An Introduction to Meta-Analysis
NJ Gogtay, UM Thatte

Introduction

Whether it is conducting a general literature search or finding an answer to a specific research question, clinicians, researchers and policy makers alike are faced with a huge volume of information that does not necessarily give them an answer to what they are seeking to find. For instance, if they are looking to answer the question is Drug A a better anti-hypertensive than Drug B, they are likely to find studies of three types – 1) Where the answer is equivocal, or 2) favors A or 3) favors B. Thus, results of a search often do not direct them to a clear, coherent or cogent result. While researchers are usually able to deal with ambiguity, practicing clinicians and policy makers can easily get befuddled with it.

Thus, a single paper that summarizes and synthesizes all relevant papers to answer a specific research question with the help of statistics would be of great value. A meta-analysis is that single paper.

Definitions and Historical Perspective

The first definition of meta-analysis was given by Gene Glass [1976] as “the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings”. Glass also called meta analysis as “an analysis of analyses”. The Greek word “meta” refers to “after” or beyond” and therefore meta-analysis go beyond individual studies. Huque [1988] defined the term as “A statistical analysis that combines or integrates the results of several independent clinical trials considered by the analyst to be combinable”. Historically, it was social scientists and statisticians in America who began to actively develop methods that would deal with large volumes of data and quantitatively synthesize them.

Why a Meta-Analysis is Needed

There are several reasons why it is commonplace to find results of studies that are asking similar research questions to be at variance with each other. This diversity that inherently exists amongst studies is called heterogeneity [see later].

These include:
1. Use of different case definitions for the disease under investigation [for instance, bleeding due to warfarin in one study may include mild bleeds only while another study may include hospitalizations and deaths due to bleeding which are severe events]
2. The study population may come from different parts of the same country or even from different countries [this would be important in infectious diseases like malaria where resistance patterns vary from country to country and within the same country],
3. The inclusion and exclusion criteria may vary and methodology to arrive at conclusions may be different [for example, peripheral smear diagnosis of malaria in one study to PCR based diagnosis in another].

One way of combining all studies on a particular topic is the traditional narrative review. This review typically combines several studies in a chronological discourse by an expert in that field. While the research question itself for the review may be well thought through, these reviews tend to be largely subjective and prone to bias as they are dependent upon the expert evaluating the studies, quality of the search and number of studies identified therein. They are also easier to carry out when the number of studies is not too many. Additional disadvantages include different researchers coming to different conclusions and lack of critical and in-depth analysis of each study included in the review. Meta-analyses, on the other hand, offer the advantage of applying objective statistical criteria, including addressing the variability between studies (heterogeneity) and thus can easily be done with the ready to use software [combined with training] regardless of the number of studies that need to be synthesized.

The Distinction between a Systematic Review and a Meta-Analysis

A term that is often used alongside a meta-analysis is “systematic review”. The two terms are also erroneously used as synonyms. A systematic review is
a type of review which answers a focused research question, and within which “meta-analyses” may or may not be a part. Systematic reviews typically have a specifically formulated research question, a clear search strategy, a pre-decided protocol that includes methods to identify which studies are to be included [or excluded based on selection criteria], quality assessment of studies and methods to analyze the ones included. When systematic reviews contain a statistical synthesis of the included studies to generate a single number [also called effect size, see later], this becomes a meta-analysis. Thus, systematic reviews can be standalone [without a meta-analysis] or include a meta-analytic component. In summary, a Meta-analysis refers to that portion of the systematic review that involves the statistical analysis. Since there is a fair amount of overlap between the two, one or both may be alluded to as appropriate within individual sections of this paper.

**Steps in a Meta Analysis**

A total of seven steps need to be followed while conducting a systematic review and/or meta-analysis. These include-

1. Formulating a research question
2. Writing the protocol and registering it in public domain
3. Identification of the studies using a clear and comprehensive search strategy
4. Selecting the right studies to be included [based on the protocol]
5. Data abstraction
6. Quality Assessment of included studies
7. Statistical analysis [including generating the Forest plot]

Each of these steps is now described in detail.

**Step 1- Formulating the Research Question**

Perhaps the most important step of clinical research in general and meta-analysis in particular, is to formulate the research question well. This is the uncertainty or lacuna that the researcher is attempting to answer. Asking the right question will lead to the right study design, an appropriate literature search strategy and statistical analysis that will generate the right research evidence that is needed to drive practice decisions. Thus, it ensures that the question will be answered in all likelihood. There are several choices available for formulating a research question and these are given below given by acronyms or mnemonics.

**I. PICOT**

A widely accepted and used acronym or mnemonic for formulating a research question is PICO or PICO[T]. It stands for:

P- Patient or Problem or Population
I- Intervention
C- Control
O- Outcome
T- Time

It essentially involves breaking down the research question into five components that ensures that the researcher and the reader are able to identify its’ individual elements.

Let us understand this with a classical meta-analysis from published literature. Lau J and colleagues [1992] performed a meta-analysis of n = 33 trials done between 1959 and 1988 that evaluated the impact of streptokinase on mortality after an acute coronary syndrome. The meta analysis showed a 21% reduction in death following the use of streptokinase.

The PICOT for this study would be framed as follows

1. In patients with acute myocardial infarction
2. Does treatment with intravenous streptokinase reduce patient waiting time
3. Compared to placebo
4. Impact mortality?

**II. ECLIPSE**

This stands for Expectation [what does the search requester want the information for], Client Group [for whom is the service intended], Location [where is the service physically situated], Impact [what constitutes success and how is this measured?], Professionals [who provide or improve this service], Service [Its nature-outpatient/inpatients/day care only and so on]. The mnemonic helps formulate research questions in the area of health policy management. For example, the Director of a major hospital may be interested in reducing the waiting time for out-patients who visit his hospital. The ECLIPSE for this study would be as follows

E- Reduce patient waiting time
C- All out-patients visiting the hospital
L- Hospital located at south end of the city
I- Impact- Reduction [by at least 15 minutes] in waiting time measured in minutes
P- All doctors or department/s who evaluate these out-patients
O- Outpatients who attend the hospital

However, not all questions are well served by the PICOT or the ECLIPSE mnemonics. Hence, other authors have proposed other models that can be used and the acronyms are listed below. These include

S P I D E R "- Sample, Phenomenon, Design, Evaluation and Research type [largely for qualitative research and/or mixed research methods
the protocol and the systematic review [with or without a meta-analytic component] are available from the Cochrane Database of Systematic Reviews [CDSR]. In 2007, the Indian Council of Medical Research [ICMR] became the first low income country to purchase national access for Indians with internet to the Cochrane Library through an agreement with the publishing partner of The Cochrane Collaboration, John Wiley and Sons Limited.\textsuperscript{11} This access continues with a recent renewal of the agreement with John Wiley and sons.\textsuperscript{12}

One feature that is unique to the Cochrane reviews is that they are dynamic and updated as and when new evidence emerges.\textsuperscript{13} Non-Cochrane systematic reviews and meta-analysis can be registered with PROSPERO [International Prospective Register of Systematic Reviews - \url{http://www.crd.york.ac.uk/Prospero/}]\textsuperscript{14} an international database that has been set up by the University of York and is free. From October 2013, PROSPERO also docks Cochrane protocols [these automatically get added to the PROSPERO database].\textsuperscript{15}

All elements that should necessarily be present in a protocol for a meta-analysis are outlined in Table 1. The Preferred Reporting Items for Systematic reviews and Meta-Analyses for Protocols (PRISMA-P 2015) is a 17-item checklist that helps authors in preparing a robust protocol.\textsuperscript{16} This can also be used by peer reviewers and editors to assess the quality of systematic reviews and/or meta-analysis protocols submitted.

### Table 1: Elements that should be included in a protocol for a meta-analysis\textsuperscript{15}

<table>
<thead>
<tr>
<th>Background</th>
<th>Describes the key contextual and conceptual factors relevant to the review question and provide the justification for the review.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The research question using the PICOT format</td>
<td>Clear search strategy including databases that will be searched for identifying the research evidence.</td>
</tr>
<tr>
<td>Describe inclusion and exclusion criteria</td>
<td>Describe how studies will be shortlisted for final inclusion.</td>
</tr>
<tr>
<td>Describe process of Data extraction</td>
<td>Pre-specify the tool/s to be used for assessment of quality of the included studies.</td>
</tr>
<tr>
<td>Describe how results will be synthesized</td>
<td>The choice of model (random effects/fixed)</td>
</tr>
</tbody>
</table>

[that involve a combination of quantitative and qualitative research], SPICE\textsuperscript{9} - This stands for Setting [where?], Perspective [for whom?], Intervention [what is being tested?], Comparison [versus what?] and Evaluation [with what result?]. This mnemonic is believed to work well in the context of social sciences research and COPES - Client Oriented Practical Evidence Search (COPES) which addresses problems seen in day to day practice.\textsuperscript{10}

### Step 2- Writing and registering the study protocol

The protocol for a systematic review and/or meta-analysis should clearly state the rationale, objectives, search strategy, methods, end points and quality checks that would be used. The PRISMA [Preferred Reporting Items Systematic Reviews and Meta-Analysis] guidelines recommend registration of the protocol à priori. Registration ensures that the protocol [and the methodology within] is accessible to all [much like registration of clinical trials before they are initiated] and will also prevent duplication by another author. For systematic reviews and/or meta-analysis done with and for the Cochrane group, both

*tailored for each individual database. Sensitivity of a strategy refers to identification of as many *potentially relevant* articles as possible. Specificity refers to picking up the *definitely relevant* articles. All search strategies should aim at *maximizing sensitivity* so as not to miss articles that are likely to be relevant.

Commonly searched databases include National Library of Medicine [Medline], Experta Medica Database [EMBASE], Biosciences Information Service [BIOSIS], Cumulative Index to Nursing and Allied Health Literature [CINAHL], Health Services Technology, Administration and Research [HEALTHSTAR], and Cochrane’s central register of controlled trials. Boolean operators [AND, OR, NOT] should be used along with search terms to narrow or broaden the search. All databases have filters [for example type of article, language of publication, dates of publication, age of participants and so on] and these should be used to narrow down the results to those articles likely to be relevant to the research question. In addition, the search should also include evaluating the cross-references from the articles retrieved. Use of controlled vocabulary [subject headings only] may result in a sub optimal yield. Therefore, uncontrolled vocabulary for example, variations such as abbreviations, generic name, terms used internationally, differential spellings used in another country and so on should also be used in the search.\textsuperscript{17}

Given that negative results are often not published, the search strategy should also include [to the extent possible] unpublished data, thesis/project reports that may be available on Institutional or University websites, conference proceedings and abstracts and telephonic/email contact with trialists and experts in that field. Developing a search strategy is an iterative process - that is a
process of continual assessment and refinement

Citation managers- Once the results from the search are available, it is useful to export them into a citation manager. The advantage of these is that as they are electronic, preclude manual errors, eliminate duplicates, save time and also back up search results. Zotero and Mendeley are two citation managers that are available free for use. EndNote and RefWorks are paid software. Citation managers also incorporate an array of reference styles and in the event that the paper is rejected by one journal, it is easy to change the formatting and style of referencing for another journal.

Step 4- Selecting the right studies to be included - narrowing the results of a search strategy to a final number

The next step is to read the title and abstract of each reference obtained and eliminate those that are not relevant. Subsequently, we obtain full texts of potentially relevant articles [those likely to pass the selection criteria]. The focus while reading the full text should remain on the methods and results section rather than the Introduction.

Step 5- Data Extraction

Once the final list is ready, from each article, depending upon the protocol, we extract the relevant information-case/disease definitions used, key variables, study design, outcome measures, nature of participants; therapeutic area, year of publication; results; setting and so on. These will now need to be fed into the software for analysis [Revman, see later].

Step 6- Quality assessment of included studies

Once the number of studies to be included is firmed, it is important to assess their quality. This is because a flawed study is in fact worse than no study at all. Several methods are available to assess quality of studies, each with its own merits and demerits. These include among others, the Jadad score, the CONSORT statement, and the Cochrane Back Review Group criteria. We describe the Jadad score here as an illustrative example. It is a 5 point score where one point each is allocated to randomization, description of the method used for generating the random sequence, whether or not blinded, description of the method used for blinding/masking and clear cut information on drop outs and withdrawals. One point each is deducted if randomization is described but the method therein is inappropriate and if blinding is described but again the method for blinding is inappropriate [flawed]. Its strength lies in brevity and thereby ease of use. For example, Boussagen [2013] conducted a meta-analysis of RCTs that evaluated all cause mortality and deaths from cardiovascular events related to intensive glucose lowering treatment in people with type 2 diabetes. Quality of the RCTs was assessed using the Jadad score. Studies with a score of more than 3 were indicative of high quality. The overall meta-analysis using all studies irrespective of the quality (as assessed by their Jadad score) showed limited benefits of intensive glucose lowering treatment. This was confirmed by evaluating only studies with a Jadad score of more than 3 which also showed that intensive treatment was NOT associated with any benefit.

Step 7 - Statistical analysis of included studies

Understanding what Effect Size is

One term that is frequently used in meta-analysis [and subsequently used in this paper] is “effect size” which represents the basic unit of a meta-analysis. We have seen this earlier in the article on sample size calculation. When we compare two interventions [say A and B], we are seeking to find the difference between them. Meta – analysis is also about A vs. B comparisons. Simply put, the effect size is the difference between A and B and the “size” or “magnitude” of this difference. This is a standardized metric that expresses the difference between two groups- usually an experimental and a control group. The effect size can be expressed as any one of these metrics- odds ratio, risk ratio, standardized mean difference, person time data and so on.

Statistical synthesis of data-Once data from all the shortlisted studies is ready, it is fed into Revman (see later). The two commonly used methods for analysis are Mantel-Haenszel [Fixed effects model, see below] and DerSimonian-Lard [Random effects model, see below]. Both methods essentially provide a single number or summary statistic along with 95% Confidence Intervals, which is the goal of any meta-analysis.

Allocating weights to the different studies –As the ultimate goal of any meta-analysis is to estimate one overall effect after pooling all the studies; one way of doing it is to simply add all effect sizes and compute their mean. However, each study in a meta- analysis is actually different from the other. Hence, we allocate a “weight” to each study- in other words, we give more weight to some studies and less to the others and compute a “weighted mean”. How do we decide how much weight each study should get? This is driven by two key factors- the sample size of the study [bigger the better] and the outcomes in each study [the more the better].

Fixed and random effects models-In the fixed effects model, we assume that the effect size in all included studies is identical and any difference between them is a result of differing sample sizes and associated variability, and hence the term “fixed effects”. Thus, when we allocate weights to the studies [see below], the studies with smaller sample sizes get a lower weight and the larger studies a higher weight. In the
random effects model on the other hand, we assume that each study is unique and therefore will have its own effect size. Here, unlike the fixed-effects model, the studies with smaller sample sizes are not discounted by giving them lower weights as each study is special and is believed to make an equally important contribution to the overall analysis. The random effects model is based on the assumption that if a large number of studies for the same research question using the pre-set selection criteria, the true effect sizes for all these studies would be distributed about “a” mean. The studies included in the meta-analysis are believed to represent a “random” sample from this larger number. Hence the term “random effects”. Thus, the weights allocated in the random effects model are more balanced [relative to the fixed effects model]

In the former, the only source of uncertainty lies within the study itself, whereas in the latter model, we also take the between study variance into account. Thus, the fixed-effects model will have narrower confidence intervals and the random effects model wider confidence intervals. Conventionally, the choice of the model must be decided before beginning the analysis and described in the protocol. However, Revman [see later] can give the summary effect and the 95% CI with both models at the same time and thus both are often presented in published

![Forest Plot](https://example.com/forest_plot.png)
Testing for Heterogeneity—An important issue in meta-analysis apart from looking at the significance of treatment effects is to look at the extent to which studies included are similar to [or dissimilar] to each other. In other words, we need to assess consistency [or inconsistency!] across studies and the method to do this is called the test for heterogeneity.27 The extent of heterogeneity will significantly impact the conclusions of the meta-analysis.

Two statistics are used to assess this- the Cochran’s Q [or the Q-test] and the I² [I square]. The former is a less used metric as it has poor power [ability to detect a difference] when the number of studies is few. The I² statistic describes heterogeneity as a percentage. For example, if the I² value is 50%, it means that 50% as a percentage. For example, if the I² statistic describes heterogeneity the number of studies is few. The study showed that PMRT significantly reduced the risk of locoregional recurrence and was associated with a minor overall survival benefit.

The elements of a Forest plot are:

Column 1 - This is the column on the far left that identifies the study by the first author’s name and year of publication.

Column 2 - This describes the experimental intervention. The sub columns here describe number of events for the desired outcome of interest and the total number of patients [n and N respectively].

Column 3 - This describes the control group. The sub columns here, similar to Column 2, describe number of events for the desired outcome of interest and the total number of patients [n and N respectively].

Column 4 - This gives the weight allocated to individual studies and is described as a percentage

Column 5 - The summary measure [risk ratio in this case] is described for each individual study along with the 95% confidence intervals [CIs]. The model used [Mantel Hanzael (MH) random effects in this case] is also mentioned here.

Column 6 - This is the graphical depiction of the summary effect along with 95% CI around a central line

Let us now understand other parts of the Forest plot.

- The central vertical line - This indicates the line of no effect [when the two interventions being studied are not different from each other].
- The squares and the horizontal lines that cut the “squares” pertain to the summary statistics of individual studies [risk ratio in this example] and the horizontal lines that run through them indicate the 95% CI of the risk ratio

- The “diamond” – This is located at the bottom of all studies. This could fall on either side of the central line or fall in the middle and “cut” it. It represents the summation of all studies and the horizontal edges of the diamond indicate the 95% CI of the summation. If the diamond falls on the line it indicates no difference between the two groups. If it falls on the left it favors the experimental intervention and if it falls on the right it favors the control group.

- The lower left corner of the Forest plot - This gives the F statistic, the measure of heterogeneity along with its p value. In this case the p=0.24 indicating a lack of significant variance between the studies). This is followed by a second p value for the effect size of this meta-analysis (in this case it is p<0.00001 which indicates there is a significant difference between the two interventions studied. Note- the second p value relates to the diamond that can fall on the central line or to its left or to its right.

Revman

The software that is used for both statistical analysis and maintaining systematic reviews that are done by the Cochrane group is called Revman [Version 5 with latest major version being 5.3].29 It is freely downloadable for use for academic meta-analysis. Once the data is entered, Revman generates a Forest Plot.

Criticisms of Meta-Analyses

Several critics have pointed out that meta-analyses may be flawed. These criticisms have been summarized and eloquently answered by Borenstein.30 These are outlined in Table 3.
Conclusions

Meta-analyses are extremely important in today’s world of Evidence Based Medicine as they have the ability to use powerful statistical tools and software to combine studies with identical research questions [those that have similar designs, selection criteria and patient populations]. Their utility lies in the fact that individually, these studies may be small and underpowered to pick up treatment differences, but when combined in a meta-analysis; answer a well-formulated question to guide Evidence based clinical practice. There are some key challenges though in any meta-analysis. The first is the adequacy of the literature search and the subsequent data abstraction. The second is how similar [or dissimilar] are the studies that have been put together and thus looking at heterogeneity [the $I^2$ value] and the choice of the model used [fixed and random effects] is important. The others are the quality of the studies and the presence [or lack thereof] of publication bias. Both researchers carrying out the meta-analysis and readers who evaluate and use them should bear all of the above in mind as decision making in clinical practice is influenced by them.

References


