How to perform a systematic review and/or a meta-analysis?

¿Cómo realizar una revisión sistemática o un metaanálisis?

A. Zarazaga Monzón
General Surgery and Emergencies Service. University Hospital La Paz. Madrid

Abstract
Evidence based medicine practice aims to apply the best clinical scientific evidence available for decision taking in individualized patient care. But frequently, achieved responses are disappointing or excellent evidences cannot be applied to specific patients, forcing the clinician to act, adopting a Bayesian attitude, based on his own experience. There are continuously improved and more complex methodological tools for the research of bibliographic resources, but their use is more specialized and difficult. The logical attitude would be to leave them in the hands of methodologists, while clinicians’ would collaborate in the project applying their own experience. First, identifying the problem and formulating specific goals by means of the structured clinical questions and then, after having received the responses, evaluating the quality and feasibility, and elaborating appropriate recommendations. In this paper, addressed to clinicians, the developing of a consensus meeting, in which they may play a greater role, is emphasized. Without forgetting secondary roles, but irreplaceable, of clinicians to develop in earlier stages of whatever systematic reviews.

Keywords: meta-analysis, systematic reviews, consensus meetings.

Introduction
This article address to general practitioners—and not methodologists—, insists in the importance of the application of the “clinical competence” as the essential part in the development of the systematic reviews. It does not pretend, neither it can, to go in depth in the methodological process, therefore a panoramic approach is carried out leading the interested reader, based on some of the innumerable sources that are widely documented and that exist both in printing and digital format. The practice of the evidence-based medicine (EBM) consists of the conscious, explicit and judicious use of the best available clinical scientific evidence in order to take decisions about the care of special patients. However, we forget frequently during its application about the importance of the application in clinical experience. The practice of the EBM requires, by definition, the integration of both concepts: clinical “experience” and the best available external clinical evidence. The different components are defined below as some terms are distorted occasionally.

Randomized and controlled clinical trial
It is a prospective study that tries to compare the effect and the value of one or more interventions versus control conducted in patients with a medical condition. In the current medical practice, a randomized and controlled clinical trial that faces a therapeutic with another one is the accepted way through which the treatment usefulness is judged. It constitutes the best scientific test in order to support the efficacy of the therapeutic interventions. In order to evaluate the quality of the randomized and controlled clinical trials (RCCT), there are defined criteria as the
scale of Jadad,\textsuperscript{1} that assesses the randomization and masking of the trial. Some consensed\textsuperscript{2} and updated\textsuperscript{3,4} categorizations are also used that point out the different levels of the trials scientific evidence and indicate the applicable strength to the consistent recommendations regarding to this quality.

An incorrect presentation of the RCCT, even with good results, might entail mistaken and confusing interpretations and complicate the comparison of the trials. With the aim of avoiding these problems, different standards have been developed for the publication of the results. An international group was created in the 90’s the Consolidated Standards of Reporting Trials Group (CONSORT),\textsuperscript{6} made up of biostatisticians, clinical investigators and publishers of medical journals;\textsuperscript{6} this group issued methodological recommendations for the performance of RCCT reports.\textsuperscript{6-8} The Declaration CONSORT is made up of 22 points and a flow chart, facilitating the later interpretation and critical evaluation (table 1).

When a RCCT does not achieve a sufficient statistical power so to draw relevant conclusions, it is essential to choose other biostatistical tools as the meta-analyses (MA) and the systematic reviews (SR), which try to integrate independent trials.

**Meta-analysis**

It is a statistical analysis that combines or integrates the results of different independent clinical trials, that analyst considers “combinable”.\textsuperscript{9-10} Therefore, it consists in the statistical integration of independent studies.\textsuperscript{11} It is the essential tool in the search and integration of evidence, though it has also detractors. Nevertheless, the MA presents the inconvenience that it requires the integration of randomized clinical trials for its preparation, and they are not always available for all the questions that we want to study in a review. Therefore, occasionally, due to the excessive heterogeneity of the results, its low methodological quality or the lack of valid and/or relevant data, it cannot and might not be used methodologically in a meta-analysis.

**Systematic reviews**

They search answers to one or several questions with systematic and explicit methods, in order to identify, select and evaluate the relevant investigation critically and to then collect and analyze the study data included in the review (collaboration Cochrane).\textsuperscript{12} They might also refer to any bibliographic review that uses clearly defined criteria and methods, prepared by means of strategies that avoid the turns and that includes a section of material and methods. A SR might or not include a meta-analysis and therefore they might be a component of the SR; however, it constitutes a broader investigation process as in front of determined questions or the impossibility of performing a quantitative aggregation of the data, non-randomized trials shall incorporate themselves to the analysis, which will constitute the better existing evidence in this case. In order to assess the quality of the non-randomized trials and their presentation of results, a list of specific verification has been developed: the Transparent Reporting of Evaluations with Nonrandomized Designs Statement (TREND).\textsuperscript{13-14} It is a complement of the CONSORT,\textsuperscript{6} which was prepared by the RCCT as we have seen above.

The reviews offers us the possibility of being informed, without being necessary the spending of excessive time. One of the most efficient manners of finding the scientific evidence is to search a systematic review of the RCCT, as it provides reliable results about which we can draw conclusions and take decisions.\textsuperscript{15} The SR offer us clear advantages: a) they present less turns in the identification and selection of studies; b) the conclusions are more reliable and precise, c) the assimilation of a great quantity of information might be faster; d) they reduce the time among the findings and their clinical application; e) they might compare results of several studies in order to determine their applicability and concordance; f) they allow identifying the reasons of the heterogeneity, being able to formulate new hypothesis and g) the quantitative SR (meta-analysis) increase the precision of the global results.

For its credibility it comes together with the correct performance, which requires:

**A clearly defined protocol**

1. To specify the problem that is being evaluated and to define the investigation questions previously,
2. To specify the inclusion (and exclusion) criteria of the studies.
3. To formulate a bibliographic search plan.
4. To select and recover the studies that complies with the inclusion criteria.
5. Critical assessment of the scientific quality of these original studies.
6. Integration or combination of the findings in a systematic manner.
7. Formulate the conclusions and/or recommendations based on the evidence quality and the clinical experience.

**A correct information analysis**

There are different options for the obtaining of data in the SR:
- The clinical trial and/or meta-analysis model.
- The qualitative structured review.
- The Bayesian model of confidence profile.
- The collection model and statistical combination.

The preferential model will be the RCCT and/or the meta-analysis that is spite of being a complex technique and the statistical methodology that is used is difficult or at least specialized, the results that it offers are the most objective and strongest ones. However, when the meta-analysis seems to us inappropriate, we can resort to the qualitative structured review (QSR). Unlike the other, it has a simple methodology; it is reasonably reproducible and is usable at the absence of RCCT, which might be used for any clinical matter. It requires a thorough methodology, simple but strict, and draws up qualitative conclusions subject to subjective opinions.\textsuperscript{16,17} The protocol of the QSR is similar to the model systematic review with some methodological variant.
Table 1. List of CONSORT verification for the drawing up of RCCT

<table>
<thead>
<tr>
<th>Section and subject</th>
<th>Point #</th>
<th>Description</th>
<th>Informed on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title and summary</td>
<td>1</td>
<td>How the participants are assigned to the interventions (random assignment, randomization or assigned randomly)</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
<td>Scientific history, explanation and reasoning</td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>3</td>
<td>Screening criteria of the participants, as well as the devices and services where the data have been collected</td>
<td></td>
</tr>
<tr>
<td>Interventions</td>
<td>4</td>
<td>Specify details of the interventions for each group, and also when and how they have really been administered</td>
<td></td>
</tr>
<tr>
<td>Objectives</td>
<td>5</td>
<td>Specify the objectives and the hypothesis</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>6</td>
<td>Define clearly the primary and secondary measurements of the results and, when necessary, to describe any method used to increase the quality of the measurements (e.g. multiple observations, previous training of the observers and consultancies)</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>7</td>
<td>How was the sample size determined and if applicable, add the explanation of any intermediate analysis and the ending rules</td>
<td></td>
</tr>
<tr>
<td>Randomization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generation of the sequence</td>
<td>8</td>
<td>Method used to generate the randomized assignment sequence, including the details of any restriction (e.g. blocks or stratification)</td>
<td></td>
</tr>
<tr>
<td>Blind distribution</td>
<td>9</td>
<td>Method used to implement the randomized assignment sequence (e.g. numbered containers or central telephone directory), clarifying if the sequence was masked until the interventions were assigned</td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td>10</td>
<td>Who generated the assignment sequence, who screened the patients and who assigned the participants to the groups</td>
<td></td>
</tr>
<tr>
<td>Blind (masking)</td>
<td>11</td>
<td>If the participants, those who administered the intervention and who evaluated the results were blind to the assignment of the groups. If so, how was the success evaluated as regards to the blind process (masking)</td>
<td></td>
</tr>
<tr>
<td>Statistical methods</td>
<td>12</td>
<td>Statistical methods used to compare the groups in their primary results; additional analysis methods, such as subgroups analysis or adjusted analysis</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flow of participants</td>
<td>13</td>
<td>Flow of participants through each phase (the use of the chart is earnestly recommended). To specify in each group which is the number of the randomly assigned participants, who received the assigned treatment, completing the study protocol analyzed for the primary results. To describe the deviations to the designed study protocol, together with the reasons</td>
<td></td>
</tr>
<tr>
<td>Recruitment</td>
<td>14</td>
<td>Dates that define the recruitment and follow-up periods</td>
<td></td>
</tr>
<tr>
<td>Baseline data</td>
<td>15</td>
<td>Baseline demographic data and clinical characteristics of each group</td>
<td></td>
</tr>
<tr>
<td>Analyzed numbers</td>
<td>16</td>
<td>Number of participants (denominator) in each group included in each analysis, and to specify if it was: “for intention-to-treat” analysis. To determine the results in absolute numbers when feasible (e.g. 10/20, not 50%)</td>
<td></td>
</tr>
<tr>
<td>Results and estimation</td>
<td>17</td>
<td>For each primary and secondary result, a summary of results of each group and the estimated effect of the sample and its precision (e.g. confidence interval of 95%)</td>
<td></td>
</tr>
<tr>
<td>Auxiliary analysis</td>
<td>18</td>
<td>Add multiplicity informing about any analyzed analysis, included analysis of subgroups and adjusted analysis, besides the pre-specified and exploratory analysis</td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td>19</td>
<td>All the important adverse events or side effects of each intervention group</td>
<td></td>
</tr>
<tr>
<td>Comments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpretation</td>
<td>20</td>
<td>Interpretation of the results taking into account the study hypothesis, the sources of potential turn or imprecision and the risks associated to the multiplicity of analysis and results</td>
<td></td>
</tr>
<tr>
<td>External validity</td>
<td>21</td>
<td>External validity of the study findings</td>
<td></td>
</tr>
<tr>
<td>Global evidence</td>
<td>22</td>
<td>General interpretation of the results in the context of the current evidence</td>
<td></td>
</tr>
</tbody>
</table>

RCCT: randomized and controlled clinical trial.
As from the bibliographic review, struc-
tset of guidelines
methodological point of view (Oxman)
investigation of drugs from this comment.
of decisions and investigation of results), though we exclude the
retain relevant data, applicable to the daily clinical practice (taking
considering the modern multidisciplinary medicine, we con-
eration with the aim of answering to those questions to which a
summarized in tables of evidence with some qualitative consid-
statistical determination, but it also evaluates other publications,
uses the first one, where we count with RCCT and MA with sta-
the SR, both from the methodological point of view (Oxman) and the drawing-up.
The Quality of Reporting of Meta-Analyses Group (QuEChERS) developed a
of the SR and the performed MA. An alternative method, pro-
posed by the Critical Appraisal Skills Programme, of the English
Health Service, helps to undertake a critical reading of the SR. There is a group named CASP in Spain (Critical Appraisal Skills Programme), from which we can download information. There are specific tools, as the QUADAS Statement, which present a process diagram that has to be followed in order to assess and develop the SR on studies of diagnostic precision.

Computer tools
In order to facilitate the storage and use of data, there are computer tools that speed up the revision and analysis procedure. Some of them are of free availability, as the “RevMan” (Review Manager). It is a software used by the Cochrane Collaboration in order to prepare and keep the SR and is provided by the Information Management System (IMS) of the Nordic Cochrane Center. The GRADE Pro is another software developed by The Grading of Recommendations Assessment, Development and Evaluation Working Group. It provides a strict approach of the quality qualification of clinical evidence in the preparation of clinical practice guidelines.

The correct performance of a SR is not only a matter of good will and “great desire” but it requires the use of specific and sophisticated tools, a strict methodology, have broad statistical and information knowledge and time to carry it out. But ignore now this line of reasoning and let us follow another direction. We still count with a SR and some MA correctly performed by methodological experts; frequently we, the general practitioners, do not find the replies to our doubts. Even versus a SR we can draw different conclusions and take different decisions. We count with the information, but we have not generated the knowledge. Therefore, the external clinical evidences provided by the reviews might make up, but never replace, the clinical competence, and this latter is the one that finally has to decide if the evidences are applicable to the patient, in our environment, and how they should be integrated in a clinical decision. We, the general practitioners should achieve this final touch though “the consensus”.

Consensus meeting (CM)
The consensus methods combine the set of evidences obtained as from a strict and structured bibliographic review with the opinion of a panel of experts. The main objective is to obtain a sort of agreement in those areas of uncertainty where the absence of controlled and conclusive clinical trials makes recommendable a qualitative approach to the problem solution. There are two types of consensus meetings:

- **Informal consensus.** Based almost exclusively in the opinion of experts. There is not a screening about the scientific quality of the provided bibliography and it is not systemized. It only provides recommendations without scientific evidence level and provides some information about the procedure in which these latter are sustained.

- **Formal consensus.** As from the bibliographic review, structured by independent advisors (methodologists), a panel of ex-

The methodology is based on the order of the questions appearance, ordering the possible alternatives of replies supported by the bibliography, and indicating the magnitude of the evidence, according to quality and reliability that has been detected. Then it is essential to define the doubtful aspect for each question, to identify the level of uncertainty of each reply, according to the obtained evidence level and the aspects that need a later investigation.

Another usable model, though more extended, is the “Bayesian model” of confidence profile. The Bayesian methods with a different interpretation of probability concept constitute an alternative to the traditional statistics focused on the contrast of the hypothesis. They incorporate external information to the study, and with it and the observed data, a probability distribution is estimated for the magnitude —effect— that is being investigated. Though there are “enthusiasts” that sustain that there is an increasing interest for the use of Bayesian techniques when inferring and taking decisions, the truth is that the reality of the medical publications does not reflect, not less, such pretended situation, being only a minority who publish results analyzed from a Bayesian point of view; moreover, there are not too many readers with education to understand them. However, it might be an adequate tool in appropriate hands and in the correct environment.

Finally, there is a mixed model (open to the three others), the “collection model and statistical combination”. It combines and uses the first one, where we count with RCCT and MA with statistical determination, but it also evaluates other publications, classifying them according to the quality range. The data are summarized in tables of evidence with some qualitative consideration with the aim of answering to those questions to which a quantitative methodology cannot be applied. Personally, and taking into account the modern multidisciplinary medicine, we consider it as the most useful and flexible procedure in order to obtain relevant data, applicable to the daily clinical practice (taking of decisions and investigation of results), though we exclude the investigation of drugs from this comment.

There are criteria to assess the quality of the SR, both from the methodological point of view (Oxman) and the drawing-up. There are two types of consensus meetings:
experts discusses about a previously prepared questionnaire about a determined medical practice and submits recommendations. At the end, a final document is prepared with evidence and strength level scores about the recommendations and is disclosed as a clinical guidelines in some cases.

The evidence about the efficiency of a determined technique in the treatment of a group of sick persons with a determined pathology has to be looked for in a CM as well as the publication of final recommendations (according to the evidence levels and the experience of the experts). The consensus combines the external evidence (based on tests) and the consensus competence of the experts (experience and reflection). The consensus decisions might be approved unanimously, or expressing the conflicting opinion separately, when they exist.

In this type of consensus, the level of recommendation shall depend on the two parameters:
1. From the level of evidence of the consulted sources.
2. From other provided and consensed concepts by the panel of experts.
   - Personal experience.
   - Applicability in the environment.
   - Obtained results: cost and / or own procedure risk versus benefit.

The sources used in the CM are the ones obtained through a structured or systematic review. We observe the development of a CM in figure 1. Some questions are set out, and the bibliography search is performed with search strategies and previously well-defined inclusion/exclusion criteria. The bibliography is selected and collected according to different levels of evidence. Each question, if the obtained evidence is sufficient to reply the search, shall pass to the panel of clinical experts which shall do the recommendations. If not, another bibliography shall be searched to reply to the set out question at a lower evidence level, though keeping a minimum acceptable level, in order to generate consistent recommendations.

The performance of a CM needs a laborious planning. Figure 2 shows an organization chart model, with the preparation phases: financing, subject choice, creation of an Organizing Committee, constitution of different groups or committees (that shall include experts of all the specialties taking part in the analyzed process), as well as the later stages on the meeting itself.

**Conclusions**

The available and used tools are many and varied for the “search of evidence” and it is the competence of the methodologist to know and apply them adequately. The general practitioner has an important part of the process: to define the objectives (problems or doubts that have to be solved) by means of a well-formulated “structured clinical question” that includes their three components: groups of patients, intervention/s and results. The participation of the clinical practitioner is essential in this first phase. “A close answer to a well-formulated problem is much more valuable than an exact reply to a close problem”.

---

**Figure 1. Methodological process that has to be followed in a systematic review**
With this information, the methodologists shall search in different sources, some of them unreachable and/or unknown for the clinical practitioner and shall choose the acceptable bibliography that replies to all or some of our questions, bibliography that should be classified per levels of scientific quality. Considering the results (replies), and being these results classified per evidence levels, the clinical practitioner recovers protagonism again and together with other experts of all the specialties taking part in the analyzed process, shall issue conclusions and shall set out recommendations, after analyzing the accepted tests and his own experience. We would have applied the recommended resources for the practice of the EBM in this process: to use the best external evidence together with the consensus skill of recognized clinical practitioners (experts).

Finally, for those who want to start in the review methodology, they should be addressed to *The Cochrane Collaboration*. There is a wide basic information about the SR in their web server, and its free access learning material, together with the *Cochrane Manual* describe the creation process of a SR in detail.

**Declaration of potential conflict of interests**

The author took part as expert in five consensus multidisciplinary systematic reviews, between 2000 and 2004, which were published in specialized journals. Moreover, he led and took part in Doctorate Courses and Continuous Medical Education related to the development subject in the article. As the data of previous publications have not been incorporated, but his personal experience, there are no conflicts of interest.

**References**


