META-ANALYSIS

M. Buyse
Executive Director, International Drug Development Institute, Brussels, Belgium

Keywords: Meta-analysis, clinical trials, heterogeneity, interaction, fixed effects models, random effects models, stratification

Contents

1. Introduction
2. Types of meta-analyses
3. Statistical principles of meta-analysis
3.1. Estimation
3.2. Stratification
3.3. Heterogeneity
3.4. Interaction
4. Statistical models for meta-analysis
4.1. Fixed Effects Models
4.1.1. Test for Treatment Effect
4.1.2. Estimate of Overall Treatment Effect
4.1.3. Test for Heterogeneity
4.1.4. Test for Interaction
4.1.5. Test for Trend
4.2. Random Effects Models
4.3. Bayesian Models
5. An Example of a meta-analysis
5.1. Tests for Treatment Effect and Heterogeneity
5.2. Graphical Display
5.3. Tests for Trend and Interaction
6. Further topics in meta-analysis
6.1. Weighted, Cumulative, and Prospective Meta-analysis
6.2. Further Uses of Meta-analysis
6.3. Meta-regression
Glossary
Bibliography
Biographical Sketch

Summary

Meta-analysis is the statistical process of combining information from several studies addressing the same question. Although meta-analysis can in principle be used to combine the results of all types of studies used in epidemiology and medicine, it has gained popularity in the context of randomized clinical trials, in which interest focuses on the (possibly small) difference in outcome between a group of patients receiving some experimental therapy and a group of patients receiving the best available standard therapy. A meta-analysis yields a more precise and more unbiased estimate of the true

treatment effect than any component trial. Moreover, when the results of various studies seem to conflict with each other, i.e. when there is heterogeneity of the treatment effect across the studies or across different patient populations, meta-analysis provides a means of presenting the information, of testing whether the apparent heterogeneity is statistically significant, and of analyzing putative sources of heterogeneity. Meta-analysis can use data extracted from the literature (an approach that is not fully informative and often seriously biased), summary statistics from all relevant trials, or individual patient data from these trials (an approach that may be considered the gold standard of meta-analysis). Different modeling approaches can be used to carry out a meta-analysis. In all cases, the analysis is stratified by trial, so that patients from one trial are never compared directly with patients from another trial. Fixed effects models are most commonly used to estimate and test the statistical significance of an overall treatment effect, to test for heterogeneity between the trials included in the meta-analysis, and to test for interactions between the effect of treatment and clinically relevant prognostic factors.

1. Introduction

Meta-analysis (also called "quantitative synthesis", or "overview") is the statistical process of combining information from several studies addressing the same question. Meta-analysis was initially proposed to combine the published results of studies in education and sociology. Its potential soon became obvious in epidemiology and medicine, where several studies are generally performed to investigate a clinical question. For instance, in the development of an experimental therapy, it is typical to carry out a series of clinical trials, most of which randomize the new therapy against one that is considered the standard of care for the condition under study (see *Modern Biometry*). In such a situation, clinicians informally combine pieces of knowledge coming from these various trials. Meta-analysis is a systematic and quantitative approach to this very combination process, and as such, it is an essential tool for evidence-based medicine.

Today, meta-analysis has become extraordinarily common in the health sciences. The number of papers using "meta-analysis" as a keyword in Medline was 21 in 1985, 323 in 1990, 605 in 1995, and 1200 in 2000. Although meta-analysis can in principle be used to combine the results of all types of studies used in epidemiology and medicine, such as case-control studies, cohort studies, longitudinal studies, and prevention, diagnostic, screening or therapeutic trials (see *Data Collection and Analysis in Biometrics*), it has gained enormous popularity in the context of randomized clinical trials, in which interest focuses on the difference in outcome between a group of patients receiving some experimental therapy and a group of patients receiving the best available standard therapy. The difference expected between these randomized groups is often small. The detection of small, but medically worthwhile treatment effects, requires as many observations as possible, and therefore the combination of all trials addressing the same question has a better chance of being conclusive than any of the trials taken in isolation. Due to the larger sample size, a meta-analysis yields a more powerful statistical test and an increased precision of the treatment effect under consideration. In addition, individual trials may be subject to various sources of bias, and a meta-analysis may then provide a less biased picture of the true treatment effect by looking at the

totality of the information, rather than at some selected subset of it: thus, in the equation below, meta-analysis reduces both the random error and the bias (or systematic error):

$$observed\ treatment\ effect\ =\ true\ treatment\ effect\ +\ bias\ +\ random\ error\ .\qquad (1)$$

Finally, when the results of various studies seem to conflict with each other, i.e. when there is heterogeneity of the treatment effect across the studies or across different patient populations, meta-analysis provides a means of presenting the information, of testing whether the apparent heterogeneity is statistically significant, and of analyzing putative sources of heterogeneity. This chapter assumes that the meta-analysis bears on randomized clinical trials, but similar principles apply to the combination of results from other types of studies, although the combination of studies having different designs and/or different outcome measures raises specific issues that are beyond the scope of this chapter.

2. Types of Meta-analyses

The reliability of a meta-analysis depends crucially on the sources that were used to gather the data from the various studies being combined. It is convenient to distinguish three broad sources of data:

1. the medical literature (meta-analysis based on literature data, or MAL). A search is undertaken to find all publications of clinical trials addressing the question of interest. When this approach is adopted, it is important to adopt a search strategy that will uncover all trials. Search strategies primarily rely on databases such as MEDLINE, EMBASE, BIOSIS, and other more specialized or geographically focused databases. The published results of all trials are then combined based on the information available in the publication, such as the value of a test statistic or its associated p-value, the proportions of events in each treatment group, the survival estimates read off the published curves in each treatment group at some meaningful time point, etc.
2. summary data (meta-analysis based on summary data, or MAS). Relevant summary data, such as the number of events of interest and the number of patients treated in each treatment group, are obtained from the principal investigators of all trials, whether published or unpublished. The search for unpublished trials can be based on databases of trial protocols as well as on the abstracts of conferences on the medical specialty of interest. This search is often difficult and requires the active involvement of opinion leaders and experts in the field.
3. individual patient data (meta-analysis based on patient data, or MAP). Data on all individual patients are obtained from the principal investigators of all trials, whether published or unpublished. The data required on each patient typically include the patient and center identification, the date of randomization, the treatment assigned by randomization

There are advantages and drawbacks to each type of meta-analysis:

1. MAL is often flawed by the presence of publication bias, whereby clinical trials

that happen to show a large (or significant) treatment benefit tend to be published faster than the others. When this is the case, a MAL will yield an overestimate of the true treatment benefit. Sometimes the statistical methods used in the publications introduce bias through exclusion of some patients (e.g. those considered impossible to evaluate or insufficiently treated), and in this case too may a MAL produce a misleading result. Another major drawback of MAL, aside from the possibility of bias, is that the data available in the publications are inadequate to perform meaningful calculations. If the outcome of interest is survival, for instance, published survival curves do not provide sufficient information, in general, for the meta-analysis to be possible without strong, unwarranted and untestable assumptions.

2. MAS is an alternative that may be of interest to avoid the biases inherent in MAL: publication bias can be avoided by including all trials, whether published or not, while exclusion bias can be avoided by considering all randomized patients. Moreover, simple summary statistics can be obtained even for outcomes that are time-related (e.g. the hazard ratio calculated from a life table analysis). MAS requires that the principal investigators of all trials be contacted, which takes time but ensures the relevance of the questions addressed by the meta-analysis, and the interpretation of its results.
3. MAP is often considered the gold standard of meta-analysis in medicine, but it requires the analyst to obtain detailed data sets from the principal investigators of all trials. This is a long and difficult process that can take several years. However, many important advantages ensue: the latest follow-up can be included on all patients; the quality of individual patient data can be controlled and questions raised in case of doubt (for instance, if the randomization sequence appears suspicious); and more detailed analyses can be performed if individual patient data are available (e.g. subgroup analyses, prognostic factor analyses, analyses of temporal patterns, etc.).

3. Statistical Principles of Meta-analysis

3.1. Estimation

Historically, the first techniques proposed to combine the results of several experiments were based on the combination of their p-values or standardized test statistics (such as chi-square test statistics). These techniques do not explicitly take into consideration the size of the trials (number of events, duration of follow-up, etc.), and they only provide an indication of the statistical significance of an overall treatment effect, regardless of its magnitude. Such techniques are too limited to be of use in medicine, where estimation of the treatment effect is just as important for decision-making as testing the hypothesis that the effect is real and not merely due to the play of chance.

The treatment effect can be estimated in various ways, even in the simplest situation where the outcome of interest is some untoward event, such as lack of response to treatment, progression of the disease, occurrence of some toxicity, or the patient's death. If in all trials the comparison is between two groups of patients randomly allocated to an experimental treatment ("Treatment") or to a control treatment ("Control"), the data can be summarized as in Table 1.

	Treatment	***Control***	***Total***
Number of failures	F_t	F_c	F
Number of successes	S_t	S_c	S
Number of patients	N_t	N_c	N
Failure rate	$r_t = F_t / N_t$	$r_c = F_c / N_c$	$r = F / N$

Table 1: Summary data of a trial assessing the effect of an experimental treatment, as compared to a control treatment, on the incidence of an untoward event

Within each trial, the effect of treatment can be estimated in terms of the absolute difference in failure rates,

$$risk\ difference = r_t - r_c, \tag{2}$$

in terms of the ratio of failure rates,

$$relative\ risk = \frac{r_t}{r_c}, \tag{3}$$

or in terms of the ratio of the odds of failure,

$$odds\ ratio = \frac{[r_t / (1 - r_t)]}{[r_c / (1 - r_c)]}. \tag{4}$$

Other measures of treatment effect may be needed for more complex outcomes. For time-to-event outcomes such as time to disease progression or time to death, which are frequently used in medical experiments, the hazard ratio is the measure of choice. It generalizes the relative risk (Eq. (3) above) by estimating the ratio of instantaneous failure rates over time (see *Survival Analysis*).

-
-
-

Bibliography

Abrams K, Jones DR, Sheldon TA, Song F, Sutton AJ. (2000). *Methods for Meta-analysis in Medical Research.* New York: John Wiley. [Excellent overview of meta-analysis in medicine]

Cooper H., Hedges L.V. (1994). *Meta-analysis: Cumulating Research Findings across Studies.* Russell Sage Foundation. [Introductory book on meta-analysis in general, with historical perspective]

Cucherat M. (1997). *Méta-Analyse des Essais Thérapeutiques*. Paris: Masson. [Very comprehensive textbook on meta-analysis of clinical trials, available only in French]

Davey-Smith G. (2001). *Clinical Meta-analysis.* New York: John Wiley. [Meta-analysis for clinicians]

Early Breast Cancer Trialists' Collaborative Group (1990). *Treatment of Early Breast Cancer. Volume I: Worldwide Evidence 1985-1990*. Oxford: Oxford University Press. [Landmark publication on the role of meta-analysis in the evaluation of adjuvant therapy for breast cancer. Provides basic methodology for meta-analysis without mathematical jargon]

Stangl DK, Berry DA. (2000). *Meta-analysis in Medicine and Health Policy.* New York: Marcel Dekker. [Bayesian techniques in meta-analysis]

Whitehead A. (2002). *Meta-analysis of Controlled Clinical Trials*. Chichester: John Wiley. [The most comprehensive textbook on meta-analysis].

Biographical Sketch

Marc E. Buyse holds degrees in engineering (Brussels University), business administration (Cranfield School of Management), statistics (Brussels University) and biostatistics (Harvard School of Public Health). He is the Executive Director of the International Drug Development Institute in Brussels, and holds a teaching and research position at the Center for Statistics, Hasselt Univesity, Diepenbeek, Belgium. He is past President of the International Society for Clinical Biostatistics, the Quetelet Society (Belgian Region of the International Biometric Society), and board member of the Society for Clinical Trials and the Meta-Analysis Group in Cancer. His interests include drug development, clinical epidemiology, clinical trials, meta-analysis, validation of surrogate markers, statistical methods for the detection of fraud, and clinical oncology.