Pain Physician 2009; 12:819-850 • ISSN 1533-3159

Evidence-Based Medicine

Evidence-Based Medicine, Systematic Reviews, and Guidelines in Interventional Pain Management: Part 6. Systematic Reviews and Meta-Analyses of Observational Studies

Laxmaiah Manchikanti, MD[1], Sukdeb Datta, MD[2], Howard S. Smith, MD[3], and Joshua A. Hirsch, MD[4]

From: [1]Pain Management Center of Paducah, Paducah, KY; [2]Vanderbilt University Medical Center, Nashville, TN; [3]Albany Medical College, Albany, NY; and [4]Massachusetts General Hospital and Harvard Medical School, Boston, MA.

Dr. Manchikanti is Medical Director of the Pain Management Center of Paducah, Paducah, KY. Dr. Datta is Director, Vanderbilt University Interventional Pain Program, Associate Professor, Dept. of Anesthesiology, Vanderbilt University Medical Center, Nashville TN. Dr. Smith is Associate Professor and Academic Director of Pain Management for Albany Medical College Department of Anesthesiology, Albany, NY. Dr. Hirsch is Chief of Minimally Invasive Spine Surgery, Depts. of Radiology and Neurosurgery, Massachusetts General Hospital and Associate Professor of Radiology, Harvard Medical School, Boston, MA.

Address Correspondence:
Laxmaiah Manchikanti, MD
2831 Lone Oak Road
Paducah, KY 42003
E-mail: drlm@thepainmd.com

Disclaimer: There was no external funding in preparation of this manuscript. Conflict of Interest: Dr. Datta receives research support from Sucampo Pharmaceuticals and an honorarium from Smith and Nephew. Dr. Hirsch is a consultant for Cardinal Healthcare, a minor shareholder in Medtronic and Cardinal Healthcare, serves on the Steering Committee for KAVIAR trial (volunteer position), and on the Data and Safety Monitoring Board (DSMB), and CEEP trial (volunteer position).

Manuscript received: 08/23/2009
Revised manuscript received:09/004/2209
Accepted for publication: 09/10/2009

Free full manuscript: www.painphysicianjournal.com

Observational studies provide an important source of information when randomized controlled trials (RCTs) cannot or should not be undertaken, provided that the data are analyzed and interpreted with special attention to bias. Evidence-based medicine (EBM) stresses the examination of evidence from clinical research and describes it as a shift in medical paradigm, in contrast to intuition, unsystematic clinical experience, and pathophysiologic rationale. While the importance of randomized trials has been created by the concept of the hierarchy of evidence in guiding therapy, much of the medical research is observational. The reporting of observational research is often not detailed and clear enough with insufficient quality and poor reporting, which hampers the assessment of strengths and weaknesses of the study and the generalizability of the mixed results. Thus, in recent years, progress and innovations in health care are measured by systematic reviews and meta-analyses. A systematic review is defined as, "the application of scientific strategies that limit bias by the systematic assembly, clinical appraisal, and synthesis of all relevant studies on a specific topic." Meta-analysis usually is the final step in a systematic review.

Systematic reviews and meta-analyses are labor intensive, requiring expertise in both the subject matter and review methodology, and also must follow the rules of EBM which suggests that a formal set of rules must complement medical training and common sense for clinicians to integrate the results of clinical research effectively. While expertise in the review methods is important, the expertise in the subject matter and technical components is also crucial.

Even though, systematic reviews and meta-analyses, specifically of RCTs, have exploded, the quality of the systematic reviews is highly variable and consequently, the opinions reached of the same studies are quite divergent. Numerous deficiencies have been described in methodologic assessment of the quality of the individual articles. Consequently, observational studies can provide an important complementary source of information, provided that the data are analyzed and interpreted in the context of confounding bias to which they are prone. Appropriate systematic reviews of observational studies, in conjunction with RCTs, may provide the basis for elimination of a dangerous discrepancy between the experts and the evidence.

Steps in conducting systematic reviews of observational studies include planning, conducting, reporting, and disseminating the results. MOOSE, or Meta-analysis of Observational Studies in Epidemiology, a proposal for reporting contains specifications including background, search strategy, methods, results, discussion, and conclusion. Use of the MOOSE checklist should improve the usefulness of meta-analysis for authors, reviewers, editors, readers, and decision-makers.

This manuscript describes systematic reviews and meta-analyses of observational studies. Authors frequently utilize RCTs and observational studies in one systematic review; thus, they should also follow the reporting standards of the Quality of Reporting of Meta-analysis (QUOROM) statement, which also provides a checklist. A combined approach of QUOROM and MOOSE will improve reporting of systematic reviews and lead to progress and innovations in health care.

Key words: Observational studies, evidence-based medicine, systematic reviews, meta-analysis, randomized trials, case-control studies, cross-sectional studies, cohort studies, confounding bias, QUOROM, MOOSE

Pain Physician 2009; 12:819-850

Evidence-based medicine and comparative effectiveness research have been described as analogous to religion and politics – meaning different things to different people. With the Obama health care reform hotly debated, in the center of which is the comparative effectiveness research, the recent articles published in the New England Journal of Medicine evaluating the effectiveness of vertebroplasty for painful osteoporotic vertebral fractures (1,2) and accompanying editorial (3) raise further questions with regards to the ability of randomized trials to effectively determine the efficacy of an intervention. The accompanying editorial by Weinstein (3) elegantly describes the evidence in the context of comparative effectiveness research as proposed in the American Recovery and Reinvestment Act. Weinstein describes that although clinical trials are an integral part of comparative effectiveness research, from a safety and effectiveness standpoint, data from clinical trials combined with those from registries or other large longitudinal databases are necessary to provide the best evidence.

Over the past decade, 2 major movements have emerged in medicine, both intended to improve patient care (4). The medical humanism movement seeks to understand the patient as a person, focusing on individual values, goals, and preferences with respect to clinical decisions. The second movement is evidence-based practice, which aims to put medicine on a firm scientific footing; experts evaluate the best available data and develop clinical guidelines designed to standardize procedures and therapies. While Obama health care seeks to combine both with a universal coverage and cost containment – the goal is unattainable as these 2 polarizing issues are poised to collide rather than coalesce.

Even though the trials of vertebroplasty (1,2) provide the best available scientific evidence for an informed choice, it remains to be seen whether there will be a paradigm shift in the treatment of vertebral compression fractures with vertebroplasty (3). The history shows us otherwise with lack of effect of such randomized trials on practice patterns with coronary artery revascularization (5-8), intraarticular facet joint injections (9), lumbar interlaminar epidurals (10), and various other treatments (11-15). The growth has not been hindered. In fact, growth has been exponential for some interventions (11,16-19). In the case of vertebroplasty, observational studies suggested that there is an immediate and sustained reduction in pain after this procedure is performed (20). Further, one randomized, open trial involving 34 patients (21) and 2 quasi-experimental, open, controlled, before-after studies that compared vertebroplasty with conservative treatment (22,23) also showed similar results. Further, multiple systematic reviews provided favorable evidence (24-28).

While many aspects of evidence-based medicine including randomized trials, systematic reviews, meta-analyses, and clinical guidelines indicate signs of progress in the effort to keep pace with health care innovations, the medical profession continues to struggle with conflicts of humanism and evidence-based practice. While it is hypothesized that shared decision-making will close the gap, it appears that it may actually widen the gap. Evidence-based medicine is considered as a shift in medical paradigm, which acknowledges that intuition, unsystematic clinical experience, and pathophysiologic rationale are insufficient grounds for clinical decision-making (29-35). The hierarchy of strength of evidence for treatment decisions varies from N of 1 randomized controlled trials (RCTs) on the top, followed by systematic reviews of randomized trials, and ranging all the way down to unsystematic clinical observations (35). Consequently, systematic reviews and meta-analyses are considered popular evidence-based tools for randomized, observational, and diagnostic accuracy studies and they are quite often used to answer complex research questions across many different research domains (36,37).

1.0 An Introduction to Systematic Reviews and Meta-Analyses

The history of systematic reviews has been described in Part 2 (31). Historically, the philosophy dates back to 1747 (38). During the 1960s and 1970s, early systematic review methods were advanced by social scientists (39). The term "systematic review" was coined long before evidence-based medicine (40).

The terminology used to describe systematic reviews and meta-analyses has been described in Part 3 (32). In short, a systematic review utilizes explicit methodology of clearly formulated questions and methods to identify, select, and critically appraise relevant research and then collect and analyze the data from the studies that are included in the review.

Meta-analysis was described in 1904 (41). Meta-analysis incorporates the statistical pooling of data across studies to generate a summary in the form of pooled estimates of effects (42,43). Thus, a meta-analy-

sis ideally starts with an unbiased systematic review that incorporates articles chosen using predetermined inclusion criteria, with meta-analysis constituting the final step in a systematic review, even though meta-analysis may be performed without systematic reviews.

Both systematic reviews and meta-analyses, despite their differences, have many similarities and represent a continuum providing clinicians, researchers, policy-makers, and patients with a synthesis of an unmanageable and exponentially increasing number of manuscripts by linking and correlating huge amounts of information with identification of beneficial or harmful interventions (43,44).

Systematic reviews are labor intensive and require expertise in both the subject matter and review methodology. Systematic reviewers should/must follow the rules of evidence-based medicine which suggests that a formal set of rules must complement medical training and common sense for clinicians to interpret the results of clinical research effectively. Thus, knowing the tools of evidence-based practice is necessary, but not sufficient for delivering the higher quality of patient care. Consequently, expertise in one area or another is not enough and may lead to inaccurate conclusions, in turn leading to inappropriate applications of the results (45-47). Consequently, expertise in the subject matter and review methodology are not only important and crucial but essential. A systematic review is defined as, "the application of scientific strategies that limit bias by the systematic assembly, critical appraisal, and synthesis of all relevant studies on a specific topic" (37,48,49).

Generally systematic reviews and meta-analyses are performed for randomized trials (50-56). However, systematic reviews also may be performed for observational studies and diagnostic accuracy studies (24-28,57-86). However, observational studies have been considered inferior and inadequate in providing evidence (33), generally based on the lack of understanding or politics, leading to the impressions and characterization of evidence-based medicine as a stick by which policy-makers might impact clinical practice (87-94). Further, the ability of randomized trials to determine effectiveness of new treatments applicable to everyday practice and systematic reviews and meta-analyses of randomized trials has been noted to be, at times, methodologically flawed (50,95-99). In fact, in an 1987 survey of 86 English language meta-analyses (50) assessing each publication on 23 items from 6 content areas considered important in the conduct and reporting of meta-analysis of randomized trials, only 24 or 28% of the 86 meta-analysis reported that all 6 content areas had been addressed. An example is that of 4 systematic reviews examining the cardiovascular effects of vitamin E supplements (100-103). Even though, consistency was observed in 3 of the 4 systematic reviews, and, despite the number of studies of these systematic reviews varied from 7 to 84, there was no association between vitamin E and any cardiovascular endpoint (100-102). Further, the fourth review (103) conducted a dose-response analysis for which high doses of vitamin E were shown to significantly increase the risk of all-cause-mortality by 9% to 14%.

The Quality of Reporting of Meta-analyses (QUOROM) statement was developed with details of reporting recommendations published in 1999 to improve the quality of reporting (104). Unfortunately, there continues to be considerable evidence that key information is often poorly reported in systematic reviews, thus diminishing their potential usefulness (105-109). In fact, the reporting evidence has been updated in the form of PRISMA statement for Preferred Reporting Items for Systematic Reviews and Meta-analyses of studies (95,96). Sampson et al (97) attempted to identify validated or evaluated search reporting instruments used in reporting systematic review searches and to compare reported and recommended searching practices. They concluded that there was no clear consensus regarding optimum reporting of systematic review search methods and commonly recommended items show suboptimal reporting.

Most questions in medical research are investigated in observational studies (110-116). Observational studies are also more likely to provide an indication of daily medical practice (117). Consequently, it has been proposed that observational studies and RCTs can be viewed as expressions in the setting of modern clinical research of the steps of observation and experimentation that form the basis of the scientific methodology (116). However, reporting of observational research is often not detailed and clear enough to assess the strengths and weaknesses of the investigations (106,107,110,118-129). Further, multiple deficiencies with overwhelming heterogeneity (130), methodological quality (131,132), deficiencies in statistical methods (133), and other deficiencies (134-136) have been widely reported.

An observational study is defined as an etiologic or effectiveness study, a cross-sectional study, a case

series, a case-control design, a design with historical controls, or a cohort design. The studies of risk factors generally cannot be randomized because they relate to inherent human characteristics or practices and exposing subjects to harmful risk factors is unethical (120-122). Apart from that, observational data may be needed to assess the effectiveness of an intervention in a community as opposed to the special setting of a controlled trial (137); design of diagnostic studies is based on observation rather than randomization (34,138-143).

2.0 Why Systematic Reviews and Meta-Analyses of Observational Studies?

The necessity for systematic reviews and meta-analyses of RCTs has been described (32). It has been stated that systematic reviews are a vital link in the great chain of evidence that stretches from the laboratory bench to the bedside (144). Further, a systematic review provides the mechanism to identify studies with weak designs because their results can be biased, often overestimating the benefits of the treatment being studied (145-149). The results of a single study often apply only to a certain kind of patient or a particular setting, a systematic review of many studies can provide information relevant to a broad range of patients at different treatment doses and in different treatment settings. In addition, a systematic review serves multiple purposes including reduction of a large amount of information to a manageable size, helping to determine whether the results are consistent from study to study and to generalize the results, serves the purpose of reducing the cost as it is less expensive and quicker to conduct than to embark on a new study, may also reduce the delay between publication of research findings and the implementation of new effective treatment strategies, it also combines information from individual studies so that its overall sample size is greater than that of any one study, which leads to an increase in the power of the investigation, and finally, a systematic review limits bias and improves the reliability and accuracy of recommendations because of its formalized and thorough method of investigation (36).

3.0 Why Quality Systematic Reviews?

Assessment of methodologic quality is crucial in all types of studies – moreso for observational studies. Multiple systems and tools have been developed to assess the methodologic quality of observational studies which is the essential part of a systematic review (138).

Empiric research on the quality of systematic reviews has shown that not all systematic reviews are truly systematic (150,151), that the quality of systematic reviews is highly variable (95,99,123-136,152,153), and that the Cochrane reviews, on average, may be more rigorous and better reported than journal reviews (151,154). Even then, some studies have shown deficiencies in Cochrane reviews with methodological problems (155,156). While there are numerous descriptions about systematic reviews and meta-analyses of randomized trials, there is a paucity of literature with systematic reviews and meta-analyses of observational studies (105,120). Similar to Consolidated Standards of Reporting Trials (CONSORT) (157,158), Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines (110,111) have been described to report observational studies. In addition, SQUIRE (Standards for Quality Improvement Reporting Excellence) (126) and STREGA (Strengthening of Reporting of Genetic Association Studies – an extension of the Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement (127,128) have been published. It has been described that coherent and transparent decision rules are needed for deciding when only to include RCTs, when to include non-RCTs, and when to include other types of evidence (159). It is important to include observational studies as RCTs are limited. In a review (160), it was shown that a random 1% sample of meta-analysis published by Cochrane Collaboration in 2003, 6 of 16 reviews included 2 studies or fewer. In addition, 158 of 183 analyses conducted in 7 additional studies were limited to 2 or fewer studies. Thus, addition of information from observational studies, systematic reviews, and meta-analyses may aid in clinical reasoning and establish a more solid foundation for causal inferences. Shrier et al (161) found that advantages including both observational studies and RCTs in a meta-analysis could outweigh the disadvantages in many situations and that observational studies should not be excluded a priori. Thus, it could reduce false inferences based solely on RCTs. In fact, Shrier (162) in a systematic review which included RCTs and observational studies led to a different conclusion that stretching immediately before exercise would not reduce injury. In contrast, the prevailing opinion prior to 1999 that stretching immediately before exercise was of benefit, a recommendation mostly based on 4 small RCTs (161). In addition, a subsequent large RCT

that directly addressed the questions supported Shrier's hypothesis (163), as have other systematic reviews and meta-analyses (164,165).

Shrier et al (161) demonstrate the reasons to include observational studies in systematic reviews. These include the fact that the findings of RCTs sometimes contradict the findings of highly publicized observational studies but sometimes RCTs also contradict the findings of highly publicized RCTs (166). MacLehose et al (167) evaluated in a systematic review the effect sizes derived from randomized and non-randomized studies. MacLehose et al (167) concluded that discrepancies for high quality studies were small but that discrepancies for low quality studies were large. Sacks et al (168) examined the inclusion of studies with historical controls versus RCTs and found that historical control studies produce effect estimates of much larger magnitude. Concato et al (169) found similar estimates of effect for meta-analysis based on RCTs versus high-quality cohort studies. Benson and Hartz (170) found similar results between meta-analysis based on RCTs and on cohort studies performed after 1984.

Ioannidis et al (171) found discrepancies in only 8% of the topics covered by prospective studies. Furlan et al (172) showed that the discrepant results between cohort and RCT studies regarding low back pain were almost all attributable to the quality of the studies and to homogeneity. Thus, well conducted observational studies will yield similar estimates of effect compared with RCTs when the bias created by the potential limitations exclusive to observational studies is small in magnitude compared with the variability and/or bias created by choice of study population, types of subjects willing to enter a study, quality of data acquired, and other random effects. Therefore it is important to study the discrepancies that occur between studies (whether due to the study design or otherwise) because they provide information that can be used for appropriate clinical reasoning and causal inferences (161). The important point is that either qualitatively or quantitatively assesses the probability of bias due to lack of randomization. Even then, some argue against the inclusion of observational studies because the researcher needs to know the estimate of the effect for a particular topic and not whether the observational studies agree with RCTs on average (173,174). This is in contradiction for evidence-based medicine to other sectors, including economic appraisals and bayesian decision theoretic approaches, in which decision-making is a function of both probabilities and utility or loss of function (175,176). Further, including observational studies in a systematic review without a meta-analysis presents fewer problems than including them in one with meta-analysis (161).

Wang and Schoenbaum (135) describing opportunities and limitations in assessing treatment effects by using observation analysis in the context of the American Recovery and Reinvestment Act of 2009 and comparative effectiveness research described that application of observational, quasi-experimental, and other non-experimental methods may also be important in this endeavor. However, they also noted that such methods are inherently susceptible to various types of potential bias and thus present special challenges in the search and generalizable evidence.

In a review about observational research, Bluhm (134) argues that what matters is whether the treatment and control groups are similar with respect to potential confounding factors, not whether they got that way through randomization. In addition, they describe other advantages of non-randomized studies including longevity and size of the trial compared to randomized trials. In summary, it was described that even though the clean randomized trials identified as "best evidence" on the hierarchy of evidence are important and useful, they have limitations. Further, pragmatic, non-randomized studies are a necessary part of the evidence base for medicine, both because they are able to provide information about a larger and more diverse population of patients, and because they are more likely to follow patients at outcome over a long period of time. However, the quality of randomized and non-randomized studies depends in part on whether potentially confounding factors, such as age, gender, or the presence of comorbid conditions, occur in roughly the same proportions in the treatment and the control groups. Thus, it was concluded that there is no reason to maintain a hierarchy of evidence that favors randomized or non-randomized studies. Instead, it was proposed that we must think in terms of a continuum of study designs from clean to pragmatic, with large databases falling at the most pragmatic end of the spectrum. We should acknowledge that clean and pragmatic studies provide equally important kinds of evidence to inform clinical decision-making.

Greene (177) after discussing various limitations of both types of studies, supported an integrative approach that targets the use of observational studies

and RCTs at different stages of the research process based on their respective strengths and weaknesses and seeks to maximize the information gained by joint evaluation of both types of evidence. Recently, an extension of STROBE statement was published as Strengthening the Reporting of Genetic Association Studies, also known as STREGA (127). The authors concluded that despite increasing recognition of a multitude of issues related to reporting, the quality of reporting needs to be improved (178-182).

Morshed et al (133) highlighted the special analytic considerations required for proper reporting and interpretation of observational studies. Lu (132) reviewed study designs, challenges, and strategies to reduce confounding. It was highlighted that there was an expanding body of literature using observational designs, partly because observational studies are less resource intensive than RCTs, as they often use electronic health care data that have already been collected, which have become more available in the last decade, leading to accumulation of large databases. They also noted that observational studies highlighted the increased risk of cardiovascular events with rofecoxib (183,184).

Simunovic et al (131) discussed methodological issues in systematic reviews and meta-analyses of observational studies in orthopedic research and concluded that sometimes, observational studies represent the best available evidence. However, they also warn that observational studies may overestimate treatment or exposure effects subject to selection, information, and confounding biases.

Maguire et al (130) reported overwhelming heterogeneity in systematic reviews of observational anti-epileptic studies. They discussed that systematic reviews of observational studies are prone to significant heterogeneity and bias, which cannot be adequately explained by reported study characteristics. However, they suggested that reporting standards for observational studies of anti-epileptic drugs could be improved by following guidelines for reporting nonrandomized studies of interventions.

4.0 Methodologic Quality Assessment of Systematic Reviews

The Evidence-based Practice Center's (EPC) Partner's Guide (185) from the Agency for Healthcare Research and Quality (AHRQ) states that systematic reviews are only as complete and useful as the evidence that exists on a particular topic or the scope and nature of the evidence questions that guide the review. Along with an explanation of systematic reviews and meta-analyses, empiric research on the quality of systematic reviews has shown that not all systematic reviews are truly systematic (150,151,186-189). Often, systematic reviews seem to ignore the basic principles of evidence-based medicine and the very different hierarchies necessary for issues of diagnosis, prognosis, and therapy leading to substantial variability of the results (152,153) leading to the conclusion that along with an increase of systematic reviews and meta-analyses, the ignorance also continues to increase and it is mandatory that reviewers follow appropriate rules (54,55,189-197).

Oxman et al (193) provided guidance on synthesis of evaluation with 2 instruments critically appraising systematic reviews (138,194). West et al (138) reviewed different instruments for critical appraising systematic reviews and found 20 systems concerned with the appraisal of systematic reviews or meta-analysis, including one scale, 10 checklists, and 9 guidance documents, and identified 7 domains that they considered important to appraise: study question, search strategy, inclusion and exclusion criteria, data extraction, study quality, data synthesis and analysis, and funding or ownership (Table 1). Another review (194) identified 240 quality assessment instruments for systematic reviews, RCTs, and observational studies, and nearly 50 evidence grading systems. Following this critical and extensive review, the AMSTAR (A Measurement Tool to Assess Systematic Reviews) 2005 was selected as the best instrument for appraising systematic reviews (Table 2) (194,195).

5.0 How to Conduct Systematic Reviews and Meta-analyses

Guidance has been provided for reading, writing, and interpreting systematic reviews and meta-analyses (95-99,123-133,153,177,198-203). Two important objectives in conducting systematic reviews are to summarize the evidence on a specific clinical question and to critically evaluate the quality of the primary studies (43,44,143,144,204-208).

Stroup et al (120) proposed a checklist containing specifications for reporting of Meta-analysis of Observational Studies in Epidemiology (MOOSE), including descriptions of background, search strategy, methods, results, discussion, and conclusion.

Cochrane methodology (207) suggestions include formulation of the question, identification of

Table 1. *Domains in the Agency for Healthcare Research and Quality (AHRQ) criteria for evaluating systematic reviews.*

DOMAIN	ELEMENTS*
Study question	• **Question clearly specified and appropriate**
Search strategy	• ***Sufficiently comprehensive and rigorous with attention to possible publication biases*** • *Search restrictions justified (e.g., language or country of origin)* • Documentation of search terms and databases used • Sufficiently detailed to reproduce study
Inclusion and exclusion criteria	• **Selection methods specified and appropriate, with *a priori* criteria specified if possible**
Interventions	• **Intervention(s) clearly detailed for all study groups**
Outcomes	• **All potentially important harms and benefits considered**
Data extraction †	• Rigor and consistency of process • Number and types of reviewers • Blinding of reviewers • Measure of agreement or reproducibility • Extraction of clearly defined interventions/exposures and outcomes for all relevant subjects and subgroups
Study quality and validity	• ***Assessment method specified and appropriate*** • Method of incorporation specified and appropriate
Data synthesis and analysis	• ***Appropriate use of qualitative and/or quantitative synthesis, with consideration of the robustness of results and heterogeneity issues*** • Presentation of key primary study elements sufficient for critical appraisal and replication
Results	• **Narrative summary and/or quantitative summary statistic and measure of precision, as appropriate**
Discussion	• **Conclusions supported by results with possible biases and limitations taken into consideration**
Funding or sponsorship	• ***Type and sources of support for study***

* Elements appearing in italics are those with an empirical basis. Elements appearing in bold are those considered essential to give a system a Yes rating for the domain.
† Domain for which a Yes rating required that a majority of elements be considered.

Adapted from West S et al. *Systems to Rate the Strength of Scientific Evidence, Evidence Report, Technology Assessment No. 47.* AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality, 2002. www.thecre.com/pdf/ahrq-system-strength.pdf (138).

relevant studies, study selection, study quality assessment, collection of data, analysis and interpretation of results, presentation of results, and finally, improvement and update of the reviews. Others (208) also described methodology similar to that of Cochrane which include formulating a question, finding relevant studies, selecting and assessing those studies, summarizing and synthesizing relevant study results, interpreting the review findings, and updating the review.

Table 2. A measurement tool to assess systematic reviews (AMSTAR), 2005.

1. Was an 'a priori' design provided?	The research question and inclusion criteria should be established before the conduct of the review.	Yes	No	Can't answer	Not applicable
2. Were there duplicate study selection and data extraction?	There should be at least two independent data extractors and the consensus procedure for disagreements should be reported.	Yes	No	Can't answer	Not applicable
3. Was a comprehensive literature search performed?	At least two electronic sources should be searched. The report must include years and databases (e.g., Central, EPOC, and MEDLINE). Key words and/or MeSH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.	Yes	No	Can't answer	Not applicable
4. Was the status of publication (i.e., grey literature) used as an exclusion criterion?	The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status.	Yes	No	Can't answer	Not applicable
5. Was a list of studies (included and excluded) provided?	A list of included and excluded studies should be provided.	Yes	No	Can't answer	Not applicable
6. Were the characteristics of the included studies provided?	In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions, and outcomes. The ranges of characteristics in all the studies analyzed (e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases) should be reported.	Yes	No	Can't answer	Not applicable
7. Was the scientific quality of the included studies assessed and reported?	'A priori' methods of assessment should be reported (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.	Yes	No	Can't answer	No applicable
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.	Yes	No	Can't answer	Not applicable
9. Were the methods used to combine the findings of studies appropriate?	For the pooled results, a test should be done to ensure the studies were combinable, to assess the homogeneity (i.e., Chi-squared test for homogeneity, I2). If heterogeneity exists, random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?).	Yes	No	Can't answer	Not applicable
10. Was the likelihood of publication bias assessed?	An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot) and statistical tests (e.g., Egger regression test).	Yes	No	Can't answer	Not applicable
11. Was the conflict of interest stated?	Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.	Yes	No	Can't answer	Not applicable

Source: Oxman AD et al. *Improving the use of research evidence in guideline development*: 8. Synthesis and presentation of evidence. Health Res Policy Syst 2006; 4:20 (195).

5.1 Formulating a Question

Clearly framed questions are essential for determining the structure of a systematic review or meta-analysis of observational studies similar to RCTs (208-210). A properly formulated question will guide much of the review process, including strategies for locating and selecting studies or data, for critically appraising their relevance and validity, and for analyzing variations among their results.

5.1.1 Key Components of a Question

A well formulated question consists of several key components that provide criteria for selecting studies (53,211). Thus, components of a question should specify the types of participants, types of interventions or exposures, and the types of outcomes that are of interest. The types of studies also should be specified (prospective or case-controlled, etc.). However, it is not necessary to provide equal preci-

sion for each component. Rather, it should be based on the importance.

5.1.1.1 Types of Participants

Inclusion criteria for types of participants must be clear with definitions of the disease or condition that is of interest, such as facet joint pain, discogenic pain, or radicular pain. Next, various demographics must be addressed based on interest of the study if it is a special population group determined by age, sex, race, education, etc., or the presence of a particular condition such as low back pain or lower extremity pain. Finally, the setting is also important such as office practice, ambulatory surgery center, hospital outpatient setting, community setting, or inpatient setting.

In a systematic review, multiple restrictions must be identified, related either to population characteristics or settings (207). Public policy requires studies in specific types of population such as work related injury population, Medicare population, or a disease-specific population. However, it is not justifiable to focus a review on a particular subgroup of people based on some irrelevant factor based on personal interest or bias without an underlying identifiable biological or sociological justification.

5.1.1.2 Types of Interventions

It is crucial to define the interventions in formulating a question, along with the specifications of the interventions that are of interest. Observational studies can be cohort studies, cross-sectional studies, or case-controlled studies. The observations are either in controlled settings or audits, etc. Observational studies generally measure effectiveness, the degree of beneficial effect in clinical practice. In contrast to the explanatory trials that are most commonly conducted in academic settings measuring the efficacy, requiring major funding, observational studies provide the results of benefit of the treatment produced in a routine clinical practice setting.

5.1.1.3 Types of Outcomes

One of the key components of a well-formulated question is the delineation of particular outcomes that are of clinical interest. While pain relief is utilized as the primary outcome, it is becoming increasingly clear that other outcomes including functional status, opioid intake, return to work, or patient satisfaction are increasingly becoming important.

5.1.1.4 Types of Study Designs

STROBE describes the reporting criteria for 3 types of observational studies. These include case-control study, cross-sectional study, and cohort study. However, some studies incorporate more than one design. The most commonly utilized designs for therapeutic interventions are cohort or case-control designs (212).

5.1.2 Importance of a Question

Properly focused questions determine the initial strategies related to the condition being studied, intervention being assessed, and the population being studied. Thus, questions that the review addresses may be broad or narrow in scope, with each one of them associated with their own advantages and disadvantages. While the questions may be refined based on the data which is available during the review, it is essential to guard against bias and modifying questions, as post-hoc questions are more susceptible to the bias than those asked a priori and data-driven questions can generate false conclusions based on spurious results. However, any such change to the protocol that results from revising the question for the review should be documented clearly and limitations described.

5.2 Finding Relevant Studies

Finding the relevant studies, while a complex and time-consuming process, can seriously affect the results and introduce significant bias. The objective of the search is to generate as comprehensive a list as possible of primary studies, both published and unpublished, which may be suitable to answer the question in the review (213-216). Identification of all relevant non-randomized trials by a thorough, unbiased, and comprehensive search strategy is crucial. Consequently, the validity of a systematic review can be determined by the comprehensiveness of the search used to capture the relevant studies. In addition, multiple other issues related to the results, including the level of precision of the effect estimate and minimization of the bias in a systematic review are dependent on the volume of valid information included in the review. The recent analysis of Cochrane reviews of interventional pain management and other extensively quoted reviews (56,217-219) showed a lack of appropriate criteria and absence of many key manuscripts. The same was true of highly outspoken critics of interventional pain management reviews (220,221).

5.2.1 Searching for the Studies

Various types of search criteria have been recommended. Unfortunately, many develop their own criteria which can be significantly inadequate due to lack of understanding of the importance of searching. The most commonly utilized search of MEDLINE, also known as a quick and dirty search, is considered inadequate. Studies have shown that only 30% to 80% of all known published RCTs were identifiable using MEDLINE. Variations in the journals indexed in databases indicate a need to search more than one database to ensure optimal coverage of published literature, in subject, scope, and language of report (222-225). Some subject areas have been shown to require a more comprehensive selection of sources and unrestricted language searching in order to avoid substantial bias and increase the precision, generalizability, and applicability of the findings. However, there is evidence to indicate that exclusion of studies in languages other than English from reviews might make no significant impact to the overall estimates of the effects of treatments (226-230). However, reviewers should at least at minimum add EMBASE to MEDLINE in the search strategy (223). The overlap of EMBASE and MEDLINE has been estimated to be 10% to 87% depending on the topic under investigation (231-235). Thus, at minimum for interventional pain management subjects, an electronic search strategy generally includes 3 databases (MEDLINE, EMBASE, Cochrane Library), a hand search of references for eligible trials, direct contact with the corresponding authors of eligible trials asking for additional published or unpublished trial information (236), and 3 sets of terms which include terms to search for the health condition of interest, terms to search for interventions evaluation, and terms to search for the types of study design.

Sampson et al (97) evaluated search reporting instruments used in reporting systematic review searches and compared the recommended searching practices. They concluded that there was no clear consensus regarding optimal reporting of systematic review search methods and commonly recommended items show suboptimal reporting. They also identified 11 instruments and 18 distinct search-related items addressed by these instruments.

5.3 Study Selection

After the completion of the search for the relevant studies, the studies should be assessed for the relevance to the question posed in the review. The selection process should be explicit and should be conducted in such a way as to minimize risk of errors of judgment and bias (237-239). Quality assessment of primary studies is used at various stages in the review process. As a second step, an explicit and standardized method for selecting studies from among all of those identified and then assessing the selections is a key part of the systematic review. Such a method serves the dual purpose of choosing the highest quality studies and also demonstrates the selection and assessments have been as free from bias as possible (240-245). Study selection criteria must be predetermined and inclusion or exclusion of the studies is made according to the appropriate quality assessment criteria.

Inclusion and exclusion criteria should follow logically from the review questions and they should be defined in terms of the population, their interventions, the outcomes, and the study designs of interest (32). Consequently, the studies included must meet all of the inclusion criteria but none of the exclusion criteria. In general, the inclusion criteria must specify the type of the study design (246).

Study selection also is a multi-stage process; initially encompassing liberal criteria until the citations are generated and searching, varying to more stringent criteria with retrieving the full text of all potential irrelevant citations and reviewing them. Further, a list of excluded studies detailing the reasons for each exclusion may be appropriate (104,105).

5.4 Quality Assessment of Observational Studies

Quality is a construct which includes study quality, the degree to which a study employs measures to minimize biases; focusing on internal validity or methodologic quality, bias or systematic error, a tendency to produce results that depart systematically from the true results, whereas unbiased results are internally valid; internal validity, the degree to which the results of a study are likely to approximate to the truth and which is a prerequisite for external validity; and finally external validity, generalizability or applicability, the extent to which the effects observed in a study are applicable outside of the study – in routine clinical practice. The information gained from quality assessment is crucial in determining the strength of references, inferences, and in assigning grades to recommendations generated with a review. Quality assessment can be used at various stages in a review,

starting with the study selection to data synthesis and interpretation.

The quality assessment of observational studies is crucial as it is for all other types of studies (42,247-254). Despite a paucity of the literature, numerous publications dealt with methodologic quality assessment of observational studies (33,138). West et al (138) in the AHRQ evidence report of technology assessment provided pertinent evidence to rating the quality of individual pertinent guidance to rating the quality of individual articles including observational studies. They assessed 19 systems relating to observational studies or investigations and developed 5 key domains to arrive at a set of high performance scales of checklists pertaining to observational studies which included comparability of subjects, exposure or intervention, outcome measurement, statistical analysis, and funding or sponsorship.

Sanderson et al (255) in a systematic review of tools for assessing quality and susceptibility to bias in epidemiology identified a number of useful assessment tools. They concluded that tools should be rigorously developed, evidence-based, valid, reliable, and easy to use.

5.4.1 Validity

The validity of a study is the extent to which its design and conduct are likely to prevent systematic errors or bias. An important issue that should not be confused with validity is precision. Precision is the measure of the likelihood of chance effect leading to random errors. This is reflected in CI around the estimate of effect from each study and the weight given to the results of each study when an overall estimate effect or weighted average is derived; however, more precise results are given more weight.

5.4.2 Assessment of Bias

The quality of any study within the hierarchy of evidence depends on the confidence that the trial design, conduct, and analysis has minimized or avoided biases in its treatment comparisons (131).

It has been described that well documented and performed prospective observational studies are less biased and are therefore considered higher-quality evidence than retrospective studies. This is because the predictor variable is measured before the outcome, thus establishing a time sequence of events and preventing predictor measurements from being influenced by knowledge of the outcome (256). Consequently, retrospective studies are considered as more prone to overestimating the treatment effect due to confounding and selection bias (257). In fact, observational studies may represent the highest form of evidence in some cases. In a recent systematic review that examined return to function at limb salvage or early amputation for the treatment of severe lower limb injury, the highest available evidence consisted of small prospective, cohort, and case-control studies (258). In addition, the best available evidence of the effect on mortality of delay in surgical treatment for hip fracture comes from prospective observational studies (259). In a systematic review of quality of reporting confounding bias in observational intervention studies Groenwold et al (124) concluded that the quality of reporting of confounding in articles on observational medical intervention studies was poor. They analyzed 174 articles and the potential for confounding bias was reported in the majority of studies (98%). Details on the selection and inclusion of observed confounders were reported in 10% and 51% respectively. Further, the potential for unobserved confounding was reported in 60% and 9% commented on the potential effect of such remaining confounding. Thus, they provided a mediocre score for the quality of reporting of confounding of a median score of 4 points with an interquartile range of 3 to 5.

Lu (132) reviewed study designs, challenges, and strategies to reduce confounding in observational studies. Table 3 illustrates major challenges of observational studies and strategies to reduce confounding as described by Lu and Rochon et al (132,260).

5.4.3 Confounding

Selection bias can result in confounding. A factor can confound an association only if it differs between the intervention and comparison groups (132). For a variable to confound an association it must be associated with both the intervention and outcome and its relation to the outcome should be independent of its association with the intervention. Table 3 illustrates the definition and strategies to reduce confounding. Confounding occurs when the difference in baseline characteristics between the study groups result in difference in the outcome between the groups apart from those related to the intervention and their investigation (261). Confounding can cause over- or under-estimation of the true relationship and may even change the direction of the observed effect.

Table 3. *Major challenges of observational studies.*

Selection bias*: a systematic error in creating intervention groups, causing them to differ with respect to prognosis. The groups differ in measured or unmeasured baseline characteristics because of the way in which participants were selected for the study or assigned to their study groups.
Confounding*: a situation in which the estimated intervention effect is biased because of some difference between the comparison groups apart from the planned interventions such as baseline characteristics, prognostic factors, or concomitant interventions. For a factor to be a confounder, it must differ between the comparison groups and predict the outcome of interest.
Strategies to reduce confounding Design phase Restriction: inclusion to the study is restricted to a certain category of a confounder (e.g., male). Matching of controls to cases to enhance equal representation of subjects with certain confounders among study groups. Analytical phase Stratification: the sample is divided into subgroups or strata on the basis of characteristics that are potentially confounding the analysis (e.g., age). Statistical adjustments Regression: estimates the association of each independent variable with the dependent variable (the outcome) after adjusting for the effects of other variables. Propensity score: a score that is the conditional probability of exposure to an intervention given a set of observed variables that may influence the likelihood of exposure. Instrumental variable: a pseudo-randomization method that divides patients according to levels of a covariate that is associated with the exposure but not associated with the outcome.

*Definitions by the CONSORT statement from Rochon et al. Reader's guide to critical appraisal of cohort studies: 1. Role and design. BMJ 2005; 330: 895-897 (260).
Source: Lu CY. Observational studies: A review of study designs, challenges and strategies to reduce confounding. *Int J Clin Pract* 2009; 63:691-697 (132).

5.5 Statistical Methodology

Observational studies can provide an important complementary source of information, provided that the data are analyzed and interpreted in the context of the confounding bias to which they are prone (133). The concepts of statistical methodology in observational studies include the relationship between a study sample and the target population and the 2 primary forms of statistical analysis: estimation and hypothesis testing. The concept of bias, confounding in particular, is considered as an obstacle to drawing valid conclusions from an observational study.

5.5.1 Populations and Distributions

The analysis of any clinical study is based on the principle of taking a random or representative sample of subjects in order to draw an inference about a larger population of similar individuals called the target population (133). However, going from a population to a sample leads to some degree of uncertainty or margin of error because of the need to rely on the use of estimation without knowledge of the entire population. To quantify this uncertainty, researchers rely on mathematically defined probability distributions such as normal distribution or continuous data and binomial distribution for categorical data. These distributions are based on parameters such as mean and standard deviation. If the assumption is made that the observed data are a sample from a population with a distribution that has a known theoretical form, then it is reasonable to use parameters of the distribution (those observed) to calculate probabilities of different values occurring. This parametric approach to statistics is wide ranging and ubiquitous in medical research. Consequently, unrealistic assumptions may not generate valid results. Thus, when data deviate from a so-called "normal" pattern, non-parametric or distribution-free methods should be used (133).

5.5.2 Estimation and Hypothesis Testing

Statistical analyses are of 2 general types: estimation and hypothesis testing. A primary objective of any types of observational studies is to provide some numerical value that expresses the probability or average of a measured outcome. This is expressed as a proportion or mean or the relative effect associated with a specific treatment or prognostic factor, which is also expressed as a relative risk or odds ratio.

In contrast, estimation typically involves the calculation of a point estimate of disease or outcome prevalence (typically expressed as a probability, rate, or mean) or effect (typically expressed as an odds ratio, relative risk, or risk difference). Estimation gives the quantity, typically in the form of a confidence interval (CI), which informs how large an error might be made with an estimated effect. Thus, interpretation of a 95% CI would include the range of values that contains the true population mean with a probability of 0.95%.

5.5.3 Analytical Techniques

Analytical methods for observational studies vary widely and are chosen according to the type of study that is being performed (133). Most case series require very basic descriptive statistics, such as probabilities or simple averages. Therapeutic and prognostic studies strive to give unconfounded estimates of association and therefore incorporate more elaborate techniques, each with relative strengths and weaknesses. Because therapeutic studies can be thought of as a special case of the prognostic study in which we are only interested in the effect associated with one risk factor (such as specific treatment), the analytical methods used to control confounding are similar and will be presented together.

In analysis of case series, the first issue relates to the target population, which must be definable and the study sample must be representative (133). Next, the intervention must be reproducible so that an interventionalist with adequate training can expect similar results if the procedure is faithfully replicated (133). Third, the outcomes that are measured should be clinically important, and, finally, follow-up should be as complete as possible to limit loss of precision and to avoid selection bias (133). When these criteria are met, a simple descriptive statistic such as risk (number of new cases per number at risk), rate (number of new events per unit of time), or mean (numerical average), along with CIs (generated from a statistical model of a probability distribution such as the binomial for risk data or Poisson for rates) can set an important benchmark for providers and be very helpful in providing information with regard to patient expectations (133).

A classic example of a case series, which is used as a gold standard, involves a well-defined series of 940 operatively treated displaced acetabular fractures and a follow-up period of more than 33 years (262). This report showed that of the 567 hips that were operated on within 21 days, 73.7% were assessed as perfect reductions. Between 3 weeks and 4 months after injury, the probability of a perfect reduction among 150 hips decreased to 64.7%. If the data is projected to 95% CIs, the perfect reduction would have been 70% to 77.3% prior to 3 weeks, and 56.5% to 72.3% between 3 weeks and 4 months after injury. Even though, as many as 18% of patients were either lost to follow-up or had incomplete data, the results have been reproduced in case series reported by others (263-265).

In the analysis of therapeutic and prognostic studies, to avoid confounding and bias, matching, stratification, and multivariable regression can be used. Matching is a strategy whereby confounders are identified and subjects in the treatment group are matched on the basis of these factors so that, in the end, the treatment groups are "the same" with regard to these factors. Matching can either be done on a one-to-one basis or on the basis of frequencies and subjects can be matched with respective single confounder or multiple confounders. Matching is used in both prospective and retrospective observational designs, including case control studies. For example, one study examined the impact of small incisions (< 5 cm) on a variety of outcomes including blood loss, operative time, and postoperative complications, in patients undergoing primary total hip arthroplasty (266). In this study, to ensure that the group of patients who received a small incision was as homogenous as possible with the comparator group of patients who received a standard-size incision, the authors used a matched-pair cohort design by matching 60 patients in each group on a variety of potentially confounding factors, including age, sex, body mass index, diagnosis, prosthesis, type of fixation, anesthesia, pre-anesthetic status, surgical approach, and positioning. The results showed no significant differences in outcome between the 2 techniques.

While matching is an effective way of balancing multiple confounders, it is also associated with several important limitations, including difficulty of finding exact matches between the 2 groups of patients and matching eliminates substantial numbers of subjects due to an inability to match all subjects, resulting in a decrease sample size and power (133).

Stratification is also related to matching and provides another means by which to control confounding. Potentially confounding variables are identified and the cohort is grouped by levels of this factor. The analysis is then performed on each subgroup within which the factor remains constant, thereby removing

the confounding potential of the factor. While stratification allows for control over a confounding factor, it also facilitates investigation into whether the effect of interest is constant across levels of the factor by which stratification is undertaken. Stratification is a useful strategy when there are only 1 or 2 risk factors or confounders, but it quickly becomes unmanageable and difficult to interpret when there are multiple confounders with multiple levels each.

The use of multivariate regression for the adjustment of multiple confounding factors is one of the most commonly used analytical techniques in therapeutic and prognostic studies. Regression analysis is based on modeling the mathematical relationships between 2 or more variables that give an approximate description of the observed data.

Propensity score analysis (267) is an approach to controlling for confounding through the generation of a score that "summarizes" the confounding by multiple variables. This form of analysis is a 2-stage approach in which, first, rather than modeling the outcome as a function of multiple risk factors, the probability of being treated is modeled, taking into account any possible confounding variables. This probability, usually generated by a logistic regression model, is the propensity score, and ranges from 0 to 1. Once the propensity score is generated for each subject, it can be used to match them (usually within some narrow range) or perform stratified analysis on levels of the propensity score or it can be inserted into multivariable regression, along with the treatment variable for use in estimating the outcome.

5.5.3.1 Interpretation in Reporting of Results

The reporting or interpretation of results from observational studies must be tempered with the limitations implicit both in the data and in the methods applied to the analysis of those data (133). Matching and stratification provide a means to limit confounding by another factor by holding its level constant in the analysis. Conventional multivariable adjustment offers the power to adjust for multiple confounders at the same time, advancing the pursuit of potential causal relationships. Still, multiple other criteria are required to establish causation (268). Further, it has been stated that multivariable adjustments cannot give causation unless factors such as appropriate temporal ordering of predictors and outcome are ensured and there are no unaccounted-for confounders missing from the analysis (133).

Other important limitations to the validity of observational and also randomized studies include missing data and loss to follow-up or censoring. Missing data and censoring are a form of selection bias in that those with complete data or follow-up may differ systematically in their association with outcome from those without complete data or follow-up. In the most benign sense, data missing at random should only lessen the precision or power of a study (133). However, this may also result in substantial bias of estimates if missing data is large. Numerous methods have been described to account for missing data, the most robust of which is multiple imputations (269). In dealing with the problem of patients lost to follow-up, sensitivity analysis (assigning all of those with incomplete follow-up to one or the other outcome) can at least put boundaries around the range of effect that may have been witnessed had complete follow-up been achieved.

5.6 Data Collection

Data collection is a bridge between what has been reported by primary investigators and what is ultimately reported by the authors of the systematic review. The data collection must be directly linked to the formulated review question and planned assessment of included studies. Data collection also provides a format for the historical record of the multitude of decisions and changes to decisions that occur throughout the review process. Finally, the data collection format is the data repository from which the analysis will emerge. Key components of data collection form should include essential information and also methodologic quality assessment criteria for observational studies and systematic reviews of observational studies. The essential data collection elements for systematic reviews are illustrated in Tables 1 and 2.

5.7 Summarizing and Synthesizing Relevant Study Results

The primary goals of a systematic review are to summarize the findings of the best studies available and the evidence (270,271). Thus, systematic reviews should provide a written summary of each of the relevant studies, often as a table of summaries. If a quantitative synthesis of results is described, the statistical method of meta-analysis is employed, and a summary result is produced, but this is not always necessary or appropriate. Larger studies that provide more precise estimates of treatment effects are routinely given

more weight in the meta-analysis calculation. One of the most reliable forms of systematic review involves collaborating researchers pooling individual patient data from different studies. While not common, this method has been used in a number of studies. Very few studies have been produced in interventional pain management with meta-analysis.

Data synthesis in systematic reviews or meta-analyses can be achieved through a descriptive or non-quantitative synthesis, complemented by the use of formal statistical techniques (271). Thus, an integral part of the data synthesis is to investigate whether the effects are consistent across the included studies and if not, to investigate the reasons for the differences, in addition to generating a summary of the effects of interventions.

5.7.1 Descriptive or Non-Quantitative Synthesis

The objective of a descriptive or non-quantitative review is to correlate and present the extracted data in a manner such that information of the characteristics such as population, interventions, outcomes, study quality, and results of the studies included in the review are summarized in a meaningful way. When this evidence is presented in a tabular format, it allows the readers to look at the evidence, its methodological rigor, and the differences between the studies. Thus, the descriptive overview is an essential part of the data on which an understanding of the data, planning, and quantitative data synthesis, and preventing errors in its interpretation.

The process of carrying out the descriptive part of the data synthesis should be explicit and rigorous (242,272). Effectiveness of health care interventions is dependent on the information that relates to a large number of factors including the recipients of intervention, who delivers it, and finally, how and in what context it is delivered. The key elements of the data synthesis involving a descriptor approach may include multiple characteristics including population, intervention, settings where the technology was applied, environmental, societal, and cultural factors that may influence compliance, nature of the outcome measures used, their relative importance and robustness, the validity of the evidence, the sample sizes, and results of the studies included in the review.

Data synthesis involves computation of an average effect where the results of each study are weighed according to some measures of the study's importance. Each study's weight usually relates to its sample size, the quality of the study, and the resulting precision of the estimate of effect.

5.7.2 Quantitative Synthesis

Meta-analysis is not always possible when necessary data to perform meta-analysis cannot be obtained and it may not be appropriate when the data are sparse or when the studies are too heterogenous to be sensibly combined. The meta-analysis is performed generally to increase the power, to improve precision, and to answer the questions not posed by the individual studies, and to settle controversies arising from conflicting studies or to generate new hypothesis (270). Once it is established that a meta-analysis is possible and appropriate, reviewers should make choices about comparisons to be made, outcome measures to be used in the synthesis, and effect measures utilized in quantifying the intervention.

5.7.3 Interpretation of the Review Findings

The final step in the systematic review is the interpretation of the results based on the question that was formulated in explaining how well results have answered it. Even though, the results of a systematic review should stand on their own, many are faced with the decision to the authors' conclusions for help in interpreting the results (273,274). This leads to different interpretation of the results of the same studies in systematic reviews, which is fairly common in all disciplines of medicine.

5.7.4 Strength of Evidence

Evaluation of strength of evidence is one of the important functions of a systematic review or meta-analysis. This also incorporates the addressing of many important methodological limitations of the included studies and the methods used in the review that might affect the practical decisions about health care or future research. The integral part of the strength of evidence of an intervention is based on the quality of the studies included and their assessment. Authors of systematic reviews should look at if they have followed the reporting guidelines of observational studies as described by STROBE statement (111).

5.7.5 Level of Evidence

West et al (138) published systems to rate the quality of individual articles, as well as systems for grading the strength of a body of evidence. The National Health and Medical Research Council (NHMRC)

of Australia considered scientific data to be at the core of evidence-based approaches to clinical or public health issues, emphasizing that evidence needs to be carefully gathered and collated from a systematic literature review of each particular issue in question (275). Consequently, grading the quality of individual studies and rating of the strength of the body of evidence are both crucial elements. Strength of evidence is based on various factors including the size, credibility, and robustness of the combined studies of a given topic. However, systems for grading the strength of a body of evidence are less uniform and consistent than those rating the study quality (138). Selecting the evidence to be used in grading systems depends on the reason for measuring evidence strength, the types of studies that are being summarized, and the structure of the review panel. Domains for rating the overall strength of a body of evidence are listed in Table 4 (275). Table 5 shows panel ratings of available evidence supporting guideline statements developed by AHRQ (formerly AHCPR) (276), that is now outdated. The American Society of Interventional Pain Physicians (ASIPP) guidelines utilized a graded strength of evidence over the years as illustrated in Table 6, which was changed to quality of evidence developed by USPSTF (Table 7) (277).

5.7.6 Grading Recommendations

Guyatt et al (278) developed grading system based on the philosophy that guidelines panels should make recommendations to administer or not to administer an intervention on the basis of a trade-off between benefits on one hand and the risks, burdens, and potential costs on the other. In contrast to many grading recommendations, they provided recommendations which included 2 levels: strong and weak as illustrated in Table 8. Guideline panels must consider a number of factors in grading recommendations including 1) methodologic quality of evidence reporting estimates of likely benefit and likely risk, inconvenience, and costs, 2) importance of the outcome, 3) magnitude of the treatment effect, 4) estimate of treatment effect, 5) risks associated with therapy, 6) burden of therapy, 7) risk of target event, 8) costs, and finally 9) circumstances, patients' or societal values. In the section on methodologic quality of supporting evidence strong

Table 4. *Criteria for rating the overall strength of a body of evidence.*

Domain	Definition
Quality	• The quality of all relevant studies for a given topic, where "quality" is defined as the extent to which a study's design, conduct, and analysis has minimized selection, measurement, and confounding biases
Quantity	• The magnitude of treatment effect
	• The number of studies that have evaluated the given topic
	• The overall sample size across all included studies
Consistency	• For any given topic, the extent to which similar findings are reported from work using similar and different study designs

Adapted from How to use the evidence: Assessment and application of scientific evidence. National Health and Medical Research Council, Canberra, Commonwealth of Australia, 2000, pp 1-84 (275).

Table 5. *Panel ratings of available evidence supporting guideline statements.*

A	Strong research-based evidence (multiple relevant and high-quality scientific studies).
B	Moderate research-based evidence (one relevant high-quality scientific study or multiple adequate scientific studies*).
C	Limited research-based evidence (at least one adequate scientific study* in patients with low back pain).
D	Panel interpretation of information that did not meet inclusion criteria as research-based evidence.

* Met minimal formal criteria for scientific methodology and relevance to population and specific method addressed in guideline statement.

Note: These criteria were derived from Bigos SJ et al. Acute low back problems in adults. Clinical Practice Guideline No.14, AHCPR Publication No. 95-0642. Rockville, Maryland. U.S.A., Agency for Health Care Policy and Research, Public Health Service, U.S., Department of Health and Human Services, December, pp. 1-60, 1994 (276). AHCPR was extinguished by Congress in 1995, changing AHCPR to AHRQ. Acute Low Back Pain Guidelines (276) provide a disclaimer "not for patient care."

Table 6. *Designation of levels of evidence as used in evidence-based guidelines by the American Society of Interventional Pain Physicians.*

Level I	Conclusive: Research-based evidence with multiple relevant and high-quality scientific studies or consistent reviews of meta-analyses
Level II	Strong: Research-based evidence from at least one properly designed randomized, controlled trial; or research-based evidence from multiple properly designed studies of smaller size; or multiple low quality trials
Level III	Moderate: a) Evidence obtained from well-designed pseudorandomized controlled trials (alternate allocation or some other method); b) evidence obtained from comparative studies with concurrent controls and allocation not randomized (cohort studies, case-controlled studies, or interrupted time series with a control group); c) evidence obtained from comparative studies with historical control, 2 or more single-arm studies, or interrupted time series without a parallel control group
Level IV	Limited: Evidence from well-designed non-experimental studies from more than one center or research group; or conflicting evidence with inconsistent findings in multiple trials
Level V	Indeterminate: Opinions of respected authorities, based on clinical evidence, descriptive studies, or reports of expert committees

Table 7. *Quality of evidence developed by USPSTF.*

I	Evidence obtained from at least one properly randomized controlled trial
II-1	Evidence obtained from well-designed controlled trials without randomization
II-2	Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group
II-3	Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidenc
III	Opinions of respected authorities, based on clinical experience descriptive studies and case reports or reports of expert committees

Adapted from the U.S. Preventive Services Task Force (USPSTF) (277).

recommendations 1B with moderate quality evidence derived from exceptionally strong evidence from observational studies. Obviously, observational studies or case series encompass the evidence in other categories including 1C strong recommendation with low quality evidence and of course, all weak recommendations with high or low quality evidence.

5.7.7 Applicability

Applicability or generalizability of the results of a systematic review is crucial. Decisions about applicability depend on knowledge of particular circumstances in which decisions about health care are being made, however authors of systematic reviews should cautiously approach the issue of applicability and should not assume that their own circumstances, or the circumstances reflected in the included studies, are necessarily the same as those of others. However, authors of systematic reviews may assist with recommendations about applicability by drawing attention to the spectrum of circumstances to which the evidence is likely to be applicable (274).

5.7.8 Limitations

The interpretation may also discuss the trade-offs between benefits and harms, and, less often, costs. The cost effective analysis or economic evaluation are important for policy decisions.

5.8 Updating Reviews

Updating and improving access to the reviews is crucial in modern medicine. The Cochrane Collaboration requires that reviews consider updating each synthesis every 2 years in some cases. ASIPP also requires updating of these reviews every 2 to 4 years. The emergence of important new evidence from a fresh study can mean that updating is needed even sooner. The requirements for updates illustrated that a qualitative or quantitative signal for updating occurred for 57% of reviews with a median duration of survival free of a signal for updating of 5.5 years (279). However, approximately 23% of reviews required a review within 2 years and 15% within one year. It was also shown that in 7% of the reviews they required revision at the time of the publication. Longevity and

Table 8. Grading recommendations.

Grade of Recommendation/ Description	Benefit vs Risk and Burdens	Methodological Quality of Supporting Evidence	Implications
1A/strong recommendation, high-quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	RCTs without important limitations or overwhelming evidence from observational studies	Strong recommendation, can apply to most patients in most circumstances without reservation
1B/strong recommendation, moderate quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from observational studies	Strong recommendation, can apply to most patients in most circumstances without reservation
1C/strong recommendation, low-quality or very low-quality evidence	Benefits clearly outweigh risk and burdens, or vice versa	Observational studies or case series	Strong recommendation but may change when higher quality evidence becomes available
2A/weak recommendation, high-quality evidence	Benefits closely balanced with risks and burden	RCTs without important limitations or overwhelming evidence from observational studies	Weak recommendation, best action may differ depending on circumstances or patients' or societal values
2B/weak recommendation, moderate-quality evidence	Benefits closely balanced with risks and burden	RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from observational studies	Weak recommendation, best action may differ depending on circumstances or patients' or societal values
2C/weak recommendation, low-quality or very low-quality evidence	Uncertainty in the estimates of benefits, risks, and burden; benefits, risk, and burden may be closely balanced	Observational studies or case series	Very weak recommendations; other alternatives may be equally reasonable

Adapted from Guyatt G et al. Grading strength of recommendations and quality of evidence in clinical guidelines. Report from an American College of Chest Physicians task force. *Chest* 2006; 129:174-181 (278).

survival for interventional pain management topics may be shorter than other subjects, however, it has been reported that the improvement in quality with subsequent reporting was seen only in certain individual items with no overall improvement seen with updating and methodologic quality (280).

6.0 Reporting of Systematic Reviews

The MOOSE statement (120), a proposal for reporting of Meta-analysis of Observational Studies in Epidemiology describes specifications of reporting including background, search strategy, methods, results, discussion, and conclusion. Use of the checklist is expected to improve the usefulness of meta-analysis for authors, reviewers, editors, readers, and decision-makers. The checklist is illustrated in Table 9. However this document has not provided a flow diagram. The QUOROM flow diagram (Fig. 1) shows progress through the stages of systematic reviews.

6.1 Title

The title should identify the report as a systematic review or meta-analysis of observational studies.

6.2 Abstract

The structured abstract must provide a series of headings pertaining to the design, conduct, and analysis of a trial with standardized information appearing under each heading. It has been shown that structured abstracts are of higher quality than the more traditional descriptive abstracts (158,281-284) and they also allow readers to find information more easily (283,285). These headings should include or incorporate objectives showing the clinical question explicitly; data sources showing the databases and other information sources; review methods showing the selection criteria; methods of validity assessment, data extraction, and study char-

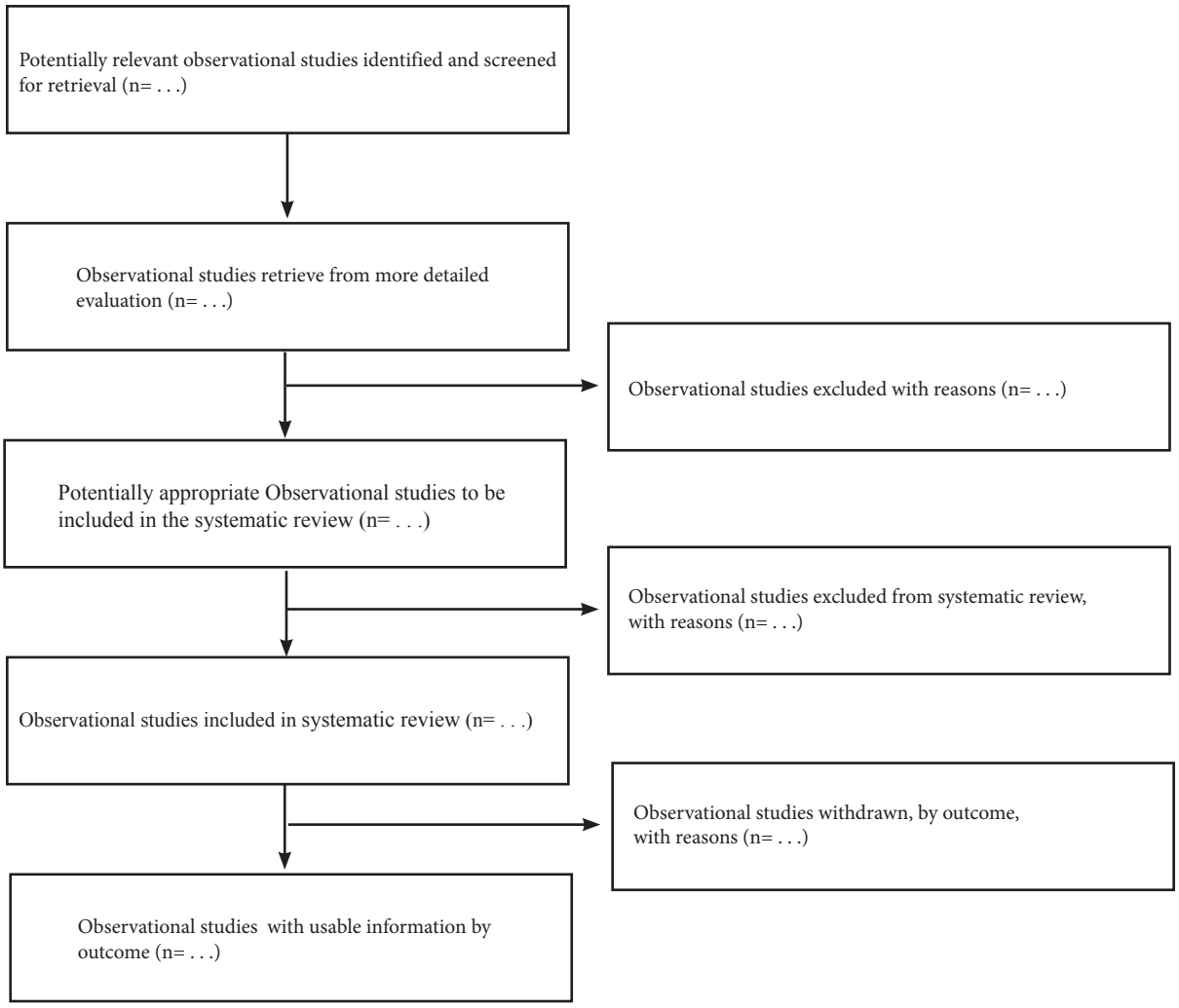

Fig. 1. Progress through the stages of asystematic review of observational studies.

acteristics; quantitative data synthesis in sufficient detail to permit replication; results; characteristics of the observational studies included and excluded; quantitative and qualitative findings, and subgroup analysis available; and the conclusion with the main results (32).

6.3 Introduction or Background

The introduction includes the scientific background and an explanation of rationale. Typically the introduction includes reporting of the background with a free-flowing text, without a structured format, in which the authors explain the scientific background of the clinical problem, biological rationale for the intervention, and rationale for the systematic review. In addition, the introduction or background should provide an appropriate explanation for how the systematic review might work and the research involving people should be based on a thorough knowledge of the scientific literature (286,287). The authors should

Table 9. *A proposed reporting checklist for authors, editors, and reviewers of meta-analyses of observational studies.*

Reporting of background should include
Problem definition
Hypothesis statement
Description of study outcome(s)
Type of exposure or intervention used
Type of study designs used
Study population
Reporting of search strategy should include
Qualifications of searchers (eg, librarians and investigators)
Search strategy, including time period included in the synthesis and keywords
Effort to include all available studies, including contact with authors
Databases and registries searched
Search software used, name and version, including special features used (eg, explosion)
Use of hand searching (eg, reference lists of obtained articles)
List of citations located and those excluded, including justification
Method of addressing articles published in languages other than English
Method of handling abstracts and unpublished studies
Description of any contact with authors
Reporting of methods should include
Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested
Rationale for the selection and coding of data (eg, sound clinical principles or convenience)
Documentation of how data were classified and coded (eg, multiple raters, blinding, and interrater reliability)
Assessment of confounding (eg, comparability of cases and controls in studies where appropriate)
Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results
Assessment of heterogeneity
Description of statistical methods (eg, complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated
Provision of appropriate tables and graphics
Reporting of results should include
Graphic summarizing individual study estimates and overall estimate
Table giving descriptive information for each study included
Results of sensitivity testing (e.g., subgroup analysis)
Indication of statistical uncertainty of findings
Reporting of discussion should include
Quantitative assessment of bias (eg, publication bias)
Justification for exclusion (eg, exclusion of non–English-language citations)
Assessment of quality of included studies
Reporting of conclusions should include
Consideration of alternative explanations for observed results
Generalization of the conclusions (i.e., appropriate for the data presented and within the domain of the literature review)
Guidelines for future research
Disclosure of funding source

Source: Stroup DF et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; 283:2008-2012 (120).

also explicitly explain if the systematic review is limited to the review of the individual articles or if the meta-analysis is planned.

As shown in Table 9, the MOOSE recommends a format which includes problem definition, hypothesis statement, description of study outcome(s), type of exposure or intervention used, type of study designs used, and study population.

6.4 Methods

Methods include searching, selection, validity assessment, data extraction, study characteristics, and data synthesis.

6.4.1 Searching

Reporting of the search strategy should include qualification of the searchers, specification of data-

bases used, search strategy and index terms, use of many specific features (i.e., explosion), search software used, use of hand searching in contact with authors, use of material and languages other than English, use of unpublished material, and exclusion criteria used. Published research shows that the use of electronic databases may find only half of all the relevant studies and contacting authors may be useful (288), although this result may not be true for all topic areas (289). Search strategies and finding the relevant studies are described in section 5 of how to conduct systematic reviews in meta-analysis.

6.4.2 Selection

The authors should clearly describe the inclusion and exclusion criteria with the definition of the population, intervention, principle outcomes, and study design (290). Precise details of the population, setting and locations, interventions, outcomes, and objectives must be clearly described. MOOSE as shown in Table 9, illustrates that reporting of methods should include description of relevance, appropriateness of studies assembled for assessing the hypothesis to be tested. Further, rationale for the selection and coding of data, documentation of how data were classified and coded is essential in the reporting.

6.4.3 Validity Assessment

The multiple criterion process used to assess the validity must be described. These may include appropriate allocation, they may include questions such as those utilized in AHRQ quality assessment criteria for observational studies (33,138) such as study question, study population, comparability of subjects, exposure or intervention, outcomes measures, statistical analysis, and the results. The MOOSE statement describes that reporting of methods should include assessment of confounding (i.e., comparability of case and control studies where appropriate, assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors or study results).

6.4.4 Data Extraction

Data extraction should be described clearly whether it was completed independently or in duplicative (32).

6.4.5 Study Characteristics

Under this section, the type of study design, participants' characteristics, details of intervention, outcome definitions, and the assessment of clinical heterogeneity must be described. Along with assessment of the study quality, assessment of heterogeneity must be reported (120).

6.4.6 Data Synthesis

The principle measure of effect, relative risk, method of combining results, statistical testing, and confidence intervals, handling of missing data, how statistical heterogeneity was assessed, a rationale for any prior sensitivity and subgroup analysis, and any assessment of publication bias should be clearly documented and reported. Methods for quantitative and qualitative data synthesis must be reported.

6.4.7 Statistical Analysis

Description of statistical methods should be provided with a completed description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or accumulative meta-analysis in sufficient detail to be replicated.

6.5 Results

The results section includes trial flow, study characteristics, and quantitative data synthesis. The MOOSE statement (120) describes that the reporting of results should include graphics summarizing the individual study estimate and overall estimate, tables giving descriptive information to each study included, results of sensitivity testing or subgroup analysis, and indication of statistical uncertainty of the findings.

6.5.1 Study Flow

A study flow figure should be inserted which shows how the literature was searched and inclusion/exclusion criteria were met as shown in Figure 1.

6.5.2 Study Characteristics

Authors should present descriptive data for each study, along with sample size, intervention, dose, duration, and follow-up periods, etc. This data may be presented in the form of a table providing descriptive information for each study included.

6.5.3 Quantitative Data Synthesis

Results should show the principle measures of effect, statistical testing, and confidence intervals; handling of missing data; results of statistical heterogeneity; results of subgroup analysis if performed; and

the results of publication bias if they were assessed. Further, it should be reported on agreement on the selection and validity assessment in the form of simple summary results for each treatment group in each trial for each primary outcome; data needed to calculate effect sizes and confidence intervals; and intention-to-treat analysis with tables of counts, means, and standard deviations or proportions (32). Data may be provided in graphic format summarizing individual study and overall estimates along with indication of statistical uncertainty of findings (120).

6.5.4 Level of Evidence

The level of evidence may be presented based on the conditions and the results, however, this is not required based on MOOSE statement.

6.5.5 Recommendations

Grading of recommendations may be provided which is not a recommendation of the MOOSE statement. Further, cost-effectiveness analysis may also be provided which is not a requirement of the MOOSE statement.

6.5.6 Discussion

The discussion should summarize key findings; discuss clinical inferences based on internal and external validity; interpret the results in light of the totality of available evidence; describe potential biases in the review process such as publication bias; and suggest a future research agenda. MOOSE specifically describes that discussion should include quantitative assessment of bias, (i.e., publication bias, justification for exclusion such as, exclusion of non-English language citation), and assessment of quality of included studies. It is essential to utilize the recommendations of QUOROM (104) and MOOSE (120) specifically if systematic reviews or meta-analysis combined both randomized and observational studies.

Significant guidance has been provided with journals encouraging a structure to the authors' discussion of the results (291-293). The Annals of Internal Medicine (292) recommends that authors structure the discussion section as follows:

1) A brief synopsis of the key findings
2) Consideration of possible mechanisms and explanation
3) Comparison with relevant findings from other published studies
4) Limitations of the present study and methods used to minimize and compensate for those limitations
5) A brief section that summarizes the clinical and research implications of the work, as appropriate

However, it is of particular importance to discuss the weakness and the limitations of the study as described by multiple authors and required by some journals including Pain Physician (158,281,294-296). It is also essential to describe the differences between statistical significance and clinical importance. Further, the major feature of descriptions in recent years with personal biases begs for detailed disclosure of the authors and organizations conducting systematic reviews and guidelines.

It is essential to include a comprehensive conclusion, the MOOSE statement (120) includes that the conclusion should include consideration of an alternative explanation for observed results, generalization of the conclusion (i.e., appropriate for the data presented and within the domain of the literature review), guidelines for future research, and disclosure of funding source.

7.0 Discussion

Assessment of health care interventions can be misleading unless investigators ensure unbiased comparisons in all aspects from conducting the studies to reporting of individual studies, systematic reviews and meta-analyses, evidence synthesis, and guideline preparation. In interventional pain management settings, results of clinical trials, both randomized and observational, along with multiple systematic reviews, have been ruled ineffective based on flawed methodology in the evidence synthesis. Poorly executed systematic reviews tend to exaggerate treatment effects both negative and positive with important biases. Thus, it is essential to produce high-quality research, which the result of consistently eliminates bias and shows significant effect size.

The design, implementation, and reporting of a systematic review or meta-analysis requires methodologic, as well as clinical expertise including meticulous effort, high index or suspicion for unanticipated difficulties in bias, potentially unnoticed problems, and methodological deficiencies; and skills to report the findings appropriately with close attention to minimize bias. It is crucial that sound methodology encompassing adequate reporting and conduct of the review which rests on the footing of sound science. This will limit the exposure of readers to specula-

tion. Interventional pain specialists must understand the differences between multiple types of reviews – systematic, meta-analysis, narrative, focused, health technology assessments, and types of methodologic quality assessment, and levels of evidence and grading of recommendations.

Acknowledgments

The authors wish to thank Vidyasagar Pampati, MSc, statistician; Sekar Edem for assistance in search of literature; and Tonie M. Hatton and Diane E. Neihoff, transcriptionists; for their assistance in the preparation of this manuscript. We would also like to thank the editorial board of Pain Physician for their constructive comments and guidance in improving this manuscript.

References

1. Buchbinder R, Osborne RH, Ebeling PR, Wark JD, Mitchell P, Wriedt C, Graves S, Staples MP, Murphy B. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *N Engl J Med* 2009; 361:557-568.
2. Kallmes DF, Comstock BA, Heagerty PJ, Turner JA, Wilson DJ, Diamond TH, Edwards R, Gray LA, Stout L, Owen S, Hollingworth W, Ghdoke B, Annesley-Williams DJ, Ralston SH, Jarvik JG. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *N Engl J Med* 2009; 361:569-579.
3. Weinstein JN. Balancing science and informed choice in decisions about vertebroplasty. *N Engl J Med* 2009; 361:619-621.
4. Hartzband P, Groopman J. Keeping the patient in the equation--humanism and health care reform. *N Engl J Med* 2009; 361:554-555.
5. Murphy ML, Hultgren HN, Detre K, Thomsen J, Takaro T. Treatment of chronic stable angina. A preliminary report of survival data of the randomized Veterans Administration cooperative study. *N Engl J Med* 1977; 297:621-627.
6. Eleven-year survival in the Veterans Administration randomized trial of coronary bypass surgery for stable angina. The Veterans Administration Coronary Artery Bypass Surgery Cooperative Study Group. *N Engl J Med* 1984; 311:1333-1339.
7. Eighteen-year follow-up in the Veterans Affairs Cooperative Study of Coronary Artery Bypass Surgery for stable angina. The VA Coronary Artery Bypass Surgery Cooperative Study Group. *Circulation* 1992; 86:121-130.
8. Peduzzi P, Kamina A, Detre K. Twenty-two-year follow-up in the VA Cooperative Study of Coronary Artery Bypass Surgery for Stable Angina. *Am J Cardiol* 1998; 81:1393-1399.
9. Carette S, Marcoux S, Truchon R, Grondin C, Gagnon J, Allard Y, Latulippe M. A controlled trial of corticosteroid injections into facet joints for chronic low back pain. *N Engl J Med* 1991; 325:1002-1007.
10. Carette S, Leclaire R, Marcoux S, Morin F, Blaise GA, St-Pierre A, Truchon R, Parent F, Levesque J, Bergeron V, Montminy P, Blanchette C. Epidural corticosteroid injections for sciatica due to herniated nucleus pulposus. *N Engl J Med* 1997; 336:1634-1640.
11. Deyo RA, Mirza SK, Turner JA, Martin BI. Overtreating chronic back pain: Time to back off? *J Am Board Fam Med* 2009; 22:62-68.
12. Weinstein JN, Lurie JD, Tosteson TD, Skinner JS, Hanscom B, Tosteson AN, Herkowitz H, Fischgrund J, Cammisa FP, Albert T, Deyo RA. Surgical vs nonoperative treatment for lumbar disk herniation. The Spine Patient Outcomes Research Trial (SPORT): Observational cohort. *JAMA* 2006; 296:2451-2459.
13. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, Abdu WA, Hilibrand AS, Boden SD, Deyo RA. Surgical vs nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT): A randomized trial. *JAMA* 2006; 296:2441-2450.
14. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Blood E, Hanscom B, Herkowitz H, Cammisa F, Albert T, Boden SD, Hilibrand A, Goldberg H, Berven S, An H; SPORT Investigators. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *NEJM* 2008; 358:794-810.
15. Weinstein JN, Lurie JD, Tosteson TD, Zhao W, Blood EA, Tosteson AN, Birkmeyer N, Herkowitz H, Longley M, Lenke L, Emery S, Hu SS. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. Four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am* 2009; 91:1295-1304.
16. Manchikanti L, Singh V, Pampati V, Smith HS, Hirsch JA. Analysis of growth of interventional techniques in managing chronic pain in the medicare population: A 10-year evaluation from 1997 to 2006. *Pain Physician* 2009; 12:9-34.
17. Manchikanti L. Health care reform in the United States: Radical surgery needed now more than ever. *Pain Physician* 2008; 11:13-42.
18. Manchikanti L, Hirsch JA. Issues in health care: Interventional pain management at the crossroads. Health Policy Update. *Pain Physician* 2007; 10:261-284.
19. Manchikanti L, Giordano J. Physician payment 2008 for interventionalists: Current state of health care policy. *Pain Physician* 2007; 10:607-626.
20. Ploeg WT, Veldhuizen AG, The B, Sietsma MS. Percutaneous vertebroplasty as a treatment for osteoporotic vertebral compression fractures: A systematic review. *Eur Spine J* 2006; 15:1749-1758.
21. Voormolen MH, Mali WP, Lohle PN, Fransen H, Lampmann LE, van der Graaf Y, Juttmann JR, Jansssens X, Verhaar HJ. Percutaneous vertebroplasty compared with optimal pain medication treatment: Short-term clinical outcome of patients with subacute or chronic painful osteoporotic vertebral compression fractures: The VERTOS Study. *AJNR Am J Neuroradiol* 2007; 28:555-560.
22. Diamond TH, Champion B, Clark WA. Management of acute osteoporotic vertebral fractures: A nonrandomized trial comparing percutaneous vertebroplasty with conservative therapy. *Am J Med* 2003; 114:257-265.
23. Alvarez L, Alcarez M, Pérez-Higueras A, Granizo JJ, de Miguel I, Rossi RE, Quiñones D. Percutaneous vertebroplasty: Functional improvement in pa-

tients with osteoporotic compression fractures. *Spine (Phila Pa 1976)* 2006; 31:1113-1118.

24. Taylor RS, Taylor R, Fritzell P. Balloon kyphoplasty and vertebroplasty for vertebral compression fractures: A comparative systematic review of efficacy and safety. *Spine (Phila Pa 1976)* 2006; 31:2747-2755.
25. Taylor RS, Fritzell P, Taylor RJ. Balloon kyphoplasty in the management of vertebral compression fractures: An updated systematic review and meta-analysis. *Eur Spine J* 2007; 16:1085-1100.
26. Hulme PA, Krebs J, Ferguson SJ, Berlemann U. Vertebroplasty and kyphoplasty: A systematic review of 69 clinical studies. *Spine (Phila Pa 1976)* 2006; 31:1983-2001.
27. Felder-Puig R, Piso B, Guba B, Gartlehner G. Orthopade. Kyphoplasty and vertebroplasty for the management of osteoporotic vertebral compression fractures: A systematic review. *Orthopade* 2009; 38:606-615.
28. Gill JB, Kuper M, Chin PC, Zhang Y, Schutt R. Comparing pain reduction following kyphoplasty and vertebroplasty for osteoporotic vertebral compression fractures. *Pain Physician* 2007; 10:583-590.
29. Manchikanti L, Boswell MV, Giordano J. Evidence-based interventional pain management: Principles, problems, potential, and applications. *Pain Physician* 2007; 10:329-356.
30. Manchikanti L. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 1: Introduction and general considerations. *Pain Physician* 2008; 11:161-186.
31. Manchikanti L, Hirsch JA, Smith HS. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 2: Randomized controlled trials. *Pain Physician* 2008; 11:717-773.
32. Manchikanti L, Benyamin RM, Helm S, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 3: Systematic reviews and meta-analysis of randomized trials. *Pain Physician* 2009; 12:35-72.
33. Manchikanti L, Singh V, Smith HS, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 4: Observational studies. *Pain Physician* 2009; 12:73-108.
34. Manchikanti L, Derby R, Wolfer LR, Singh V, Datta S, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 5. Diagnostic accuracy studies. *Pain Physician* 2009; 12:517-540.
35. Guyatt G, Drummond R. Part 1. The basics: Using the medical literature. 1A. Introduction: The philosophy of evidence-based medicine. In: Guyatt G, Rennie D (eds). *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. AMA Press, Chicago, 2002, pp 3-12.
36. Petrie A, Bulman JS, Osborn JF. Further statistics in dentistry Part 8: Systematic reviews and meta-analyses. *Br Dent J* 2003; 194:73-78.
37. Crowther MA, Cook DJ. Trials and tribulations of systematic reviews and meta-analyses. *Hematology Am Soc Hematol Educ Program* 2007; 2007:493-497.
38. Lind J. *A Treatise of the Scurvy in Three Parts. Containing an inquiry into the Nature, Causes and Cure of that Disease, together with a Critical and Chronological View of What Has Been Published on the Subject*. A. Millar, London, 1753.
39. Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res* 1976; 5:3-8.
40. Guyatt G, Cairns J, Churchill D, Haynes B, Hirsh J, Irvine J, Levine M, Nishikawa J, Sackett D, Brill-Edwards P, Gerstein H, Gibson J, Jaeschke R, Kerigan A, Neville A, Panju A, Detsky A. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992; 268:2420-2425.
41. Pearson K. Report on certain enteric fever inoculation statistics. *Br Med J* 1904; 3:1243-1246.
42. Pai M, McCulloch M, Gorman JD, Pai N, Enanoria W, Kennedy G, Tharyan P, Colford JM. Systematic reviews and meta-analyses: An illustrated, step-by-step guide. *Natl Med J India* 2004; 17:86-95.
43. Egger M, Smith GD, Altman DG (eds). *Systematic Reviews in Health Care. Meta-Analysis in Context*. BMJ Publishing Group, London, 2001.
44. *Undertaking Systematic Reviews of Research on Effectiveness. CRDs Guidance for Carrying Out or Commissioning Reviews*. CRD Report Number 4 (2nd), CRD Centre for Reviews and Dissemination, University of York, York, UK. March 2001. www.york.ac.uk/inst/crd/report4.htm
45. 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; 48:167-171.
46. Mulrow C, Langhorne P, Grimshaw J. Integrating heterogeneous pieces of evidence in systematic reviews. *Ann Intern Med* 1997; 127:989-995.
47. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal anti-inflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10:31-56.
48. Wright RW, Brand RA, Dunn W, Spindler KP. How to write a systematic review. *Clin Orthop Relat Res* 2007; 455:23-29.
49. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd ed. Churchill Livingstone, Edinburgh, UK, 2000.
50. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316:450-455.
51. Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: An update. *Mt Sinai J Med* 1996; 63:216-224.
52. Moher D, Tricco AC. Issues related to the conduct of systematic reviews: A focus on the nutrition field. *Am J Clin Nutr* 2008; 88:1191-1999.
53. Richardson WS, Wilson MS, Nishikawa J, Hayward RSA. The well-built clinical question: A key to evidence based decisions. *ACP J Club* 1995; A12-13.
54. American College of Occupational and Environmental Medicine Low Back Disorders Chapter. In: *Occupational Medicine Practice Guidelines: Evaluation and Management of Common Health Problems and Functional Recovery of Workers,* Second Edition. American College of Occupational and Environmental Medicine, Elk Grove Village, 2007.
55. American College of Occupational and Environmental Medicine. Chronic Pain Chapter (revised 2008). In: *Occupational Medicine Practice Guidelines: Evaluation and Management of Common Health Problems and Functional Recovery of Workers,* Second Edition. American College of Occupational and Environmental Medicine, Elk Grove Vil-

lage, 2008.

56. Staal JB, de Bie RA, de Vet HC, Hildebrandt J, Nelemans P. Injection therapy for subacute and chronic low back pain: An updated Cochrane review. *Spine (Phila Pa 1976)* 2009; 34:49-59.
57. Manchikanti L, Boswell MV, Singh V, Benyamin RM, Fellows B, Abdi S, Buenaventura RM, Conn A, Datta S, Derby R, Falco FJE, Erhart S, Diwan S, Hayek SM, Helm S, Parr AT, Schultz DM, Smith HS, Wolfer LR, Hirsch JA. Comprehensive evidence-based guidelines for interventional techniques in the management of chronic spinal pain. *Pain Physician* 2009; 12:699-802.
58. Boswell MV, Trescot AM, Datta S, Schultz DM, Hansen HC, Abdi S, Sehgal N, Shah RV, Singh V, Benyamin RM, Patel VB, Buenaventura RM, Colson JD, Cordner HJ, Epter RS, Jasper JF, Dunbar EE, Atluri SL, Bowman RC, Deer TR, Swicegood JR, Staats PS, Smith HS, Burton AW, Kloth DS, Giordano J, Manchikanti L. Interventional techniques: Evidence-based practice guidelines in the management of chronic spinal pain. *Pain Physician* 2007; 10:7-111.
59. Atluri S, Datta S, Falco FJE, Lee M. Systematic review of diagnostic utility and therapeutic effectiveness of thoracic facet joint interventions. *Pain Physician* 2008; 11:611-629.
60. Falco FJE, Erhart S, Wargo BW, Bryce DA, Atluri S, Datta S, Hayek SM. Systematic review of diagnostic utility and therapeutic effectiveness of cervical facet joint interventions. *Pain Physician* 2009; 12:323-344.
61. Datta S, Lee M, Falco FJE, Bryce DA, Hayek SM. Systematic assessment of diagnostic accuracy and therapeutic utility of lumbar facet joint interventions. *Pain Physician* 2009; 12:437-460.
62. Singh V, Manchikanti L, Shah RV, Dunbar EE, Glaser SE. Systematic review of thoracic discography as a diagnostic test for chronic spinal pain. *Pain Physician* 2008; 11:631-642.
63. Wolfer L, Derby R, Lee JE, Lee SH. Systematic review of lumbar provocation discography in asymptomatic subjects with a meta-analysis of false-positive rates. *Pain Physician* 2008; 11:513-538.
64. Manchikanti L, Dunbar EE, Wargo BW, Shah RV, Derby R, Cohen SP. Systematic review of cervical discography as a diagnostic test for chronic spinal pain. *Pain Physician* 2009; 12:305-321.
65. Manchikanti L, Glaser S, Wolfer L, Derby R, Cohen SP. Systematic review of lumbar discography as a diagnostic test for chronic low back pain. *Pain Physician* 2009; 12:541-560.
66. Conn A, Buenaventura R, Datta S, Abdi S, Diwan S. Systematic review of caudal epidural injections in the management of chronic low back pain. *Pain Physician* 2009; 12:109-135.
67. Parr AT, Diwan S, Abdi S. Lumbar interlaminar epidural injections in managing chronic low back and lower extremity pain: A systematic review. *Pain Physician* 2009; 12:163-188.
68. Benyamin RM, Singh V, Parr AT, Conn A, Diwan S, Abdi S. Systematic review of the effectiveness of cervical epidurals in the management of chronic neck pain. *Pain Physician* 2009; 12:137-157.
69. Buenaventura RM, Datta S, Abdi S, Smith HS. Systematic review of therapeutic lumbar transforaminal epidural steroid injections. *Pain Physician* 2009; 12:233-251.
70. Helm S, Hayek S, Benyamin RM, Manchikanti L. Systematic review of the effectiveness of thermal annular procedures in treating discogenic low back pain. *Pain Physician* 2009; 12:207-232.
71. Smith HS, Chopra P, Patel VB, Frey ME, Rastogi R. Systematic review on the role of sedation in diagnostic spinal interventional techniques. *Pain Physician* 2009; 12:195-206.
72. Frey ME, Manchikanti L, Benyamin RM, Schultz DM, Smith HS, Cohen SP. Spinal cord stimulation for patients with failed back surgery syndrome: A systematic review. *Pain Physician* 2009; 12:379-397.
73. Epter RS, Helm S, Hayek SM, Benyamin RM, Smith HS, Abdi S. Systematic review of percutaneous adhesiolysis and management of chronic low back pain in post lumbar surgery syndrome. *Pain Physician* 2009; 12:361-378.
74. Patel VB, Manchikanti L, Singh V, Schultz DM, Hayek SM, Smith HS. Systematic review of intrathecal infusion systems for long-term management of chronic non-cancer pain. *Pain Physician* 2009; 12:345-360.
75. Rupert MP, Lee M, Manchikanti L, Datta S, Cohen SP. Evaluation of sacroiliac joint interventions: A systematic appraisal of the literature. *Pain Physician* 2009; 12:399-418.
76. Hayek SM, Helm S, Benyamin RM, Singh V, Bryce DA, Smith HS. Effectiveness of spinal endoscopic adhesiolysis in post lumbar surgery syndrome: A systematic review. *Pain Physician* 2009; 12:419-435.
77. Hirsch JA, Singh V, Falco FJE, Benyamin RM, Manchikanti L. Automated percutaneous lumbar discectomy for the contained herniated lumbar disc: A systematic assessment of evidence. *Pain Physician* 2009; 12:601-620.
78. Manchikanti L, Boswell MV, Singh V, Derby R, Fellows B, Falco FJE, Datta S, Smith HS, Hirsch JA. Comprehensive review of neurophysiologic basis and diagnostic interventions in managing chronic spinal pain. *Pain Physician* 2009; 12:E71-E120.
79. Manchikanti L, Boswell MV, Datta S, Fellows B, Abdi S, Singh V, Benyamin RM, Falco FJE, Helm S, Hayek S, Smith HS. Comprehensive review of therapeutic interventions in managing chronic spinal pain. *Pain Physician* 2009: 12:E123-E198.
80. Abdi S, Datta S, Trescot AM, Schultz DM, Adlaka R, Atluri SL, Smith HS, Manchikanti L. Epidural steroids in the management of chronic spinal pain: A systematic review. *Pain Physician* 2007; 10:185-212.
81. Datta S, Everett CR, Trescot AM, Schultz DM, Adlaka R, Abdi S, Atluri SL, Smith HS, Shah RV. An updated systematic review of diagnostic utility of selective nerve root blocks. *Pain Physician* 2007; 10:113-128.
82. Trescot AM, Chopra P, Abdi S, Datta S, Schultz DM. Systematic review of effectiveness and complications of adhesiolysis in the management of chronic spinal pain: An update. *Pain Physician* 2007; 10:129-146.
83. Hansen HC, McKenzie-Brown AM, Cohen SP, Swicegood JR, Colson JD, Manchikanti L. Sacroiliac joint interventions: A systematic review. *Pain Physician* 2007; 10:165-184.
84. Buenaventura RM, Shah RV, Patel V, Benyamin RM, Singh V. Systematic review of discography as a diagnostic test for spinal pain: An update. *Pain Physician* 2007; 10:147-164.
85. Sehgal N, Dunbar EE, Shah RV, Colson JD. Systematic review of diagnostic utility of facet (zygapophysial) joint injections in chronic spinal pain: An update. *Pain Physician* 2007; 10:213-228.
86. Boswell MV, Colson JD, Sehgal N, Dunbar E, Epter R. A systematic review of

therapeutic facet joint interventions in chronic spinal pain. *Pain Physician* 2007; 10:229-253.

87. Hotopf M. The pragmatic randomized controlled trial. *Adv Psychiatr Treat* 2002; 8:326-333.
88. Williams DD, Garner J. The case against "the evidence": A different perspective on evidence-based medicine. *Br J Psychiatry* 2002; 180:8-12.
89. Hotopf M, Lewis G, Normand C. Putting trials on trial: The costs and consequences of small trials in depression: A systematic review of methodology. *J Epidemiol Community Health* 1997; 51:354-358.
90. Thornley B, Adams C. Content and quality of 2,000 controlled trials in schizophrenia over 50 years. *BMJ* 1998; 317:1181-1184.
91. Hotopf M, Churchill R, Lewis G. Pragmatic randomized controlled trials in psychiatry. *Br J Psychiatry* 1999; 175:217-223.
92. Miles A, Charlton B, Bentley P, Polychronis A, Grey J, Price N. New perspectives in the evidence-based healthcare debate. *J Eval Clin Pract* 2000; 6:77-84.
93. Boswell MV, Giordano J. Evidence-based or evidence-biased: The need to re-appraise and re-align levels of information with stakeholder values. *Pain Physician* 2009; 12:283-286.
94. Holmes D, Murray SJ, Perron A, Rail G. Deconstructing the evidence for evidence-based discourse in health sciences: Truth, power and fascism. *Int J Evidence-Based Healthcare* 2006; 4:180-186.
95. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Ann Intern Med* 2009; 151:W65-W94.
96. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* 2009; 339:b2535.
97. Sampson M, McGowan J, Tetzlaff J, Cogo E, Moher D. No consensus exists on search reporting methods for systematic reviews. *J Clin Epidemiol* 2008; 61:748-754.
98. Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, Thorlund K. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008; 61:587-865.
99. Tseng TY, Dahm P, Poolman RW, Preminger GM, Canales BJ, Montori VM. How to use a systematic literature review and meta-analysis. *J Urol* 2008; 180:1249-1256.
100. Vivekananthan DP, Penn MS, Sapp SK, Hsu A, Topol EJ. Use of antioxidant vitamins for the prevention of cardiovascular disease: Meta-analysis of randomized trials. *Lancet* 2003; 361:2017-2023.
101. Eidelman RS, Hollar D, Hebert PR, Lamas GA, Hennekens CH. Randomized trials of vitamin E in the treatment and prevention of cardiovascular disease. *Arch Intern Med* 2004; 164:1552-1556.
102. Shekelle PG, Morton SC, Jungvig LK, Udani J, Spar M, Tu WJ, Suttorp M, Coulter I, Newberry SJ, Hardy M. Effect of supplemental vitamin E for the prevention and treatment of cardiovascular disease. *J Gen Intern Med* 2004; 19:380-389.
103. Miller ER 3rd, Pastor-Barriuso R, Dalal D, Riemersma RA, Appel LJ, Guallar E. Meta-analysis: High-dosage vitamin E supplementation may increase all cause mortality. *Ann Intern Med* 2005; 142:37-46.
104. 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; 354:1896-1900.
105. Moher D, Simera I, Schulz KF, Hoey J, Altman DG. Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. *BMC Med* 2008; 6:13.
106. Dixon E, Hameed M, Sutherland F, Cook DJ, Doig C. Evaluating meta-analyses in the general surgical literature: A critical appraisal. *Ann Surg* 2005; 241:450-459.
107. Hemels ME, Vicente C, Sadri H, Masson MJ, Einarson TR. Quality assessment of meta-analyses of RCTs of pharmacotherapy in major depressive disorder. *Curr Med Res Opin* 2004; 20:477-484.
108. Wen J, Ren Y, Wang L, Li Y, Liu Y, Zhou M, Liu P, Ye L, Li Y, Tian W. The reporting quality of meta-analyses improves: A random sampling study. *J Clin Epidemiol* 2008; 61:770-775.
109. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007; 4:e78.
110. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Ann Intern Med* 2007; 147:W163-W194.
111. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Ann Intern Med* 2007; 147:573-577.
112. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004; 328:39-41.
113. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; 312:1215-1218.
114. Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of study designs in four major US journals of obstetrics and gynecology. *Gynecol Obstet Invest* 2001; 51:8-11.
115. Scales CD Jr., Norris RD, Peterson BL, Preminger GM, Dahm P. Clinical research and statistical methods in the urology literature. *J Urol* 2005; 174:1374-1379.
116. Greene T. Are observational studies "just as effective" as randomized clinical trials? *Blood Purif* 2000; 18:317-322.
117. Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ* 2006; 174:635-641.
118. Hempel CG. *Philosophy of Natural Science*. Prentice-Hall, Englewood Cliffs, 1966.
119. Popper KR. *The Logic of Scientific Discovery*. Cambridge University Press, London, 1958.
120. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; 283:2008-2012.
121. Peipert JF, Phipps MG. Observational studies. *Clin Obstet Gynecol* 1998;

41:235-244.

122. Lipsett M, Campleman S. Occupational exposure to diesel exhaust and lung cancer: A metaanalysis. *Am J Public Health* 1999; 89:1009-1017.
123. Egger M, Altman DG, Vandenbroucke JP; of the STROBE group. Commentary: Strengthening the reporting of observational epidemiology the STROBE statement. *Int J Epidemiol* 2007; 36:948-950.
124. Groenwold RH, Van Deursen AM, Hoes AW, Hak E. Poor quality of reporting confounding bias in observational intervention studies: A systematic review. *Ann Epidemiol* 2008; 18:746-751.
125. Talmon J, Ammenwerth E, Brender J, de Keizer N, Nykänen P, Rigby M. STARE-HI--Statement on reporting of evaluation studies in Health Informatics. *Int J Med Inform* 2009; 78:1-9.
126. Ogrinc G, Mooney SE, Estrada C, Foster T, Goldmann D, Hall LW, Huizinga MM, Liu SK, Mills P, Neily J, Nelson W, Pronovost PJ, Provost L, Rubenstein LV, Speroff T, Splaine M, Thomson R, Tomolo AM, Watts B. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: Explanation and elaboration. *Qual Saf Health Care* 2008; 17:i13-i32.
127. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Zou GY, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman M, Zecevic M, King R, Infante-Rivard C, Stewart A, Birkett N. STrengthening the REporting of Genetic Association studies (STREGA): An extension of the STROBE Statement. *Ann Intern Med* 2009; 150:206-215.
128. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Zou GY, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman M, Zecevic M, King R, Infante-Rivard C, Stewart A, Birkett N; STrengthening the REporting of Genetic Association Studies. STrengthening the REporting of Genetic Association Studies (STREGA): An extension of the STROBE statement. *PLoS Med* 2009; 6: e22.
129. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Zou GY, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman M, Zecevic M, King R, Infante-Rivard C, Stewart AF, Birkett N. Strengthening the reporting of genetic association studies (STREGA): An extension of the strengthening the reporting of observational studies in epidemiology (STROBE) statement. *J Clin Epidemiol* 2009; 62:597-608.
130. Maguire MJ, Hemming K, Hutton JL, Marson AG. Overwhelming heterogeneity in systematic reviews of observational anti-epileptic studies. *Epilepsy Res* 2008; 80:201-212.
131. Simunovic N, Sprague S, Bhandari M. Methodological issues in systematic reviews and meta-analyses of observational studies in orthopaedic research. *J Bone Joint Surg Am* 2009; 91:87-94.
132. Lu CY. Observational studies: A review of study designs, challenges and strategies to reduce confounding. *Int J Clin Pract* 2009; 63:691-697.
133. Morshed S, Tornetta P 3rd, Bhandari M. Analysis of observational studies: A guide to understanding statistical methods. *J Bone Joint Surg Am* 2009; 91:50-60.
134. Bluhm R. Some observations on observational research. *Perspect Biol Med* 2009; 52:252-263.
135. Wang PS, Schoenbaum M. Invited Commentary: Assessing treatment effects by using observational analyses--opportunities and limitations. *Am J Epidemiol* 2009; 170:286-287.
136. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *J Clin Epidemiol* 2008; 61:344-349.
137. Mann CC. Can meta-analysis make policy? *Science* 1994; 266:960-962.
138. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L. *Systems to Rate the Strength of Scientific Evidence*, Evidence Report, Technology Assessment No. 47. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Health Care Research and Quality, 2002. www.thecre.com/pdf/ahrq-system-strength.pdf
139. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061-1066.
140. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995; 274:645-651.
141. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:25.
142. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clin Chem* 2003; 49:1-6.
143. Eden J, Wheatley B, McNeil B, Sox H. Developing trusted clinical practice guidelines. In: *Knowing What Works in Health Care: A Roadmap for the Nation*. National Academies Press, Washington, DC, 2008, pp 121-152.
144. Mulrow C, Cook D (eds). *Systematic Reviews: Synthesis of Best Evidence for Health Care Decisions*. ACP Press, Philadelphia, 1998.
145. Chalmers TC, Matta RJ, Smith H, Kunzler AM. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med* 1977; 297:1091-1096.
146. Juni, P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282:1054-1060.
147. Shulz K, Chalmers I, Hayes R, Altman D. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412.
148. Moher D, Pham B, Jones A, Cook D, Jadad A, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998 352:609-613.
149. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007; (2):MR000012.
150. Petticrew M, Song F, Wilson P, Wright

K. Quality-assessed review of health care interventions and the Database of Abstracts of Reviews of Effectiveness (DARE). *Int J Technol Assess Health Care* 1999; 15:671-678.

151. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: A comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998; 280:278-280.
152. Mulrow CD. The medical review article: State of the science. *Ann Intern Med* 1987; 106:485-488.
153. McAlister FA, Clark HD, van Walraven C, Straus SE, Lawson FM, Moher D, Mulrow CD. The medical review article revisited: Has the science improved? *Ann Intern Med* 1999; 131:947-951.
154. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, Stevens R. Systematic reviews and meta-analyses of treatment of asthma: Critical evaluation. *BMJ* 2000; 320:537-540.
155. Olsen O, Middleton P, Ezzo J, Gøtzsche PC, Hadhazy V, Herxheimer A, Kleijnen J, McIntosh H. Quality of Cochrane reviews: Assessment of sample from 1998. *BMJ* 2001; 323:829-831.
156. Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper journals. *Eval Health Prof* 2002; 25:116-129.
157. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134:663-694.
158. Moher D, Schulz KF, Altman D, for the CONSORT Group. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; 285:1987-1991.
159. Oxman AD. The Cochrane Collaboration in the 21st century: Ten challenges and one reason why they must be met. In: Egger M, Smith GD, Altman DG (eds). *Systematic Reviews in Health Care. Meta-analysis in Context*. BMJ Publishing Group, London, United Kingdom, 2001, pp 459-473.
160. Shrier I. Cochrane Reviews: New blocks on the kids. *Br J Sports Med* 2003; 37:473-474.
161. Shrier I, Boivin JF, Steele RJ, Platt RW, Furlan A, Kakuma R, Brophy J, Rossignol, M. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; 166:1203-1209.
162. Shrier I. Stretching before exercise does not reduce the risk of local muscle injury: A critical review of the clinical and basic science literature. *Clin J Sport Med* 1999; 9:221-227.
163. Pope RP, Herbert RD, Kirwan JD, Graham BJ. A randomized trial of preexercise stretching for prevention of lower-limb injury. *Med Sci Sports Exerc* 2000; 32:271-277.
164. Thacker SB, Gilchrist J, Stroup DF, Kimsey CD Jr. The impact of stretching on sports injury risk: A systematic review of the literature. *Med Sci Sports Exerc* 2004; 36:371-378.
165. Herbert RD, Gabriel M. Effects of stretching before and after exercising on muscle soreness and risk of injury: Systematic review. *BMJ* 2002; 325:468.
166. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; 294:218-228.
167. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess* 2000; 4:1-154.
168. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72:233-240.
169. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; 342:1887-1892.
170. Benson K, Hartz AJ. Observational studies and randomized trials. *N Engl J Med* 2000; 342:1878-1886.
171. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, Contopoulos-Ioannidis DG, Lau J. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001; 286:821-830.
172. Furlan AD, Tomlinson G, Jadad A. Why randomized trials and non-randomized studies of the same interventions agree or disagree. Presented at the 4th Canadian Cochrane Symposium, Montreal, Quebec, Canada, March 12, 2005.
173. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating nonrandomized intervention studies. *Health Technol Assess* 2003; 7:1-173.
174. Kunz R, Oxman AD. The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998; 317:1185-1190.
175. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Deciding on the best therapy. In: Sackett DL, Haynes RB, Guyatt GH (eds). *Clinical Epidemiology: A Basic Science for Clinical Medicine*, 2nd ed. Little, Brown and Company, Toronto, Canada, 1991, pp 187-248.
176. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986; 5:1-30.
177. Greene T. Randomized and observational studies in nephrology: How strong is the evidence? *Am J Kidney Dis* 2009; 53:377-388.
178. Bogardus ST Jr, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: The need for methodological standards. *JAMA* 1999; 281:1919-1926.
179. Peters DL, Barber RC, Flood EM, Garner HR, O'Keefe GE. Methodologic quality and genotyping reproducibility in studies of tumor necrosis factor -308 G-->A single nucleotide polymorphism and bacterial sepsis: Implications for studies of complex traits. *Crit Care Med* 2003; 31:1691-1696.
180. Clark MF, Baudouin SV. A systematic review of the quality of genetic association studies in human sepsis. *Intensive Care Med* 2006; 32:1706-1712.
181. Lee W, Bindman J, Ford T, Glozier N, Moran P, Stewart R, Hotopf M. Bias in psychiatric case-control studies: Literature survey. *Br J Psychiatry* 2007; 190:204-209.
182. Yesupriya A, Evangelou E, Kavvoura FK, Patsopoulos NA, Clyne M, Walsh MC, Lin BK, Yu W, Gwinn M, Ioannidis JP, Khoury MJ. Reporting of human genome epidemiology (HuGE) association studies: An empirical assessment. *BMC Med Res Methodol* 2008; 8:31.
183. Mamdani M, Juurlink DN, Lee DS, Rochon PA, Kopp A, Naglie G, Austin PC, Laupacis A, Stukel TA. Cyclo-oxygen-

ase-2 inhibitors versus non-selective non-steroidal anti-inflammatory drugs and congestive heart failure outcomes in elderly patients: A population-based cohort study. *Lancet* 2004 363:1751-1756.

184. Ray WA, Stein CM, Daugherty JR, Hall K, Arbogast PG, Griffin MR. COX-2 selective non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease. *Lancet* 2002; 360:1071-1073.
185. Agency for Healthcare Research and Quality. Evidence-based Practice Centers Partner's Guide. January 2005. www.ahrq.gov/clinic/epcpartner. Accessed August 11, 2009.
186. Manchikanti L, Singh V, Derby R, Helm S, Trescot AM, Staats PS, Prager JP, Hirsch JA. Review of occupational medicine practice guidelines for interventional pain management and potential implications. *Pain Physician* 2008; 11:271-289.
187. Manchikanti L, Singh V, Helm S, Trescot AM, Hirsch JA. A critical appraisal of 2007 American College of Occupational and Environmental Medicine (ACOEM) practice guidelines for interventional pain management: An independent review utilizing AGREE, AMA, IOM, and other criteria. *Pain Physician* 2008; 11:291-310.
188. Manchikanti L, Singh V, Derby R, Schultz DM, Benyamin RM, Prager JP, Hirsch JA. Reassessment of evidence synthesis of occupational medicine practice guidelines for interventional pain management. *Pain Physician* 2008; 11:393-482.
189. Levin JH. Prospective, double-blind, randomized placebo-controlled trials in interventional spine: What the highest quality literature tells us. *Spine J* 2009; 9:690-703.
190. Manchikanti L, Shah RV, Datta S, Singh V. Critical evaluation of interventional pain management literature provides inaccurate conclusions. *Spine J* 2009; 9:706-708.
191. Chou R. Generating evidence on spinal cord stimulation for failed back surgery syndrome: Not yet fully charged. *Clin J Pain* 2008; 24:757-758.
192. North RB, Shipley J, Taylor RS. Generating evidence on spinal cord stimulation for failed back surgery syndrome: Not yet fully charged. *Clin J Pain* 2008; 24:757-758 (Letter to the Editor). *Clin J Pain* 2009; 25:650; author reply 651.
193. Oxman AD. Systematic reviews: Checklists for review articles. *BMJ* 1994; 309:648-651.
194. Proposed Evaluation Tools for COMPUS [www.ccohta.ca/compus/compus_pdfCOMPUS_Evaluation_Methodology_draft_e.pdf]. Canadian Coordinating Office for Health Technology Assessment, Ottawa, November 29, 2005.
195. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. *Health Res Policy Syst* 2006; 4:20.
196. Royle P, Waugh N. Literature searching for clinical and cost effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. *Health Technology Assessment* 2003, 7(34): www.ncchta.org/execsumm/summ734.htm.
197. Coulter ID. Evidence summaries and synthesis: Necessary but insufficient approach for determining clinical practice of integrated medicine? *Integr Cancer Ther* 2006; 5:282-286.
198. Klassen TP, Jadad AR, Moher D. Guides for reading and interpreting systematic reviews: 1. Getting started. *Arch Pediatr Adolesc Med* 1998; 152:700-704.
199. Lyman GH, Kuderer N. The strengths and limitations of meta-analyses based on aggregate data. *BMC Med Res Methodol* 2005; 5:14.
200. Jadad AR, Moher D, Klassen TP. Guides for reading and interpreting systematic reviews: II. How do the authors find the studies and assess their quality? *Arch Pediatr Adolesc Med* 1998; 152:812-817.
201. Moher D, Jadad AR, Klassen TP. Guides for reading and interpreting systematic reviews: III. How did the authors synthesize the data and make their conclusions? *Arch Pediatr Adolesc Med* 1998; 152:915-920.
202. Lohr KN. Rating the evidence of scientific evidence: Relevance for quality improvement programs. *Int J Qual Health Care* 2004; 16:9-18.
203. Lohr KN, Eleazar K, Mauskopf J. Health policy issues and applications for evidence-based medicine and clinical practice guidelines. *Health Policy* 1998; 46:1-19.
204. Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. www.cochrane-handbook.org.
205. Glasziou P, Irwig L, Bain C, Colditz G. *Systematic Reviews in Health Care. A Practical Guide*. University Press, Cambridge, 2001.
206. Higgins JPT, Green S (eds). Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 (updated September 2006). In: *The Cochrane Library*, Issue 4, 2006, John Wiley & Sons, Ltd. Chichester, UK.
207. Higgins JPT, Green S (eds). Formulating the problem. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 (updated September 2006), Section 4. In: *The Cochrane Library*, Issue 4, 2006. John Wiley & Sons, Ltd., Chichester, UK.
208. Jackson GB. Methods for integrative reviews. *Rev Educ Res* 1980; 50:438-460.
209. Cooper HM. The problem formulation stage. In: Cooper HM (ed). *Integrating Research. A Guide for Literature Reviews*. Sage Publications, Newbury Park, 1984, pp 19-37.
210. Hedges LV. Statistical considerations. In: Cooper H, Hedges LV (eds). *The Handbook of Research Synthesis*. Russell Sage Foundation, New York, 1994; 29-38.
211. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997; 127:380-387.
212. Browner WS. Methods. In: *Publishing and Presenting Clinical Research*. Lippincott Williams & Wilkins, Philadelphia, 2006, pp 27-44.
213. Glanville J. Stage II – Conducting the review, Phase 3 – Identification of research. In: *Undertaking Systematic Reviews of Research on Effectiveness*. CRDs guidance for carrying out or commissioning reviews. CRD Report Number 4 (2nd), CRD Centre for Reviews and Dissemination, University of York, York, UK, March 2001. www.york.ac.uk/inst/crd/report4.htm.
214. Goodman C. *Step 3: Formulate Plan for Literature Search and Step 4: Conduct Literature Search and Retrieval.* Swedish Council on Technology Assessment in Health Care, 1993.
215. Higgins JPT, Green S. Locating and selecting studies. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 (updated September 2006); Section 5. In: *The Cochrane Library*, Issue 4, John Wiley & Sons, Ltd., Chiches-

ter, UK, 2006.

216. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. In: Mulrow C, Cook D (eds). *Systematic reviews. Synthesis of Best Evidence for Health Care Decisions.* ACP Press, Philadelphia, 1998, pp 67-79.
217. Nelemans PJ, DeBie RA, DeVet HCW, Sturmans F. Injection therapy for subacute and chronic benign low back pain. *Spine (Phila Pa 1976)* 2001; 26:501-515.
218. Peloso PMJ, Gross A, Haines T, Trinh K, Goldsmith CH, Burnie SJ; Cervical Overview Group. Medicinal and injection therapies for mechanical neck disorders. *Cochrane Database Syst Rev* 2007; 3:CD000319.
219. Armon C, Argoff CE, Samuels J, Backonja MM; Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Assessment: Use of epidural steroid injections to treat radicular lumbosacral pain: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 2007; 68:723-729.
220. Manchikanti L, Jasper J, Singh V. Cochrane Review by Nelemans et al. *Spine (Phila PA 1976)* 2001; 26:2641-2643.
221. Carragee EJ, Hurwitz EL, Cheng I, Carroll LJ, Nordin M, Guzman J, Peloso P, Holm LW, Côté P, Hogg-Johnson S, van der Velde G, Cassidy JD, Haldeman S, Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Treatment of neck pain: Injections and surgical interventions: Results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)* 2008; 33:S153-S169.
222. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309:1286-1291.
223. Lefebvre C, Eisinga A, McDonald S, Paula N. Enhancing access to reports of randomized trials published worldwide – the contribution of EMBASE records to the Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library. *Emerg Themes Epidemiol* 2008; 5:13.
224. McDonald S, Taylor L, Adams C. Searching the right database. A comparison of four databases for psychiatry journals. *Health Libr Rev* 1999; 16:151-156.
225. Turp JC, Schulte JM, Antes G. Nearly half of dental randomized controlled trials published in German are not included in MEDLINE. *Eur J Oral Sci* 2002; 110:405-411.
226. Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomized trials published in languages other than English in systematic reviews. *Health Technol Assess* 2003; 7:1-90.
227. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. *J Clin Epidemiol* 2005; 58:769-776.
228. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: Empirical study. *Int J Epidemiol* 2002; 31:115-123.
229. Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003; 7:1-76.
230. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, Liberati A, Linde K, Penna A. Completeness of reporting of trials published in languages other than English: Implications for conduct and reporting of systematic reviews. *Lancet* 1996; 347:363-366.
231. Odaka T, Nakayama A, Akazawa K, Sakamoto M, Kinukawa N, Kamakura T, Nishioka Y, Itasaka H, Watanabe Y, Nose Y. The effect of a multiple literature database search – A numerical evaluation in the domain of Japanese life science. *J Med Syst* 1992; 16:177-181.
232. Smith BJ, Darzins PJ, Quinn M, Heller RF. Modern methods of searching the medical literature. *Med J Aust* 1992; 157:603-611.
233. Rovers JP, Janosik JE, Souney PF. Cross over comparison of drug information online database vendors: Dialog and MEDLARS. *Ann Pharmacother* 1993; 27:634-639.
234. Ramos-Remus C, Suarez-Almazor M, Dorgan M, Gomez-Vargas A, Russell AS. Performance of online biomedical databases in rheumatology. *J Rheumatol* 1994; 21:1912-1921.
235. Royle P, Bain L, Waugh N. Systematic reviews of epidemiology in diabetes: Finding the evidence. *BMC Med Res Methodol* 2005; 5:2.
236. Hopewell S, Clark M, Lefebvre C, Scherer R. Handsearching still a valuable element of the systematic review. *Evid Based Dent* 2008; 9:85.
237. Khan KS, Kleijnen J. Stage II – Conducting the review, Phase 4 – Selection of studies. In: *Undertaking Systematic Reviews of Research on Effectiveness.* CRDs guidance for carrying out or commissioning reviews. CRD Report Number 4 (2nd), CRD Centre for Reviews and Dissemination, University of York, York, UK, March 2001. www.york.ac.uk/inst/crd/report4.htm.
238. Khan KS, ter Riet G, Popay J, Nixon J, Kleijnen J. Stage II — Conducting the review, Phase 5 — Study quality assessment. In: *Undertaking Systematic Reviews of Research on Effectiveness.* CRDs guidance for carrying out or commissioning reviews. CRD Report Number 4 (2nd), CRD Centre for Reviews and Dissemination, University of York, York, UK, March 2001. www.york.ac.uk/inst/crd/report4.htm.
239. Higgins JPT, Green S. Assessment of study quality. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 (updated September 2006); Section 6. In: *The Cochrane Library*, Issue 4. John Wiley & Sons, Ltd., Chichester, UK, 2006.
240. Oxman A, Guyatt G. The science of reviewing research. *Ann NY Acad Sci* 1993; 703:125-134.
241. Oxman AD, Stachenko SJ. Meta-analysis in primary care: Theory and Practice. In: Tudiver F, Bass MJ, Dunn EV, Norton PG (eds). *Assessing Interventions: Traditional and Innovative Research Methods for Primary Care*. Sage Publications, Newbury Park, 1992, pp 191-207.
242. Slavin B. Best evidence synthesis: An intelligent alternative to meta-analysis. *J Clin Epidemiol* 1995; 48:9-18.
243. Goodman C. Step 2: Specify inclusion criteria for studies. Swedish Council on Technology Assessment in Health Care, 1993.
244. Clarke M, Oxman AD (eds). Locating and selecting studies. Cochrane Reviewers' Handbook 4.1.5 [updated April 2002]; Section 5. In: *The Cochrane Library*, Issue 4, 2002. Oxford Update Software. Updated quarterly.
245. Cooper H, Ribble RG. Influences on the outcome of literature searches for integrative research reviews. *Knowledge* 1989; 10:179-201.
246. Eysenck HJ. Meta-analysis and its problems. *BMJ* 1994; 309:789-792.

247. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: A systematic review. *Ann Intern Med* 2004; 140:189-202.
248. Plasencia CM, Alderman BW, Baron AE, Rolfs RT, Boyko EJ. A method to describe physician decision thresholds and its application in examining the diagnosis of coronary artery disease based on exercise treadmill testing. *Med Decision Making* 1992; 12:204-212.
249. Guyatt GH, Rennie D. *Users' Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice*. The Evidence-Based Medicine Working Group. AMA Press, Chicago, 2002.
250. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323:42-46.
251. Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, Lau J. Challenges in systematic reviews of diagnostic technologies. *Ann Inter Med* 2005; 142:1048-1055.
252. Carragee EJ, Lincoln T, Parmar VS, Alamin T. A gold standard evaluation of the "discogenic pain" diagnosis as determined by provocative discography. *Spine (Phila Pa 1976)* 2006; 31:2115-2123.
253. Derby R, Kim BJ, Chen Y, Seo KS, Lee SH. The relation between annular disruption on computed tomography scan and pressure-controlled diskography. *Arch Phys Med Rehabil* 2005; 86:1534-1538.
254. Carragee EJ, Tanner CM, Yang B, Brito JL, Truong T. False-positive findings on lumbar discography. Reliability of subjective concordance assessment during provocative disc injection. *Spine (Phila Pa 1976)* 1999; 24:2542-2547.
255. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *Int J Epidemiol* 2007; 36:666-676.
256. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing Clinical Research,* 3rd Ed. Lippincott Williams and Wilkins, Philadelphia, 2007.
257. Hartz A, Marsh JL. Methodologic issues in observational studies. *Clin Orthop Relat Res* 2003; 413:33-42.
258. Busse JW, Jacobs CL, Swiontkowski MF, Bosse MJ, Bhandari M; Evidence-Based Orthopaedic Trauma Working Group. Complex limb salvage or early amputation for severe lower-limb injury: A meta-analysis of observational studies. *J Orthop Trauma* 2007; 21:70-76.
259. Shiga T, Wajima Z, Ohe Y. Is operative delay associated with increased mortality of hip fracture patients? Systematic review, meta-analysis, and meta-regression. *Can J Anaesth* 2008; 55:146-154.
260. Rochon PA, Gurwitz JH, Sykora K, Mamdani M, Streiner DL, Garfinkel S, Normand SL, Anderson GM. Reader's guide to critical appraisal of cohort studies: 1. Role and design. *BMJ* 2005; 330:895-897.
261. Mamdani M, Sykora K, Li P, Normand SL, Streiner DL, Austin PC, Rochon PA, Anderson GM. Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. *BMJ* 2005; 330:960-962.
262. Letournel E, Judet R, Elson R. *Fractures of the Acetabulum*, 2nd Ed. Springer, New York, 1993.
263. Moed BR, Carr SE, Watson JT. Open reduction and internal fixation of posterior wall fractures of the acetabulum. *Clin Orthop Relat Res* 2000; 377:57-67.
264. Matta JM. Fractures of the acetabulum: Accuracy of reduction and clinical results in patients managed operatively within three weeks after the injury. *J Bone Joint Surg Am* 1996; 78:1632-1645.
265. Liebergall M, Mosheiff R, Low J, Goldvirt M, Matan Y, Segal D. Acetabular fractures. Clinical outcome of surgical treatment. *Clin Orthop Relat Res* 1999; 366:205-216.
266. Ciminiello M, Parvizi J, Sharkey PF, Eslampour A, Rothman RH. Total hip arthroplasty: Is small incision better? *J Arthroplasty* 2006; 21:484-488.
267. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70:41-55.
268. Hill AB. The environment and disease: Association or causation? *Proc R Soc Med* 1965; 58:295-300.
269. Rubin DB, Schenker N. Multiple imputation in health-care databases: An overview and some applications. *Stat Med* 1991; 10:585-598.
270. Deeks JJ, Higgins JPT, Altman DG (eds). Analysing and presenting results. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]; Section 8. In: *The Cochrane Library*, Issue 4, 2006. John Wiley & Sons, Ltd., Chichester, UK.
271. Deeks J, Khan KS, Song F, Popay J, Nixon J, Kleijnen J. Stage II – Conducting the review, Phase 7 – Data synthesis. In: *Undertaking Systematic Reviews of Research on Effectiveness.* CRDs guidance for carrying out or commissioning reviews. CRD Report Number 4 (2nd), CRD Centre for Reviews and Dissemination, University of York, York, UK, March 2001. www.york.ac.uk/inst/crd/report4.htm.
272. Marcus SH, Grover PL, Revicki DA. The method of information synthesis and its use in the assessment of health care technology. *Int J Technol Assess Health Care* 1987; 3:497-508.
273. Khan KS, ter Riet G, Kleijnen J. Stage III — Reporting and dissemination, Phase 8 — The report and recommendations. In: *Undertaking Systematic Reviews of Research on Effectiveness.* CRDs guidance for carrying out or commissioning reviews. CRD Report Number 4 (2nd), CRD Centre for Reviews and Dissemination, University of York, York, UK, March 2001. www.york.ac.uk/inst/crd/report4.htm.
274. Deeks JJ, Higgins JPT, Altman DG (eds). Interpreting results. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]; Section 9. In: *The Cochrane Library*, Issue 4, 2006. John Wiley & Sons, Ltd., Chichester, UK.
275. How to use the evidence: Assessment and application of scientific evidence. National Health and Medical Research Council, Canberra, Commonwealth of Australia, 2000, pp 1-84.
276. Bigos SJ, Boyer OR, Braen GR, Brown K, Deyo R, Haldeman S, Hart JL, Johnson EW, Keller R, Kido D, Liang MH, Nelson RM, Nordin M, Owen BD, Pope MH, Schwartz RK, Stewart DH, Susman J, Triano JJ, Tripp LC, Turk DC, Watts C, Weinstein JN. Acute low back problems in adults. Clinical Practice Guideline No.14, AHCPR Publication No. 95-0642. Rockville, Maryland. U.S.A., Agency for Health Care Policy and Research, Public Health Service, U.S., Department of Health and Human Services, December, 1994, pp 1-60.
277. Berg AO, Allan JD. Introducing the third U.S. Preventive Services Task Force. *Am J Prev Med* 2001; 20:S3-S4.

278. Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, Raskob G, Lewis SZ, Schünemann H. Grading strength of recommendations and quality of evidence in clinical guidelines. Report from an American College of Chest Physicians task force. *Chest* 2006; 129:174-181.

279. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007; 147:224-233.

280. Shea B, Boers M, Grimshaw JM, Hamel C, Bouter LM. Does updating improve the methodological and reporting quality of systematic reviews? *BMC Med Res Methodol* 2006; 6:27.

281. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA* 2006; 295:1152-1160.

282. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P; CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Ann Intern Med* 2008; 148:295-309.

283. Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association*. *CMAJ* 1994; 150:1611-1615.

284. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990; 113:69-76.

285. Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: Are structured abstracts more efficient? *Journal of Information Science* 1996; 22:349-356.

286. World Medical Association declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *JAMA* 1997; 277:925-926.

287. Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996; 313:1390-1393.

288. McManus RJ, Wilson S, Delaney BC, Fitzmaurice DA, Hyde CJ, Tobias RS, Jowett S, Hobbs FD. Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. *BMJ* 1998; 317:1562-1563.

289. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials. *Pediatrics* 1989; 84:374-380.

290. Khan KS, Daya S, Collins JA, Walter S. Empirical evidence of bias in infertility research: Overestimation of treatment effect in crossover trials using pregnancy as the outcome measure. *Fertil Steril* 1996; 65:939-945.

291. Horton R. The rhetoric of research. *BMJ* 1995; 310:985-987.

292. *Annals of Internal Medicine*. Information for Authors. Available at www.annals.org.

293. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999; 318:1224-1225.

294. Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editorial process: Characterizing the evolution of a clinical paper. *JAMA* 1998; 280:227-228.

295. Kiviluoto T, Sirén J, Luukkonen P, Kivilaakso E. Randomised trial of laparoscopic versus open cholecystectomy for acute and gangrenous cholecystitis. *Lancet* 1998; 351:321-325.

296. *Pain Physician*. Information for Authors. Available at www.painphysicianjournal.com/2009/january/2009;12;1-7.pdf