

Focus on: Contemporary Methods in Biostatistics (III)

Systematic Reviews and Meta-Analysis: Scientific Rationale and Interpretation

Ignacio Ferreira González,^{a,b,*} Gerard Urrútia,^{b,c} and Pablo Alonso-Coello^{b,c}

^aUnidad de Epidemiología, Servicio de Cardiología, Área del Cor, Hospital Vall d'Hebron, Barcelona, Spain

^bCIBER de Epidemiología y Salud Pública (CIBERESP), Spain

^cCentro Cochrane Iberoamericano-Servei d'Epidemiologia Clínica i Salut Pública, Institut d'Investigació Biomèdica Sant Pau, Barcelona, Spain

Article history:

Available online 30 June 2011

Keywords:

Systematic review

Meta-analysis

Evidence-based medicine

ABSTRACT

Systematic reviews represent a specific type of medical research in which the units of analysis are the original primary studies. They are essential tools in synthesizing available scientific information, increasing the validity of the conclusions of primary studies, and identifying areas for future research. They are also indispensable for the practice of evidence-based medicine and the medical decision-making process. However, conducting high quality systematic reviews is not easy and they can sometimes be difficult to interpret. This special article presents the rationale for carrying out and interpreting systematic reviews and uses a hypothetical example to draw attention to key-points.

© 2011 Sociedad Española de Cardiología. Published by Elsevier España, S.L. All rights reserved.

Revisiones sistemáticas y metaanálisis: bases conceptuales e interpretación

RESUMEN

Las revisiones sistemáticas son investigaciones científicas en las cuales la unidad de análisis son los estudios originales primarios. Constituyen una herramienta esencial para sintetizar la información científica disponible, incrementar la validez de las conclusiones de estudios individuales e identificar áreas de incertidumbre donde sea necesario realizar investigación. Además, son imprescindibles para la práctica de una medicina basada en la evidencia y una herramienta fundamental en la toma de decisiones médicas. Sin embargo, la realización de una revisión sistemática de calidad no es una tarea sencilla, como en ocasiones tampoco lo es su interpretación. En este artículo especial se presentan las bases conceptuales para la realización y la interpretación de revisiones sistemáticas, poniendo especial énfasis en los puntos clave durante su ejecución mediante un ejemplo hipotético.

© 2011 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Palabras clave:

Revisión sistemática

Metaanálisis

Medicina basada en la evidencia

Abbreviations

MA: meta-analysis

RCT: randomized clinical trial

SR: systematic review

INTRODUCTION

It has been a tiring week. You sit down to look back calmly on the decisions you've made. These included requesting an operation for a patient with three-vessel disease, deciding whether to treat an 82 year old patient with inferior infarction of 70 min duration with fibrinolytic therapy or move her to your center for primary angioplasty, and deciding on anticoagulation treatment for an outpatient with atrial fibrillation.

* Corresponding author: Unidad de Epidemiología, Servicio de Cardiología, Hospital Vall d'Hebron, Pg. Vall d'Hebron 119-129, 08035 Barcelona, Spain.

E-mail address: nacho@ferreiragonzalez.com (I. Ferreira González).

Although you are reasonably sure your decisions were based on the best available evidence, you have some lingering doubts. Perhaps studies have been published which could lead your decisions to be questioned? Or perhaps different studies of the same intervention have produced different results? It's true that you have not had much time for reading over the past few months. To quickly clear up your doubts, you realize you need a concise, current, and rigorous summary of the best available evidence regarding the decisions you had to take. In other words, you need a systematic review (SR).¹

SRs are considered to be the most reliable source in informing medical decision-making,² which may explain their increasing popularity and the large rise in the number of SRs published in recent years.² However, performing a high-quality SR is not easy. There are rules governing the way they should be carried out and, as with other designs, recommendations on how results should be presented. These quality control guidelines have been developed by international, multidisciplinary groups of experts which include authors of SRs, methodologists, clinicians, and editors.²⁻⁴ This article presents the underlying rationale for performing and interpreting SRs and uses a hypothetical example to highlight key-points in their execution.

CONCEPT AND NOMENCLATURE

SRs are scientific investigations in which the unit of analysis is the original primary studies. These are used to answer a clearly formulated question of interest using a systematic and explicit process. For that reason, SRs are considered to be secondary research (“research-based research”). On the other hand, reviews which do not follow a systematic process (narrative reviews) cannot be considered to constitute a formal research process, but are simply a type of scientific literature based primarily on opinion.

From a formal point of view, SRs summarize the results of primary research using strategies to limit bias and random error.⁵ These strategies include:

- Systematic and exhaustive searching for all potentially relevant articles.
- The use of explicit and reproducible criteria to select articles which are eventually included in the review.¹
- Describing the design and implementation of the original studies, synthesizing the data, and interpreting the results.

Although SRs are a tool for synthesizing information, it is not always possible to present the results of the primary studies briefly. When results are not combined statistically, the SR is called a qualitative review. In contrast, a quantitative SR, or meta-analysis (MA) is an SR which uses statistical methods to combine the results of two or more studies.¹

An SR should not to be confused with an MA. The first is always possible, while the second is only sometimes possible. However, when conditions allow, MAs provide very useful, manageable information regarding the effect of a treatment or intervention, both in general and in specific patient groups. In addition, MAs make it possible to estimate the effect of an intervention more precisely and to detect moderate but clinically important effects that may have gone undetected in the primary studies. Typically, MAs combine aggregate data from published studies, but sometimes individual data from patients in different studies can be combined. This is called individual patient data meta-analysis and is considered the gold standard in SR.⁶

It should be noted that, in contrast to narrative reviews, SRs use a systematic method to search for all potentially relevant studies and apply explicit, reproducible, previously defined criteria to select the articles included in the final review. It is these features which give SRs their scientific character, in contrast to narrative reviews. Table 1 shows the difference between the two types of review.

As in clinical trials, a protocol should be developed prior to carrying out an SR.⁷ This will help the researchers to give due consideration to the most appropriate methods for use in the review and will also prevent decisions being taken *a posteriori* based on the results. The first international register of protocols for systematic reviews, apart from the Cochrane SRs, was recently

published under the name of PROSPERO (<http://www.crd.york.ac.uk/prospere/>).

STAGES IN A SYSTEMATIC REVIEW

Briefly, a SR consists of the following steps:

- Definition of the clinical question of interest and the inclusion and exclusion criteria for studies.
- Identification and selection of relevant studies.
- Extraction of data from primary studies.
- Analysis and presentation of results.
- Interpretation of results.

Definition of the Clinical Question of Interest

The first step is to correctly formulate the clinical question of interest. In general, this should be explicit and structured so as to include the following key components:⁸

- The specific population and context. For example, elderly patients (over 75 years) admitted for acute myocardial infarction with ST elevation.
- The exposure of interest. This could be a risk factor, a prognostic factor, an intervention or treatment, or a diagnostic test. In the case of an intervention, treatment or diagnostic test a control exposure is usually defined at the same time. For example, primary angioplasty (intervention) versus fibrinolysis (control).
- Events of interest. For example, total mortality, cardiovascular mortality, readmission for acute coronary syndrome, revascularizations, etc.

From these elements, you might frame the question as follows: compared with fibrinolysis, does primary angioplasty reduce mortality and myocardial infarction in patients over 75 years of age? Once the question of interest has been defined and circumscribed, it is easier to establish the inclusion and exclusion criteria for primary studies. An ill-defined research question, on the other hand, leads to confused decision-making about which studies may be relevant in answering the question.

In many cases not easy to decide what the specific research question should be. It is however clear that it should be clinically relevant. If questions are too vague (e.g. is primary angioplasty useful in acute myocardial infarction?), they will be of little help to the clinician when making a decision about a particular patient. Exposures or patient characteristics which may affect the event of interest should also be taken into account. For example, it is not uncommon for patients over 75 years of age to be treated with oral anticoagulation, which could affect the expected event of interest. The study population could therefore be restricted to patients who are not receiving oral anticoagulation. However, overly specific inclusion criteria may limit the applicability of the results. Another

Table 1
Differences Between Systematic and Narrative Reviews.

Characteristic	Narrative review	Systematic review
Question of interest	Not structured, not specific	Structured question, well-defined clinical problem
Article search and sources	Not detailed and not systematic	Structured and explicit search
Selection of articles of interest	Not detailed and not reproducible	Selection based on explicit criteria uniformly applied to all articles
Assessing the quality of the information	Absent	Structured and explicit
Synthesis	Often a qualitative summary	Qualitative and quantitative summary
Inferences	Sometimes evidence-based	Normally evidence-based

option is to define a broad question that makes clinical sense and go on from there to explore more specific questions. For example, we could include all patients with acute myocardial infarction and then use exploratory analysis to determine the effect of the intervention in those treated with oral anticoagulants and those who are not. However, this strategy can lead to problems which are similar to those found in subgroup analysis.⁹ Finally, using inclusion criteria which are too broad means running the risk that the analysis will make no real clinical or biological sense.¹⁰

At this stage, it is important to decide which study designs to include in the review. The decision depends on the type of question we are trying to answer. If the idea is to evaluate the efficacy of an intervention, as in the previous example, we clearly need to include randomized clinical trials (RCT), if any are available. The same is true for the assessment of the reliability and safety of a diagnostic test. For SRs of community or public health interventions or when interventions are assessed over the long-term (particularly as regards their safety), then observational studies are more relevant. Occasionally, we may find that there are no RCTs available on a given intervention and, in that case as well, we will need to analyze observational studies.

Figure 1 shows, in simplified form, the elements of the question of interest in our example and the inclusion and exclusion criteria for the SR deriving from it.

Identifying and Selecting Studies

This stage consists of several parts:

1. Identifying potential articles
 - Deciding on restrictions regarding the publication language.
 - Deciding on the sources for obtaining primary studies.
 - Obtaining the titles and abstracts of potential primary studies.
2. Selecting potential articles
 - Applying the inclusion and exclusion criteria to the titles and abstracts obtained.
 - Obtaining potential articles based on eligible titles and abstracts and applying the inclusion and exclusion criteria.
 - Evaluating the level of agreement in study selection.

Identifying Potential Articles

As with any research study, the presence of errors in data extraction can invalidate the results of a SR. So as to minimize random error and bias, it is extremely important to obtain as many

primary studies as possible on the question of interest. If studies are omitted, it can introduce bias if the final sample of selected articles is not representative. Two decisions need to be made at this point: firstly, whether to restrict the language of publication and whether to include studies which are not published in medical journals.

With regard to language, for practical purposes only publications in English and the native language of the author of the SR are usually included. Nevertheless, studies have shown that research quality is not necessarily related to the language of publication.¹¹ Restricting the language of publication may also bias the results of the SR by excluding potentially relevant studies.¹²

For reasons of practicality, it seems reasonable to include only studies published in medical journals. One could argue that the fact that these studies have been through a peer review process makes them the most reliable.¹³ However, independently of quality, it is also true that studies with negative or inconclusive results are less likely to be published,^{14,15} and their exclusion may bias the results of the SR. This is called publication bias and it makes it more likely that a SR which excludes unpublished studies will overestimate the relationship between exposure and the event of interest. In extreme situations, this bias can lead to completely futile treatments being presented as effective.^{16,17}

Once a decision has been taken on the preceding two questions, the following step is crucial: where to search for primary studies? Various strategies are available:

- Electronic databases: MEDLINE, EMBASE, CENTRAL.
- Non-indexed databases: AMED, CINAHL, BIOSIS, etc.
- Hand-searching of journal content tables, minutes and summaries of scientific meetings, and books.
- Lists of references and citations: Science Citation Index and similar.
- Records of ongoing studies (e.g., clinicaltrials.gov).
- Contact with pharmaceutical companies.
- Contact with fellow experts in the field of interest.

For obvious reasons, the most widely used strategy today is to perform the search in electronic databases. This is not, however, a simple strategy. Although there is some overlap between databases, many journals included in one database are not included in others. MEDLINE, for example, only indexes about 5600 of the more than 16 000 existing biomedical journals, and most of those are in English. EMBASE indexes over 1000 journals which are not included in MEDLINE, many of them European. Other databases further complement MEDLINE and EMBASE, as efforts have been made to register literature on unpublished studies (gray literature).¹⁸

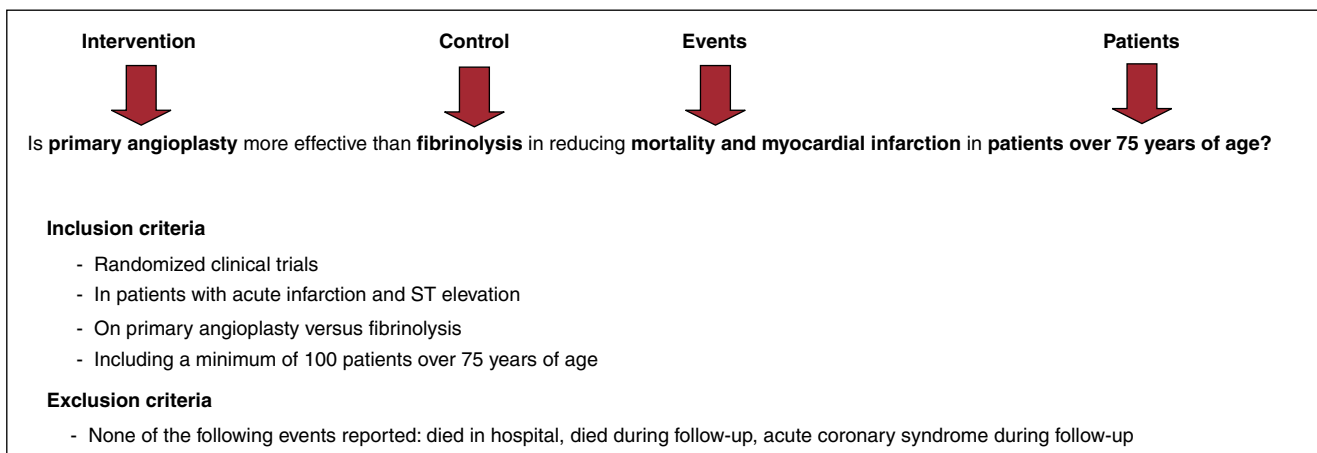


Figure 1. Example of the elements making up the question of interest for a systematic review, and inclusion and exclusion criteria for the primary studies.

For practical reasons, the identification of potentially eligible articles is often undertaken by identifying titles and abstracts. However, each database has a particular structure and uses more or less specific criteria for indexing and keywords. For example, in MEDLINE, the specific index of controlled vocabulary is called MeSH (Medical Subject Headings). The assistance of a documentalist who is expert in SRs is crucial at this stage. The final search strategy should also be described in the publication of the SR, to verify its reproducibility.

Depending on the topic of interest, restricting the search for studies exclusively to electronic databases may not be optimal. Sometimes it is convenient to include a complementary strategy to identify unpublished studies. This phase is one of the most laborious and often requires hand searching of journal abstracts or conference proceedings, contact with experts in the field or with pharmaceutical companies, etc. Clearly, this involves a significant investment of time and financial resources. Recognizing this, the Cochrane Collaboration has led an international initiative to develop a register of controlled trials, formerly known as the Cochrane Controlled Trials Register, and now called CENTRAL.¹⁹ This resource, which is constantly updated, contains hundreds of thousands of records or citations of studies published in indexed journals as well as supplements (usually conference proceedings) and is of undoubted interest in identifying RCTs.²⁰

Applying search strategies in several electronic databases usually generates a large number of references many of which are likely to be duplicated between databases. It is therefore useful at this stage to use software for the automated management of bibliographic citations, such as ProCite or Reference Manager.

Selection of Potential Articles

Once we have a first list of titles and abstracts, we need to make an initial selection or screening of potentially eligible articles. A study selection form based on explicit, understandable criteria is useful at this stage.²¹ A simple example for the research question referred to above is shown in Figure 2. The selection process usually begins with a review of titles and abstracts; if there are doubts about the suitability of an article, the full text should be reviewed.

In order to increase the reliability and thoroughness of the process, study selection should be carried out by two independent reviewers. It is equally important to measure the degree of agreement between reviewers by calculating the kappa statistic for each of the items on the selection form. Simply put, this statistic measures the degree of agreement between reviewers above that expected by chance.^{22,23} Where reviewers disagree about whether to include an article or not, a third, senior researcher is usually appointed to act as arbitrator and to make the final decision.

Finally, it is important to note that the entire process of identifying and selecting studies should be accurately reported using a flow diagram of articles identified in each phase, articles eliminated, and the reasons for their elimination (Fig. 3).

Extracting Data From Primary Studies

In this phase, information should be extracted from each study as reliably as possible and should therefore be done in duplicate if

<p>Study reference: _____</p> <p>Selection level (check all that apply):</p> <p>Title ___ Abstract ___ Text ___</p> <p>Selection criteria:</p> <p>Patients</p> <p>Are patients with acute infarction and elevated ST studied? Yes/No</p> <p>Are patients over 75 years of age included? Yes/No</p> <p>Intervention</p> <p>Did at least one of the treatment groups receive primary angioplasty? Yes/No</p> <p>Control</p> <p>Did at least one of the treatment groups receive fibrinolysis? Yes/No</p> <p>Events</p> <p>Are any of the following events included as individual or combined outcome variables: overall mortality, cardiovascular mortality, non-fatal myocardial infarction, new revascularizations? Yes/No</p> <p>Design</p> <p>Is it a randomized clinical trial? Yes/No</p> <p>ACTION (ONLY include if the answer to ALL of the above was 'yes'):</p> <p>Include ___ Exclude ___ Doubtful ___</p>
--

Figure 2. Hypothetical example of a form for selecting potential articles for a systematic review.

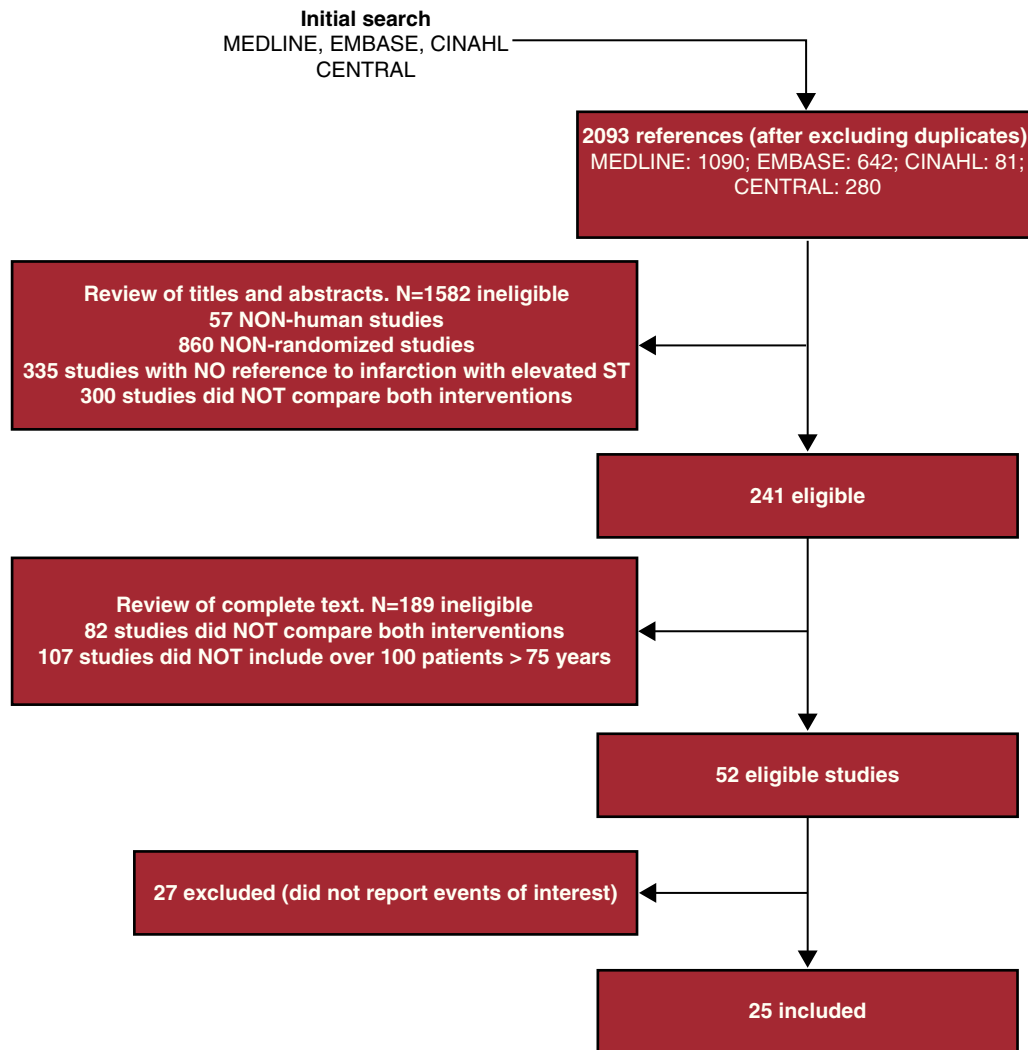


Figure 3. Example of the article selection process.

possible. If that is not feasible, an alternative is to have another reviewer carry out an independent audit of a random sample of studies.

The information to be extracted from the primary studies should be agreed on during the design phase of the study. In general, it should be data that allows the study hypothesis to be accepted or rejected. In summary, the data extraction form usually includes:²⁴

- A. Information on patients, the intervention of interest, the control intervention, and study design.
- B. Information on the results.
- C. Information on the methodological quality of the study.

Point A covers all information that may impact the outcome, and may vary between studies. In our example, an RCT which included patients with a mean age of 65 years and a median door-to-balloon time of 63 min would not be the same as another trial in which patient mean age was 75 years and door-to-balloon time was 96 min. All of these differences could explain a difference in the magnitude of the intervention effect between studies, i.e. they could help to explain the heterogeneity of the effect. The challenge is to strike a balance between the completeness of the information collected while, at the same time, avoiding

including too much unnecessary information that would overload the review.

The information in B corresponds to the extraction of results. The format chosen will depend on how the events of interest are defined: whether as a dichotomous variable or as a continuous variable. In the first case, it will sometimes be easy to get the number and percentage of patients in both branches of the study in which the event of interest occurred. On other occasions, the results may be expressed as a measure of association or impact, such as relative risk, relative risk reduction, absolute risk reduction, odds ratio or hazard ratio. Figure 4 shows an example of a hypothetical form for the extraction of results. Fortunately, it is easy to get the number and approximate percentage of patients with the event of interest from any of the usual measures of association or impact. In the event that the outcome variable is continuous, such as the ejection fraction, the information of interest is the mean and standard deviation in each treatment group.

Finally, the data extraction form should include information on the methodological quality of the studies included, as this is closely related to the magnitude of the effect. There is controversy regarding the best way to reflect the methodological quality of a study.²⁵ Some argue for the use of quality rating scales and several of these have been developed,²⁶ mostly for RCTs. Some are generic

Study reference: _____

	Primary angioplasty Total no randomized: _____ Total no analyzed: _____		Fibrinolysis Total no randomized: _____ Total no analyzed: _____		Measures of association or impact			
	N	%	N	%	RR (95% CI)	RRR (95% CI)	ARR (95% CI)	HR (95% CI)
Total mortality								
CV mortality								
Readmissions for ACS								
New revascularizations								

Figure 4. Hypothetical example of a form for extracting results from studies included in a systematic review. ACS, acute coronary syndrome; ARR, absolute risk reduction; CI, confidence interval; CV, cardiovascular; HR, hazard ratio; RR, relative risk; RRR, relative risk reduction.

and others specific to certain clinical areas. However, it has been shown that using one scale or another can lead to substantial variation in the results of an MA,²⁷ so that none of them are totally reliable. Recently, a new system to measure the methodological quality of studies included in SRs has been developed called GRADE (Grading of Recommendations Assessment, Development, and Evaluation).²⁸ This system, which was developed and agreed upon by a group of international leaders in the implementation of clinical practice guidelines, offers some interesting advantages over the others. Basically:

- The quality of the evidence, which is classified as high, moderate, low and very low, is reported separately from the grade of recommendation (strong or weak recommendation).
- The values and preferences of patients are recognized and incorporated.
- It provides a clear and pragmatic level of recommendation (strong or weak) for clinicians, patients and managers.
- It explicitly assesses the importance to patients of the outcome variables for the therapeutic alternatives considered

Some authors advocate collecting and evaluating the methodological elements of each individual study, instead of using the controversial scales.²⁹ The information collected will depend on the study design. In the case of RCTs, which are the most common design in reviews of interventions, aspects of design and implementation most frequently related to the risk of bias are: the way randomization is concealed, the method of masking interventions, tracking losses, the type of analysis, or early termination of the trial because of an apparent benefit. **Figure 5** shows a hypothetical example of a data collection form for methodological elements of RCTs.

Analysis and Presentation of Results

Summarizing and presenting the results of primary studies obtained using a systematic, reproducible methodology constitutes, in itself, a qualitative SR. The next step is to combine the results of primary studies using statistical methods, i.e., meta-analysis (MA).

Conceptually, MAs are used to combine the results of two or more similar studies on a particular intervention, as long as the same outcome variables are used. Meta-analysis does not provide a simple arithmetic average of the results of different studies, but a weighted average. In other words, it gives greater weight to studies with a larger information load, i.e. studies that are larger and / or

have a higher number of events. On combining the results, a different weight is assigned to each study and it is that which provides the weighted average. Because MA also takes into account intra- and inter-study variability when combining results, the validity of the conclusions is further enhanced. If there is a large amount of variation between the results of the studies included (heterogeneity) it may not be appropriate to combine the results statistically. In that case, only the results and the characteristics of individual studies should be presented, without further statistical treatment. It can be useful to present the results graphically and the study characteristics in tables.

Basically, two types of models can be used to statistically combine results, those being fixed effects models and random effects models. The first assumes that the treatment effect is constant in all studies, while the second assumes that the effect is randomly distributed among the different studies. In other words, the fixed effects model assumes that there is only one source of variability in the results (the study), while the random effects model introduces a second source of variation between studies. The practical consequence of this is that the random effects model tends to produce more conservative estimates (wider confidence intervals) of the combined effect. Which model we use depends on our view of similarities and differences between the studies to be merged, although it is common to use both approaches.

But what do we mean when we say that there is “heterogeneity” between studies, and how is that heterogeneity measured? Basically, it means that, after weighting the studies, the results of individual studies (the effect of the intervention) differ more than would be expected by chance. In other words, the effect of the intervention “differed” between studies either because of differences in the study design used, in the methods used to collect information, or the type of analysis used, and/or the characteristics of the study population. For example, suppose that one clinical trial of an antihypertensive drug included 70% African Americans and another study, using the same drug, included only 10%. Now imagine that the drug has a very powerful antihypertensive effect in African American populations and no effect in the rest of the population. Clearly, the first study will show the intervention has a positive effect while the second study will show it has no effect. It would not be appropriate to combine the results of the two studies because it would show a “mean” effect for the intervention which would hide a much richer and complex reality, and would lead to confusion.

Various statistical tests are available to quantify heterogeneity. The most common are the Q, H and I² statistics. The easiest to interpret is the I², which indicates the proportion of observed

<p>1. Allocation concealment of random assignment:</p> <ul style="list-style-type: none"> • Central randomization (via telephone call, Web page, independent investigator or through the pharmacy) • Coded containers with the study and control medication • Sequentially-numbered, opaque, sealed envelopes • Envelopes used, but without specifying the three above-mentioned elements • Pseudo-randomization (coin toss, birth-date, day of the week, etc.) • Masking of randomization mentioned but not described • Not reported • Lack of randomization masking is reported <p>2. Masking or blinding of the intervention (check that which applies):</p> <p>Patients</p> <ul style="list-style-type: none"> • Health care personnel administering the medication • Investigators collecting the data • Investigators evaluating events • Investigators carrying out the statistical analysis <p>3. Participants lost to follow-up:</p> <p>N=_____ %=_____</p> <ul style="list-style-type: none"> • Not reported <p>4. Data analysis</p> <p>Intention to treat (patients are analyzed in the intervention arm to which they were randomized)</p> <ul style="list-style-type: none"> • Per protocol (patients are analyzed according to the intervention they eventually received) • Both • Not reported <p>5. The study was interrupted early because of a benefit</p> <ul style="list-style-type: none"> • Yes • No

Figure 5. Hypothetical example of a form to assess the quality of randomized clinical trials included in a systematic review.

(between-study) variability in the intervention effect that is due to heterogeneity between studies, i.e. non-random variability. It is generally held that a proportion of 25% indicates that there is little heterogeneity, 50% shows moderate heterogeneity, and 75% a high degree of heterogeneity.³⁰

The results of MAs are presented graphically using forest plots. This type of graph shows data from individual studies together with a representation of the statistical weight of each study in relation to confidence intervals and standard error of the mean. For example, suppose that in our example we combine eight studies to analyze the effect on cardiovascular mortality of angioplasty compared to fibrinolysis in patients aged over 75 years. Figure 6A presents the combined analysis of eight studies in the form of a forest plot. It can be seen that the overall effect is inconclusive and, moreover, that there is significant heterogeneity between studies ($I^2 = 90\%$).

A simplistic interpretation would be that primary angioplasty is not superior to fibrinolysis in reducing cardiovascular death in patients over 75 years of age. However, the high heterogeneity between studies should make us suspect that this result is unrealistic. In fact, visually it can be seen that angioplasty showed a superior effect to fibrinolysis in studies 2, 4, 5 and 6, but was

inferior in studies 1, 3, 7 and 8. A careful reading of these surveys indicates that mean door-to-balloon time was <90 min in all studies in which angioplasty showed a superior effect, but that it was inferior to fibrinolysis in all studies with a door-to-balloon time over 140 mins. In this case it would be justifiable to perform a subgroup analysis in which studies with a door-to-balloon time <90 min and those with longer door-to-balloon times are analyzed separately. Figure 6B shows the results. It can be seen that heterogeneity is significantly reduced in each subgroup and that the results are consistent, with primary angioplasty showing a significant benefit for studies with lower door-to-balloon times (bottom of Fig. 6B) but a smaller effect than fibrinolysis in studies with longer door-to-balloon times (top of fig. 6B).

Finally, it is important to point out that guidelines and recommendations for the presentation and publication of SR are available and provide checklists of items to be covered in the publication. The MOOSE guidelines detail the specific items that SRs of observational studies should include,⁴ while the PRISMA guidelines (which replaced the QUORUM recommendations) cover MAs of RCTs.³ They can be consulted on-line at <http://www.consort-statement.org>

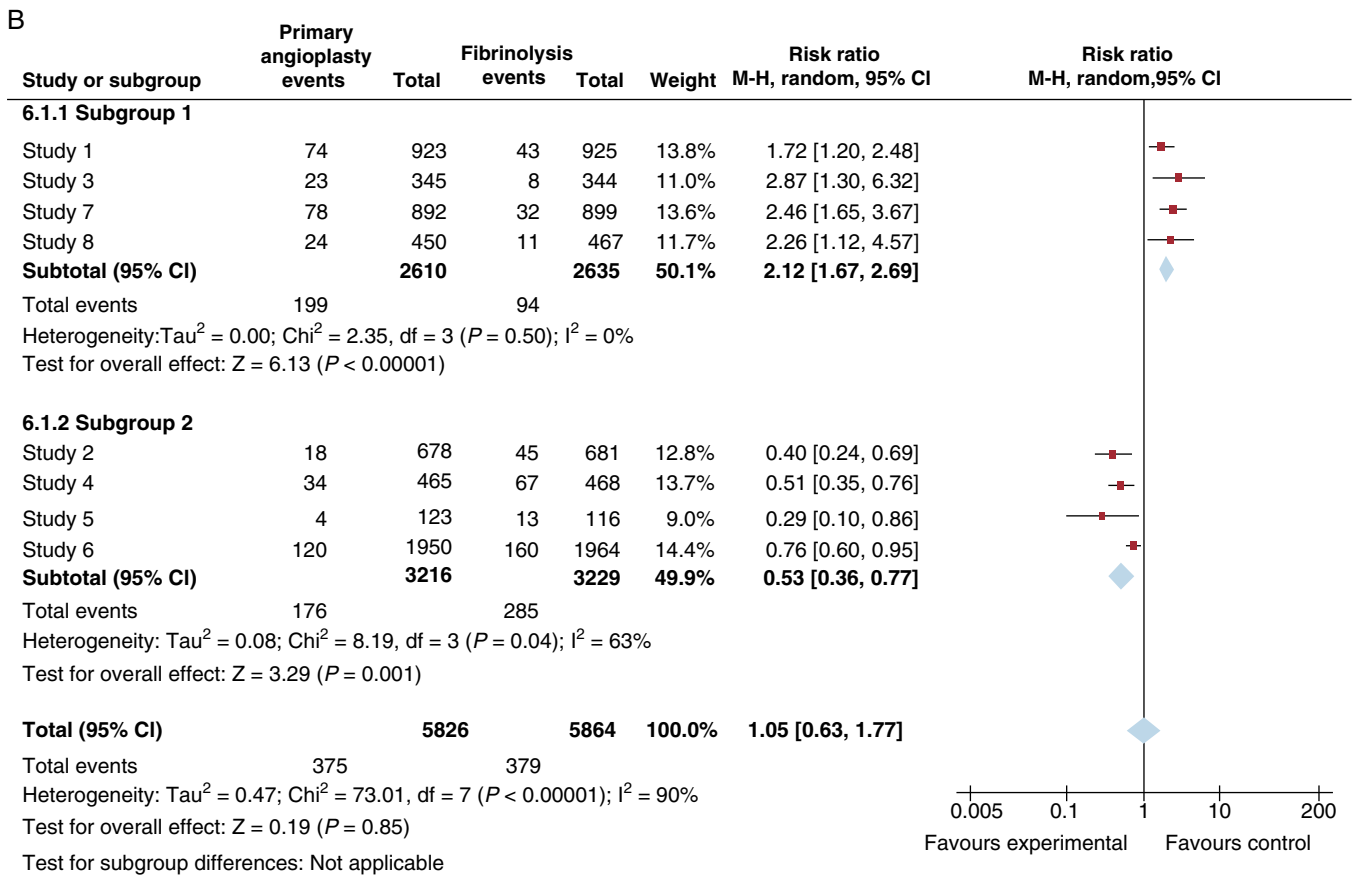
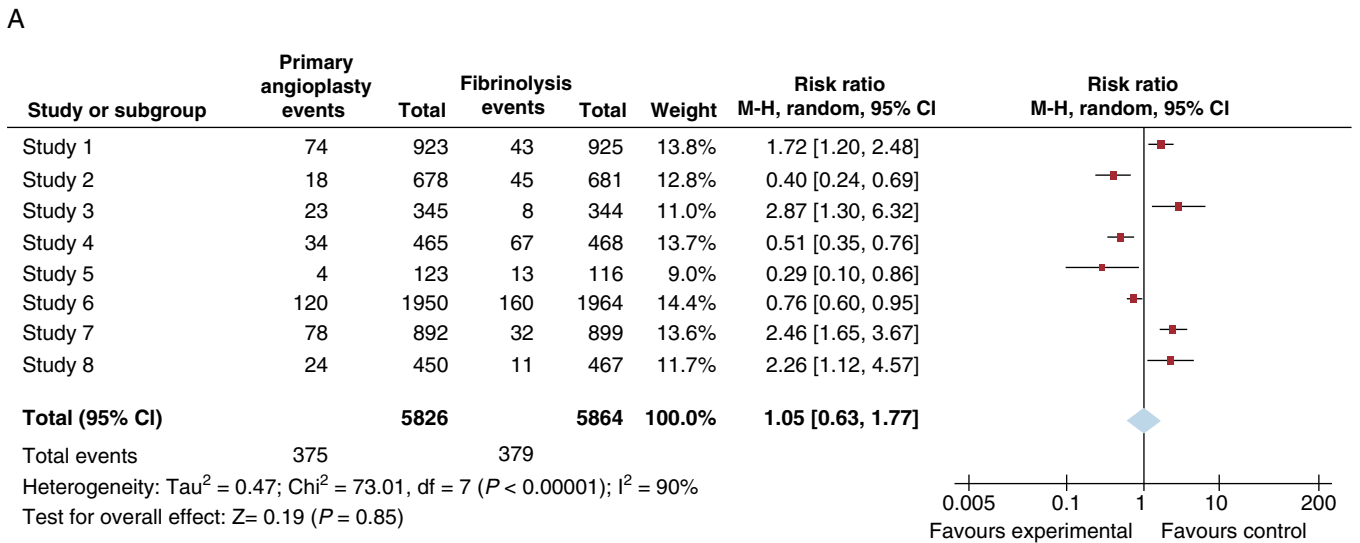


Figure 6. Meta-analysis of the effect of primary angioplasty versus fibrinolysis on total mortality in patients with myocardial infarction. A, including all studies. B, analysis of subgroups based on door-to-balloon time; upper figure, studies in which mean door-to-balloon time exceeded 140 min; lower figure, studies in which the mean door-to-balloon time was <90 min. CI: confidence interval.

Interpreting the Results

Finally, the SR ends with the interpretation of results. This includes a discussion of the limitations of the study (of the review) as well as any potential biases of the original studies, and potential biases that could affect the SR itself. It is also important to discuss the consistency of the findings and their applicability, and to propose recommendations for future research on the topic of interest.

SYSTEMATIC REVIEWS AND CLINICAL PRACTICE

As much as the findings of an RS are consistent and convincing, in the end it is the clinician who must make decisions regarding a particular patient. The findings of a SR should not be taken as fixed and unchanging standards resulting from an “evidentialist orthodoxy”. In other words, we must adapt the findings of an SR to the patient, and not vice versa. In this regard, before making a

decision on a patient based on an SR, we recommend making the following reflections.³¹

Are the Findings Applicable to My Patient?

The SR has shown that primary angioplasty is superior to fibrinolysis in elderly patients. But imagine my particular patient has a condition that was an exclusion criterion in all clinical trials included in the SR, such as a creatinine clearance rate <30 ml / min. In this case, my patient would not be represented by the specific clinical trials included in the SR.

Is the Intervention Feasible in My Patient?

There may be regional differences in the availability and/or experience of applying a particular technique, differences which should be taken into account when applying the intervention to an individual patient.

What is the Risk-Benefit for My Patient?

Even if the intervention is feasible and applicable, the specific risks for a particular patient should be taken into account and this is an aspect which is generally poorly represented by clinical trials.

What Are My Patient's Particular Values and Preferences?

Accustomed to making decisions based markers of myocardial necrosis, electrocardiograms, and other complementary tests, we can succumb to an overly paternalistic form of medicine, whereby we treat the patient with the best intentions, but without taking his or her point of view into account.

CONCLUSIONS

SRs are an essential tool in synthesizing available scientific information; they enhance the validity of the findings from individual studies and identify areas of uncertainty where research is needed. They are essential to the practice of evidence-based medicine. However, SRs should be implemented following a strict methodology and applying quality control to avoid biased conclusions. Ultimately, it is the clinician who must make decisions regarding a particular patient; SRs are just one more tool to be used, judiciously, in making those decisions.

CONFLICTS OF INTEREST

None declared.

REFERENCES

- Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997;126:376–80.
- Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med.* 2007;4:e78.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62:e1–34.
- 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;283:2008–12.
- Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol.* 1995;48:167–71.
- Stewart LA, Tierney JF, Clarke M. Reviews of individual patient data. In: Higgins JP, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1* (updated September 2008). The Cochrane Collaboration; 2008.
- Straus S, Moher D. Registering systematic reviews. *CMAJ.* 2010;182:13–4.
- Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club.* 1995;123:A12–3.
- Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ.* 2010;340:c117.
- Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. *Epidemiol Rev.* 1992;14:154–76.
- Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet.* 1996;347:363–6.
- Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol.* 1995;48:159–63.
- Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med.* 1987;6:315–28.
- Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci.* 1993;703:135–46.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet.* 1991;337:867–72.
- Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med.* 1987;6:11–29.
- Egger M, Smith GD. Misleading meta-analysis. *BMJ.* 1995;311:753–4.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ.* 1994;309:1286–91.
- Dickersin K, Larsen K. Section V: Establishing and maintaining an International register of RCTs. In: *The Cochrane Collaboration Handbook*. London: BMJ Publishing Group; 2010.
- Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *Int J Technol Assess Health Care.* 2003;19:591–603.
- Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. *Ann Intern Med.* 1997;127:531–7.
- Streiner DL, Norman GR. Reliability. In: *Health measurement scales: a practical guide to their development and use*. New York: Oxford University Press; 1995.
- Altman DG. Measuring agreement. In: *Practical statistics for medical research*. London: Chapman and Hall; 1991.
- Guyatt GH, Rennie D, Meade MO. Summarizing the evidence. In: *Users' guides to the medical literature*. New York: McGraw Hill; 2009.
- Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care.* 1996;12:195–208.
- Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials.* 1995;16:62–73.
- Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA.* 1999;282:1054–60.
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336:924–6.
- Juni P, Altman DG, Egger M. Assessing the quality of randomized controlled trials. In: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care: metaanalysis in context*. London: BMJ Publishing Group; 2001.
- Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21:1539–58.
- McAlister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group; 2001.