View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

International Journal of Surgery 7 (2009) 291–296

Contents lists available at ScienceDirect

International Journal of Surgery

journal homepage: www.theijs.com

Review

Meta-analysis: A practical decision making tool for surgeons

Sukhmeet S. Panesar [a], Mohit Bhandari [b], Ara Darzi [c], Thanos Athanasiou [c,*]

[a] National Patient Safety Agency (NPSA), 4–8 Maple Street, W1T 5HD London, United Kingdom
[b] 293 Wellington Street North, Suite 110, McMaster University, Hamilton, Ontario, Canada L8S4L8
[c] Division of Surgery, Oncology, Reproductive Biology and Anaesthetics, Imperial College London, 1030, 10th Floor, Queen Elizabeth the Queen Mother Wing (QEQM), St. Mary's Campus, Praed Street, W2 1NY London, United Kingdom

ARTICLE INFO

Article history:
Received 2 April 2009
Received in revised form
7 May 2009
Accepted 11 May 2009
Available online 21 May 2009

Keywords:
Evidence
Meta-analysis
Systematic review
Surgery

ABSTRACT

Background: The exponential rise in published medical research on a yearly basis demands a method to summarise best evidence towards its application to patient care in clinical practice. A robust meta-analysis is a valid tool. It is often considered to be a simple process of pooling results from different studies. This is not true. It appears that surgeons lack a reference guide to help them conduct and appraise a meta-analysis.

Methods: This paper provides a structural framework to perform a meta-analysis. It guides the surgeon on a journey from identification of the correct clinical question to data analysis and through to producing a structured report. Statistical methods are discussed briefly as most commercial software calculates most results in the background. An example of a recent meta-analysis is given. However, important caveats are mentioned as there are limitations of the meta-analytical technique.

Conclusion: Whereas meta-analyses of homogeneous studies are the highest form of evidence, poorly conducted meta-analyses create confusion and serve to harm the patient. Surgeons practising their art in an era of evidence-based surgery need to understand the principles of meta-analyses.

© 2009 Surgical Associates Ltd. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Historically, surgical decisions have been made based on personal experience, unquestioned use of methods suggested by senior colleagues and recommendations from surgical authorities. The progress of absorbing higher forms of evidence into the surgical knowledge base has been slow. The proportion of systematic reviews and randomised controlled trials (RCTs) in leading surgical journals stands at 5%.[1]

To ensure the best possible outcomes for patients, clinicians are increasingly required to implement best practices and continual quality improvement processes in the clinical environment. This inextricably involves the application of the best available knowledge, usually in the form of scientific research, to guide clinical decision making. Hence, the use of clinical research is no longer an option but a necessity.

However, with increasing pressures of being a practising surgeon and the reduction in the number of working hours,[2] two problems remain. One is the ability to synthesise and apply the best evidence to improve patient care, bearing in mind that the average clinician would have to read 19 original articles each day in order to keep up with advances in his chosen field.[3] Conflicting or inconclusive results across individual research studies attributable to a statistical play of chance or poorly designed study methodology add further challenges to uncovering best evidence for clinicians.

Evidence synthesis is a highly valuable tool in health care which can save clinicians valuable time and can also allow the best evidence to be identified. Systematic reviews stand at the top the hierarchical pyramid of evidence (level 1 evidence), when the primary studies are all high quality (randomised trials) and homogeneous in their results.[4] Meta-analysis refers to the term applied to systematic reviews with a quantitative combination of results across two or more studies.[5] Meta-analysis has become a widely popular method of evidence synthesis.[6] The increased use of meta-analysis in the surgical literature has not been accompanied by resources to ensure methodological rigor in its conduct nor critical appraisal guides to assist clinicians in its interpretation. Our aim is to provide a road-map for the essential steps involved in conducting such a meta-analytical study.

* Corresponding author. Tel.: +44 20 7886 7619.
E-mail addresses: sukhmeet.panesar@npsa.nhs.uk (S.S. Panesar), bhandam@mcmaster.ca (M. Bhandari), a.darzi@imperial.ac.uk (A. Darzi), t.athanasiou@imperial.ac.uk (T. Athanasiou).

1743-9191/$ – see front matter © 2009 Surgical Associates Ltd. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijsu.2009.05.004

1.1. The science of meta-analysis

A valid meta-analysis requires the same careful planning as any other research study. The goals of a meta-analysis include a relevant clinical question, explicit eligibility criteria, a comprehensive search for best evidence, the rationale quantitative summary of the data, and an explanation of variability between studies.

1.2. Defining the objectives of the study

A well built clinical research question is the critical first step of a meta-analysis. The acronym widely used to remember a good clinical question is P.I.C.O. For a therapeutic intervention, this includes a description of the study population, the intervention, a comparison group, and the outcome.

1.3. Defining the population of studies to be included

A discrete and objective statement of inclusion and exclusion criteria for studies should be provided. *A priori* inclusion and exclusion criteria help to eliminate selection bias in the meta-analysis. Any inclusion criteria must include:

- *The study type* – it must be decided from the onset whether only randomised controlled trials (RCTs) or observational studies will be included, although this remains debatable.[7,8] A meta-analysis is only as good as the quality of the primary studies incorporated into it.[9]
- *Patient characteristics* – these include age, gender, ethnicity, presenting condition, co-morbidities, duration of illness and method of diagnosis.
- *Treatment modalities* – for the condition in question, the allowable treatment type, dosage, duration and conversion from one treatment to another should be addressed.

1.4. Defining the outcome measures

The protocol of the meta-analysis should explicitly specify the key outcomes. Only one set of results from a single study should be included even if multiple publications are available in order to avoid duplication of the data set. Usually in these circumstances, the data included will be from the latest publication or the paper with the most complete data on the outcome measures of interest. The outcome should generally be the most relevant and patient-important.

1.5. Locating all relevant studies

This involves using a structured search strategy involving databases such as NLH Medline, PubMed, EMBASE, CINAHL and Google scholar. There are different search strategies for the various databases and effective use must be made of MeSH headings, synonyms, and the 'related articles' function in PubMed. Beyond databases, researchers can find potentially relevant papers in the proceedings handbooks of medical meetings, bibliographies of journal articles, textbooks, and experts in the field. These form part of the 'grey literature' an often useful but missed source of data.

1.6. Screening, evaluation and data abstraction

A rapid review of the abstracts of the papers will eliminate those that are fit for exclusion because of inadequate study design, specific population, or duration of treatment or date of the study. When available written information is insufficient for the meta-analysis, efforts must be made to contact the principal investigator to obtain the needed information to reduce the effects of publication bias.[10]

Usually it is recommended that two or more independent observers then extract the data from the studies using a pre-designed data extraction form to avoid errors. Patient demographics, baseline characteristics and clinical outcomes of interest should be extracted and a table created, that shows all variables and their values from all included studies.

The validity of the included studies should be assessed. While several quality indices exist for the evaluation of surgical trials, the use of a single index remains controversial. Juni et al. have reported that different systems can lead to different interpretations of study quality. Nevertheless, scores may be used to develop thresholds for inclusion of studies.[11,12] Blinding observers to the names of the authors and their institutions, the names of the journals, sources of funding, and acknowledgments can lead to more consistent scores.[11]

1.7. Choose and standardise the outcome of measure

Individual results have to be expressed in a standardised format in order to compare the studies. If the end point is continuous such as the length of hospital stay after bypass surgery, the mean difference (weighted mean difference, WMD) between the treatment and control groups is used. The size of a difference, however, is influenced by the underlying population characteristics. Differences are therefore often presented in units of standard deviation.

These data are presented in a forest plot as shown in Fig. 1. We have chosen to look at the age-stratified mortality in off-pump cardiopulmonary bypass (OPCAB) surgery versus on pump cardiopulmonary bypass (ONCAB) surgery.

If the end point is binary or dichotomous, such as mortality or no mortality, then the odds ratio (OR) or relative risk or risk ratio (RR) is calculated.

The OR is the probability that a particular event will occur to the probability that it will not occur, and can be any number between zero and infinity. Risk describes the probability with which a health outcome (usually an adverse event) will occur. Measures of relative effect express the outcome in one group relative to that in the other. For treatments that increase the chances of events, the odds ratio will be larger than the risk ratio, so the tendency will be to misinterpret the findings in the form of an overestimation of treatment effect. For treatments that reduce the chances of events, the odds ratio will be smaller than the risk ratio, so that again misinterpretation overestimates the effect of treatment. This error in interpretation is unfortunately quite common in published reports of individual studies and systematic reviews.[13]

Absolute measures, such as the absolute risk reduction or the number of patients needed to be treated (NNT) to prevent one event are more helpful when applying results in clinical practice.[14] The NNT can be calculated as 1/risk difference (RD).

1.8. Statistical methods for calculating overall effect

This involves calculating overall effect by combining the data. The results from small studies are more subject to the play of chance and should therefore be given less weight. Methods used for meta-analysis use a weighted average of the results, in which the larger trials have more influence than the smaller ones.

2. Fixed and random effects models

Two models can be used to assess the way in which the variability of the results between the studies.[15]

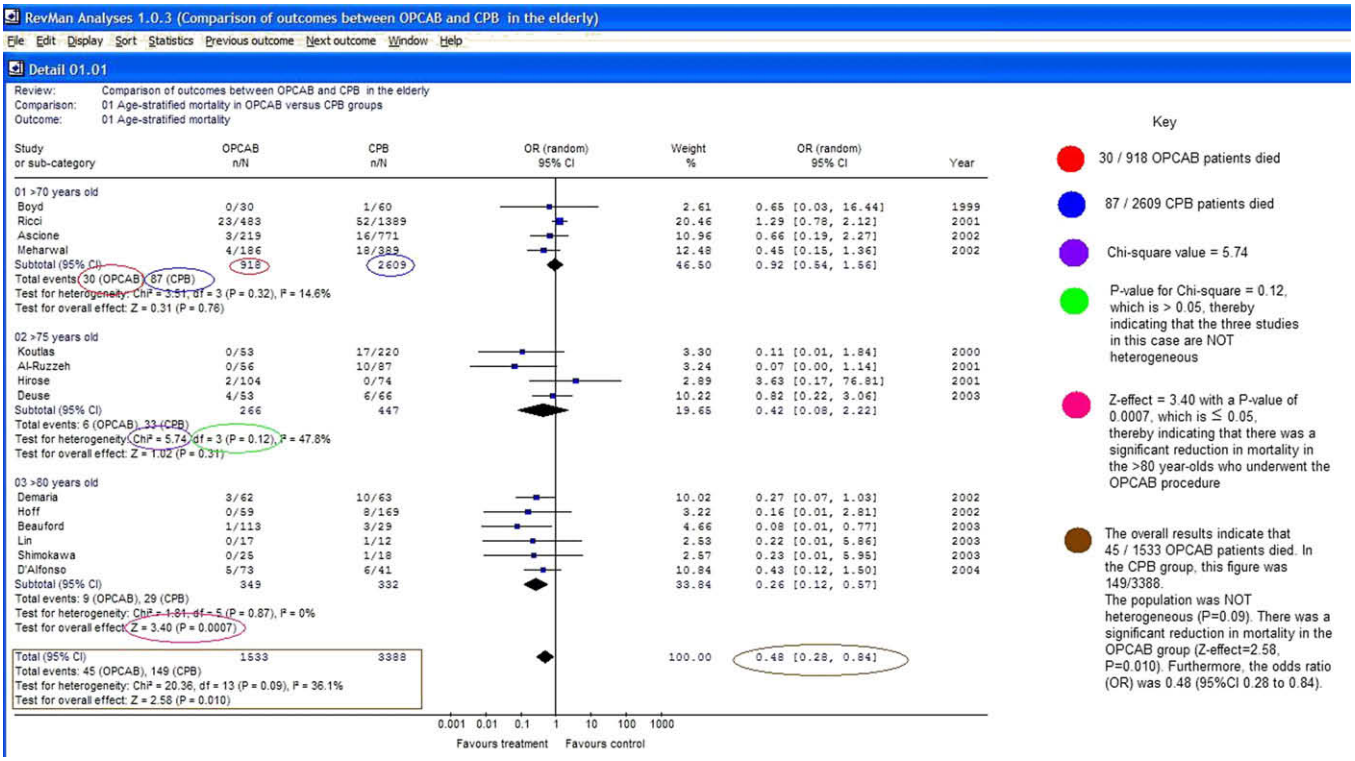

Fig. 1. Example of a Forest plot. [28]

2.1. Fixed effects model

The "fixed effects" model considers that this variability is exclusively due to random variation, i.e. if all the studies were infinitely large they would give identical results.

Methods of fixed effect meta-analysis are based on the mathematical assumption that a single common (or 'fixed') effect underlies every study in the meta-analysis, i.e. in a meta-analysis of odds ratios, we would assume that every study is estimating the same odds ratio. Under this assumption, if every study were infinitely large, every study would yield an identical result.[16] Thus, the summary measure is a simple weighted average and can be easily interpreted as an estimate of a single population outcome measure. The 95% CI will reflect only the variability between patients; hence, with this class of methods, the 95% CI will be very narrow with *more power to reject the null hypothesis*. The fixed effects analysis may be justified when the test for heterogeneity is not significant; i.e. when there is no evidence of major differences among studies. In the fixed effects analysis, the methods used to analyze binary outcomes are the general inverse-variance method, the Mantel–Haentzel method[17,18] and the Peto method.[15]

2.2. Random effects model

The "random effects" model assumes a different underlying effect for each study and takes this into consideration as an additional source of variation, which leads to somewhat wider confidence intervals than the fixed effects model.[19] Effects are assumed to be randomly distributed, and the central point of this distribution is the focus of the combined effect estimate. When there is some statistical heterogeneity, as detected by a statistically significant heterogeneity test, it will be implausible to assume that the 95% CI or imprecision of the summary outcome reflects only between-patient variability. Therefore, the fixed effects model will not fit the observed data well as the 95% CI will be too narrow. In the random effects analysis, it is assumed that all the studies are fundamentally different and that the outcome of a study will estimate its own unique outcome, which differs from that of the other studies. Hence, each study outcome is not assumed to fluctuate around a fixed, common population outcome but to fluctuate around its own true value. It is assumed, however, that each of these true values is drawn "randomly" from some underlying probability distribution; i.e. that of a "superpopulation", commonly assumed to be a normal distribution; hence, the name "random" effects analysis. That is, under a random effects assumption, not only is each study performed on a sample drawn from a different population of patients but that each of these populations is still taken randomly from a common "superpopulation". A random effects analysis makes the assumption that individual studies are estimating different treatment effects; hence, the 95% CI in a random effects analysis, reflecting the overall variability in the data will be wider than that of a fixed effects analysis because of both inter-patient variability and inter-study variability.[16]

Both these individual statistical methods have their limitations and a substantial difference in the combined effects calculated by fixed and random effect models will be seen only if studies are markedly heterogeneous.

2.3. Heterogeneity between study results

Sometimes, the variance between the overall effect sizes in each study might not be due to random sampling variation but instead could be due to the presence of other factors inherent within individual studies. This effect size variation due to slightly different study designs is termed heterogeneity. If the result of each study differs greatly from each other and is deemed to be largely due to heterogeneity, then it may not be appropriate to conduct a meta-analysis in the first place. If a test for homogeneity shows

homogeneous results then the differences between studies are assumed to be a consequence of sampling variation, and a fixed effects model is appropriate. If, however, the test shows that significant heterogeneity exists between study results then a random effects model is advocated. If there is excess heterogeneity, then not even the random effects model could compensate for this and the viability of the meta-analysis should be questioned. A major limitation with heterogeneity tests is that these statistical tests will lack power to reject the null hypothesis of homogeneous results even if substantial differences between studies exist. After significant heterogeneity has been discovered, the causes and sources of this need to be explored in detail.[16]

Various statistical tests are used to assess for heterogeneity. A useful statistic is $I^2 = [(Q - \mathrm{df}/Q)] \times 100$, where Q is the chi-square statistic and df is its degree of freedom. This statistic defines variability along a scale-free range as a percentage from 0 to 100%. This describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance).[16] Typically, I^2 values above 40% suggest substantial heterogeneity and challenge the rationale to pool data across studies.

2.4. Funnel plots

Funnel plots aid to graphically reveal the presence of publication bias – being a scatter plot function of study effect size and estimated effect size.[20,21] The premise being that larger studies will have result estimates that are more precise while the contrary for small studies is true. In an ideal setting where all possible published and unpublished trials are available, individual studies would form a symmetrical inverted funnel with more precise results from larger trials bunched up at the top and less precise results from smaller trials scattered symmetrically across. This is shown in Fig. 2.

An asymmetrical distribution can occur usually represented by a deficiency in a certain region of the funnel that can be attributed to the skewed distribution of studies created from publication or reporting bias.[20]

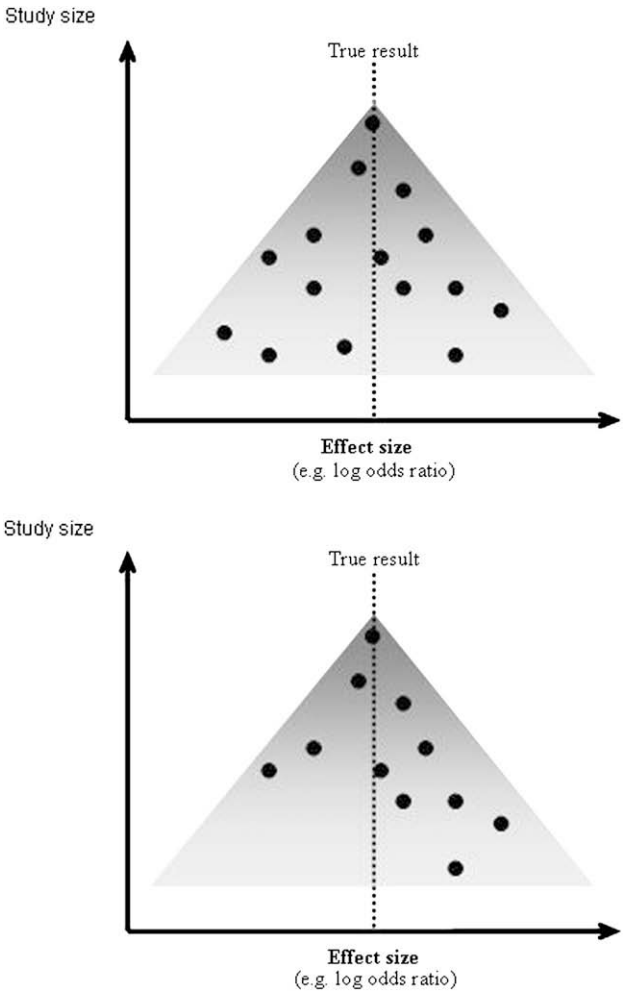

Fig. 2. Funnel plots.

2.5. Forest plots

Results from each trial, together with their confidence intervals, can be graphically displayed in a useful manner on a forest plot. A black square and a horizontal line, which corresponds to the point estimate and the 95% confidence intervals of the outcome measure respectively, represent each study. The dotted vertical line corresponds to no effect of treatment (e.g. an odds ratio or relative risk of 1.0). If the confidence interval includes 1, then the difference in the effect of experimental and control treatment is not significant at nominally tolerated levels ($P > 0.05$). The size (or area) of the black squares reflects the weight of the study in the meta-analysis while the diamond represents the combined odds ratio, calculated using a fixed effects model, at its centre with the 95% confidence interval being represented by its horizontal.[22]

Most of the studies, if they are homogenous in design and population would have overlapping confidence intervals. However, if the confidence intervals of two studies don't overlap at all, there is variation between the two studies which is not likely due to chance and likely due to the presence of heterogeneity. Other than graphically using a forest plot, a numerical method could be achieved via use of the (χ^2) chi-squared test.[16]

2.6. Sensitivity analysis

This procedure assesses the robustness of the findings of the meta-analysis. Both fixed and random effects models should be used. The methodological quality of studies should be assessed on one of the existing scoring scales or an arbitrary quality scale. The meta-analysis can be repeated for high and low quality studies. Significant results are more likely to get published than non-significant ones, i.e. publication bias, which can distort the results of a meta-analysis. Its presence can be identified by stratifying the analysis by study size. Smaller effects can be significant in larger studies. If publication bias is present, it is expected that, of the published studies, the larger ones will report the smallest effects and the smallest ones may report the largest effects. If exclusion of the smaller studies does not significantly affect the overall estimate of the meta-analysis, the sensitivity analysis confirms the validity of the meta-analysis, i.e. it is not affected by exclusion of trials of poor quality.[23]

2.7. Subgroup analysis

The principal aim of meta-analysis is to produce an estimate of average effect seen in trials of a particular treatment or intervention. The clinician must make a decision as to whether his or her patient is comparable to the patient group in the meta-analysis. The meta-analysis may show superiority of one intervention over another in only particular subgroups, when this analysis is undertaken, but no difference between the two groups when the subgroups are not analysed separately. Subgroup analysis could also be used to explain heterogeneity by determining which component of the study design may be contributing to treatment

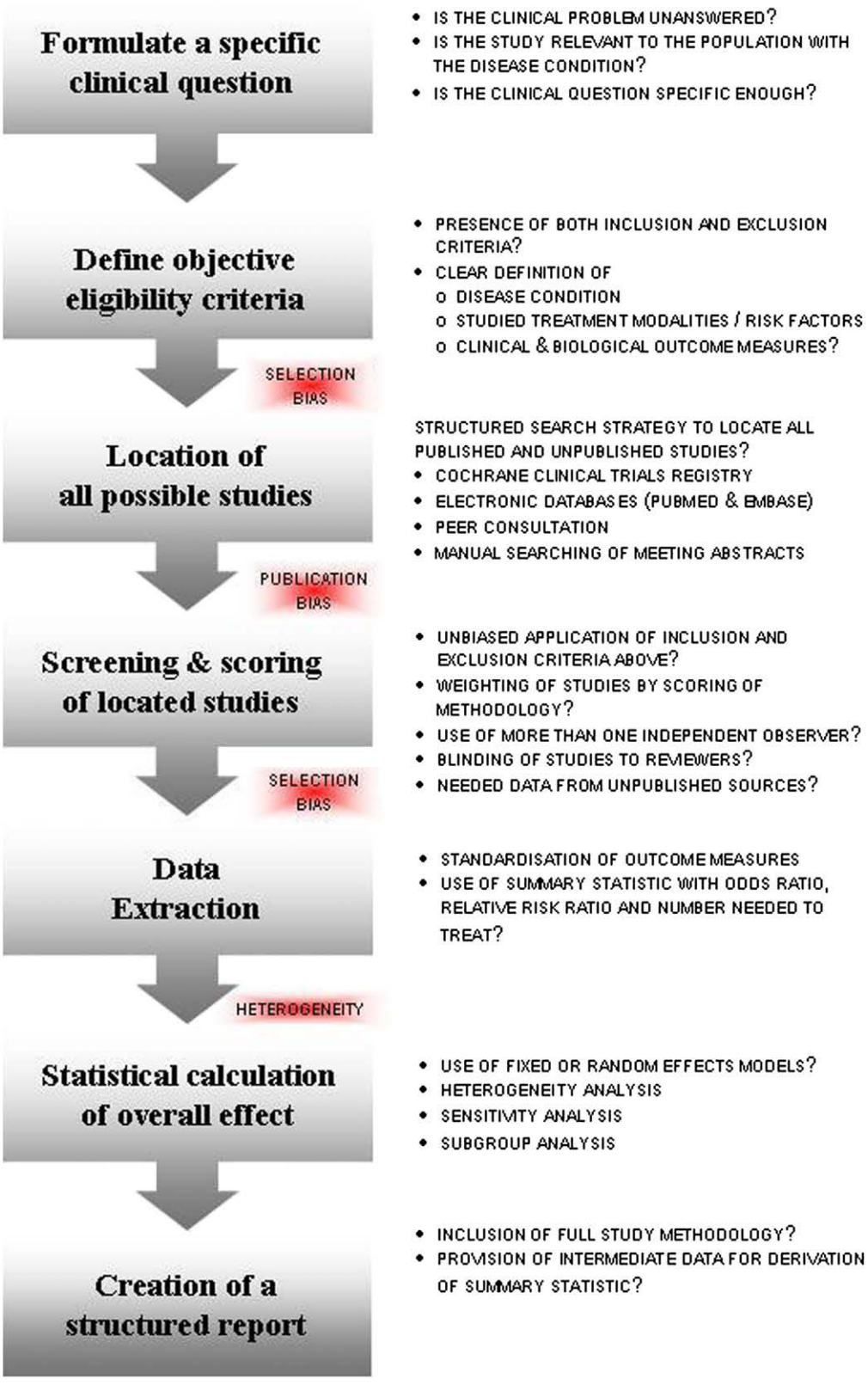

Fig. 3. Overall pathway of systemic review/meta-analysis.

effect.[24] Subgroup analyses should always be interpreted with caution.

The steps involved in performing a meta-analysis are outlined in Fig. 3.

3. Discussion

In this paper, we have provided a broad outline of the procedures for performing a meta-analysis in a systematic fashion. However, this process should be performed with great care bearing in mind that poorly conducted meta-analysis is of limited value to the surgical literature.

Meta-analysis, especially surgical meta-analysis, is often an exercise in compromise. It should not be seen as an opportunity to pool results of different surgical studies that are dissimilar. For example, inclusion of only experimental trials may exclude the most contemporary literature, or an attempt to include all available evidence may introduce heterogeneity. Compromise requires qualitative value-judgements that even underpin a statistically robust meta-analysis with a comprehensive literature review. This does not necessarily demean the efficacy of meta-analysis as a tool for literature review, evidence synthesis and clinical decision making; however it is imperative that these value-judgements are explained and justified. Assessing the various dimensions of quality is important. Unfortunately, this is often not the case.

The advantages of well-conducted meta-analyses are that they allow objective appraisal of the evidence in comparison with traditional narrative reviews, provide a more precise estimate of a treatment effect, and may explain heterogeneity between the results of individual studies. The application of meta-analysis in surgery however is problematic and surgical meta-analyses are often vulnerable to bias. Patient population characteristics, technological development and evolving surgical expertise have the potential to significantly affect event rates which can bias the results of meta-analysis.[25]

There are numerous instances, where meta-analyses have pooled results from small trials with disparate results, and as a result have produced conflicting evidence. Furthermore, results have been generated that were in conflict with the results of subsequent large randomised clinical trials.[26,27] When this occurs, the reliability of the evidence is questioned resulting in poorly guided clinical decisions. As a result, doubts have been raised about the reliability of using meta-analyses to guide clinical practice. Whilst even advocates of meta-analysis argue that it is not a substitute for the clinical trails, it may be a useful guide to clinical decision makers until unequivocal experimental evidence is available. However, if meta-analysis is to continue to have a role in surgical decision making, surgeons needs to be able to perform, assess, compare and communicate the quality of meta-analyses, particularly in areas where several meta-analyses are available.

Contributorship

All authors have made substantial contributions to all of the following: (1) the conception and design of the study, or acquisition of data, or analysis and interpretation of data, (2) drafting the article or revising it critically for important intellectual content, and (3) final approval of the version to be submitted.

Conflict of interest
None declared.

Funding
None declared.

Ethical approval
Not required.

References

1. Panesar SS, Thakrar R, Athanasiou T, Sheikh A. Comparison of reports of randomized controlled trials and systematic reviews in surgical journals: literature review. *J R Soc Med* September 2006;**99**(9):470–2.
2. Royal College of Surgeons of England. Implementing the EWTD: college response. Available from: http://www.rcseng.ac.uk/publications/docs/ewtd_communicaion.html/attachment_download/pdffile [accessed 01.04.09].
3. Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine: a new journal to help doctors identify the information they need. *BMJ* 1995;**310**: 1085–6.
4. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. *JAMA* 1995;**274**:1800–4.
5. Glass GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976;**5**: 3–8.
6. Egger M, Smith GD. Meta-analysis. Potentials and promise. *BMJ* 22 November 1997;**315**(7119):1371–4.
7. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA* 2000;**283**(15):2008–12.
8. Thompson SG, Pocock SJ. Can meta-analyses be trusted? *Lancet* 1991;**338**(8775):1127–30.
9. Olkin I. Meta-analysis: reconciling the results of independent studies. *Stat Med* 1995;**14**(5–7):457–72.
10. Berman NG, Parker RA. Meta-analysis: neither quick nor easy. *BMC Med Res Methodol* 2002;**9**(2):10.
11. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;**282**(11):1054–60.
12. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;**16**(1):62–73.
13. The Cochrane Collaboration. Measures of relative effect: the risk ratio and odds ratio. In: The Cochrane Handbook. p. 111–3.
14. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988;**318**(26):1728–33.
15. Berlin JA, Laird NM, Sacks HS, et al. A comparison of statistical methods for combining event rates from clinical trials. *Stat Med* 1989;**8**(2):141–51.
16. The Cochrane Collaboration. Diversity and heterogeneity: identifying statistical heterogeneity. The Cochrane Collaboration open learning material 2002. Available from: http://www.cochrane-net.org/openlearning/HTML/mod13-3.htm [accessed 01.04.09].
17. Mantel N, Hanezsel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;**22**(4):719–48.
18. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 1985;**41**(1):55–68.
19. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**(3):177–88.
20. Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998;**316**(7124):61–6.
21. Stuck AE, Rubenstein LZ, Wieland D. Bias in meta-analysis detected by a simple, graphical test. Asymmetry detected in funnel plot was probably due to true heterogeneity. *BMJ* 1998;**316**(7129):469.
22. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ* 1997;**315**(7121):1533–7.
23. Easterbrook PJ, Berlin JA, Gopalan R, et al. Publication bias in clinical research. *Lancet* 1991;**337**(8746):867–72.
24. Davey Smith G, Egger M, et al. Meta-analysis. Beyond the grand mean? *BMJ* 1997;**315**(7122):1610–4.
25. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;**320**:157.
26. Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, Lau J. Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA* 1996;**276**:1332–8.
27. Borzak S, Ridker PM. Discordance between meta-analyses and large-scale randomized, controlled trials. Examples from the management of acute myocardial infarction. *Ann Intern Med* 1995;**123**:873–7.
28. Panesar SS, Athanasiou T, Nair S, Rao C, Jones C, Nicolaou M, Darzi A. Early outcomes in the elderly: a meta-analysis of 4921 patients undergoing coronary artery bypass grafting – comparison between off-pump and on-pump techniques. *Heart* 2006;**92**:1808–16.