META-ANALYSIS: THE WAY FORWARD IN MEDICAL DISCOVERY

Akinyemi J.O. MSc (Medical Statistics), B Tech (Comp. Sc.)

Correspondence:
Mr. J.O. Akinyemi
Department of Epidemiology,
Medical Statistics and Environmental Health,
Faculty of Public Health,
College of Medicine,
University of Ibadan, Ibadan, Nigeria
E-mail: odunjoshua@yahoo.com, jakinyemi@comui.edu.ng
Mobile: +234-8055676325

Introduction
The increasing volume of information published in biomedical literatures poses an enormous challenge to evidence-based healthcare and scientific discoveries. It is common for important issues in medical research to be addressed in several studies. The idea of summarizing a set of studies is not new in medical literature; review articles have long had an important role in helping practitioners keep up to date and make sense of the many studies on any given topic. Meta-analysis goes a step further by using statistical procedures to combine the results of several studies.

Definition
Several definitions exist in the literature. However, Glass who developed the technique defined it as the statistical analysis of a large collection of analysis results for the purpose of integrating the findings. Another author defined it as a statistical procedure that integrates the results of several independent studies considered to be “combinable.” According to Medical Subject Headings (MESH), meta-analysis is a quantitative method of combining the results of independent studies (usually drawn from published literature) and synthesizing summaries and conclusions which may be used to evaluate therapeutic effectiveness, plan new studies, etc., with application chiefly in the areas of research and medicine. Well conducted meta-analyses allow a more objective appraisal of the evidence than traditional narrative reviews, provide a more precise estimate of a treatment effect, and may explain heterogeneity between the results of individual studies. Ill conducted meta-analyses, on the other hand, may be biased owing to exclusion of relevant studies or inclusion of inadequate studies. Misleading analyses can generally be avoided if a few basic principles are observed. This review article discusses these principles, along with the steps in performing meta-analysis. It concludes by highlighting the future and role of meta-analysis in medical discoveries.

Evolution in Medicine
The first meta-analysis was a combination of studies (with small sample sizes) of typhoid vaccine effectiveness performed by Karl Pearson in 1904 in an attempt to overcome the problem of reduced statistical power. Although meta-analysis is widely used in epidemiology and evidence-based medicine today, a meta-analysis of a medical treatment was not published until 1955. The term was coined by Glass in 1976, even though some meta-analytic methods have been in use for almost fifteen years in Education and Psychology. The concept made its way into medicine as researchers began to incorporate the idea. At the onset, the concept was not popular among medical scientists until mid 1980s when a group of clinicians and statisticians at Oxford University initiated the process of giving it scientific prominence. The Oxford group took the approach of gathering all studies, published and unpublished, and excluding those that used different endpoints. They focused on studies of therapeutic issues. Their conclusions were applied by others to clinical practice, to further ascertain validity. By 1985, there was a book on statistical methods for meta-analysis. In addition, there was a 1985 publication, “Findings for Public Health from Meta-analysis” which clarifies the difference between meta-analysis and traditional literature review. Since then, the technique has grown in leaps and bounds with application in different areas of medicine and other specialties. Its statistical methodology is always reviewed and constantly improved to accommodate new realities. For the results of a meta-analysis to be meaningful, a great deal of thought and planning are needed. Protocols for the reporting of meta-analysis results were developed for Randomized Clinical Trials (RCTs) (Quality of Reports of Meta-analysis [QUOROM]) and observational studies (Meta-analysis of Observational Studies in Epidemiology [MOOSE]). These guidelines were developed to provide proper procedures for conducting a meta-analysis and to standardize the methods of reporting it. Using these 2 protocols as a guide, the steps necessary to perform a meta-analysis include the following: (1) define the research question, (2) perform the literature search, (3) select the studies, (4) extract the data, (5) analyze the data, and (6) report the results.

Define the Research Question
A meta-analysis begins with a question. Common questions addressed in meta-analyses are whether one...
The reasons for the inclusion of items in the meta-analysis, the researchers need to define the problem or question of interest. The investigators should also have a good understanding of the problem and the subject matter. The study population baseline data (e.g., age, race, gender, diagnosis, length of illness), the study outcomes, treatment or intervention, and type of study designs to be used (e.g., restricted to RCTs or include observational studies such as prospective or retrospective studies) also should be defined.

Perform the Literature Search

Once the research question has been defined, a systematic search of the literature can begin. This is a critical step in the meta-analysis and often the most difficult part. The initial search of the literature should be broad so that as many studies as possible are gathered. During the selection phase, some of the initial studies will be weeded out using the inclusion criteria.

The literature search begins with searching electronic databases of published studies such as MEDLINE, EMBASE, CINAHL, etc. MEDLINE is maintained by the National Library of Medicine and contains more than 13 million citations dating back to 1966. EMBASE is a database produced by the publisher Elsevier BV and contains data from 1974 to the present. Although EMBASE and MEDLINE overlap in their coverage of the literature, EMBASE has better coverage of European journals. CINAHL covers literature related to nursing and allied health from 1982 to the present. The researchers should search more than just MEDLINE to ensure a comprehensive search. For example, a report found that approximately only half of all RCTs presented as abstracts are subsequently published on MEDLINE.

It is necessary to use other sources to access many of these unpublished studies. A good source for unpublished clinical trials is the Cochrane Central Register of Controlled Trials, which is a database of controlled trials. The database was set up to provide a source of data for systematic reviews and contains more than 300,000 references to RCTs. Other suggestions for locating studies include searching reference lists from the gathered reports, manually searching journals with lists of abstracts presented at meetings, or searching the Internet. Contacting experts in the field or networking with colleagues also could be a source of studies, although this mode of data gathering is seldom used.

Select the Studies

Once the literature search is complete, it is time to select which studies to include in the meta-analysis. The inclusion and exclusion criteria for studies need to be defined at the beginning, during the design stage of the meta-analysis. Factors determining inclusion in the analysis are study design, population characteristics, type of treatment or exposure, and outcome measures. The inclusion and exclusion criteria should be part of the meta-analysis protocol. One should keep track of the studies included and excluded at each step of the selection to document the process. The QUOROM guidelines for reporting a meta-analysis request that investigators provide a flow diagram of the selection process. The flow diagram lists the number of studies excluded and included at each stage of the selection process and the reasons for exclusion. The selection process involves reviewing the titles and abstracts of all articles identified through the literature search.

Many of the studies will be excluded at this stage based on the exclusion criteria. The remaining studies will be read to determine their suitability for inclusion. The validity of a meta-analysis depends on the quality of the studies included, and an assessment of quality is a necessary part of the process. The researcher wants to include as many studies as possible, but reduce the number of studies with low quality data; however, restricting the meta-analysis to only perfect studies may leave the researcher with little data. A variety of checklists and scales have been developed to assess quality in RCTs. Checklists provide guidelines as to what should be reported in an RCT, whereas scales are a way of quantifying the level of bias in an RCT. For example, a scale will assign a score based on a specific characteristic of an RCT (e.g., presence of adequate concealment of patient assignment to treatment groups), but a checklist does not assign scores. Although quality needs to be assessed in some way, caution should be used when using these scales or checklists. The reasons for the inclusion of items in a scale or checklist often are not given and the score assigned to scales can be arbitrary. The relatively imprecise scoring scheme in some of the scales may change the results of a meta-analysis.

There are several options available to deal with study quality once it has been ascertained. A cut-off value for the quality score can be used to exclude or include studies. Another choice is to use the quality scores to weight study results in the analysis. MOOSE reporting guidelines, however, recommend using a sensitivity analysis rather than weighting for quality scores. Sensitivity or subgroup analysis allows comparisons between studies of different quality. For example, studies can be separated into high versus low quality and then the meta-analysis can be repeated for each group. Results then can be compared between the 2 groups. A method that is being used increasingly is meta-regression. Quality scores or some measure of study quality (e.g., assignment to a treatment group) are entered into a regression model as an explanatory variable. This method allows the researchers to estimate the effect of quality on the results of the meta-analysis.

Extract the Data

The type of data to be extracted from each study should be determined in the design phase and a standardized form is constructed to record the data. Examples of data commonly extracted include study design, descriptions of study groups (e.g., number in each group, age, gender), diagnostic information, treatments, length of follow-up evaluation, and
outcome measures. Two independent reviewers will be instructed on the appropriate data to collect. For example, how will age be recorded on the abstract form? Will the standard deviation or standard error be used in the analysis? If data are missing, they should be recorded on the form. If too much data are missing, the study may need to be excluded. It is recommended that the reviewers be blinded to the investigators’ names but it is not essential. Data to be extracted are identified before beginning the meta-analysis to avoid data dredging or a fishing expedition. The difficulty with data extraction is that studies often use different outcome metrics, which make combining the data awkward. The data should be converted to a uniform metric for pooling. For example, data reported may be continuous (e.g., blood pressure) or binary (e.g., high blood pressure vs low blood pressure). A meta-analysis estimating the effect of a medication on blood pressure may find some studies reporting blood pressure as a continuous outcome whereas in other studies the outcome is reported only as high or low blood pressure. In this case it would be necessary to convert continuous blood pressure measurements into categories of high or low blood pressure to standardize the data into one format. Similarly, some studies report the standard deviation and others report standard error. Again, it is necessary to convert one into the other to make the data uniform. Although it is difficult to resist combining the data, if combining data is not possible because different metrics are used then it is best to leave the analysis as a systematic review.

Analyze the Data
A statistician who is familiar with meta-analysis should be consulted to help plan this type of project and to participate in analyzing the data. Detailed instructions for data analysis exist. A meta-analysis calculates a weighted average of the study effect that is pooled from the selected studies. The weight is directly proportional to the precision of the effect estimate and usually the inverse of the variance (square of the standard error) of the effect estimate. Therefore, larger studies will have more influence over the summary estimate than smaller studies. A summary estimate is calculated by multiplying each study’s weight by its effect estimate and adding these values together. This sum then is divided by the sum of the study weights. There are 2 statistical models used in a meta-analysis: fixed effects and random effects. The fixed-effects model assumes that the true effect of treatment is the same for every study. Because there is no heterogeneity between study results, only within-study variability is taken into account. Given the degree of variation or heterogeneity among studies, this assumption may be unreasonable. The random-effects model is often more realistic because it assumes that the true effect estimate for each study does vary. Sources of variation may include differences in patient population or treatment methods. The random-effects model will produce an estimate with wider confidence intervals, but the summary estimates for both models will be similar if there is not a great deal of heterogeneity among studies. A statistical test for heterogeneity can be used, but this test has low statistical power in most cases. Power refers to the ability of a statistical test to reject the hypothesis being tested (null hypothesis) when it is false. The null hypothesis states that there is no heterogeneity or variation among the studies. Low power for the heterogeneity test means that we are unable to reject the null hypothesis of no heterogeneity even when important heterogeneity exists. For example, the studies used in the meta-analysis may in reality vary considerably, but the low power makes the heterogeneity test non-significant. This would lead the researcher to the incorrect conclusion that the amount of variation among the studies is low. The best choice may be to always use the random-effects model or to use both models and compare the results. Statistical packages are available to calculate summary estimates using either model. If heterogeneity can be explained, then it should be included in the model. For instance, we may observe that some of the variation in studies can be explained by gender. In that case, separate summary estimates can be calculated for males and for females. Or, meta-regression models can be used to explain heterogeneity, but a large number of studies are needed when investigating multiple effects.

Report the Results
Detailed guidelines for the reporting of meta-analyses for RCTs were described in the QUOROM statement. Similar guidelines were developed for observational studies by the MOOSE group. These articles should be consulted during the design phase to ensure that these reporting procedures are used and that proper data are collected and presented in the report. Similar to a research report, a meta-analysis report should include a title, abstract, and introduction, and methods, results, and discussion sections. The title should identify the report as a meta-analysis. The introduction should indicate the clinical question of interest, the hypothesis being tested, the types of treatment or exposure being studied, the study designs to be included, and a description of the study population. The methods section should describe the literature search, specifically the databases used, and if the search was restricted in any way (e.g., English language only). The selection process for articles, quality assessment, methods of data abstraction, and synthesis also should be described in this section. The results section should include a flow chart of studies included in each step of the selection process, a figure displaying the results from each individual study such as a forest plot, results of heterogeneity testing, overall summary statistic and its 95% confidence interval, and results of a sensitivity analysis and meta-regression, if performed. For sensitivity analysis, several features of a meta-analysis can be altered to assess the robustness of the results, such as excluding questionable or unpublished studies. The sensitivity analysis may include an analysis weighted by a quality score for each study. The discussion section should summarize the key findings and identify possible sources of bias and heterogeneity. A forest plot, the figure with the effect estimate from each study and their associated confidence intervals along with the
summary estimate, is an important part of the report. Studies can be grouped by size or by other study characteristics such as year of publication. It allows the reader to observe the heterogeneity of the studies included. If the confidence intervals for effect estimates are not overlapping, indicating a great deal of study variation, a meta-analysis may not be appropriate. In this case, it is necessary to explore the reasons for the variation among the studies, which may lead to the discovery of associations between the study design or patient groups and the study outcome.

A funnel plot is used as a way to assess publication bias in a meta-analysis. The funnel plot is a scatter plot of each study’s effect estimate (e.g., odds ratio or mean difference) on the x-axis against the measure of the study’s precision on the y-axis. The overall sample size can be used on the y-axis but often an inverse of the standard error is used. If a publication bias is not present, the plot will resemble an inverted funnel. Large studies should have smaller variation and therefore a more precise effect estimate, whereas small studies should have larger variation and therefore a less-precise estimate. One expects the effect estimates for small studies will have wider scatter at the bottom of the plot and larger studies will have less scatter at the top of the plot. If small studies with negative or null results tend not to be published, one would see asymmetry in the funnel plot from the left bottom of the plot containing few or no data points. On the other hand, if fewer studies with non-statistically significant odds ratios were included in the literature search, it would result in an asymmetric funnel plot. Funnel plots can be inspected visually but interpretation can differ from person to person. Statistical tests such as the rank correlation test developed by Beggs and Mazumdar are available to assess the symmetry of the plot. The correlation test, however, should be used with caution in small meta-analyses because the power of the test depends on the number of studies included. In a small meta-analysis (25 studies), the correlation test will have low statistical power so a non-significant test will not rule out bias in the literature search.

A meta-analysis is a statistical method of combining results from multiple studies to determine the overall impact of a treatment or exposure. If performed properly, using the steps above, a meta-analysis can be a powerful research tool. Although meta-analyses are considered to have the highest level of evidence and are cited more often than other study designs, there are still lingering questions regarding its validity when compared with well-conducted clinical trials. The results from a meta-analysis can be used to plan a large RCT to test a treatment effect; however, a report comparing the results from meta-analyses and subsequent RCTs found only fair agreement. The distrust of meta-analyses and the lack of agreement with RCTs do not imply that the meta-analysis should be abandoned. It does, however, point out the limitations and biases involved with the meta-analysis and shows the need for conducting a thoughtful, well planned meta-analysis with minimal bias. When performed appropriately, however, a meta-analysis can lend evidence to many of the difficult decisions clinicians made in their daily practice.

Recent Development and Future of Meta-analysis

In a review of recent developments in meta-analysis, Sutton et al noted that there is considerable research activity in the field of meta-analysis. Meta-analysis methodologies are being developed for concepts such as prospective meta-analysis, meta-analysis of individual patient data, etc. There are meta-analysis techniques for complex evidence synthesis which involve models that incorporate evidence on multiple parameters and/or that specifically model data from different study designs. There are developments in meta-analyses of studies on effects of interventions, aetiology, diagnosis and screening. A recent publication of recommendations for reporting tumor marker prognostic studies was reputed to be a good initiative that will aid the meta-analysis of such studies in the future.

There are notable developments in software for meta-analysis as well. However, due to the fact that meta-analysis of summary data needs a unique set of analysis tools, the large developers of general statistical software have been reticent about providing the required routines. Fortunately, users have developed collections of macros, e.g. for SAS and, most comprehensively, for STATA. Stand-alone packages have also been developed, the most sophisticated of which is probably the commercial Comprehensive Meta Analysis. The Cochrane Collaboration software, RevMan, continues to be developed and a new freely downloadable Excel add-in MIX offers excellent functionality and educational content for those on a tight budget. Sutton et al reported that they found MetaDiSe very useful for carrying out the specialized analyses required for diagnostic tests meta-analyses. There is little doubt that the development of metaanalytic techniques will continue into the future. Multiple treatment comparisons will receive greater attention as it has potential to address the relative benefits of competing treatments, and address questions such as the probability that a particular treatment is superior to all the alternatives.

CONCLUSION

The potential of meta-analysis for discovery was demonstrated in the recent discovery of ten new genes related to human growth which was published in the latest issue of Nature Genetics. There is no doubt that meta-analyses have many positive attributes. Busy physicians have difficulty keeping abreast of the huge volume of medical literature, and some may not possess the analytic skills to resolve the often non-definitive or conflicting findings. Meta-analysis provides an attractive solution to this problem. By examining the totality of data available about an issue, meta-analysis can identify inadequacies in existing data and point to areas of needed research, reduce the potential for erroneous findings occurring by chance, and more...
accurately define the benefit and possible adverse effects of management strategies. In fact, by combining smaller datasets, meta-analysis may establish unrecognized outcomes, may provide evidence of statistical significance where it was previously absent, or may eliminate any possible bias in individual studies. Although, there are weaknesses such as publication bias, citation bias, etc; in spite of these, meta-analysis has a promising future for biomedical research and development.

REFERENCES

44. DeMaria A.N., Meta-analysis. JACC. 2008;52; 3; 237–238.