



Review Article

Synthesizing Quantitative Evidence for Evidence-based Nursing: Systematic Review



Eui Geum Oh, PhD, RN *

College of Nursing, Mo-Im Kim Nursing Research Institute, Yonsei University, Seoul, South Korea

ARTICLE INFO

Article history:

Received 6 May 2016

Accepted 9 May 2016

Keywords:

evidence-based practice

meta-analysis

systematic review

SUMMARY

As evidence-based practice has become an important issue in healthcare settings, the educational needs for knowledge and skills for the generation and utilization of healthcare evidence are increasing. Systematic review (SR), a way of evidence generation, is a synthesis of primary scientific evidence, which summarizes the best evidence on a specific clinical question using a transparent, a priori protocol driven approach. SR methodology requires a critical appraisal of primary studies, data extraction in a reliable and repeatable way, and examination for validity of the results. SRs are considered hierarchically as the highest form of evidence as they are a systematic search, identification, and summarization of the available evidence to answer a focused clinical question with particular attention to the methodological quality of studies or the credibility of opinion and text. The purpose of this paper is to introduce an overview of the fundamental knowledge, principals and processes in SR. The focus of this paper is on SR especially for the synthesis of quantitative data from primary research studies that examines the effectiveness of healthcare interventions. To activate evidence-based nursing care in various healthcare settings, the best and available scientific evidence are essential components. This paper will include some examples to promote understandings.

Copyright © 2016, Korean Society of Nursing Science. Published by Elsevier. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Evidence-based practice (EBP) is an important element in patient safety and quality of healthcare. The landmark report from the Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century*, described a healthcare system that is highly disintegrated and lacking care coordination [1]. The Roundtable on Evidence-Based Medicine from the Institute of Medicine has been called to help transform the way evidence on clinical effectiveness is generated and used to improve health and healthcare. Participants have set a goal that, by the year 2020, 90.0% of clinical decisions will be supported by accurate, timely, and up-to-date clinical information, and will reflect the best available evidence [2]. Quickly following these reports, recommendations were made in the *Quality Chasm* series that underscored the centrality of EBP as a solution in redesigning care that is effective and safe [1–3]. The EBP movement was significantly accelerated by these reports,

and key recommendations were made, which were (a) to provide services based on scientific knowledge to all who could benefit [1], (b) to educate all healthcare professionals to deliver evidence-based care [3] and (c) to assess effectiveness of clinical services to provide unbiased information about what really works in healthcare [2].

EBP is a problem-solving approach to patient care that incorporates the conscientious use of current best evidence available from well-designed research studies, clinical expertise and assessment. Patient values and preferences are also incorporated within a caring context [4–7]. According to Joanna Briggs Institute (JBI) model of EBP, evidence-based healthcare is a cyclical process. Global healthcare needs, as identified by clinicians or patients/consumers, are addressed through the generation of research evidence that is effective, feasible, appropriate and meaningful to specific populations, cultures and settings. This evidence is collected and the results are appraised, synthesized and transferred to service delivery settings and health professionals who utilize it and evaluate its impact on health outcomes, health systems and professional practice. Therefore, in order to work in and use healthcare systems globally, one should consider evidence generation, different forms of evidence in a formal assessment called a

* Correspondence to: Eui Geum Oh, PhD, RN, College of Nursing, Mo-Im Kim Nursing Research Institute, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul, 03722, South Korea.

E-mail address: euigeum@yuhs.ac

systematic review (evidence synthesis), dissemination of information in appropriate, relevant formats to inform health systems, health professionals and consumers (evidence transfer), and effective implementation of evidence and evaluation of its impact on healthcare practice (evidence utilization) [8].

As implementation of EBP has become a core competency among healthcare professional in promoting high-quality, patient-centered healthcare, the need for acquisition of skills in evidence generation and utilization has increased. Several organizations have contributed to the preparation of systematic reviews, including the National Institute of Health and Clinical Excellence in the UK, the Evidence-based Practice Center Program, funded by the Agency for Healthcare Research and Quality in the United States, the JBI, and the International Campbell and Cochrane Collaborations, with the latter being the largest single producer of systematic reviews in healthcare, with more than 5,200 published by the end of 2015 (Cochrane Collaboration). Often, healthcare providers and policy makers are overwhelmed with an unmanageable amount of information, including evidence from healthcare research. There is a need for time, skills and resources to find, appraise and interpret this evidence and to incorporate it into healthcare decisions.

Therefore, the purposes of this paper are to introduce concept and characteristics of SR, and to explain major steps in synthesizing quantitative evidences suggested by the JBI and Cochrane methodology in conducting SR.

Definition and characteristic of SR

SR, which is also called “research synthesis” is an attempt to integrate empirical data for the purpose of uncovering international evidence and producing statements about that evidence to guide decision making. SR requires explicit and exhaustive reporting of the methods used in synthesis [9,10]. The characteristics of SR are (a) protocol-driven process, (b) clearly stated set of objectives with predefined eligibility criteria for studies, (c) explicit and reproducible methodology, (d) systematic search that attempts to identify all studies that would meet the eligibility criteria, (e) assessment of the validity of the findings of the included studies, and (f) systematic presentation, and synthesis, of the characteristics and findings of the included studies [11]. SRs have become the “gold standard” in the synthesis of literature at each evidential level, as they enable rigorous, transparent and replicable analysis of all relevant study results [12]. The most common type of clinical questions for SR is the treatment effectiveness and the synthesis of data from randomized controlled trials in quantitative studies [10]. However, other types of quantitative research such as quasiexperimental, cohort, and cross-sectional studies can be valuable for SR in nursing care.

SRs provide a reliable estimate because they are required to follow strict scientific design based on explicit, prespecified and reproducible methods. They can also provide insights as to where knowledge is lacking which provides guidance for future research [13].

Methodological process for conducting SR

Review title of SR

The title of the SR protocol should be as descriptive as is reasonable and reflect relevant information. If the review aims to examine clinical effectiveness, this should be stated in the title. If specific interventions and/or patient outcomes are to be examined, these should also be included. If possible, the setting and target population should also be stated [14], for example, “A systematic review of the effectiveness of lifestyle interventions for improving bone health in women at high risk of osteoporosis” [15]. This

example provides readers with an indication of the population, “women at high risk of osteoporosis”, the interventions, “lifestyle interventions”, and the outcome of interest, “bone health”, as well as the fact that it is a SR. Ensuring the relevant fields of the Population, Intervention, Comparison intervention, Outcomes (PICO) reminder are incorporated in the title assists peer reviewers as well as end users to identify the scope and relevance of the review.

Developing a review question

Sackett et al [7] stated that a good clinical question should have four essential factors: (a) the patient or problem in question; (b) the intervention, test, or exposure of interest; (c) comparison interventions (if relevant); (d) the outcome, or outcomes, of interest. It should be clear, directly focused on the problem at hand, and answerable by searching the medical literature [16]. The clinical question in PICO format aims to clarify its purpose and is used to define the properties of studies to be considered for inclusion in the review. PICO is used to construct a clear and meaningful question when searching for quantitative evidence [17].

Population within a specific setting within a specific time frame should be described in the clinical questions for SR [17]. There are no subgroups or exclusions described, thus all patients meeting the described criteria would be included in the analysis for each outcome. Specific reference to population characteristics, either for inclusion or exclusion should be based on a clear, scientific justification rather than based on unsubstantiated clinical, theoretical or personal reasoning [17].

Interventions of interest are nursing care, treatment, or exposure. The intervention should be described in detail, particularly if it is complicated. Consideration should also be given to whether there is a risk of exposure to the intervention in comparator groups in the included primary studies [17].

Comparison is being made with the intervention of interest. For JBI reviews of effectiveness, the comparator is the one element of the PICO mnemonic that can be either left out of the questions, or posited as a generalized statement. SR of effectiveness based on the inclusive definition of evidence adopted by JBI often seeks to answer broader questions about complex interventions [17]. Usually there are active and passive comparators. Active comparators are specific comparators of one's interest. For example, if one is interested in comparing clinical effectiveness of different types of exercise such as aerobic exercise versus nonaerobic exercise in metabolic syndrome, the nonaerobic exercise group can be an active comparator. Otherwise, passive comparator is a current status or placebo comparator. For example, if one is interested in knowing the effectiveness of exercise, the types of passive comparator can be a group with no exercise or maintaining usual daily activities of life.

Outcomes are the measures of effectiveness of the intervention. The protocol should list all the outcome measures presented in the study. The relevance of each outcome to the review objective should be apparent from the background section. Outcomes should be measurable and appropriate to the objective of the SR. It is useful to list outcomes and identify them as primary or secondary, short-term or absolute and discuss which ones will be included. It is also important to consider and include nursing outcomes for both clients' perspectives such as functional status and/or quality of life, and those of the nursing professionals such as nurses' satisfaction.

Developing a search strategy: a guide to evidence-based information retrieval

After formulating the clinical question with the format of PICO, the next step is to search for the relevant evidence that will help

answer the clinical questions. Unfortunately, studies have shown that, even in countries where hospitals have facilities for internet access allowing healthcare personnel access to a number of electronic databases, many people are not familiar with the process of carrying out efficient search [18,19]. Effective search aims to maximize the potential of retrieving relevant articles within the shortest possible time.

The first step for searching is a plan for optimal search strategies. Careful consideration is required in searching, such as different spelling (e.g., American spelling or British English spelling), synonyms, and plurals. Also, when constructing the search, it is good to break up the search question, and not to overcomplicate the search by including too many terms. Using MeSH/EMTREE terms, which are internationally acknowledged, is the way to search for best evidence.

In quantitative research, the aim is usually to establish changes in the other variable (experimental studies), and/or imply a correlation or association between variables (observational studies). For the best available evidence, study design (e.g., SRs, RCTs, quasiexperimental studies, cohort studies), or the level of evidence should be considered. For instance, if one is interested in knowing the effectiveness of exercise on metabolic syndrome, studies conducted using RCTs, experimental design, or quasiexperimental design should be included in the search.

Searching for evidence is a three-step process: (a) exploratory search; (b) implement a standardized tested search strategy within each selected database; (c) review the reference list of retrieved studies. Prior to commencing SR it is important to search the existing SR libraries to ensure that the review being planned has not already been conducted or is currently being updated. (e.g., The Cochrane Library, JBI Library of Systematic Reviews and the Centre for Reviews and Dissemination).

The comprehensiveness of searching and documenting the databases searched is a core component of the SR's credibility. In addition to databases of published research, there are several online sources of grey or unpublished literature that should be considered. Grey literature is a term that refers to papers, reports, technical notes or other documents produced and published by governmental agencies, academic institutions and other groups that are not distributed or indexed by commercial publishers. Many of these documents are difficult to locate and obtain. Rather than compete with the published literature, grey literature has the potential to complement and communicate findings to a wider audience, as well as to reduce publication bias.

Critical appraisal of retrieved evidences

Critical appraisal is a process for increasing the effectiveness of reading, by encouraging systematic assessment of reports of research evidence to see which ones can best answer clinical problems and inform “best practice” [20]. Researchers should appraise critically because combining results of poor quality research may lead to biased or misleading estimates of effectiveness [10]. Therefore, the aim of critical appraisal is to establish validity and to establish the risk of bias [10]. In critical appraisal, the primary reviewer and the secondary reviewer should discuss what is considered acceptable to the needs of the review in terms of the specific study characteristics. The reviewers should be clear on what constitutes acceptable levels of information to allocate a positive appraisal being compared with a negative appraisal, or an appraisal of “unclear” or “not applicable”. This discussion should take place before independently conducting the appraisal. Selection of tools for critical appraisal need to be based on the study design. There are a variety of gold standard checklists and tools available to assess the validity of studies. Most of these tools use a

series of criteria that can be scored as being met, not met or unclear. The decision as to whether or not to include a study can be made based on meeting a predetermined proportion of all criteria, or on certain criteria being met. It is also possible to weight the different criteria differently, for example blinding of assessors (to prevent detection bias) may be considered twice as important as blinding the caregivers (to prevent performance bias). It is important that appraisal tools are appropriate for the design of the study; this is so that the questions of the tool are specific to that study design [14]. The decisions about the scoring system and the cut-off for inclusion should be made in advance, and be agreed upon by all participating reviewers before commencing the critical appraisal.

Ranking the quality of evidence on effectiveness is to what extent the study design minimizes bias or demonstrates validity. Hierarchy of evidence is most often used, with levels of quality equated with specific study designs [10]. RCTs are the “gold standard” for primary study design upon which to base decisions on the effectiveness of healthcare interventions, but they are not necessarily appropriate, or ethical, to answer other questions [21]. In RCTs, the most commonly used tool for critical appraisal is Cochrane’s “Risk of bias” table. The standard “Risk of bias” table includes (a) random sequence generation (selection bias), (b) allocation concealment (selection bias), (c) blinding of participants and personnel (performance bias), (d) blinding of outcome assessment (detection bias), (e) incomplete outcome data (attrition bias) and (f) other bias. For each item, the table provides a description of what was reported to have happened in the study and a subjective judgement regarding protection from bias (“Yes” for a low risk of bias, “No” for a high risk of bias; “Unclear” otherwise) [22].

Data extraction

Data extraction refers to the process of identifying and recording relevant details from the original (e.g. primary) research studies that will be included in the SR. A standardized extraction tool is used to minimize the risk of error when extracting data. Other error-minimizing strategies include ensuring that both reviewers have practiced using the extraction tool and can apply the tool consistently. It is also recommended that reviewers extract data independently before conferring. This process include details of the types of data extracted from the included studies, as predetermined in the protocol. If no data was available for particular outcomes, one can consider contacting the corresponding author, or discuss it in the review. The included studies may include several outcomes; however the review should focus on extracting information related to the research questions and outcomes of interest. Information that may impact the generalizability of the review findings such as study method, setting and population characteristics should also be extracted and reported. Population characteristics include factors such as age, past medical history, comorbidities, complications or other potential confounders. JBI-MAStARI reduces errors in data extraction by using two independent reviewers and a standardized data extraction instrument [17]. The data extracted will vary depending on the review question. However, it will generally either be dichotomous or continuous in nature. Dichotomous data will include the number of participants with the exposure/intervention (n) and the total sample (N) for both control and treatment groups. Classically, this is stated as n/N ; therefore there will be two columns of data for each outcome of interest. For continuous data, the mean and standard deviation, plus sample size are extracted for each specified outcome for both the control and intervention (or exposure) group. Typically, this is expressed as “ $M(SD)/n$ ” where n is the sample size for the particular group. If only the standard error (SE) is reported, SD can be calculated from SE , as long as the sample size

(n) is known. The equation for converting from SE to SD is simply: $SD = SE \times \sqrt{n}$.

In some cases it may not be possible to extract all necessary raw data from an included study for the SR, as sometimes only aggregated data are reported, or perhaps data from two different patient populations have been combined in the data analysis, and the review needs to focus on only one of the patient populations. In these circumstances, the standard approach is to make contact with the authors of the publication and seek their assistance in providing the raw data. Most researchers are obliging when it comes to these requests provided that records are still available. If the study's authors do not respond or if the data is unavailable, this should be noted in the report. In addition to the data, conclusions that study authors have drawn based on the data are also extracted. It is useful to identify the study authors' conclusions and establish whether there is agreement with the conclusions made by the authors of the SR.

Data synthesis

Extracted data should be synthesized for the next process. A synthesis can either be descriptive (narrative synthesis) or statistical (meta-analysis). If the data is heterogeneous and presented as a narrative summary, potential sources of heterogeneity should be discussed (e.g., clinical, methodological or statistical) as well as on what basis it was determined inappropriate to combine the data statistically [10]. Also, it is important to combine the studies in an appropriate manner using methods appropriate to the specific type and nature of data that has been extracted. The methods of synthesis should be congruent with the description of the type of data for extraction. Therefore, data synthesis should describe the specific types of data and effect measures used [10]. Statistical combination of study data provides a summary estimate, using transparent rules specified in advance [23].

This allows the overall effect of a treatment/intervention to be determined. Although the ultimate aim for a quantitative SR is to combine study data in meta-analysis, this is not always appropriate or possible. Data from two or more separate studies are required to generate a synthesis. Studies to be included in the meta-analysis should be similar to each other so that generalization of results is valid. The three main areas surrounding data that should be considered when deciding whether or not to combine data are (a) clinical—are the patient characteristics similar (e.g., age, diagnoses, comorbidities, treatments)? (b) methodological—do the studies use the same study design and measure the same outcomes? (c) statistical—were outcomes measured in the same way, at the same time points, using comparable scales? When considering conducting a meta-analysis, the statistical methods and software (e.g., JBI-MASARI or Cochrane-RevMan) can be used. Also, effect size which describes the relationship between two variables and is represented by a square on a forest plot, and heterogeneity which is the amount of variation in the characteristics of included studies should be considered [10].

Reporting and evaluation of SR

There is no standardized international approach to structuring how the findings of reviews will be reported. The audience for the review should be considered when structuring and writing up the findings. Meta-view graphs represent a specific item of analysis that can be incorporated into the results section of a review. However, the results are more than the meta-view graphs. Whether it is structured based on the intervention of interest, or some other structure, the content of this section needs to present the results

with clarity using the available tools (meta-view graphs, tables, figures) supported by textual descriptions.

The results section should be framed in such a way that as a minimum, the following fields are described in the protocol as either planned for reporting, or given consideration by the reviewers in preparing their SR report as per the following example: number of studies identified; number of retrieved studies; number of studies matching preferred study design (i.e., RCTs); number and design of other types of studies; number of appraised studies; number of excluded studies and overview of reasons for exclusion; number of included studies. These results are commonly written in narrative style, and also illustrated with a flow diagram.

Systematic reviews should follow a logical, linear process. Flaws in their presentation in peer-reviewed journals led to the establishment of the Preferred Items for Reporting Systematic Reviews and Meta-analyses (PRISMA) reporting guidelines [24]. An extension of the PRISMA statement has been developed for protocols (PRISMA-P). PRISMA-P is intended to guide the development of protocols of systematic reviews and meta-analyses evaluating therapeutic efficacy. PRISMA-P is meant to be used primarily by authors preparing systematic review protocols for publication, public consumption, or otherwise. PRISMA-P is also used for journal editors and peer reviewers gauging the adequacy of review protocols for publication [25].

In addition, the “assessment of multiple systematic reviews” is also used for evaluation of methodological quality of SR [26]. It contains 11 items including the following: (a) was an a priori design provided? (b) was there duplicate study selection and data extraction? (c) was a comprehensive literature search performed? (d) was the status of publication (i.e., grey literature) used as an inclusion criterion? (e) was a list of studies (included and excluded) provided? (f) were the characteristics of the included studies provided? (g) was the scientific quality of the included studies assessed and documented? (h) was the scientific quality of the included studies used appropriately in formulating conclusions? (i) were the methods used to combine the findings of studies appropriate? (j) was the likelihood of publication bias assessed? (k) was the conflict of interest stated? [26]

Discussion

SR protocol is important because it allows reviewers (a) to plan carefully and thereby anticipate potential problems, (b) to explicitly document what is planned before they start their review; (c) to prevent arbitrary decision making with respect to inclusion criteria and extraction of data; and (d) to reduce redundancy of efforts and enhance collaboration, when available.

This paper was written to provide an introductory guide to reliably, transparently and rigorously conduct SRs on the effects of healthcare interventions. It also includes an extensive discussion and description of the parameters for meta-analysis, what the process is, what the necessary considerations related to ensuring that the meta-analysis is conducted appropriately and with transparency are. Importantly the limitations of meta-analysis are discussed, and directions on how to interpret and use the results of a meta-analysis even where the results are potentially confounded by statistical heterogeneity. These are essential considerations and the new reviewer needs to be aware that a metaview graph is not the end of the task. The graph needs to be interpreted appropriately, and further investigation may be necessary.

As this is an introductory text and/or educational article for SR, the discussion in this paper finishes not with methods for meta-analysis, but with more pragmatic directions and recommendations on how to write up a completed review report. The guidance in this section is mostly based on the processes developed and used

across Cochrane, JBI and its international collaboration. These are highly standardized methods that are associated with high quality syntheses for the new reviewer, the student, or those looking for introductory information on methods. There are also numerous other texts on this topic, however in this paper an introduction to the most acceptable and standardized procedures of SR is provided.

Sometimes it may be a challenge for nurses to generate evidence through rigorous review methods. Therefore, collaborative efforts among nurses, synthesis scientists, and information professionals are recommended for conducting a transparent and qualified SR.

Conflicts of interest

The author claims no conflicts of interest.

References

- Richardson WC, Berwick DM, Bisgard J, Bristow L, Cassel C. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press, Committee on Quality of Health Care in America, Institute of Medicine; 2001.
- Olsen L, Aisner D, McGinnis JM. Institute of Medicine roundtable on evidence-based medicine. Washington, DC: National Academies Press, Institute of Medicine; 2007.
- Knebel E, Greiner AC. Health professions education: a bridge to quality. Washington, DC: National Academies Press, Institute of Medicine; 2003.
- Melnik BM, Fineout-Overholt E. Evidence-based practice in nursing and healthcare: a guide to best practice. Philadelphia (Pennsylvania): Lippincott Williams & Wilkins; 2011.
- Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". *JAMA*. 1998;280(15):1336–8.
- van Achterberg T, Schoonhoven L, Grol R. Nursing implementation science: how evidence-based nursing requires evidence-based implementation. *J Nurs Scholarsh*. 2008;40(4):302–10. <http://dx.doi.org/10.1111/j.1547-5069.2008.00243.x>
- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 2nd ed. New York: Churchill Livingstone; 2000.
- Pearson A, Wiechula R, Court A, Lockwood C. The JBI model of evidence-based healthcare. *Int J Evid Based Healthc*. 2005;3(8):207–15. <http://dx.doi.org/10.1111/j.1479-6988.2005.00026.x>
- Aromataris E, Pearson A. The systematic review: an overview. *Am J Nurs*. 2014;114(3):53–8. <http://dx.doi.org/10.1097/01.NAJ.0000444496.24228.2c>
- Joanna Briggs Institute. Joanna Briggs Institute reviewers' manual: 2013 edition. Adelaide (Australia): University of Adelaide, Joanna Briggs Institute; 2013.
- Pearson A, Stern C, Weeks S. Translation science and the JBI model of evidence-based healthcare. Philadelphia (PA): Lippincott Williams & Wilkins; 2011.
- Howick J. Oxford Centre for Evidence-based Medicine - Levels of Evidence (March 2009). Oxford (UK): Centre for Evidence-Based Medicine, Oxford University; 2009 [cited 2016 Apr 17]. Available from: <http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>
- University of York, Centre for Reviews Dissemination. Systematic reviews : CRD's guidance for undertaking reviews in health care. York (UK): Author; 2009.
- Joanna Briggs Institute. Joanna Briggs Institute reviewers' manual: 2008 edition. Adelaide (Australia): University of Adelaide: Author; 2008.
- Oh EG, Lee JE, Yoo JY. A systematic review of the effectiveness of lifestyle interventions for improving bone health in women at high risk of osteoporosis. *JBI Libr Syst Rev*. 2012;10(30):1738–84.
- Roberts KL. Evidence-based practice: an idea whose time has come. *Collegian*. 1998;5(3):24–7.
- Joanna Briggs Institute. Joanna Briggs Institute reviewers' manual: 2014 edition. Adelaide (Australia): University of Adelaide: Author; 2014.
- Rosenberg WM, Deeks J, Lusher A, Snowball R, Dooley G, Sackett D. Improving searching skills and evidence retrieval. *J R Coll Physicians Lond*. 1998;32(6):557–63.
- Jordaan M, Jones R. Adoption of Internet technology by UK postgraduate centres: a questionnaire survey. *Health Libr Rev*. 1999;16(3):166–73.
- O'Rourke A. Critical appraisal. In: Bowling A, Ebrahim S, editors. Handbook of health research methods: investigation, measurement and analysis. Maidenhead (England): Open University Press; 2005.
- Craig JV, Smyth RL. The evidence-based practice manual for nurses. 2nd ed. New York: Churchill Livingstone; 2007.
- Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. <http://dx.doi.org/10.1136/bmj.d5928>
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Chichester (UK): John Wiley & Sons; 2009.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. <http://dx.doi.org/10.1371/journal.pmed.1000097>
- Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;349:g7647. <http://dx.doi.org/10.1136/bmj.g7647>
- 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. <http://dx.doi.org/10.5388/10.1186/1471-2288-7-10>