SYSTEMATIC REVIEW AND META-ANALYSIS: A GLOSSARY

Wong LP
Health Research Development Unit (HeRDU), Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia

ABSTRACT: Systematic review is a comprehensive review of research findings in which all of the primary studies are systematically identified, appraised and summarised using an explicit and reproducible methodology. Meta-analysis is the statistical component of a systematic review in which combinable studies are drawn together via a statistical process. Systematic reviews and meta-analyses are routinely being used in the evidence-based approach to medicine. These short notes intend to highlight important terms in systematic-review and meta-analysis. It is a beginner's guide for health care professional of any discipline involved in research or practice who seeks to gain more comprehensive understanding of important terms used in systematic review and meta-analysis. *(JUMMEC 2007; 10(1): 3-10)*

KEYWORDS: Systematic review, meta-analysis, glossary

Introduction

Important medical questions are often studied more than once by different research teams in different locations. Systematic reviews are a type of secondary research that evaluate the results of previous research, usually randomised trials that have addressed the same clinical question. A systematic review is a comprehensive survey of a topic in which all of the primary studies are systematically identified, appraised and then summarised according to an explicit and reproducible methodology. The review aims to review clearly formulated questions, using explicit methodology to minimise bias in the location, selection, critical evaluation and synthesis of research evidence. It is an objective way of assembling, assessing and summarising evidence to give a full and fair evaluation of the treatment under investigation and to provide a structured basis for evidence-based medicine. Where possible, the results of individual studies are combined in a meta-analysis. Systematic review of the literature can be applied to any form of research question. Often, a systematic review will include a meta-analysis.

Meta-analysis is a quantitative method of combining the results of research studies to provide overall summary statistics. Often, many trials lack power (i.e. adequate sample size) to achieve statistically significant results. Combining the results of similar trials in a meta-analysis may give sufficient statistical power to reach a clear and more reliable answer. A good quality meta-analysis should always be done in the context of a systematic review. These glossaries highlight important terms in systematic review and meta-analysis as a beginner's guide for researchers of any discipline.

Allocation concealment: Process used in studies that involve at least two groups receiving different interventions or treatments where allocation to treatment is done in such a way that the participants and health care providers do not know which intervention the participant is to receive. Allocation concealment aims to avoid bias during the group allocation process so that the intervention and control groups are similar. Assessment of methodological quality in systematic reviews should consider whether allocation was adequately concealed. Normally, studies would be excluded from inclusion if no allocation concealment was used or if there was uncertainty about the allocation concealment.

Apples and oranges: When examining the results of a meta-analysis, the question often asked is: Were apples combined with oranges? If the pooled studies are too dissimilar, the meta-analysis may be combining apples and oranges, rather than different types of oranges. The problem of heterogeneity arises, resulting from the

Correspondence:
Wong LP
Health Research Development Unit (HeRDU)
Faculty of Medicine, University of Malaya
50603 Kuala Lumpur, Malaysia
Email: wonglp@ummc.edu.my

basic differences that exist between trials. This includes differences such as in the quality of their designs, the eligibility criteria for inclusion or the treatments used. If the trials are too dissimilar, combining trials may increase the statistical precision of the result, but it decreases its clinical reliability. Trials should be combined only if they are sufficiently alike.

Bias: Distortion of the outcome is due to known or unknown differences between groups other than intervention. Bias in meta-analysis arises mainly from a) choice of studies included in the meta-analysis and b) how the results of selected studies are combined to produce an overall effect estimate. Selection bias refers to bias in how studies are selected for inclusion. Publication bias is inclusion of only published studies. This is because studies which do not show the intervention to be effective are often not published. Systematic reviews that fail to include unpublished studies may result in an overestimate of the true effect of an intervention. Attrition bias is systematic differences between comparison groups in withdrawals or exclusions of participants from the the study. Detection bias refers to differences in the assessment of outcomes. Performance bias refers to systematic differences in the care provided to the participants in the comparison groups other than the intervention under investigation.

Cochrane Collaboration: An international organisation that aims to make up-to-date information about the effects of healthcare available so that people can make well-informed healthcare decisions. The Cochrane Collaboration (http://www.cochrane.org) produces and disseminates systematic reviews of healthcare interventions and promotes the search for evidence in the form of clinical trials and other studies of interventions. The Collaboration encompasses an established network of 50 research groups worldwide that prepare and maintain Cochrane reviews, covering a range of medical specialties.

Confidence interval (CI): The CI combines information on the sample size and variance to put probabilistic bounds on estimates of an effect. In a Forest plot of log odds ratios and confidence intervals, the vertical line on the plot corresponds to an odds ratio of one, where treatment and control are equally effective (*see* Forest Plot). CI which includes this value implies no statistically significant effect was found. In other words, any interval not including one indicates significant effect.

Critical appraisal: Systematically finding, appraising and interpreting evidence of effectiveness. It is aimed to examine research to assess its validity, results and relevance before using it to form a decision.

Cumulative meta-analysis: The repeated performance of meta-analysis whenever a new trial becomes available for inclusion. In cumulative meta-analysis, studies are added one at a time in a specified order (e.g. according to date of publication or quality) and the results are summarised as each new study is added. In a graph of a cumulative meta-analysis, each horizontal line represents the cumulative summary of the results, rather than the results of a single study. In a cumulative meta-analysis plot, each study added increases the sample size and this should result in progressive narrowing of the confidence interval demonstrating, a change in point of estimate and shows how evidence has accumulated over time.

DerSimonian-Laird's method: Random effects model used in meta-analysis. It is based on the risk difference and weighted by the inverse of its variance (*see* random effects model).

Effect size: Refers to the size of a relationship between an exposure and an outcome. The term is applied to the measurement of the difference in the outcome between the study groups. Relative risk, odds ratio, and risk differences are measures of effect size. The effect size of a continuous variable is expressed as the standardised mean difference. Effect size can be measured in two ways: a) as a standardised difference between two means, or b) as a correlation between the independent variable classification and the individual scores on the dependent variable. This correlation is called the "effect size correlation".

Egger's plot: Used to investigate the possibility of a publication bias. It is a simple linear regression of the effect size in a study divided by its standard error on the inverse of standard error testing whether the intercept is statistically significant. The 95% confidence intervals of the regression line's *y* intercept should include zero if there is no evidence of publication bias (Figure 1).

Evidence-based medicine: A new approach to teaching the practice of medicine. It is a conscientious, explicit and judicious use of current best evidence in making decisions about certain aspects of medical practice. The practice of evidence-based medicine requires careful balancing and integration of three key components, namely, the best available evidence, clinical expertise and patient values. The goal of evidence-based medicine is to improve health care quality and patient outcomes across the health care system.

Fixed effect method: There are four widely used methods for estimating a combined effect estimate in meta-analysis for dichotomous outcomes, three fixed

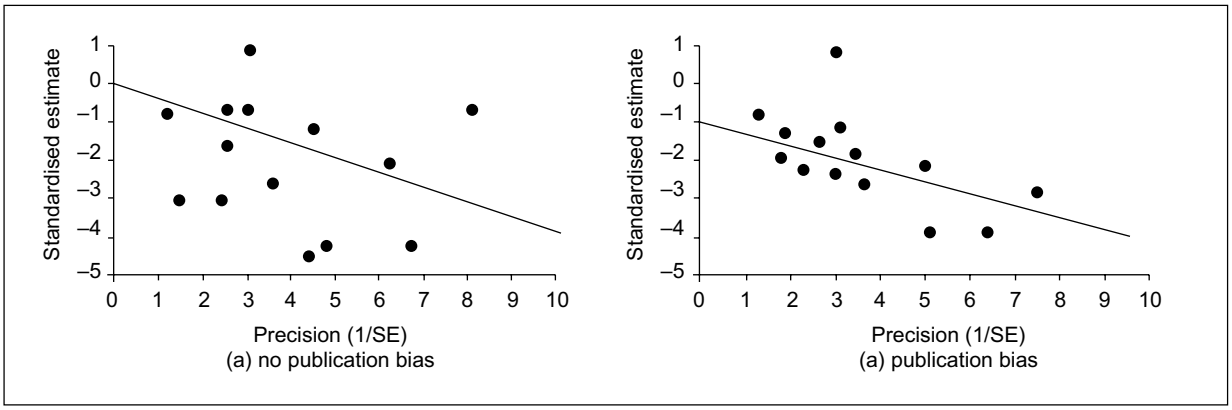

Figure 1. Regression line of estimate against precision (1/standard error)

effect methods (Mantel-Haenszel, Peto and Inverse Variance) and one random effects method (DerSimonian and Laird). The fixed effect method considers only within-study variability. Variation between the estimates of effect from each study does not affect the CI in a fixed effect model. The assumption is that studies use identical methods, patients, and measurements; that they should produce identical results; and that differences are only due to within-study variation (*see* random effect method).

Forest plot: Schematic display of the results of a meta-analysis where point estimates and 95% CIs for each study, along with the overall summary estimate and CI represented as a diamond at the bottom.The weight of each study is represented by the size of the box, indicating the estimated treatment effect. Significance is achieved if the diamond is clear of the line of no effect (Figure 2).

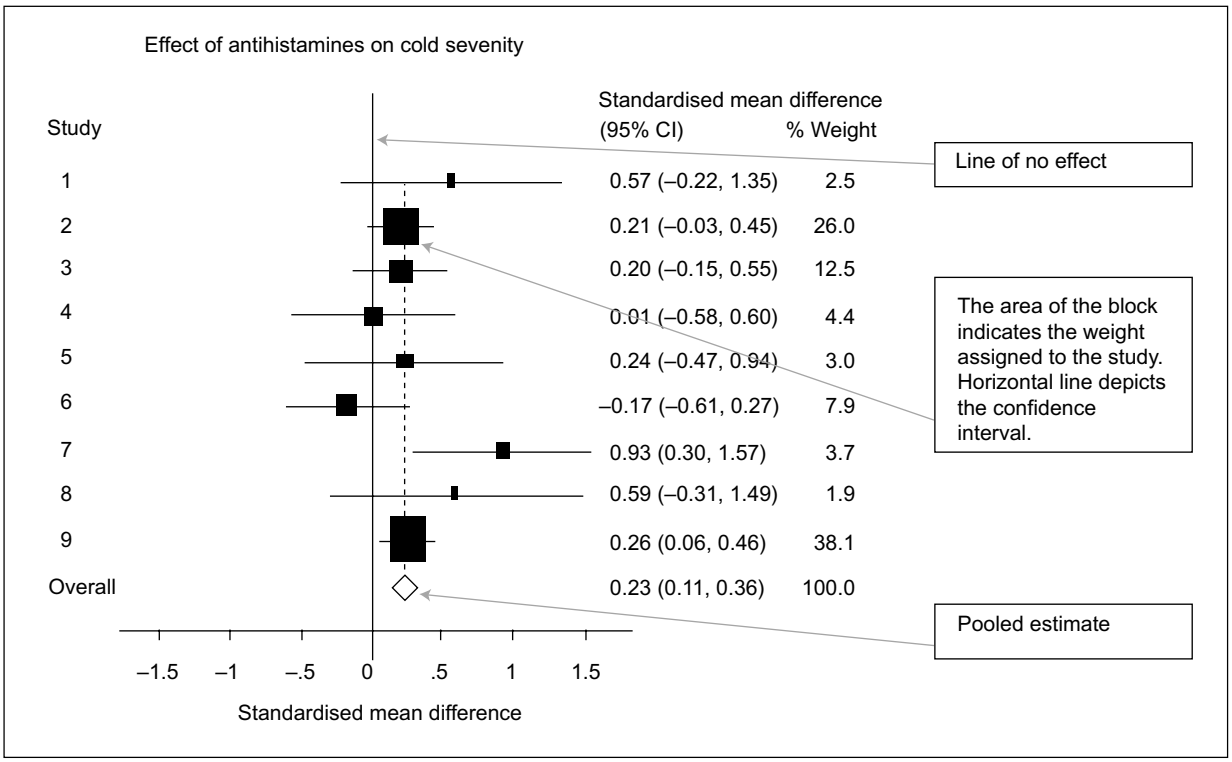

Figure 2. Forest plot meta-analysis of the effect of antihistamines on cold severity

Funnel plot: A graphical display of a plot of effect size against sample size or other indicator of precision of estimate. It is a visual tool to investigate publication bias. Funnel plots should have variance or precision (often sample size is used as a measure) on the *y*-axis. A plot should shape like an inverted funnel if there is no publication bias. Asymmetrical plots are interpreted to suggest that bias is present (Figure 3). Visual inspection of funnel plot may not be a reliable method of investigating publication bias. Egger's linear regression test and Begg's rank correlation test are two types of statistical test for the asymmetry of funnel plot.

Garbage in, garbage out: A common criticism of meta-analysis refers to meta-analyses of studies where the primary studies and the selection of studies for inclusion are poorly conducted. The major concern is the quality of the primary research included in systematic review. If invalid studies are pooled, the resulting overall estimate will also be invalid. Therefore, studies with methodological flaws should not be included in the meta-analysis.

Heterogeneity: The variation between the pooled studies. Heterogeneity in meta-analyses creates difficulty in drawing overall conclusions. There will usually be variations in patient groups, clinical settings, concomitant care and method of delivery of intervention as the trials are not conducted according to a common protocol. Variability between the results of studies can be examined by a test of homogeneity (see test of homogeneity). Common methods for investigating and dealing with sources of heterogeneity are sensitivity analysis, subgroup analysis, meta-regression, and cumulative meta-analysis.

Homogeneity: Homogeneity measures the differences or similarities between the several studies. If several studies reach nearly the same conclusion, one can combine the data with reasonable confidence. If the studies differ greatly in their outcomes, one should be more cautious about combining the data. Test of homogeneity may be used to assess homogeneity. To reach homogeneity, the authors remove the most extreme effect sizes, irrespective of whether they were extremely high or extremely low, until homogeneity is reached – if possible. Otherwise, the studies cannot be compared with one another with confidence.

Inverse of variance: The inverse variance method of performing a meta-analysis is so named because the weight given to each study is chosen to be the inverse of the variance of the effect estimate (i.e. one over the square of its standard error). Thus larger studies, which have smaller standard errors, are given more weight than smaller studies, which have larger standard errors. The choice of this weight minimises the imprecision (uncertainty) of the pooled effect estimate.

L'Abbé plot: Usually used for meta-analysis of randomised controlled trials (RCTs) where the outcome is a binary variable. The L'Abbé plot shows the proportion with the outcome in each group (e.g. outcome in the treatment group versus outcome in the control; outcome in the exposed group versus outcome in the unexposed group). If the trials are fairly

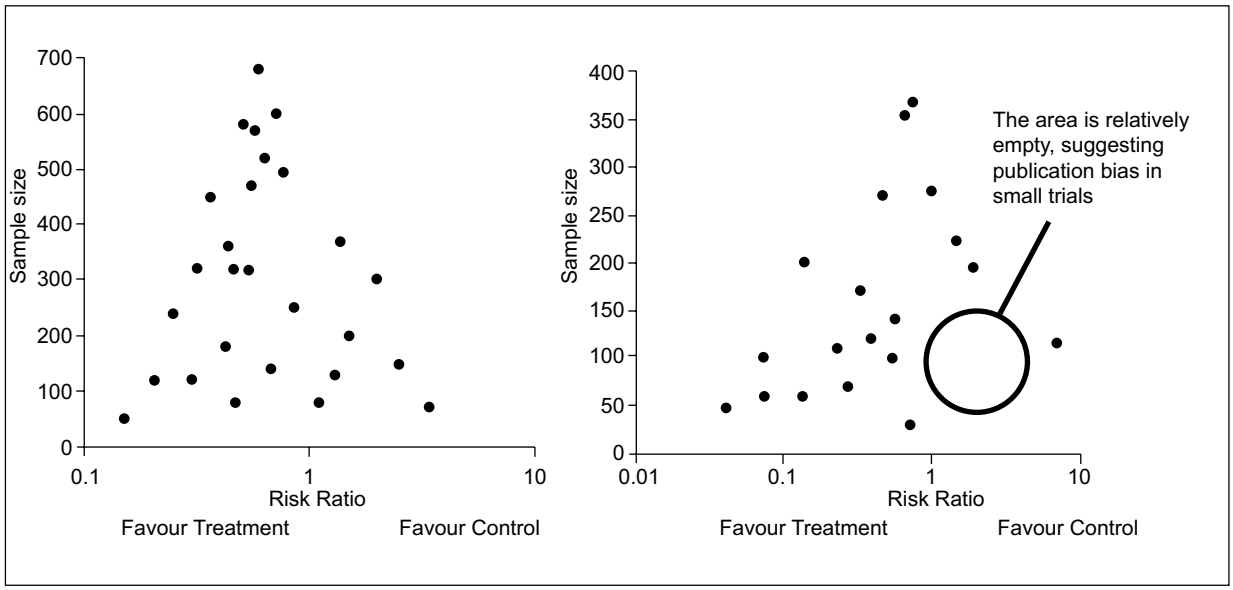

Figure 3. Funnel plot showing evidence of publication bias

homogeneous, the points should form a cloud close to a line, the gradient of which would correspond to the pooled treatment effect. Large deviations or scatter indicate possible heterogeneity (Figure 4).

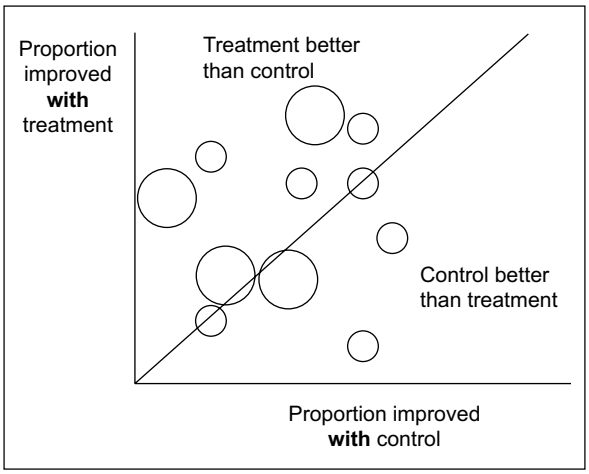

Figure 4. L'Abbé plot shows the proportion with the outcome in each group versus outcome in the control

Limitations of meta-analysis: Authors of the meta-analysis must assess the limitations of their analysis and decide what conclusions to state. Meta-analysis can be subject to data limitations and if the included studies have bias or flaws, the meta-analysis will also be flawed. Secondly, if there is a lack of consistency across studies in the composition of study population, study design or outcome measures, it will be difficult to generalise the results.

Log-odds ratio: The log of the odds ratio, used in statistical calculations in the graphical displays of odds ratios in systematic review.

Mantel-Haenszel's method: A statistical method (fixed effect) for pooling individual studies (relative risk, odds ratio and rate ratio). Mantel-Haenszel method is a method of stratified analysis of data. In Mantel-Haenszel's method, each study is considered a strata.

Meta-analysis: A method of combining independent studies that have investigated the same question and used similar study methods to produce a single estimate (Figure 5). It is often used as part of a systematic review but can be performed on studies that are not part of a systematic review. The main aim of a meta-analysis is to produce an estimate of the average effect, and the direction and magnitude of the average effect is intended as a guide in making decisions about clinical practice. The estimate of an average effect is calculated as weighted average, defined as:

$$\text{Weighted average} = \frac{\text{sum of (estimate} \times \text{weight)}}{\text{sum of weights}} = \frac{\sum T_i W_i}{\sum W_i}$$

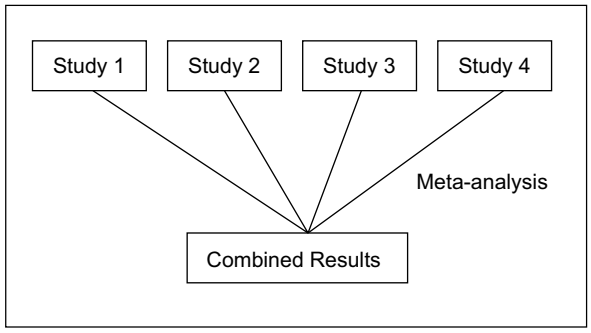

Figure 5. Combination of data from independent studies to produce a single estimate

where T_i is the treatment effect estimated in study *i*, W_i is the weight given to study *i*.

Meta-analysis of Individual Patient Data (IPD): Obtaining information or raw data on all patients included in each of the trials directly from those responsible for the trial. Collecting individual patient data (IPD) has been described as the 'gold standard' for undertaking meta-analysis. This method relies heavily on the international cooperation between the individuals and groups who have conducted relevant trials. IPD may be very time consuming and resource intensive, nevertheless, it has several advantages such as possibility of checking data consistency, provide update follow-up data, and ensure the appropriateness of the analyses.

Meta-regression: Meta-regression is an extension to meta-analysis, and a generalisation of subgroup analyses, that can be used to investigate heterogeneity of effects across studies. It examines the relationship between one or more study-level characteristics and the sizes of effect observed in the studies. Meta-regression can formally test whether there is evidence of different effects in different subgroups of trials. Figure 6 shows a weighted regression line between treatment effect and

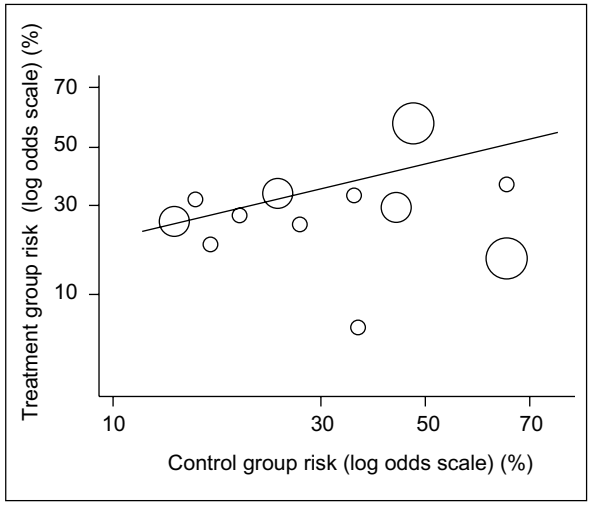

Figure 6. Meta-regression of treatment effect and underlying risk

underlying risk. Circles indicating the studies and size of circle show the weight of each study.

Narrative review: A "traditional" or "narrative" review (also known as overviews) may be no more than a subjective assessment by an expert using a select group of materials to support their conclusion and are therefore prone to bias and error. The procedure ignores search strategy, sample size, effect size and research design. The common way to review the results is to count the number of studies supporting various sides of an issue and to choose the view receiving the most votes.

Outlier: An outlier study in meta-analysis is a study that has results very different from the rest of the studies. An outlier could alter the conclusions of a meta-analysis. If a study appears to be an outlier and the inclusion of its data into the meta-analysis resulted in heterogeneity, it should be excluded from the meta-analysis.

Peto's method: Peto's method of performing meta-analysis is similar to Mantel Haenszel's but computationally simpler. It can be regarded as a modification of the Mantel Haenszel's method.

Pooled estimate: Pooled estimate is the weighted average of each individual sample's variance estimate. In a Forest plot the centre of the diamond represents the pooled point estimate (*see* Figure 2).

Publication bias: The phenomenon by which significant and positive results are more likely to be reported, and reported more prominently, than non-significant and negative results. A simple analysis of funnel plots provides a useful test for the likely presence of bias in meta-analyses. Funnel plots, plots of the trials' effect estimates against a measure of precision such as sample size, are skewed and asymmetrical in the presence of publication bias and other biases. Statistical tests for publication bias are Begg's test and Egger's test.

Qualitative meta-analysis: To appraise the quality of qualitative research results for inclusion in systematic review. Meta-synthesis of qualitative research basically follows the same, replicable procedure of a quantitative meta-analysis. However, unlike quantitative research synthesis, a qualitative meta-analysis is interpretative rather than aggregative. It involves analysis of the theories, methods and findings of qualitative research and the synthesis of these insights into new ways of thinking about a phenomenon. Instead of a statistical data analysis, the researcher analyses textual reports, creating new interpretations in the analysis process.

Random effects model: Considers both between-study and within-study variability. The assumption is that individual studies are a random sample from the universe of all possible studies. The DerSimonian Laird statistic is based on a random effects model. Random effects models are more *conservative* and generate wider confidence intervals. A random effects model is less likely to show a significant treatment effect than a fixed effects model and give wider CIs than fixed effect models.

Retrieval of studies: Includes three important steps. First, reviewer decides on the comprehensiveness of the search. Once potentially useful studies are identified, they must be obtained. Finally, the reviewer must determine which studies to include in the review. Search strategies and inclusion criteria must be clearly defined.

Search strategy: Description of the methodology to be used to locate and identify research articles pertinent to a systematic review. It includes a list of search terms, based on the subject, intervention and outcome of the review to be used when searching electronic databases, websites, reference lists and when engaging with personal contacts.

Selection bias: The introduction of error due to systematic differences in the characteristics of those selected to participate in a study, or receive an intervention. Selection bias is also used to describe a systematic error in reviews due to how studies are selected for inclusion.

Sensitivity analysis: Repetition of the analysis using different sets of assumptions to determine the impact of variation arising from these assumptions. Sensitivity analysis may examine the consistency of results across various subgroups (e.g. patient group, type of intervention or setting). Sensitivity analysis was undertaken using subgroup analysis.

Subgroup analysis: Subgroup analyses are meta-analyses on subgroups of the studies aimed to determine if the effects of an intervention vary between subgroups. Subgroups may be predefined according to many factors including: differences in subject populations, intervention, and outcome and study design.

Systematic review: A review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review. Systematic reviews are often carried out to find all the results of all the studies on a particular topic, and then the

Table 1. Systematic review versus narrative review

Characteristic	Systematic Review	Narrative Review
Search strategy	Comprehensive search of many database as well as gray literature. Uses explicit search strategy	No explicit methods for searching or reporting results
Selection criteria	Criterion-based selection, uniformly applied	Usually not specified, potentially biased
Article review or appraisal	Rigorous critical appraisal, typically using a data extraction form	Variable, depending on who is conducting the review
Assessment of study quality	Assessment of quality is always included as part of the data extraction process	May not use formal quality assessment
Synthesis	Quantitative summary (meta-analysis) if the data can be apropriately pooled; qualitative otherwise	Often a qualitative summary
Replicate	Can be replicated	Cannot be replicated

results may be combined together using meta-analysis to produce one overall result. A systematic review may, or may not, include a meta-analysis, however, a quantitative systematic review is synonymous with meta-analysis.

In contrast, a traditional or narrative review may be no more than a subjective assessment by an expert using a select group of materials to support their conclusion (Table 1).

Test of homogeneity: A statistical test to assess whether individual study results are likely to reflect a single underlying effect, as opposed to a distribution of effects (also called tests of heterogeneity). If a test of homogeneity fails to detect heterogeneity among results, it is assumed that the differences observed between individual studies are a consequence of sampling variation and simply due to chance. A major limitation of the statistical tests of homogeneity that are in use is the lack of power. Therefore, a non-significant test of heterogeneity does not necessarily exclude heterogeneity. Cochran's Q test is the standard test for testing homogeneity in meta-analysis.

Trial validity: The degree to which a result is likely to be true and free of bias. Assessment of each trial's validity is critical in systematic review. An important dimension of study quality relates to the validity of the findings generated by the study. There are two important forms of validity: internal and external validity. Internal validity is defined as the extent to which the results of a trial are valid for the conditions being studied. External validity is the extent to which results of a trial provide a correct basis for generalisations.

Weighted mean difference: A method of meta-analysis used to summarise effect size measures for continuous data where the weight given to each study is determined by the precision of its estimate and effect. The weight given to each study is determined by the precision of its estimate of effect and is equal to the inverse of the variance.

References

The short notes have been compiled and adapted from the following list.

1. Bero L, Rennie D. The Cochrane Collaboration. JAMA 1995; 274:1935-8.
2. Colditz G, Burdick España, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies. Am J Epidemiol 1995; 142:371-82.
3. Cooper H, Rosenthal R. Statistical versus traditional procedures for summarizing research findings. Psychol Bull 1980; 87:442-9.
4. Smith GD, Egger M. Incommunicable knowledge? Interpreting and applying the results of clinical trials and meta-analyses. J Clin Epidemiol 1998; 51:289-95.

5. Egger M, Smith GD, Schneider M, *et al.* Bias in meta-analysis detected by a simple graphical test. BMJ 1997; 315:629-34.
6. Egger M, Smith GD, Altman DG, eds. Systematic Reviews in Health Care: Meta-analysis in Context. London: BMJ Books; 2001.
7. Ferrer RL. Graphical methods for detecting bias in meta-analysis. Fam Med 1998; 30(8):579-83.
8. Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. J Clin Epidemiol 1995; 48:45-57.
9. Light RJ, Pillemer DB. Summing up: The Science of Reviewing Research. Cambridge: Harvard University Press; 1984.
10. Rosnow, RL, Rosenthal R. Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. Psychol Methods 1996; 1(4):331-40.
11. Sackett DL, Richardson WS, Rosenberg WMC, *et al*. Evidence-based Medicine: How to Practice and Teach EBM. 2nd Ed. London: Churchill-Livingstone, 2000.
12. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. BMJ 1996; 313:735-8.
13. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (Overviews) using updated individual patient data. Cochrane Working Group. Stat Med 1995; 14:2057-79.
14. Hunter JE, Schmidt FL. Methods of Meta-analysis: Correcting Error and Bias in Research Findings. Newbury Park, CA: Sage Publications, 1990.
15. Paterson BL, Thorne SE, Canam C, *et al.* Meta-study of Qualitative Health Research: A Practical Guide to Meta-analysis and Meta-synthesis. Thousand Oaks, CA: Sage Publications Ltd., 2001.
16. The Cochrane Collaboration (2007) http:www.cochrane.org (accessed 11 Apr 2007).
17. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. J Clin Epidemiol 2000; 53:207-16.