to the individual patient. As an example, an elite athlete is often more motivated and compliant in rehabilitation, and may have a better outcome than average with the given physical therapy or training interventions (when compared to a recreational athlete). Therefore, it is essential to merge the available evidence with the clinical evaluation and the patient's wishes (and consequent treatment planning) in order to engage an evidence-based management of the patient or athlete.

## REFERENCES

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS: Evidence based medicine: what it is and what it isn't. *BMJ*. 1996; 312(7023): 71-72.
2. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB: *Evidence-Based Medicine. How to practice and Teach*. EBM (2nd ed). London: Churchill Livingstone, 2000.
3. Black N: What observational studies can offer decision makers. *Horm Res*. 1999; 51 Suppl 1: 44-49.
4. Black N: Why we need observational studies to evaluate the effectiveness of health care. *BMJ*. 1996; 312(7040): 1215-1218.
5. US Preventive Services Task Force: *Guide to clinical preventive services, 2nd ed*. Baltimore, MD: Williams & Wilkins, 1996.
6. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F: Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997; 337(8): 536-542.
7. 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; 48(1): 81-86.
8. Bailar JC, 3rd: The promise and problems of meta-analysis. *N Engl J Med*. 1997; 337(8): 559-561.
9. Von Korff M: The role of meta-analysis in medical decision making. *Spine J*. 2003; 3(5): 329-330.
10. Tonelli M, Hackam D, Garg AX: Primer on systematic review and meta-analysis. *Methods Mol Biol*. 2009; 473: 217-233.
11. Egger M, Daey Smith G, Altman D: *Systematic Reviews in Health Care: Meta-Analysis in Context, 2nd Edition*: BMJ Books, 2001.
12. Higgins JPT, Green S: *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011). The Cochrane Collaboration (available from: <http://www.cochrane-handbook.org/>), 2008.
13. Atkins D, Fink K, Slutsky J: Better information for better health care: the Evidence-based Practice Center program and the Agency for Healthcare Research and Quality. *Ann Intern Med*. 2005; 142(12 Pt 2): 1035-1041.
14. Centre for Reviews and Dissemination: *Systematic reviews: CRD's 16 guidance for undertaking reviews in health care*. York: University of York, 2009.
15. Borenstein M, Hedges LV, Higgins JPT, H.R. R: *Introduction to Meta-Analysis*: John Wiley & Sons Ltd, 2009.
16. Clarke M: The QUORUM statement. *Lancet*. 2000; 355(9205): 756-757.
17. Moher D, Liberati A, Tetzlaff J, Altman DG: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009; 339: b2535.
18. Liberati A, Altman DG, Tetzlaff J, et al.: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009; 339: b2700.

- 
19. Sauerland S, Seiler CM: Role of systematic reviews and meta-analysis in evidence-based medicine. *World J Surg.* 2005; 29(5): 582-587.
  20. Glass GV: Primary, secondary and meta-analysis of research. *Educ Res.* 1976; 5: 3-8.
  21. Berman NG, Parker RA: Meta-analysis: neither quick nor easy. *BMC Med Res Methodol.* 2002; 2: 10.
  22. Ioannidis JP, Patsopoulos NA, Rothstein HR: Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ.* 2008; 336(7658): 1413-1415.
  23. Higgins JP, Thompson SG: Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002; 21(11): 1539-1558.
  24. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J: Assessing heterogeneity in meta-analysis: Q statistic or I<sup>2</sup> index? *Psychol Methods.* 2006; 11(2): 193-206.
  25. Stroup DF, Berlin JA, Morton SC, et al.: Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA.* 2000; 283(15): 2008-2012.
  26. Kelsey JL, Whittemore AS, Evans AS, Thompson WD: *Methods in observational epidemiology.* New York: Oxford University Press, 1996.
  27. Duchateau L, Pignon JP, Bijmens L, Bertin S, Bourhis J, Sylvester R: Individual patient-versus literature-based meta-analysis of survival data: time to event and event rate at a particular time can make a difference, an example based on head and neck cancer. *Control Clin Trials.* 2001; 22(5): 538-547.
  28. Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR: Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med.* 2008; 27(29): 6111-6136.
  29. Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG: Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials.* 2005; 2(3): 209-217.
  30. Oxman AD, Clarke MJ, Stewart LA: From science to practice. Meta-analyses using individual patient data are needed. *JAMA.* 1995; 274(10): 845-846.
  31. Stewart LA, Tierney JF: To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof.* 2002; 25(1): 76-97.
  32. Riley RD, Lambert PC, Staessen JA, et al.: Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med.* 2008; 27(11): 1870-1893.
  33. Lau J, Schmid CH, Chalmers TC: Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *J Clin Epidemiol.* 1995; 48(1): 45-57; discussion 59-60.
  34. Sutton AJ, Abrams KR: Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res.* 2001; 10(4): 277-303.
  35. Armstrong EC: The well-built clinical question: the key to finding the best evidence efficiently. *Wmj.* 1999; 98(2): 25-28.
  36. Richardson WS, Wilson MC, Nishikawa J, Hayward RS: The well-built clinical question: a key to evidence-based decisions. *ACP J Club.* 1995; 123(3): A12-13.
  37. Physiotherapy Evidence Database (PEDro): Physiotherapy Evidence Database (PEDro). *J Med Libr Assoc.* 2006; 94(4): 477-478.
  38. Greenland S, O'Rourke K: On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics.* 2001; 2(4): 463-471.
  39. Moher D, Schulz KF, Altman DG: The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet.* 2001; 357(9263): 1191-1194.
  40. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007; 370(9596): 1453-1457.
  41. Popay J, Roberts H, Sowden A, et al.: Developing guidance on the conduct of narrative synthesis in systematic reviews. *J Epidemiol Comm Health.* 2005; 59(Suppl 1): A7.