An Overview of Meta-Analysis

Statistical and Methodological Issues in Randomized Clinical Trials

Lehana Thabane, PhD
Biostatistician
Centre for Evaluation of Medicines
St. Joseph's Heathcare
105 Main Street East, Level P1
Hamilton ON L8N 1G6

thabanl@mcmaster.ca
Fax: (905)528-7386
Tel: (905)522-1155 x3720

Assistant Professor
Department of Clinical Epidemiology & Biostatistics
Faculty of Health Sciences
McMaster University
Hamilton ON

January 5, 2004

Contents

1 Introduction

1.1 Learning Objectives: What, Why and How?

- What is meta-analysis?
- Why do meta-analysis?
- How to do meta-analysis?
- How to report meta-analysis results?
- How to interpret meta-analysis results?

1.2 What is Meta-Analysis?

- *Definition 1:* Quantitative methods for combining studies have been available since early 1900s. But the term "meta-analysis" was coined after the paper by Glass (1976). Glass defined meta-analysis as

  > "...the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the causal, narrative discussions of research studies which typify our attempts to make sense of a large volume of research literature."

- *Definitions 2:* It is a collection of statistical techniques for combining studies.
- *Definitions 3:* A summary and statistical analysis of the results of several studies testing the same relationship. It is part of a Systematic Review.

1.2.1 Basic Steps of a Systematic Review and Meta-Analysis Process

1. Formulation of a purpose(s): This involves clear apriori specification of
    - question(s)/hypotheses in both biologic and healthcare terms
    - population of interest
    - intervention(s) under investigation
    - outcome(s) (could be beneficial or harmful) for the analysis
    - scope of the review
2. Identification of relevant studies
    - have comprehensive search strategies (and documented them)
    - avoid selection bias: unpublished research, language- or country-restricted search
3. Establishing inclusion and exclusion criteria
    - should be explicit and verifiable
4. Data abstraction and acquisition
    - the process should be explicit, unbiased and verifiable
    - extract data on patient characteristics, study design/methods, results, methodologic quality
    - include all relevant measures of interest (harmful and beneficial)
    - develop mechanisms to check accuracy of data abstraction process
5. Data Analysis
    - estimate heterogeneity and the sources of variation
    - estimate the size of the overall effect
    - use qualitative summaries where data are too sparse, of too low quality, or too heterogeneous to be combined
    - assess the role patient characteristics, dose levels, duration of nature of intervention

- assess the robustness of the results to methodologic quality, inclusion/exclusion criteria, publication bias, etc

6. Dissemination of results and conclusions

    - report key aspects of the review process, methods, analysis, results
    - use appropriate graphical displays to help interpret the findings
    - discuss the limitations of the review or included studies
    - discuss the implications of the review results for policy, practice and research

1.2.2 Narrative Reviews Versus Systematic Reviews

Features of narrative reviews and systematic reviews: Wells GA "Developing a protocol: The analysis plan" Cochrane Reviewers Workshop, November 22, 2002.

	NARRATIVE	SYSTEMATIC
QUESTION	Broad	Focused
SOURCES/ SEARCH	Usually unspecified; possibly biased	Comprehensive; explicit
SELECTION	Unspecified; possibly biased	Criterion-based; uniformly applied
APPRAISAL	Variable	Rigorous
SYNTHESIS	Usually qualitative	Quantitative
INFERENCE	Sometimes evidence-based	Usually evidence-based

1.2.3 Advantages of Systematic Reviews

Reference: Sutton AJ *et al*, 2000:

1. Democratization of the research and its uses: makes research findings more accessible to the general public
2. Provides knowledge base for policy makers, practitioners: decision-making on new technologies, development of practice guidelines
3. Helps to identify knowledge gaps and prevailing degrees of uncertainty:
    - it is a good learning process
    - provides good source for information for research funders: identify knowledge gaps to set priorities, avoid unnecessary duplication of past research
4. Aids the cumulative development of science:
    - successful research builds on previous efforts
    - helps research community to make sense of the past and plan for future research
    - permits international replication of research and testing of theories

1.2.4 Systematic Review: References on What, Why and How

- Cook DL, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care for Potsdam Consultation on Meta-Analysis. *J Clin Epidemiology* 1995; **48**: 167-71
- The Cochrane Collaboration. *Cochrane Reviewers' Handbook 4.1.5 and Glossary* 2002
- Sutton AJ *et al. Methods for Meta-Analysis in Medical Research.* John Wiley & Sons, New York, 2000
- Lang TA, Secic M. *How to report Statistics in Medicine.* Medical Writing & Communication, Philadelphia, 1997.
- Oxman AD, Cook DJ, Guyatt GH. User's guide to the medical literature: VI. How to use an overview. *JAMA* 1994; *272*: 1376-71
- Cook DJ, Mulrow CD, Haynes RB. Systematic Review. Synthesis of Best Evidence for Clinical Decisions. *Ann Intern Med* 1997; **126(5)**: 376-380

1.3 Why do Meta-analysis?

1.3.1 Purposes of Meta-analysis

The basic requirement of meta-analysis is that the topic under investigation should have some clinical importance and be biologically plausible. Meta-analysis may have several goals or objectives (adopted from Lang and Secic, 1997):

- to summarize a large and complex body of literature on a topic
- to resolve conflicting research reports in the literature
- to clarify or quantify the strengths and weaknesses of studies on a topic
- to document the need for a major clinical trial
- to avoid the time and expense of conducting a clinical trial
- to increase statistical power by combining many smaller studies:
  - *Statistical Power* refers to the probability of detecting a finding (eg clinically important difference) of certain size if one truly exists.
  - Note that combining studies provides a larger sample → increase in statistical power
  - Increase the evidence for, or confidence in, a finding/conclusion
- to improve the precision of an estimated treatment effect:
  - *Precision* refers to the degree of accuracy in the estimation, usually measured by the reciprocal of the variance or standard deviation of the estimate (ie Precision =1/standard deviation)
- to detect smaller treatment effects that have been reported
- to investigate variations in treatment effects through subgroup (or stratified) analysis
- to investigate or improve the generalizability of known treatment effects

2 Meta-analysis Plan

1. Provide description of the studies
   - sample sizes
   - patient characteristics
   - interventions used by each study
   - outcomes used
   - assessment of methodoligic quality of the studies
2. Describe the details of outcomes
   - type of outcome: discrete, categorical, continuous
   - level of measurement: nominal, binary, ordinal, interval
3. Provide description of effect measures

Outcome Type	Measurement level	Effect Measure
Discrete	binary	Odds ratio (OR) Relative risk (RR) Risk difference (RD) Number Needed to treat(NNT)
Continuous	interval	Mean difference (MD) Standardized mean difference (SMD)

4. State methods for pooling data
    - fixed effects model: assumes common effect for all studies
    - random effects model: assumes effect estimates vary across studies
        - within-study variation (sampling error)
        - between-study variation (heterogeneity)
5. Assessment of heterogeneity
6. Perform sensitivity and secondary analyses
    - specify subgroup or secondary analyses apriori
    - test robustness of results to key features of the studies, assumptions, decisions, etc
7. Assess publication bias: plot effect size against study size

3 Exploring the Between-Study Heterogeneity

3.1 Graphical Methods

3.1.1 Plot of Normalized Z-scores

- Calculate the z-score

$$z_i = \frac{T_i - \bar{T}}{\sqrt{V_i}} \text{ for } i = 1, 2, \cdots, k$$

where

$$\begin{aligned} T_i &= \text{Estimate of the treatment effect for Study } i \\ V_i &= \text{Variance of the Estimate } T_i \\ \bar{T} &= \frac{\sum_{i=1}^{k} W_i T_i}{\sum_{i=1}^{k} W_i} = \text{Weighted average of } T_i's \\ W_i &= \frac{1}{V_i} = \text{Weight for Study } i \end{aligned}$$

- Plot a histogram of the z-scores.
- Super-impose the plot of standard normal distribution on the histogram. The spread of the empirical distribution (histogram) greater than that of the standard normal distribution → more variation than can be expected by chance→ possible heterogeneity
- This approach is good for large number of studies
- **Example 1:** *Histogram of z-score for 34 cholesterol trials*

3.1.2 Forest Plot

- Plots of CIs of treatments effects (eg a 95% CI: $T_i \pm 1.96\sqrt{V_i}$)
- These are commonly used to display results of meta-analysis, but are quite useful in exploratory stage of a meta-analysis
- Size of the plotting symbol is used to mark the point estimate T_i and is made to be proportional to W_i, its precision
- Most precise estimates (ie those with large weights) are given large plotting symbols
- Large plotting symbols plus variability on CIs $\rightarrow$ visual judgements on the variability between study estimates
- **Example 2:** *Forest plot for 34 cholesterol trials*

3.1.3 L'Abbe Plots

- First described by L'Abbe KA, Detsky AS, O'Rourke K. Meta-Analysis in clinical research. *Annals of Internal Medicine* 1987; **107**: 224-33
- Useful for trials with binary outcomes: Scatter plot of event risk (# events in an arm/total # patients in arm) in treatment group versus event risk in control group
- If homogeneous, the points would form a 'cloud' close to a straight-line
- Large deviations or scatter from the line → possible heterogeneity
- Slope of the line corresponds to the pooled estimate
- Plotting symbols are proportional to precision of study estimates
- **Example 3:** *L'Abbe plot for 34 cholesterol trials*

3.2 Formal Test of Homogeneity

- Let $\theta_1, \theta_2, \cdots, \theta_k$ represent the treatment effects in k studies
- The test of homogeneity tests the hypothesis that the treatment effects are the same in all studies. That is

$$\begin{aligned} H_0 &: \theta_1 = \theta_2 = \cdots = \theta_k \\ H_a &: \text{At least one is different from others} \end{aligned}$$

- Essentially, this is testing whether all studies are estimating a single underlying treatment effect, θ (say) and whether the variation in the study estimates is due to chance
- *Test Statistic:*

$$Q = \sum_{i=1}^{k} W_i \left(T_i - \bar{T}\right)^2$$

- *A computationally convenient test statistic:*

$$Q = \sum_{i=1}^{k} W_i T_i^2 - \frac{\left(\sum\limits_{i=1}^{k} W_i T_i\right)^2}{\sum\limits_{i=1}^{k} W_i}$$

- Under H_0, Q has an approximate χ^2-distribution with $k-1$ degrees of freedom
- Reject H_0 if p-value of the test is less than $\alpha = 0.10$
- *Important Remarks/Caution:* Interpretation of the test is often difficult for the following reasons:
  - the statistical power of tests of heterogeneity are, in most cases, very low due to the small number of combined studies
  - when the sample sizes in each study are very large, H_0 may be rejected even when the study estimates do not really differ that much
  - likelihood of design flaws and publication biases
- **Example 4:** *Test of Homogeneity*

3.3 Possible Causes of Heterogeneity

Heterogeneity can be due to several causes:

- chance
- spurious, due to scale used to measure the treatment effect
- treatment characteristics (this can be investigated)
- patient-level covariates (sometimes possible to investigate)
- unexplainable factors
- characteristics of the design and conduct of the studies
- even apparently similar trials can differ in many ways:
  - differences in inclusion-exclusion criteria
  - other pertinent differences in baseline states of available participants despite identical selection criteria
  - variability in control or treatment interventions (eg doses, timing, brand, etc)
  - broader variability in management (eg pharmacological co-interventions, responses to intermediate outcomes, cross-overs, differences in patient care settings)
  - differences in outcome measures: variability in follow-up times, subtle differences in definition of outcomes
  - variation in analysis: handling of withdraws, drop-outs, cross-overs
  - variation in quality of designs and execution

3.4 How to Deal With Heterogeneity

There are several possible ways of dealing with heterogeneity:

- *Ignore it* and use fixed effects model (not recommended)
- *Test for it* and do not pool results if studies are significantly heterogeneous
- *Incorporate it*:
  - assume that heterogeneity is due to random differences among studies whose sources can not be identified
  - use random effects model
- *Explain it*:
  1. Fixed Effects Approach
     - assume that heterogeneity is systematic (derived from identifiable differences among studies)
     - use fixed effects model and explain heterogeneity via
       * meta-regression (controlling for the effects of study covariates)
       * subgroup analyses (eg subsets of studies or patients)
  2. Mixed Effects Approach
     - Assume that heterogeneity is partly systematic (derived from identifiable differences among studies) and partly random
     - Use mixed effects model (includes both random and fixed effects)

4 Meta-analytic Methods for Binary Outcomes

- General data-structure for binary outcomes for a single RCT:

	Failure /Dead	Success /Alive	Risk	Estimate
New Treatment	a	b	π_T	$P_T = \frac{a}{a+b}$
Control	c	d	π_C	$P_C = \frac{c}{c+d}$

- Relative measures of risk:

  1. Risk Difference (RD)= $\pi_C - \pi_T$
  2. Relative Risk (RR) = $\frac{\pi_C}{\pi_T}$
  3. Odds Ratio (OR)= $\frac{\pi_C/(1-\pi_C)}{\pi_T/(1-\pi_T)}$

- Hypotheses based on different measures

Measure	Hypothesis
RD	$H_0: \ \pi_C - \pi_T = 0$
RR	$H_0: \ \frac{\pi_C}{\pi_T} = 1$
OR	$H_0: \ OR = 1$

4.1 Fixed Effects Model

For a fixed-effects model, all k treatment effects $(\theta_1, \theta_2, \cdots, \theta_k)$ are assumed to be equal.

- That is $\theta_1 = \theta_k = \cdots = \theta_k = \theta$ where θ is assumed to be the common underlying treatment effect.
- Note that this equivalent to assuming that there is no heterogeneity (ie a test of homogeneity is not rejected; see Section 2)

4.1.1 The Mantel-Haenszel (MH) Method for OR

- *Pooled Point Estimate of OR*:

$$\bar{T}_{OR} = \frac{\sum_{i=1}^{k} a_i d_i / n_i}{\sum_{i=1}^{k} b_i c_i / n_i}$$

  where a_i, b_i, c_i and d_i are the four cells of 2×2 table for Study i and n_i is the total number of people in Study i

- On the logarithmic scale: $\ln(\bar{T}_{OR})$ is Normal with variance given by

$$V_i = \frac{\sum_{i=1}^{k} P_i R_i}{2\left(\sum_{i=1}^{k} R_i\right)^2} + \frac{\sum_{i=1}^{k} P_i S_i + Q_i R_i}{2\left(\sum_{i=1}^{k} R_i\right)\left(\sum_{i=1}^{k} S_i\right)} + \frac{\sum_{i=1}^{k} Q_i S_i}{2\left(\sum_{i=1}^{k} S_i\right)^2}$$

  where

$$\begin{aligned} P_i &= \frac{a_i + d_i}{n_i} \\ Q_i &= \frac{b_i + c_i}{n_i} \\ R_i &= \frac{a_i d_i}{n_i} \text{ and } S_i = \frac{b_i c_i}{n_i} \end{aligned}$$

- *Test Statistic:*

$$Z = \frac{\ln(\bar{T}_{OR})}{\sqrt{V_i}} \sim N(0,1)$$

- 95% *CI for OR:*

$$\exp\left\{\ln(\bar{T}_{OR}) \pm 1.96\sqrt{V_i}\right\}$$

- **Example 5:** *HM-Method Example*

4.1.2 The Peto Method for OR

- *Pooled Point Estimate of OR*:

$$\bar{T}_{OR}^{p} = \exp\left\{\frac{\sum_{i=1}^{k}(O_i - E_i)}{\sum_{i=1}^{k} v_i}\right\}$$

where

$$\begin{aligned}
O_i &= \text{Number of events in treatment group for Study } i \\
E_i &= \frac{n_{ti} \times d_i}{n_i} \\
&= \text{Expected Number of events} \\
v_i &= \frac{E_i(n_i - n_{ti})(n_i - d_i)}{n_i(n_i - 1)} \\
n_{ti} &= \text{Total number of patients in treatment group for Study } i \\
R_i &= \frac{a_i d_i}{n_i} \text{ and } S_i = \frac{b_i c_i}{n_i}
\end{aligned}$$

- On the logarithmic scale: $\ln(\bar{T}_{OR}^{p})$ is Normal with variance given by

$$V_i = \frac{1}{\sum_{i=1}^{k} v_i}$$

- *Test Statistic:*

$$Z = \frac{\ln(\bar{T}_{OR}^{p})}{\sqrt{V_i}} \sim N(0,1)$$

- 95% *(non-symmetric) CI for OR:*

$$\exp\left\{\frac{\sum_{i=1}^{k}(O_i - E_i) \pm 1.96\sqrt{\sum_{i=1}^{k} v_i}}{\sum_{i=1}^{k} v_i}\right\}$$

- **Example 6:** *Peto Method Example*

4.1.3 The Inverse Variance-weighted (W) Method

- *Pooled Point Estimate of OR*:

$$\bar{T}_{OR}^{w} = \frac{\sum_{i=1}^{k} W_i T_i}{\sum_{i=1}^{k} W_i}$$

where

$$\begin{aligned} T_i &= \ln\left(\frac{a_i d_i}{b_i c_i}\right) \\ &= \text{Estimate of } \ln(OR) \text{ for Study } i \\ V_i &= \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i} \\ &= \text{Variance of the Estimate } T_i \\ \bar{W}_i &= \frac{1}{V_i} = \text{Weight for Study } i \end{aligned}$$

- *Test Statistic:*

$$Z = \frac{\ln(\bar{T}_{OR})}{\sqrt{V_i}} \sim N(0,1)$$

- 95% *CI for OR:*

$$\exp\left\{\ln(\bar{T}_{OR}) \pm 1.96\sqrt{V_i}\right\}$$

- **Example 7:** *The Weighted Method Example*

4.1.4 Relative Merits/Demerits: Which Method to Use?

Generally, if the sample sizes of the studies are large (ie all cells ≥ 5), all the methods produce comparable results

- *The HM method*:
  1. easy to compute
  2. works well if k, the number of studies to be combined, is large, but the within-study sample size is small
  3. not recommended when the cells in individual 2×2 tables are zero (unless a continuity correction is used)
- *The Peto Method*:
  1. computationally intensive
  2. an improvement over the HM method because it can handle zero cells in individual 2×2 tables
  3. produces biased estimates of OR and corresponding variances when number of patients in treatment arms within studies are not balanced
  4. bias (underestimation) is also possible if OR is far from unity(ie large or small treatment effects)
- *The W Method*:
  1. easy to compute
  2. works well if k, the number of studies to be combined, is small, but the within-study sample sizes are large

4.2 Random Effects Model: The W Method

- The random effects model the underlying effects vary from trial to trial.
- Specifically, we assume that the estimate T_i for θ_i can be expressed as

$$T_i = \theta_i + \epsilon_i$$

  where ϵ_i is the error with which T_i estimates θ_i.

- $\text{Var}(T_i) = \tau_\theta^2 + v_i$ where τ_θ^2 is random effects variance and v_i is the sampling variance within Study i.
- Equivalently, the random effects model assumes that

$$T_i \sim N\left(\theta, \tau_\theta^2\right)$$

- The variance τ_θ^2 measures the variability between treatment effects. If $\tau_\theta^2 = 0$, there is no variability between treatments and this reduces to fixed effects model.
- The test of $\tau_\theta^2 = 0$ is equivalent to the test of homogeneity based on Q which has a χ^2-distribution with $k-1$ degrees of freedom (see Section 3.2)
- The adjusted weight for Study i is given by

$$W_i^* = \frac{1}{1/W_i + \hat{\tau}_\theta^2}$$

  where τ_θ^2 is estimate of τ_θ^2

$$\begin{aligned}
\hat{\tau}_\theta^2 &= \begin{cases} 0 & : \quad Q \leq k-1 \\ \frac{Q-(k-1)}{U} & : \quad Q \geq k-1 \end{cases} \\
U &= (k-1)\left(\bar{W} - \frac{S_W^2}{k\bar{W}}\right) \\
S_W^2 &= \frac{1}{k-1}\sum_{i=1}^{k}\left(W_i - \bar{W}\right)^2 \\
\bar{W} &= \frac{1}{k}\sum_{i=1}^{k} W_i
\end{aligned}$$

- *Remarks*: Note that $W_i^* = W_i$ if $\hat{\tau}_\theta^2 = 0$
- *Pooled Point Estimate of treatment effect*:

$$\bar{T} = \frac{\sum_{i=1}^{k} W_i^* T_i}{\sum_{i=1}^{k} W_i^*}$$

  and

$$V* = \text{Var}(\bar{T}) = \frac{1}{\sum_{i=1}^{k} W_i^*}$$

- 95% *CI for OR:*

$$\exp\left\{\ln(\bar{T}) \pm 1.96\sqrt{V^*}\right\}$$

- **Example 8:** *The Random-effects Example*

5 Meta-Analytic Methods for Continuous Outcomes

Let θ_t and θ_c be mean of outcome for treatment and control groups respectively. Different measures of treatment effect (or difference between means) include:

- mean difference:

$$T_i = \bar{Y}_{ti} - \bar{Y}_{ci} \text{ and } V_i = \text{Var}(T_i) = S_i^2 \left(\frac{1}{n_{ti}} + \frac{1}{n_{ci}} \right)$$

- effect Size:

$$T_i = \frac{\bar{Y}_{ti} - \bar{Y}_{ci}}{S_i} \text{ and } V_i = \text{Var}(T_i) = \frac{1}{n_{ti}} + \frac{1}{n_{ci}}$$

  where

$$\begin{array}{rcl} \bar{Y}_{ti} & = & \text{Sample mean for treatment group for Study } i \\ \bar{Y}_{ci} & = & \text{Sample mean for control group for Study } i \\ n_{ti} & = & \text{number of participants in treatment group for Study } i \\ n_{ci} & = & \text{number of participants in control group for Study } i \\ S_i^2 & = & \text{The pooled estimate of the population variance for Study } i \end{array}$$

- hypothesis: $H_0 : \theta_t - \theta_c = 0$ (no treatment effect)

5.1 Fixed Effects Model: The W Method

- Follows the same calculations as in Section 4.1.3.

5.2 Random Effects Model

- Follows the same calculations as in Section 4.2.

5.3 Illustrative Example 9

6 Fixed-Effects vs Random-Effects: Which Model to Use?

- Neither the fixed effects model or random effects model can be considered ideal.
  - Randomized effects model uses distributional assumptions considered unrealistic or unjustified by many.
  - Random effects models are also sensitive to publication bias.
  - Fixed effects models ignore heterogeneity.
- When should random-effects model be more preferred to (or emphasized more than) fixed effects model?
  - when there is evidence of heterogeneity that cannot be explained; however, due to lack of power of homogeneity tests, random effects model would still be worth performing even if the test of homogeneity is not significant.
- Other Options include:
  - mixed effects models (includes both fixed and random components)
  - hierarchical models
  - full Bayesian models
  - empirical Bayes models

7 Criticism of Meta-analysis

Meta-analysis is not immune to problems and challenges, most of which form on-going debate among researchers. Below are some possible problems or challenges in meta-analysis (modified from Utts JM. *Seeing Through Statistics* 2nd Edition (Chapter 24); Duxbury, New York, 1999):

- *The Simpson's Paradox*: A reversal in the direction of the relationship that occurs when different data from different sources are combined.
- *Confounding variables*: There is always potential for differences across studies that may be confounded with treatments used. For example, if studies in a meta-analysis were done in different countries, cultural differences may be confounded with treatment differences.
- *Subtle differences in treatment with same name*: See Section 3.3
- *The file drawer problem/Publication bias*: There is always a high likelihood for studies that did not achieve statistical significance not to be published. This is called the *file drawer problem* because it is assumed that this studies are filed away somewhere and not accessible to the public. Therefore statistically significant studies are more likely to be included in the meta-analysis which will result in an overestimate of the treatment effect. Potential solutions:
  - if possible contact all persons known to work in the filed to inquire about studies done, but unpublished
  - estimate how many studies it will take to reduce the treatment effect to non-significance
  - *reference:* Rosenthal R. *Meta-analytic procedures for social research*, Sage Publications; CA, 1991
- *Biased or flawed studies*: If the studies include in the meta-analysis are flawed or biased, so are the meta-analysis results!
- *Statistical significance versus practical/clinical significance*: Meta-analysis is particularly likely to find statistical significance because of the increase

in power as the sample size increases. It is important to focus on the *magnitude* of the treatment effect to assess whether the effect would be clinically significant or would be of any impact to health care.

- *False findings of "no difference"*: A statistically non-significant result may also be because one does not have enough data (or power) to detect a significant result.

  Rule of Thumb: Always determine the sample size whenever "no difference" or "no relationship" is found in a study.

8 Other Important Aspects of Meta-Analysis

- Sutton AJ *et al. Methods for Meta-Analysis in Medical Research.* John Wiley & Sons, New York, 2000: Chapters 7 & 8.

8.1 Publication Bias

- *Publication Bias:* Research with statistically significant results is often found to be more likely to be submitted, published or published faster than that with non-significant results.
- *Diagnosis:* Funnel plots are commonly used (**Example 10**).
- *Solutions:* See Chapter 7 of Sutton *et al* 2000.

8.2 Study Quality

Combining poor quality studies→ biased and potentially misleading results

- Important Features of Study Quality:
  1. *Assignment/Randomization:* Random allocation of subjects to treatment groups is considered the valid basis for comparisons
  2. *Masking/Blinding:* A study is said to be "blinded" if subjects do not know which treatment group they have allocated to; "double-blinded" if both the assessor and subject are not aware of which group the subject is in. In addition to randomization, this helps to eliminate or reduce other potential biases.
  3. *Follow-up:* Documentation of drop-outs, cross-overs, etc and how these are handled in the analysis.
  4. *Statistical Analysis:* Correct methods of analysis.
- Potential solutions include use of quality scoring systems and regression approach to adjust for quality.

9 How to report Meta-Analysis Results

9.1 An Outline

Below is suggested outline (see Sutton *et al.*, 2000, Chapter 10 for details)

- *Abstract or executive summary.*
- *Background information:* Description of the problem for which the review is needed; the purpose of the review (see Section 1.3.1).
- *Hypotheses tested/question to be addressed in the review.*
- *Methods of review*: Search strategy; assessments of relevance and validity; data extraction and synthesis; etc.
- *Details of studies included in the review*: Demographic details of patient groups; interventions and outcomes of each study; study design, quality and validity.
- *Details of studies excluded in the review*: Provide reasons for exclusion.
- *Results of meta-analysis*
    1. point estimates of each study, standard errors and corresponding CIs
    2. pooled estimate and corresponding standard error
        - fixed/random effects estimate, corresponding CI and p-value of test
    3. provide a tabular summary of the relative weight of each study
    4. provide the result of test of homogeneity: Q-value and corresponding p-value
    5. report results in absolute terms (eg absolute relative risk (ARR) or number needed to treat (NNT)). This allows clinical significance or possible impact to be assessed.
- *Report analysis of the robustness of the results*: Where there is uncertainty or missing data, sensitivity analysis should be performed to assess the robustness of the results

- *Discussion*: This should include:
  1. strength of the causal evidence
  2. potential biases in the studies and the review, and limitations they place on inferences
- *Implications of the Review*: Report the potential implications of the results for health care and future research.
- *References*: Three lists should be given:
  1. studies included in the review
  2. studies excluded in the review
  3. other references cited in the review
- *Dissemination and further research*:
  1. possible target audience
  2. main lessons of the review
  3. implications for further research

9.2 Graphical Display of Meta-Analysis Results

- *Assessment of heterogeneity*: histograms, forest and L'Abbe plots
- *Assessment of publication bias*: funnel plot
- *Distribution of effect size:* boxplots, stem-and-leaf plots
- *Pooled estimates:* forest plots, stratified forest plots

9.3 Interpreting the Results: Framing Problems

- Distinguish between clinical significance and statistical significance
  1. Clinical significance can not be expressed in terms of statistical significance levels (p-values).
  2. Clinical significance can be expressed in terms of [magnitude and size] of treatment effect or difference.
  3. In assessing equivalence, remember that effect sizes provide evidence of equivalence, but p-values do not.
- Emphasize confidence intervals and de-emphasize p-values
  1. CIs are more informative than significance levels (p-values).
  2. CIs convey the precision of the estimate of treatment. effect
- Be consistent with the use of standard deviations and standard errors:
  - standard deviations or interquartile ranges should be used as descriptive summaries
  - standard errors should be used to convey the uncertainty of estimates such as treatment effects

- Do not confuse "no evidence of effect" with "evidence of no effect"
  - Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311: 485.
- Examples of statements claiming "no effect or difference":
  1. "had no effect"
  2. "the effectiveness [of intervention A] did not differ from that of [intervention B]"
  3. "there was no difference [in outcomes]"
- Examples of *poorly worded* statements claiming "no effect or difference":
  1. "appeared to have equivalent efficacy"
  2. "may be as effective"
  3. "did not appear to be effective"
  4. "was found to be no more effective"
  5. "[the risk of outcome] was similar [between the two groups]"
  6. "is not associated with clear benefit"
- Examples of *appropriately worded* statements:
  1. "there was no statistically significant effect/difference"
  2. "there was insufficient evidence to support or refute"
- *Reference:* Alderson P, Chalmers I. Claims in abstracts of Cochrane reviews that health care interventions have 'no effect'. *BMJ* 2003; 326: 475.

9.4 Interpreting the Results: the big picture

Usually meta-analysis results are used to inform decision making process. However, there are many things to consider:

- costs of implementing recommendations based on the results
- the scope of the problem based on epidemiologcal data
- other sources of information that may not have covered by the review

It is important to look at the 'big picture' in terms of what the implications of the evidence (as presented by meta-analysis results) in terms of how they relate to the decision making process. Consider

- the strength of the evidence:
  - was the retrieval process exhaustive?
  - is there any potential bias in the results due to methodological quality?
  - weigh statistical significance versus clinical significance
- how applicable are the results to the situation at hand? Has the review covered:
  - all applicable patients?
  - all aspects of the intervention?
  - all relevant comparisons?
  - all important (harmful or beneficial) outcomes?
- what are the possible trade-offs?
  - how are the decisions made likely to impact patient/practitioner preferences or values?

10 Software for Meta-Analysis

- Commercial
  - Comprehensive Meta-Analysis: www.meta-analysis.com
  - MetaWin: www.metawinsoft.com
  - WEasyMa: www.weasyma.com
- Free-ware
  - RevMan (Review Manager) developed by the Cochrane Collaboration: www.cochrane.org
  - Meta-Analysis Version 5.3: www.statistics.com/content/freesoft/mno/meta-ana53.html
- Statistics software with meta-analytic tools or macros
  - SAS: www.sas.com
  - STATA: www.stata.com
  - WinBUGS: www.mrc-bsu.cam.ac.uk/bugs WinBUGS is a package of programs for Bayesian analysis including Bayesian meta-analysis
- Links to other meta-analysis software
  - http://www.prw.le.ac.uk/epidemio/personal/ajs22/meta
  - http://epiweb.massey.ac.nz/meta_analysis_software.htm

References

[1] Glass GV. Primary, secondary and meta-analysis of research. *Educational Researcher* 1976; **5**: 3-8.

[2] Whitehead A. *Meta-Analysis of Controlled Clinical Trials*, Wiley: New York, 2002.

[3] Stangl DK, Berry DA. *Meta-Analysis in Medicine and Health Policy*, Marcel Dekker: New York, 2000.

[4] Cook DL, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care for Potsdam Consultation on Meta-Analysis. *J Clin Epidemiology* 1995; **48**: 167-71.

[5] The Cochrane Collaboration. *Cochrane Reviewers' Handbook 4.1.5 and Glossary* 2002.

[6] Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-Analysis in Medical Research*. John Wiley & Sons, New York, 2000.

[7] Rosenthal R. *Meta-analytic procedures for social research*, Sage Publications; CA, 1991.

[8] Lang TA, Secic M. *How to report Statistics in Medicine*. Medical Writing & Communication, Philadelphia, 1997.

[9] xman AD, Cook DJ, Guyatt GH. User's guide to the medical literature: VI. How to use an overview. *JAMA* 1994; *272*: 1376-71.

[10] Cook DJ, Mulrow CD, Haynes RB. Systematic Review. Synthesis of Best Evidence for Clinical Decisions. *Ann Intern Med* 1997; **126(5)**: 376-380.

[11] L'Abbe KA, Detsky AS, O'Rourke K. Meta-Analysis in clinical research. *Ann Int Med* 1987; 107: 224-33.

[12] Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311: 485.

[13] Alderson P, Chalmers I. Claims in abstracts of Cochrane reviews that health care interventions have 'no effect'. *BMJ* 2003, 326:475.

[14] Wells GA. Developing a protocol: The analysis plan. Cochrane Reviewers Workshop, November 22, 2002.