What are the differences between a literature search, a literature review, a systematic review and a meta-analysis? And why is a systematic review considered to be so good?

CLODAGH O'GORMAN, Alan P. Macken, Walter Cullen, Jean Saunders, COLUM DUNNE, M.F. Higgins

Publication date

01-01-2013

Published in

Licence

Document Version

1

Citation for this work (HarvardUL)

What Are The Differences Between A Literature Search, A Literature Review, A Systematic Review and A Meta-Analysis? And Why Is A Systematic Review Considered To Be So Good?

CS O'Gorman[1-4], AP Macken[1-4], W Cullen[1,2], J Saunders[5], C Dunne[2,3,5], MF Higgins[6]
[1]Graduate Entry Medical School, and [2]Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
[3]The Children's Ark, University Hospital, Limerick
[4]National Children's Research Centre, Crumlin, Dublin
[5]C-Star, University of Limerick
[6]Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto

It takes time to recognise the differences between a literature search (LS), a literature review (LR), a systematic review (SR) and a meta-analysis (MA), especially as these terms are often used interchangeably by many authors. For example, a colleague said recently that she planned to do SR as part of her background for her post-graduate research thesis. She planned to have it completed within five days. After talking to her, it was clear that she did not understand the concept (or the workload!) involved in a SR. On the other hand, we all do so-called "quick and dirty" LSs every day! Those are the kind of search where you have a question, you open up your favourite search engine (PubMed, EMBASE, etc.), plug in a few key words and press "search". Usually, with this type of search, you only put more effort into the search strategy if the "quick and dirty" approach does not yield enough (or any) relevant articles or if you are doing the LR for your thesis, or research project.

So, what are the differences between LS, LR, SR and MA? Actually, they can be described along a continuum, where literature search is the most basic and SR the most complex, with a MA the statistical extension of a systematic review where appropriate. A narrative (literature) review is a review of various articles, but generally lacks explicit descriptions of systematic methods to identify articles and/or failure to critically appraise them. Let us start by discussing each term.

A **literature search** means exactly what it implies: you search the literature to answer a question, e.g., "What is Giardiasis?" You can open the book on infectious diseases, check MEDLINE or search Google in order to come across one summary article that gives you a general idea of what Giardiasis is. In experienced hands, this will take only a few minutes and you gain some superficial knowledge. (Alternatively you could spend months searching for every article written in the medical and non-medical literature on *Giardia*, but you will have still only *searched* the literature; you may not have any more information on *Giardia* than when you started).

Generally, when doing a **literature review**, a search strategy is drawn up and one or more medical databases is used to implement the search strategy. (A search strategy incorporates the issues, such as PICOT, previously discussed in our "What's the difference between PubMed and MEDLINE?" tutorial.) As part of a literature review, it is expected that the retrieved articles are reviewed (i.e., critically appraised) and not just superficially read.

A **Systematic Review** is a scientific investigation, a research article with pre-defined systematic methods that identify systematically articles relevant to the research question, appraise their quality, extract data and then synthesise the results of these articles. The original studies which make up a SR (including published and unpublished data, conference proceedings, abstracts, etc.) are the "subjects" of the scientific systematic review. A SR employs methods to limit bias and random error. In a SR, the results of primary studies will be summarized, but may or may not be statistically combined to give a final figure. When statistical methods are used to combine the results of two or more studies, this is called a **Meta-Analysis.**

Let's summarise to date:
- **Literature search**: searching the literature for some studies.
- **Literature review**: reviewing the studies which have been identified.
- **Systematic review**: systematically searching the literature to identify all relevant published and unpublished data in order to appraise their quality and summarise the overall findings. If the studies are homogeneous (similar), and of sufficient quality, then their results can be amalgamated into a **meta-analysis** in order to obtain one final result summarising all the included studies.

Why is a systematic review considered to be the highest level of evidence?
Simply because a well-designed SR will summarise good quality randomised controlled trials (or, increasingly these days, observational studies). Let's go through how you could perform a SR. If you understand the concepts and steps behind performing a SR, it will help when trying to read and appraise them.

Planning
Firstly, you have to have a question that you want to answer. And it helps if it is *your* question rather than your *supervisor*'s question, as it is likely that you will read and appraise all the articles, while your supervisor supervises! Then, the question must be framed as a research question, in the PICOT format. For this, you should try to come up with every synonym for all the elements of your PICOT. After this, some people will design their own way of searching the literature systematically, and others will use some of the pre-validated search strategies available (e.g., search strategies for retrieving randomised controlled trials from MEDLINE[1] and PubMed[2]). When planning the search, be careful not to make it too wide (or you will have a huge number of false-negative articles: imagine searching for all the treatments of a heart attack) or too narrow (where you are likely to miss many articles that would be relevant to your question). This is the planning stage, and the more you plan , the more time you can save later (if your search is not too big) or the more likely you are to find the articles that really answer your question (if your search is not too small).

Next, you should define your inclusion and exclusion criteria. For example, do you want to include only randomised controlled double blind clinical trials, or does blinding not matter? Or, if it is unlikely that you will find RCTs, will you accept cohort or case control studies? Is there an age limit appropriate to your participants for included studies? What about exclusion criteria? *Predefining* your inclusion and exclusion criteria helps to prevent against bias when reviewing your articles for inclusion in your SR.

Searching for evidence
Once you have planned your search and inclusion and exclusion criteria, now you implement your search strategy. Different databases will be more relevant to different searches. There is overlap between MEDLINE (www.ncbi.nlm.nih.gov/pubmed) and EMBASE (www.embase.com), so performing your search in both might not add a significant number of articles. However, some

European journals might be indexed in EMBASE, but not in MEDLINE. Searching in the Cochrane Controlled Trials Register (www.cochrane.org) might be very helpful for RCTs, whereas for nursing it may be best to search in CINAHL (www.ebscohost.com), and for medical education-related articles search in Academic Search Premier (www.ebscohost.com/academic/academic-search-premier).

Data extraction
At this stage, all articles that have been retrieved by the search are analysed for the predefined inclusion and exclusion criteria. In some cases, review of the title and abstract may be sufficient to exclude some articles. In other cases, the article will need to be reviewed in detail (the number of studies excluded at each of these stages are generally shown as the first figure in good SRs). Next, data extraction is performed on all studies that you think should be included. Data extraction should be exhaustive[3], because all valuable information should be retrieved from studies at this stage so that there is often no need to refer back to the original studies. It is worth noting that some studies will be excluded during the data extraction stage. Review of titles, abstracts and entire studies, and data extraction are best performed independently by two reviewers, in order to reduce bias.[4] Care should be taken to note the study design when extracting the data as results from different study designs are not always directly comparable.

Quality appraisal
The next step is to critically appraise the studies by comparing and contrasting studies that are included in the final systematic review. Comments should also be made on sources of (potential) bias and specific areas of interest such as publication bias, ethics, quality analysis of included studies and comments on specific areas of your results, as required – for example, analyses performed according to years of publication of studies might be appropriate if, say, a new treatment was released during that period. (Basically, you are looking for are good quality studies in order to form a good quality SR).

Is it appropriate to do a meta-analysis?
MA is a statistical method of combing the results of studies included in a SR so that an overall treatment effect can be ascertained. Care should be taken with MA –strict criteria should be observed, and consultation with experts may be required, to prevent heterogeneity of included studies (i.e., to prevent comparing 'apples' with 'oranges'). One of the results of MA is the generation of a "forest plot". When interpreting a forest plot, look for the following (Figures 1 and 2).

There are several types of meta-analysis. For example, random effects models are used when there is thought to be significant heterogeneity (differences) between studies, but fixed effects models are used when there is less heterogeneity, and Peto's models are used for meta-analyses with very small sample sizes. Using the wrong method in the wrong circumstances can lead to different, and therefore incorrect, conclusions. It has also been pointed out by some statisticians that it is safer to assume 'random' effects in all meta-analyses, as the tests for homogeneity
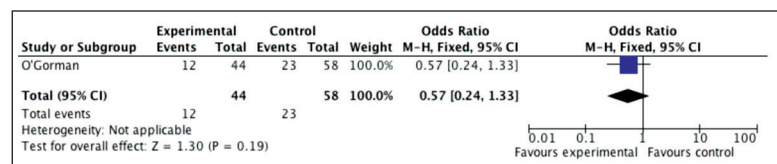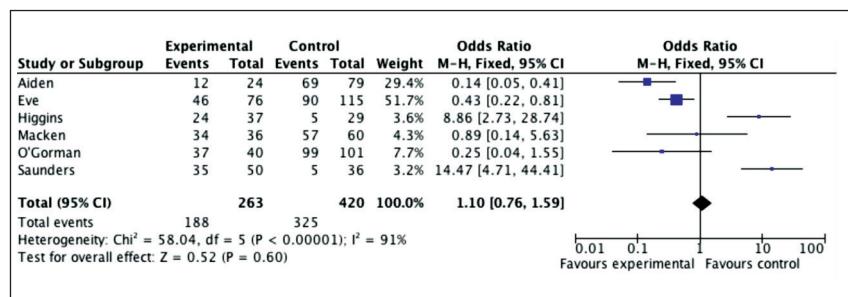

Figure 2 More complex Forest Plot
– the size of each box represents the weight of each individual study;
– the position of the box in relation to the line of no effect represents the treatment effect of each individual study;
– the size of the whiskers (the lines extending from the boxes) represent the confidence interval around the treatment effect for each individual study;
– the size and position of the diamond indicate the overall treatment effect of the meta-analysis, and the confidence interval around that treatment effect.

are of low power. Also, the effects of publication bias – where mainly positive results were published – needs to be assessed when interpreting the results.

No discussion of systematic reviews is complete without acknowledging the Cochrane Collaboration, named after Archie Cochrane, a Scottish medical researcher and pioneer of evidence-based medicine (www.cochrane.org). This is an international organization that publishes rigorous SRs evaluating the effectiveness of a wide range of health care interventions. The Cochrane Collaboration standardises, by means of expert review groups, the quality and methodology of SRs and MAs, and it encourages regular updates of its published SRs and MAs. (In fact, the symbol of the Cochrane Collaboration is the forest plot from one of the first MAs; this showed that antenatal steroids improved perinatal outcome and, thus, profoundly changed obstetrical management and reduced neonatal mortality and morbidity worldwide.[6]) This is the real strength of a SR and MA: to show an overall treatment effect which is stronger than the individual treatment effects of any included trials, even when the individual included trials show no overall effect. A further cautionary note is required here, remembering that in the past there was a certain publication bias for positive results. Some MAs, conducted using studies published during periods when this this publication bias was common, may have produced inaccurate results and treatments may have been changed to the detriment of patients. Large trials (RCTs) are always to be preferred, when possible. And, in the future, it is hoped that efforts to counteract publication bias (with the consequent publication of negative trials) will mean that future MAs will include inputs from trials with both positive and negative results.

We hope that this paper helps you to understand the methodology of SR and MA, and to appraise published data. This paper alone is not sufficient to train you to perform SR and MA, but if you are now inspired and would like to learn more, there are plenty of courses that would be very helpful. However, we do hope that in explaining the complexity of performing a systematic review we have helped you to appreciate why it is valued so as a research tool.

A more complicated MA using six studies to analyse the effect of an intervention (experiment) compared to placebo (control) showing no difference between the groups. Note however that two studies (O'Gorman and Macken) cross the line of "no effect" suggesting that there was no difference, two favour the intervention (Aiden and Eve) and two favour the control (Higgins and Saunders).

Box 1: Listed are some resources that you could access for more information. However, we learned

Figure 1 Simple Forest Plot
Forest plot showing that one study was analysed with no difference between groups. Mantel Hanzel with odds ratio crossing one.

about SRs and MAs by reading, helping to design them and doing them! An excellent resource is The Cochrane Collaboration, which regularly runs seminars on how to do a systematic review.

RESOURCES BOX

Systematic reviews in health care: meta-analysis in context[7] (book)

Systematic reviews to support evidence-based medicine: how to review and apply findings of healthcare research[8] (book)

Systematic reviews in healthcare: a practical guide[9] (book)

www.cochrane.org/resources/handbook/ [10] (web-site)

The QUORUM statement, The Lancet 1999[11] (instructions for reporting on SR and MA)

Correspondence: CS O'Gorman[1,2]
[1]Graduate Entry Medical School, and [2]Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
Email: Email: clodagh.ogorman@ul.ie

References

1. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. Br Med J 1984; 309:1286-1291.
2. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. Int J Epidemiol 2002;31:150-153.
4. Clarke MJ, Stewart LA. Obtaining data from randomized controlled clinical trials: how much do we need for reliable and informative meta-analyses? BMJ 1994; 309:1007-1010.
5. Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomised controlled trials in systematic reviews: accuracy and reliability of screening records. Statist Med 2002; 21:1635-1640.
6. Antenatal corticosteroid therapy: a meta-analysis of the randomized trials, 1972 to 1994.
7. Crowley PA. Antenatal corticosteroid therapy: a meta-analysis of the randomized trials, 1972 to 1994. Am J Obstet Gynecol 1995; 173:322-35.
8. Egger M, Davey Smith G, Altman DG. Systematic reviews in health care: meta-analysis in context. London, BMJ Publishing Group, 2005.
9. Systematic reviews to support evidence-based medicine: how to review and apply findings of healthcare research. Khan KS, Kunz R, Kleijnen J, Antes G. London, Royal Society of Press Limited, 2003.
10. Glasziou P, Irwig L, Bain C, Colditz G. Systematic reviews in healthcare: a practical guide. Cambridge, Cambridge University Press, 2001.
11. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354:1896-1900.

What are the Differences Between Common Statistical Tests?

MF Higgins[1], AP Macken[2-5], W Cullen[2,3], J Saunders[6], C Dunne[2,3,5], CS O'Gorman[2-5]
[1]Maternal-Fetal Medicine, Mount Sinai Hospital, Toronto
[2]Graduate Entry Medical School, and [3]Centre for Interventions in Infection, Inflammation & Immunity (4i), University of Limerick
[4]The Children's Ark, University Hospital, Limerick
[5]National Children's Research Centre, Crumlin, Dublin
[6]C-Star, University of Limerick

Statistics are essential components of quantitative (and qualitative) research that we all should know. We have sat through numerous lectures on the subject, and we know that we need to understand it and really should by now, even if only to not appear un-knowledgeable at journal club meetings. Still many people skip the section on statistical analysis when reading papers. Sometimes it feels that only those who have performed full time research really understand statistics - and not even then if they were lucky enough to have a statistician on their team. As part of our series of research tutorials we would now like to remove some of the mystery surrounding the art of statistics. Let's start with the raw data….

What's the difference between qualitative and quantitative data? As you begin your analysis you will always have a body of raw data which you can then use to reject or accept a null hypothesis. (Remember the Null Hypothesis? It is the chance that there is no difference between the groups being compared.) Before deciding which test you are going to use, you need to first decide what kind of data you are going to collect. Data are either qualitative (e.g., colour of hair, type of job, place of birth, *"quality" information*) or quantitative (e.g., BP readings, serum bilirubin levels, birth weight: quantities, *numbers*). While that seems relatively easy, some people will try and confuse you by referring to qualitative data as categorical or to quantitative data as numerical. We are going to keep it simple, and we suggest that you stick with the simple subtypes and then take it from there.

What's the difference between parametric and non-parametric data?
Remember the famous Gaussian curve of the normal distribution? If not, look at Figure 1, and it will immediately spring to mind again. A normal distribution is symmetrically distributed around the mean

with a bell-shaped curve. If your data are normally distributed, then you can use tests based on the normal distribution (such as the t-test: more on this later); if the data are not normally distributed (i.e., non-parametric, or skewed) then you can either transform to normal (which is not that hard) or use non-parametric tests. Transformation means using specific statistical tools to convert "not-normally" distributed data to normally distributed data, e.g., data that are positively skewed (i.e., skewed to the right) might be transformed by getting the logarithmic of each individual data in the dataset. (However this is risky as the hypothesis being tested will also change).
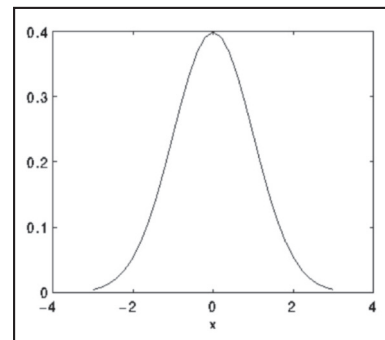

Figure 1: Gaussian curve: normally distributed population

What's the difference between average, mean, mode and median? Primary level maths taught us all the meaning of the term "average". The mean, mode and median may be different numbers but all represent the average value of data. Essentially, the mean is the arithmetic mean (the sum of all the values divided by the number of values), the median is the middle number in a series of numbers (thus, dividing the distribution in half), and the mode is the value that occurs most often (I think of it as being fashionable, or "in mode", so it is repeated most often). Here are a couple of examples, from this group of numbers, or raw data. This could be ages, or grammes of medication required to get an effect, or