J Med Sci,
Vol 42, No. 1, March 2010: 44-51

Systematic review: better or (otherwise) misleading for clinical decision?

Muhammad Bayu Sasongko[1], Hardyanto Soebono[2*]
[1]Department of Ophthalmology, [2]Department of Dermatology & Venerology, Faculty of Medicine, Gadjah Mada University, Yogyakarta, Indonesia

ABSTRACT

Systematic review is a method to combine multiple sources of evidence through an explicit and reproducible way of literature search and critical appraisal of the quality of included studies, with or without mathematical methods to synthesis these information. Since this method was first introduced more than centuries ago, systematic review has been increasingly popular and widely used particularly in the area of medicine. Systematic review is often very useful to physicians to help supporting the clinical decision making and significantly reducing their time to seek for appropriate evidence. However, despite its reproducible and systematic steps to substantially minimize the presence of biases, physicians should still be aware that systematic review is not completely biases resistant. Inclusion of poor quality studies, heterogeneity, and publication or other reporting biases are commonly evident in systematic review that may hinder the quality of the conclusion. This review summarizes the core principals of systematic review and its potential biases, and discusses when the systematic review is useful or needing careful attention.

Key words: treatment-scientific evidence- meta-analysis- critical appraisal-outcomes

INTRODUCTION

Clinical decision making can be easy but sometimes is also complicated. Physicians often experience a problematic clinical decision making when there are multiple sources of information for example from primary data, patient's preferences, individual's experience, hospital policy, and enormous amount of scientific evidence. This may become worse if sources of information to be drawn is conflicting rather than yielding similar conclusion. For example, few recent trials suggest the benefit of lipid-lowering drugs for patients with diabetes in addition to standard glucose control therapy to prevent the development of microvascular complications.[1,2] On the contrary, at the same time, there is strong evidence showing that serum lipids are not consistently associated with diabetic microvascular complications.[3] So which ones should physicians rely on?

An approach to put together this information into a single review article is increasingly important in clinical setting. Unfortunately, since reviewing is a very subjective process and depends much on the reviewer's capacity to review, perform literature search, absorb information, and generate the conclusion, medical review article is often scientifically inaccurate and less reliable for decision making. Therefore, there is a need for a clear strategy for literature identification and selection, systematic assessment of the quality of evidence, and coherent and instructive way of synthesizing data into a concise and unified piece of information more applicable to clinical practice, the so-called *systematic review*.

In the last decade, systematic review articles including, if appropriate, meta-analysis are becoming popular and notably the ground for decision making in many clinical settings. More importantly, the credit that has been given to systematic review as "the best scientific evidence" is of great interest for practicing physicians who demand representative "synoptic" evidence that helps reducing their trivial time to select and read primary reports.[4] Nevertheless, not many of these empiricism-driven physicians realize that cautious

* coresponding author: *hardyanto@ugm.ac.id*

interpretations of systematic review should still be made. First, because not only the fact that there are still a few caveats of systematic review that should be taken into consideration (e.g publication bias, heterogeneity), but second, there has also been evidence lately showing the limitations and abuses of systematic review.[5] This paper summarizes the core concepts of systematic review including its historical perspectives, principals and applications, and pitfalls in the context of clinical practice.

Historical perspectives

In the 17th century, astronomy and geodesy institution suggested the use of combination of data rather than single data to obtain a better and more representative result,[6] which was perhaps the statistical embryo of meta-analysis. In the area of medicine, the idea of compiling and summarizing relevant evidence for physicians began in the 18th century, in response to the need of busy physicians for relevant research information.[7] Two German- and English-written journals published in Leipzig, Germany (*Commentarii de Rebus in Scientia Naturali et Medicina Gestis*, 1752 – 1798) and in Edinburgh (*Medical and Philosophical Commentaries,* 1773 – 1780) reported abstracts of articles and new important books published elsewhere, for example including the book on the use of digitalis for heart disease treatment (*Account of foxglove,* 1785).[7] It was Andrew Duncan Sr. (1744 – 1828), the "anonymous" editor of *Medical and Philosophical Commentaries*, who amongst the first to critically appraise the medical practice in late 18th century in Britain. One century later, in 1904, Karl Pearson, for the first time, introduced the formal statistical techniques to combine data from different studies.[8] He proposed that many data are just too small considering the possibility of error, thus insufficient to obtain strong and generalizable conclusion, which remains the basis of systematic review/ meta-analysis these days.

In the following years, this technique was barely used in medicine. On the contrary, researchers in social science and education showed an early interest in the synthesis of research findings.[9] In 1950s, David Sackett, a clinical epidemiologist at McMaster University, Ontario, Canada, and others developed the concept of "critical appraisal" of research reports, which has been one of the most substantial components of systematic review,[10] but the term "meta-analysis" was first mentioned in the paper by a psychologist in 1976, titled "Primary, secondary, and meta-analysis of research".[11] Within less than 5 subsequent years, British physicians and epidemiologist, Archie Cochrane, became aware that there were, in fact, many physicians without proper access to reliable evidence to support their clinical practice. This was the primary foundation of thought to establish the Cochrane Collaboration that facilitates wide access in systematic reviews in the later decades.[12] Systematic review and meta-analysis in medicine have become very popular particularly in the fields of cardiovascular diseases and oncology since 1980s.[13]

The need for systematic review: limitation of a single study and narrative (non-systematic) review

In an article by Freiman and associates published in the New England Journal of Medicine in 1978,[14] seventy-one clinical trial reports were assessed whether they had sufficient power (>90%) to detect 25% and 50% difference in treatment effect. Surprisingly, most of these trials reported negative findings but interestingly, 57 of these negative trials had 25% potential improvement to show the significant treatment effect. This tells that single study often fails to achieve the required sample size to detect an association. Such study may therefore produce false negative results by showing no significant treatment effect when there is in fact such effect exists. When this is the case, conclusion made based on a single study is obviously invalid, thus combined data from multiple but comparable studies are considered.[15]

Similar scenario would have happened even if conclusion is made based on combined studies but merely according to reviewer's view without systematic approach in literature identification and quality assessment of information. Back in 1981, an article by J.R. Mitchell published in BMJ reviewed over the previous 20 years clinical trials on the use of beta-blocker in patients surviving from myocardial infarction.[16] This review suggested that there is no clear evidence if the use of beta-blocker may improve the long-term survival of post myocardial infarction patients.[16] In contrast, another review article by J.R. Hampton published in European Heart Journal within the same year and came up with opposite conclusion.[17]

"For the moment it seems perfectly reasonable to treat patients who have survived an infarction for two or three weeks with timolol,..."

Despite its observation over the previous clinical trials, if cautious attention is paid to these review articles, there was no statement as to how these authors conducted literature search, or how these authors assessed the methodological quality of reviewed studies. In light of this, an article by Mulrow reported that during mid 1980s,[18] only one of 50 published reviews in leading general medicine journals clearly specified methods of identifying, selecting, and perform the standardized methodological quality assessment of their reviewed articles. Therefore, it is not surprising that the two articles as illustrated above would have had such an opposite conclusive statement. Importantly, it also seems conceivable that without formal guidelines, reviewers would only pick whatever relevant articles stored in their filling cabinet to be reviewed and potentially ignore the significant contribution of other "un-reviewed" article in generating conclusion.

Principals and methods in conducting a systematic review

To obtain adequate power to detect a treatment effect, instead of having to recruit 10,000 participants, investigator might think to include 10 existing comparable studies of 1,000 participants, which is perhaps more feasible. Systematic review is essentially a study of studies, with a single study being considered as a participant.[19,20] A systematic review without a statistical combination of data, but merely a general summary, may be called qualitative systematic review.[21] On the other hand, a systematic review that uses statistical methods to quantitatively combine data and generate summary is called quantitative systematic review or meta-analysis.[22] Therefore, having treated a study like single participants in a single study, the result of systematic review may be violated by including irrelevant studies or excluding relevant studies, or inappropriate statistical methods for the meta-analysis, if any. This violation or bias can be reduced by following several basic principles and standardized methods in performing systematic review, which are very similar to a single study.

Developing research question, protocols, and inclusion/exclusion criteria

Conducting a systematic review should be based on relevant and applicable research question. A criterion to formulate good and representative research question has been proposed previously and widely used to date, following the PICO model[23]:

a) Patient, population, or problem (P): covering the target population or patients, age, or particular clinical condition;
b) Intervention (I): covering intervention or exposure, or any form of treatment, diagnostic test which will be applied to the target population;
c) Control or comparison intervention (C): this is only applied for studies with control group; and
d) Outcomes or expected effect (O): expected outcomes from the proposed study.

An example of well-defined research question, from the systematic review by Lago *et al*.[24] is: Is patients with pre-diabetes or diabetes (P) given thiazolidinediones (I) versus placebo or no treatment (C) have higher risk of congestive heart failure and of cardiac death (O)? In addition, few other approaches to define research question are also available.[25]

Once the study question has been well-formulated, clear and detailed objectives and protocol should be written in advanced. This may include the subgroup of interest, and detailed methods on how investigator will identify and select relevant studies, inclusion and exclusion criteria of studies to be included in the analysis, and how to extract and analyze information from included studies. The importance of having such pre-defined protocol is to avoid bias introduced by additional analysis driven by unexpected data.[21,22,26,27] It is also highly recommended that investigator have a clear check-list form containing all items specified in this protocol.[21]

METHODS FOR LITERATURE SEARCH

The easiest way to distinguish systematic review and traditional review is by following the methods for literature identification.[18, 20,28] In traditional review, authors rarely mention how they perform the literature search, thus it is almost impossible to replicate. On the contrary, in systematic review, strategies for literature search need to be clearly stated, including the

keywords, databases, conference proceedings, or even traditional mail to relevant author if any inclusion of unpublished paper was used in the review. It is very important to ensure that the literature search method is replicable and repeatable. Several works have identified the most efficient and highly sensitive search strategy for clinical trials and prognostic and diagnostic studies.[29] However, in many cases, a search term based on the PICO description is adequate to identify relevant articles to be included in the review.[30]

Several large medical-related databases have included tens of thousands of literature, including MEDLINE,[31] EMBASE, the Cochrane Central Register of Controlled Trial (CENTRAL, through the Cochrane Library), or clinical trial registry. However, other specialized databases such as Allied and Alternative Medicine (AMED), Biological Abstracts (BIOSIS), CAB Health, Cumulative Index to Nursing and Allied Health Literature (CINAHL), conference proceedings, and bibliographies of review articles, monographs studies, hand-searching of key publications, or even unpublished studies are also essential since their result may be systematically different from published studies.[22]

Selection and critical appraisal of methodological quality of studies

A large number of quality assessment methods and quality scales have been published.[21,32-36] Yet, there is no universal consensus on which criteria should investigator follows. The quality assessment might cover the design of trial, conduct and analysis, applicability, and also the quality of reporting,[37] but importantly, these aspects should be early mentioned in the protocol.[22] In general, an key element of study quality is determined by the question "how valid is the result generated by a study?", which relates to internal validity, careful design, conduct, and analysis of a study to prevent biased result, and external validity, the extent to which the result can be applied to other circumstances.[38]

Again, selecting and assessing potential studies are somewhat subjective process. Therefore it is of importance to have more than one observer checking the eligibility of potential studies and assessing the quality of those studies following exactly similar protocol. It is also beneficial to have assessors blinded from the names of the authors and their institution, names of journals, sources of funding, and acknowledgment to avoid bias due to subjective preferences. Furthermore, all of this selection and assessment procedures need to be recorded in a standardized record form.

Data extraction and analysis

In extracting the data of selected studies, it is also important to record the extraction steps in a pre-specified data extraction form and involve more than one observer to eliminate the possibility of different interpretation. A comprehensive overview of extracted data should be presented clearly in the result section of the systematic review.[39,40] Once the data have been extracted, in one hand, it is relatively straightforward to generate the conclusion either qualitatively or quantitatively if selected studies are comparable and similar. On the other hand, if the data are not sufficiently similar, a quantitative summary with particular statistical methods need to be performed to address this heterogeneity, called meta-analysis.[12,26,33]

Detailed discussion on this method is beyond the scope of this review. Briefly, meta-analysis estimates the pooled summary of effects observed in different studies.[41] This statistical pooling is based on the fixed-variance of each study (fixed effect), in which studies with larger sample size and number of events have more influence in the pooled summary, or based on the randomly distributed true treatment effect between studies (random effect).[42,43] There are several available software to perform meta-analysis such as Review Manager (RevMan) (available for free from the Cochrane Library), MetaWin (http://www.metawinsoft.com), DSTAT (http://www.erlbaum.com), or also available from general statistical packages such as STATA (http://www.stata.com), SAS (http://www.sas.com), and S-Plus (http://www.mathsoft.com).

CONCLUSION AND RECOMMENDATION

Conclusion is made based on the results of data analysis, and recommendation for future research can be constructed when available data are still insufficient or their quality is not adequate. However, this is not as easy as it seems to be. It is such ashamed that after long effort to develop and write the review, the authors only comes up with unsatisfactory conclusion and miss

the opportunity to recognize interesting gaps which remain unanswered in the review, which turns down readers. For this reason, Brown and associates proposed a recommendation to formulate research conclusion and recommendation particularly for systematic review.[44] Their recommendation was to use (E)vidence (P)opulation (I)ntervention (C)omparison (O)utcome and (T)ime stamp or EPICOT formula,[44] as the extension of PICO.

Systematic review: trustworthy?

Despite the potentials and promises of systematic review, limitations should also be noted. There are indeed systematic reviews addressing similar issues, yet their conclusions are opposite.[45,46] For example, there were two systematic reviews assessing the use of low-molecular-weight heparin for the prevention of post surgical thrombosis compared to standard heparin.[45,46] The first review by Nurmohamed *et al.*[46] suggested that there was no evidence that low molecular weight heparin is superior to standard heparin for preventing post surgical thrombosis. In contrast, the other review by Leizorovich *et al.*[45] reported that low molecular weight heparin had beneficial effect over standard heparin in preventing thrombosis. Interestingly, not only were published within the same year, 1992, by two of the most major medical journal, Lancet and BMJ, these two reviews also mentioned that they included publications within the same period, 1984 to 1991. Nevertheless, this contradictory conclusion may have arisen because the first review did not include unpublished data and had language restriction in searching the literature, while the second review, although included unpublished data and had no language restriction, it did not assess the quality of included trials. If readers were given these two reviews, which one should they choose?

Evidently, the conclusion of systematic review may be very solid or otherwise flawed depending on the quality of selected studies. If the "raw" materials are of low-quality, the conclusion generated by such review would also be highly questionable. Studies included in the systematic reviews need to be of high methodological quality, within which observed difference between treatment and control groups is confidently a result of the intervention effect, but not driven by selection bias (differences in baseline criteria between treatment and control group), performance bias (unequal provision of care between treatment and control group), decision bias (bias on assessment), or attrition bias (bias due to exclusion of patients after allocation to treatment groups).[38] There are studies showing the impact of these biases on the ultimate results of clinical trials. For example, Schulz *et al.*[47] demonstrated that trials in which participants were aware of their treatment allocation had 30 – 40% larger effects compared to trials with appropriate concealment of treatment allocation.

In addition to quality assessment of included trial, thorough identification of all relevant studies is also a key to successful systematic reviews. How important it is to include such unpublished data? A study by Stern and Simes[48] showed that among 321 studies included in this study, only 25% of those with non-significant result were published within 10 years, whereas almost all of studies with significant results were published within 10 years. Similarly, this report also showed that the median time for studies with significant results to be published was less than 5 years, while those with non-significant results took 8 years to be published.[48] These findings indicate publication and time lag biases, in which studies with positive results are more likely to be published and mostly dominate the literature for years until those with negative results, but just as important, finally get published.[49,50] Therefore, publication bias and time lag bias are very important issue and exclusion of unpublished data will cause ignorance of any important findings.[51,52]

Another most common source of bias in systematic review is language bias. It is noted that most of "prestigious" journals are published in English; hence, most investigators are encouraged to publish their results in English. However, does this mean that good quality studies are always published in English? Or is it representative enough to only include English-published studies in systematic review?

Many systematic reviews are restricted their inclusion to include only English literature.[53] However, investigators from non-English native countries may also consider publishing some of their works in local journal.[29] Importantly, there are possibilities that significant results or higher quality studies are likely to be published in English-language journal, while keeping the non-significant/ lower quality ones for local journals.[54] A report from the German literature by Egger and associates suggested that more than 60% of trials

with non-significant results were published in German as compared to only about 40% were published in English.[55] Similarly, there is also evidence that non-English trials are of lower methodological quality.[54] Therefore, bias could be introduced in systematic reviews in which only include English literature.

CONCLUSION

Systematic review has acquired substantial attention particularly among clinicians. This technique was initially developed to summarize results from different studies within the similar problems, and it may be applied even for small studies with sizeable variations. Performing systematic review therefore requires expertise in both method and the context, and very often demands collaboration between clinicians, epidemiologist, and statistician. The question must be clearly defined to maximize the relevance of included studies. Investigators must then search for every relevant study from multiple databases, reviewing bibliographies, and seeking widely for unpublished work. The collected reports then must also be carefully selected and critically appraised to meet the relevance and requirements of the systematic review, thus the main information can be abstracted in the review to answer the question.

It is increasingly clear that the essential recipe for high-quality systematic review should include the fine quality of included studies, comprehensive literature search, and abstracting key information from research reports, not merely a summary. Nevertheless, there are still many published systematic review with a range of serious problems which make is hard to trust the conclusion, including the failure of investigator to perform the literature search or critical appraisal, careless in abstracting and summarize the selected reports, and perhaps the most often, overstatement of the strength of the conclusion. Moreover, it is also not surprising to find that two or more systematic review done at the same time without compatible conclusion. These varieties of problems clearly argue against the notion that systematic review may offer "the highest level of scientific evidence" for clinical decision. On the other hand, it powerfully suggests that quality of systematic review remains needing thorough assessment; otherwise potentially jeopardize the clinical decision.

In the author's perspectives, systematic review can help, but should never replace, strong clinical reasoning about individual patient on the basis of analogy, physician's experience, clinical investigations, fundamental medical theory and knowledge, as well as researched evidence.[56-58] Awareness of treatment effectiveness derived from the systematic review is evidence or knowledge, thus would not give physicians the clinical skills about how to use that treatment in treating patients, which can be obtained from clinical training and years of practicing experience. It will lead to a bad decision if applied in an uncritical and unfeeling way.[57] Therefore, the complexity of clinical decision making can never be replaced by a single document of compiled reports, because it is a state-of-the-art in which knowledge, skills, value and research evidence are integrated in every case by case scenario.

ACKNOWLEDGMENT

Authors would like thank Prof. T. Wong, Director the Center for Eye Research Australia, University of Melbourne, Melbourne, Australia for his critical review of this manuscript.

REFERENCES

1. Chew EY, Ambrosius WT, Davis MD, Danis RP, Gangaputra S, Greven CM, et al. Effects of medical therapies on retinopathy progression in type 2 diabetes. N Eng J Med 2010;363(3):233-44.
2. Keech AC, Mitchell P, Summanen PA, O'Day J, Davis TM, Moffitt MS, et al. Effect of fenofibrate on the need for laser treatment for diabetic retinopathy (FIELD study): a randomised controlled trial. Lancet 2007;370(9600):1687-97.
3. Miljanovic B, Glynn RJ, Nathan DM, Manson JE, Schaumberg DA. A prospective study of serum lipids and risk of diabetic macular edema in type 1 diabetes. Diabetes. 2004;53(11):2883-92.
4. Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? Int J Epidemiol 2002;31(1):1-5.
5. Egger M, Smith GD, Sterne JA. Uses and abuses of meta-analysis. Clin Med 2001;1(6):478-84.
6. Plackett RL. Studies in the history of probability and statistics: VII. The principle of the arithmetic mean. Biometrika 1958;45(1):130-5.
7. Chalmers I, Trohler U. Helping physicians to keep abreast of the medical literature: medical and philosophical commentaries, 1773-1795. Ann Intern Med 2000;133(3):238-43.
8. Pearson K. Report on certain enteric fever inoculation statistics. BMJ 1904;3:1243-6.

9. Peters CC. Summary of the penn state experiments on the influence of instruction in character education. J Educ Sociol 1933;7(4):269-72.
10. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312(7023):71-2.
11. Glass GV. Primary, secondary, and meta-analysis of research. Educ Res 1976;5(10):3-8.
12. Egger M, Smith GD. Meta-analysis: potentials and promise. BMJ 1997;315(7119):1371-4.
13. Greenland S. Quantitative methods in the review of epidemiologic literature. Epidemiol Rev 1987;9:1-30.
14. Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. N Eng J Med 1978;299(13):690-4.
15. Collins R, Keech A, Peto R, Sleight P, Kjekshus J, Wilhelmsen L, et al. Cholesterol and total mortality: need for larger trials. BMJ 1992;304(6843):1689.
16. Mitchell JR. Timolol after myocardial infarction: an answer or a new set of questions? BMJ 1981;282(6276):1565-70.
17. Hampton JR. The use of beta blockers for the reduction of mortality after myocardial infarction. Eur Heart J 1981;2(4):259-68.
18. Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987;106(3):485-8.
19. Mulrow CD. Rationale for systematic reviews. BMJ 1994;309(6954):597-9.
20. Mulrow CD, Cook DJ, Davidoff F. Systematic reviews: critical links in the great chain of evidence. Ann Intern Med 1997;126(5):389-91.
21. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the potsdam consultation on meta-analysis. J Clin Epidemiol 1995;48(1):167-71.
22. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ 1997;315(7121):1533-7.
23. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995;123:A-12.
24. Lago RM, Singh PP, Nesto RW. Congestive heart failure and cardiovascular death in patients with prediabetes and type 2 diabetes given thiazolidinediones: a meta-analysis of randomised clinical trials. Lancet 2007;370(9593):1129-36.
25. Wildridge V, Bell L. How CLIP became ECLIPSE: a mnemonic to assist in searching for health policy/ management information. Health Info Libr J 2002;19(2):113-5.
26. Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. Epidemiol Rev 1992;14:154-76.
27. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116(1):78-84.
28. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126(5):376-80.
29. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994;309(6964):1286-91.
30. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. Control Clin Trials. 1985;6(4):306-17.
31. Greenhalgh T. How to read a paper. The Medline database. BMJ 1997;315(7101):180-3.
32. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA 1998;280(3):278-80.
33. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. J Clin Epidemiol 1992;45(3):255-65.
34. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995;16(1):62-73.
35. Berlin JA, Rennie D. Measuring the quality of trials: the quality of quality scales. JAMA 1999;282(11):1083-5.
36. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998;51(12):1235-41.
37. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869.
38. Elwood JM. Critical appraisal of epidemiological studies and clinical trials. New York: Oxford University Press, 2007.
39. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354(9193):1896-900.
40. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.
41. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88.
42. Hunter JE, Schmidt FL. Fixed effects vs. random effects meta-analysis models: implications for cumulative research knowledge. Int J Select Assess 2000;8(4):275-92.
43. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-58.
44. Brown P, Brunnhuber K, Chalkidou K, Chalmers I, Clarke M, Fenton M, et al. How to formulate research recommendations. BMJ 2006;333(7572):804-6.
45. Leizorovicz A, Haugh MC, Chapuis FR, Samama MM, Boissel JP. Low molecular weight heparin in prevention of perioperative thrombosis. BMJ 1992;305(6859):913-20.

46. Nurmohamed MT, Buller HR, Dekker E, Hommes DW, Rosendaal FR, Briet E, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. Lancet 1992;340(8812):152-6.
47. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273(5):408-12.
48. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ 1997;315(7109):640-5.
49. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. JAMA 1992;267(3):374-8.
50. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279(4):281-6.
51. Dickersin K. How important is publication bias? A synthesis of available data. AIDS Educ Prev 1997;9(1 Suppl):15-21.
52. Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, et al. Should unpublished data be included in meta-analyses? Current convictions and controversies. JAMA 1993;269(21):2749-53.
53. Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? J Clin Epidemiol 1995; 48(1):159-63.
54. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996;347(8998): 363-6.
55. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997;350(9074):326-9.
56. McDonald CJ. Medical heuristics: the silent adjudicators of clinical practice. Ann Intern Med 1996;124(1 Pt 1):56-62.
57. Naylor CD. Grey zones of clinical practice: some limits to evidence-based medicine. Lancet 1995;345(8953):840-2.
58. Tanenbaum SJ. What physicians know. N Eng J Med 1993;329(17):1268-71.