Evidence-Based Medicine

Evidence-Based Medicine, Systematic Reviews, and Guidelines in Interventional Pain Management: Part 3: Systematic Reviews and Meta-Analyses of Randomized Trials

Laxmaiah Manchikanti, MD[1], Ramsin Benyamin, MD[2], Standiford Helm II, MD[3], and Joshua A. Hirsch, MD[4]

From: [1]Pain Management Center of Paducah, Paducah, KY; [2]Millennium Pain Center, Bloomington, IL; [3]Pacific Coast Pain Management Center, Laguna Hills, CA; 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, and Associate Clinical Professor of Anesthesiology and Perioperative Medicine, University of Louisville, Louisville, KY.
Dr. Benyamin is President, Millennium Pain Center, Clinical Associate Professor, Department of Surgery, College of Medicine, University of Illinois, Urbana-Champaign, IL.
Dr. Helm is a Medical Director, Pacific Coast Pain Management Center, Laguna Hills, CA.
Dr. Hirsch is Chief of Minimally Invasive Spine Surgery, Depts. of Radiology and Neurosurgery, Massachusetts General Hospital and Assistant Professor of Radiology, Harvard Medical School, Boston, MA.

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

Disclaimer: There was no external funding in the preparation of this manuscript.
Conflict of interest: None.

Manuscript received: 12/5/2008
Accepted for publication: 12/19/2008

Free full manuscript:
www.painphysicianjournal.com

In recent years, progress and innovations in healthcare are measured by evidence-based medicine (EBM), systematic reviews, and meta-analyses. 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." In contrast, meta-analysis is the statistical pooling of data across studies to generate pooled estimates of effects. 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 suggest that a formal set of rules must complement medical training and common sense for clinicians to interpret the results of clinical research effectively. While expertise in the subject matter is crucial, expertise in review methods is also particularly important.

Despite an explosion of systematic reviews and meta-analyses, the empiric research on the quality of systematic reviews has shown that not all systematic reviews are truly systematic, having highly variable quality, deficiencies in methodologic assessment of the quality of the included manuscripts, and bias. Even then, systematic review of the literature is currently the best, least biased, and most rational way to organize, cull, evaluate, and integrate the research evidence from among the expanding medical and healthcare literature. However, a dangerous discrepancy between the experts and the evidence continues to persist in part because multiple instruments are available to assess the quality of systematic reviews or meta-analyses.

Steps in conducting systematic reviews include planning, conducting, reporting, and disseminating the results. The Quality of Reporting of Meta-analysis (QUOROM) statement provides a checklist and a flow diagram. The checklist describes the preferred way to present the abstract, introduction, methods, results, and discussion sections of the report of an analysis. This review describes various aspects of systematic reviews and meta-analyses of randomized trials with a special focus on interventional pain management.

Key words: Randomized trials, pragmatic trials, evidence-based medicine, systematic reviews, meta-analyses, guidelines, bias, interventional pain management, Quality of Reporting of Meta-analysis (QUOROM), Cochrane reviews

Pain Physician 2009; 12:1:35-72

Evidence-based medicine (EBM), systematic reviews, meta-analyses, and clinical guidelines in medicine in general and in interventional pain management in particular are signs of progress in the effort to keep pace with health care innovations, which continue to grow and constantly add to broader and more complex health care interventions and systems. EBM is considered as a shift in medical paradigms, which acknowledges that intuition, unsystematic clinical experience, and pathophysiologic rationale are insufficient grounds for clinical decision-making (1-3). 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 (3). Consequently, systematic reviews of randomized trials take the highest priority as N of 1 RCTs are extremely rare. Systematic reviews and meta-analyses are increasingly popular evidence-based tools and are often used to answer complex research questions across many different research domains (4,5).

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" (5-7). Systematic reviews are labor intensive and require expertise in both the subject matter and review methods. Systematic reviewers must follow the rules of EBM 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 highest 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 (8-10). While expertise in the subject matter is crucial, expertise in review methods is also particularly important.

Meta-analysis, in contrast to a systematic review, is the statistical pooling of data across studies to generate a summary (pooled estimates of effects) (11-13). Generally, a meta-analysis is the final step in a systematic review (11). A meta-analysis should ideally start with an unbiased systematic review that incorporates articles chosen using predetermined inclusion criteria. However, sometimes meta-analyses are done without an initial systematic review.

While meta-analysis and systematic review are not synonymous (4,11,12), they have many similarities and represent a continuum. Systematic reviews and meta-analyses are considered to be the best sources of evidence (12-14). Consequently, systematic reviews and meta-analyses provide 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. Numerous organizations, private and public, for-profit and not-for-profit, have been involved in evidence synthesis (15-37).

Health care providers and other decision-makers depend on systematic reviews and meta-analyses as information resources in which bias has been reduced by the systematic identification, appraisal, synthesis, and, if relevant, statistical aggregation of all relevant studies on a specific topic according to a predetermined and explicit method (8,16,25,26,38-53). However, like any research enterprise, particularly one that is observational, systematic reviews and meta-analyses of evidence can be flawed. In a 1987 survey of 86 English language meta-analyses (38) assessing each publication on 23 items from 6 content areas considered important in the conduct and reporting of meta-analyses of randomized trials, only 24 (28%) of the 86 meta-analyses reported that all 6 content areas had been addressed. An updated survey, which included subsequently published meta-analyses showed little improvement in the rigor with which they were reported (39). Moher and Tricco (40) described issues related to the conduct of systematic reviews with a focus on the nutrition field and made recommendations for improving systematic review conduct. They found multiple variations in the systematic reviews, for example 4 systematic reviews examining the cardiovascular effects of vitamin E supplements (54-57). Surprisingly, all the systematic reviews had similar questions, even though variations were apparent, such as one review focusing on effectiveness (55), another one focusing on efficacy (56), one review searching multiple databases (56), and another (54) searching only one database. They also utilized different inclusion and exclusion criteria, had language limitations, and outcome differences were also evident. The number of studies in these systematic reviews varied from 7 to 84, yet consistency was observed in 3 of these systematic reviews, with no association between vitamin E and any cardiovascular endpoint (54-56). However, the fourth review (57), which 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%. Ultimately, 2 of the systematic reviews (54,56) concluded that vitamin E had no benefit with respect to cardiovascular events, one systematic review (55) concluded that vitamin E had neither benefit nor harm with respect to cardiovascular events, and the fourth review (57) concluded that there was a dose-response relation between vitamin E and increased risk of all-cause mortality, concluding that vitamin E at high doses is harmful.

Several publications have described the science of reviewing research (44,57) differences among narrative reviews, systematic reviews, and meta-analyses (44), and how to carry them out (8,25,27), critically appraise (58-62), and apply (28) systematic reviews and meta-analyses in practice. Due to the expanding nature of systematic reviews and meta-analyses, multiple publications and guidelines have been published (11,43).

The purpose of this report is to provide guidance for clinical research for the interventional pain physician by focusing on the methodology of conducting a systematic literature review and meta-analysis.

1.0 What Is the Terminology of Reviews?

The terminology used to describe systematic reviews and meta-analyses has evolved over time. There are multiple types of reviews and analysis available in the medical literature. Other types of reviews, such as narrative reviews, do not use the explicit methods. A systematic review consists of a clearly formulated question and explicit methods to identify, select, and critically appraise relevant research and then collects and analyzes the data from the studies that are included in the review. A meta-analysis is the use of statistical techniques in a systematic review, which integrates the results of included studies. Thus, a systematic review does not necessarily include a meta-analysis and could be systematically reviewed alone or in combination with meta-analysis. All other types of reviews may be susceptible to bias (63). A properly conducted systematic review or meta-analysis is much more resource and labor intensive than a narrative review (64). Table 1 illustrates differences between a systematic review and methods of the other types of reviews (11).

Table 1. *Comparison of traditional and systematic reviews.*

Components of a review	**Traditional, narrative reviews**	**Systematic reviews**
Formulation of the question	Usually address broad questions	Usually address focused questions
Methods section	Usually not present, or not well-described	Clearly described with pre-stated criteria about participants, interventions, and outcomes
Search strategy to identify studies	Usually not described; mostly limited by reviewers, abilities to retrieve relevant studies; usually not reproducible and prone to selective citation	Clearly described and usually exhaustive; transparent, reproducible and less prone to selective citation
Quality assessment of identified studies	Usually all identified studies are included without explicit quality assessment	Only high-quality studies are included using pre-stated criteria; if lower-quality studies included, the effects of this are tested in subgroup analyses
Data extraction	Methods usually not described	Usually undertaken by more than one reviewer onto pre-tested data forms; attempts often made to obtain missing data from authors of primary studies
Data synthesis	Qualitative description employing the vote counting; approach, where each included study is given equal weight, irrespective of study size and quality	Meta-analysis assigns higher weights to effect measures from more precise studies; pooled, weighted effect measures with confidence limits provide power and precision to results
Heterogeneity	Usually dealt with in a narrative fashion	Heterogeneity dealt with by graphical and statistical methods; attempts are often made to identify sources of heterogeneity
Interpreting results	Prone to cumulative systematic biases and personal opinion	Less prone to systematic biases and personal opinion

Source: Pai M et al. Systematic reviews and meta-analyses: An illustrated, step-by-step guide. *Natl Med J India* 2004; 17:86-95 (11).

2.0 An Introduction to Systematic Reviews

The history of synthesizing research is inextricably bound up in the history of EBM — the global movement to use the best evidence about what does and does not work in health care. James Lind, a Scottish naval surgeon, who is credited with having produced one of the early records of a scientific trial and having written one of the first systematic reviews of evidence, provides modern medicine with the history of systematic reviews (65,66).

On board the Salisbury on May 20, 1747, Lind (66) took 12 patients with scurvy, whose cases "were as similar as I could have them." He divided them into 6 groups of 2 and administered different treatments to each pair of sufferers. The 6 treatments were cider, elixir vitriol, vinegar, seawater, a combination of oranges and lemons, and mixture of garlic, mustard seed, and balsam of Peru. Six days later, Lind's findings were clear. "The result of all my experiments was that oranges and lemons were the most effectual remedies for this distemper at sea" (67). The results of this were published 6 years later acknowledging the need to review the existing literature on scurvy systematically and to discard the weaker forms of evidence. Lind (67) wrote, "As it is no easy matter to root out prejudices … it became requisite to exhibit a full and impartial view of what had hitherto been published on the scurvy . . . by which the sources of these mistakes may be detected. Indeed, before the subject could be set in a clear and proper light, it was necessary to remove a great deal of rubbish." Thus, gathering the published research, getting rid of the "rubbish," and summarizing the best of what remains is essentially the science of systematic reviews. Through the early decades of the twentieth century, scientists working in diverse areas from environmental air quality to physics and agriculture employed rudimentary techniques of research synthesis. In 1904 Karl Pearson published a landmark review of the evidence about the effects of vaccines against typhoid (68).

During the 1960s and 1970s, early systematic review methods were advanced by social scientists (69). Even though the importance of evidence synthesis in medicine was recognized in the 1970s (70), the widespread use of these systematic reviews and meta-analyses did not occur until 2 decades later (71). The stimulating aspect which potentially contributed to this "movement" was evidence that the judgements and opinions of experts were often biased. Thus, the term "systematic review" was coined long before EBM (72). In 1971, Archie Cochrane (70), a British epidemiologist, persuasively advocated the scientific evaluation of commonly used medical therapies through RCTs. By 1979, Cochrane was suggesting that the results of RCTs of the same intervention be systematically summarized. A few years later in 1984, Richard Light and David Pillemer (73) published the pioneering work in the recent history of research synthesis. Three years later, Cynthia Mulrow (74) delivered her damning assessment of the quality of 50 reviews published in the world's leading medical journals during 1985 and 1986. She concluded that these reviews were often subjective, scientifically unsound, and inefficient with only one of the 50 reviews clearly specifying methods of identifying, selecting, and validating included information. Subsequently, in 1993, Oxman and Guyatt (75) published their critique of the poor quality of review articles, based on an assessment of 36 published reviews.

Governments in a number of countries have started subsidizing systematic reviews and also many health care organizations in the 1990s started producing systematic reviews in the public and private sectors (45,71,76-78). Further, the Agency for Health Care Research and Quality (AHRQ) and the U.S. Department of Health and Human Services designated research groups in the United States and Canada as evidence-based practice centers (EPCs) (77). These centers conduct "systematic, comprehensive analyses and syntheses of the scientific literature to develop evidence reports and technology assessments on clinical topics that are common, expensive, and present challenges to decision makers" (77). In addition, in 1999, the Healthcare Financing Administration (HCFA), now the Centers for Medicare and Medicaid Services (CMS) announced that it would require such systematic reviews of the evidence before making any major national coverage decisions, though this policy, in theoery, does not affect the many coverage decisions about therapies made at the regional or state level (78). But, in practice terms, regional and state level decisions (local coverage decisions or LCD's) are based on evidence-based medicine, systematic reviews, and clinical studies.

3.0 An Introduction to Meta-Analyses

Karl Pearson in 1904 summarized and synthesized the results of 11 studies in a landmark review of the effects of vaccine against typhoid, thereby anticipating the development of the meta-analysis, the statistical method used to pool the results of different studies (68). Three years later, in the United States, Joseph Goldberger abstracted and pooled data from 26 of the 44 studies (71). Since its introduction, meta-analysis has been established as an influential branch of health services research, with hundreds of meta-analyses published in the medical literature each year (79). The statistical basis of meta-analysis started in astronomy with intuition and experience suggesting that the combination of data might be better than attempts to select amongst them (80). In 1976, the psychologist Gene Glass coined the term "meta-analysis" in a paper entitled "Primary, Secondary and Meta-analysis of Research" (69). Meta-analysis is becoming increasingly popular in modern medicine (27,81,82).

Meta-analysis has been credited with overcoming the problem first identified by Pearson (68), that "any of the groups . . . are far too small to allow of any definite opinion being formed at all, having regard to the size of the probable error involved." Even though the size of trials published in medical journals has been increasing ever since 1948, many trials fail to detect, or exclude with certainty, a modest but clinically relevant difference in the effects of 2 therapies. Essentially, small trials may prove contradictory with their conclusions and confuse those seeking guidance. The meta-analytic approach may overcome this problem by combining trials evaluating the same intervention in a number of smaller, but comparable, trials. Further, meta-analysis may highlight areas where there is a lack of adequate evidence and thus identify where further studies are needed. In fact, a period of starvation is common practice after gastrointestinal surgery, but a meta-analysis (83) of RCTs of this practice concluded that giving patients nothing by mouth may do more harm than good, and that a large trial is required to clarify this issue. Meta-analysis offers a sounder basis for subgroup analyses, particularly if they are based on individual participant data (84,85).

4.0 What Is the Quality of Systematic Reviews?

There has been an explosion of systematic reviews and meta-analyses as shown in Fig. 1. Empiric research on the quality of systematic reviews has shown that not all systematic reviews are truly systematic (16,46), that the quality of systematic reviews is highly variable (44,47), and that the Cochrane reviews, on average, may be more rigorous and better reported than journal reviews (46,48). However, recent studies also have shown deficiencies even in Cochrane reviews with methodological problems (49,50). Further, it has been shown that among evaluation of 240 systematic reviews from journals, only 48% assessed their quality (51); in the evaluation of 480 systematic reviews in DARE, only 52% assessed quality (16); and in the evaluation of 50 systematic reviews on asthma, only 28% reported validity assessment criteria (48). This indicates a lack of evaluation of the quality of primary studies, which sets apart systematic reviews from traditional reviews. Further, among meta-analyses, heterogeneity is a common finding (45). Empiric work on meta-analyses also has shown that evaluation of heterogeneity is not universally done and that only approximately 45% to 68% of reviews tested for heterogeneity (16,48,52). The results from meta-analyses are not always trustworthy (86-97) led to research into the numerous ways in which bias may be introduced, and the development of methods to detect the presence of such bias.

Moher et al (86) in evaluation of epidemiology and reporting characteristics of systematic reviews concluded that the quality of their reporting was inconsistent, and the readers should not accept systematic reviews uncritically. Delaney et al (87) in a systematic evaluation of the quality of meta-analyses in the critical care literature concluded that overall quality of the reports of meta-analyses available to critical care physicians was poor. Consequently, they suggested that the physicians should critically evaluate these studies prior to considering applying the results of these studies in their clinical practice. McElvenny et al (88) in evaluation of meta-analyses in occupational epidemiology concluded that controversy remains over the definition and validity of meta-standardized mortality ratios, heterogeneity in exposure, and multiple other issues. Dixon et al (89) in critical appraisal and assessment of the meth-

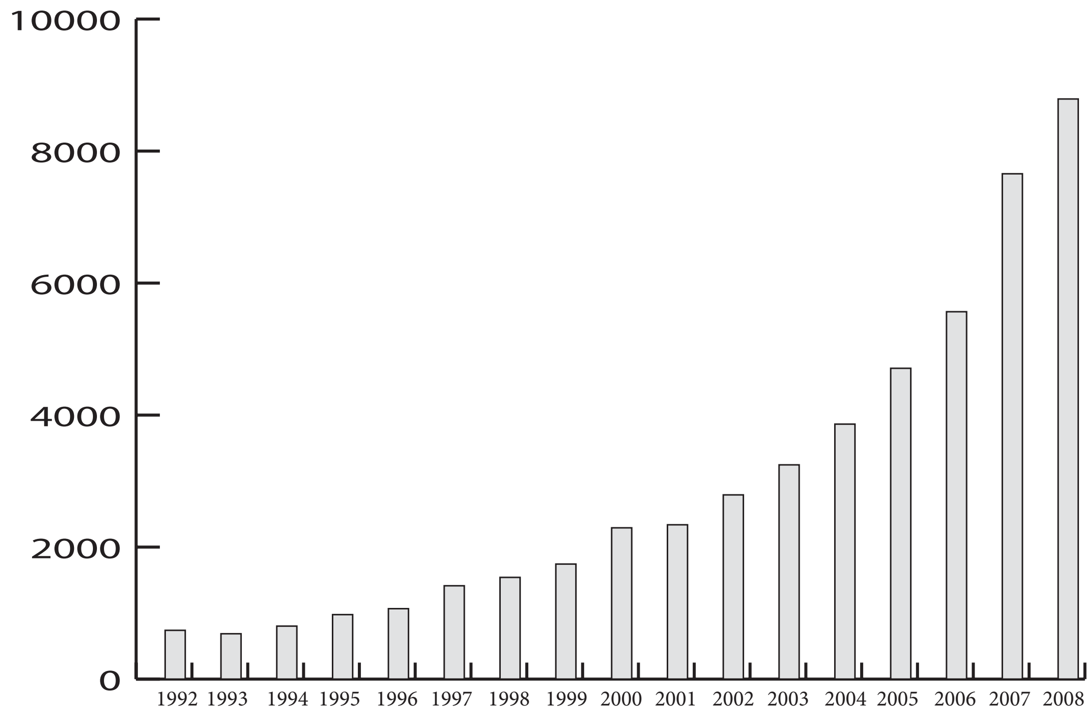

Fig. 1. *Growth of systematic reviews and meta-analyses listed on PubMed.*

odologic quality of meta-analyses of general surgery topics published in peer-reviewed journals concluded that there were frequent methodologic flaws and the quality of these reports limit the validity of the findings and the inferences that can be made about primary studies reviewed.

Lyman and Kuderer (98) in evaluation of the strengths and limitations of meta-analyses based on aggregate data concluded that individual patient data offers advantages, and when feasible, should be considered the best opportunity to summarize the results of multiple studies. In addition, they also concluded that aggregate patient data meta-analysis continues to be the mainstay of systematic reviews utilized by the U.S. Preventive Services Task Force (USPSTF), the Cochrane Collaboration, and many professional societies to support clinical practice guidelines.

5.0 Why Systematic Reviews and Meta-analyses?

Numerous reasons have been described to systematically synthesize the literature. Appropriate synthesis of evidence is essential for health care providers, consumers, researchers, and policy-makers who are inundated with unmanageable amounts of information, which is inconclusive, confusing, and many times biased. Consequently, a systematic review of the literature is currently the best, least biased, and most rational way to organize, cull, evaluate, and integrate the research evidence from among the expanding medical and health care literature (65). The results of a systematic review can help distinguish therapies and interventions that work from those that are useless or harmful and can replace guesswork with more reliable estimates of how well things function. A systematic review can also identify what is known and what is unknown, giving guidance for further research. Light and Pillemer (73) stated that, "without a clear picture of where things stand now, simply adding one new study to the existing morass is unlikely to be very useful . . . for science to be cumulative, an intermediate step between the past and future research is necessary: synthesis of existing evidence." Cynthia Mulrow (44) emphasizes that reviewing systematically is a search for the whole truth rather than just one part of it, and is thus a fundamentally scientific activity (99). Further, she emphasized that we need systematic reviews to efficiently integrate valid information and provide a basis for rational decision-making (100). In 1998, Mulrow stated that systematic reviews are a vital link in the great chain of evidence that stretches from the laboratory bench to the bedside (101). She also stated that there are now millions of studies in health care literature, and systematic reviewing helps separate the insignificant, unsound, or redundant deadwood from the salient and critical studies that are worthy of reflection (102). In addition, it is also important to identify studies with weak designs, because their results are frequently biased and misleading, often overestimating the benefits of the treatment being studied (103-107). While the results of a single study often apply only to a certain kind of patient or a particular policy setting, a systematic review of many studies can provide information relevant to a broad range of patients at different treatment doses in different treatment settings.

In summary, a systematic review serves various purposes (4):

- A systematic review reduces a large amount of information to a manageable size;
- A systematic review may help determine whether the results are consistent from study to study and to generalize the results.
- A systematic review is less expensive and quicker to conduct than to embark on a new study.
- A systematic review may reduce the delay between publication of research findings and the implementation of new effective treatment strategies.
- The systematic review 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.
- A systematic review limits bias and improves the reliability and accuracy of recommendations because of its formalized and thorough method of investigation.

6.0 Dangerous Discrepancies Between Experts and Evidence

A dangerous discrepancy between experts and evidence was noted in July 1992, when Antman et al (63) published the results of a comparison of results of a meta-analysis of RCTs and recommendations by clinical experts in the treatment of myocardial infarction. Using cumulative meta-analysis, Antman et al looked at the latest accumulated evidence for every year between 1960 and 1990 about the effectiveness of commonly used treatments to reduce the risk of heart attack — including thrombolytic therapy, prophylactic lidocaine, Class I anti-arrhythmics, and several others. Following this they compared the latest results to what experts, opinion leaders, or thought leaders were recommending in books and review articles in that year to see whether they were recommending routine use, specific use for certain patients, or no use at all (65). This study found major discrepancies between the accumulating evidence and the experts' recommendations. In most instances where studies showed treatments to be effective, experts' recommendations lagged several years behind the evidence. The most notable example was thrombolytic drugs which were not recommended by more than half of the experts until 13 years after the cumulative evidence showed them to be effective. Even more disappointing, it took 6 years after the first published meta-analysis showed these drugs to be effective before a majority of experts recommended their routine or specific use. Second, with regards to the use of lidocaine to prevent ventricular fibrillation, the study showed that most experts over a 25-year period recommended use of the drug, even though controlled studies provided no evidence that it reduced deaths. In a third example, a small number of experts were still recommending long-term use of anti-arrhythmic drugs, even though the widespread use of anti-arrhythmic drugs was documented to cause numerous deaths (108,109). Thus, this systematic review provided a compelling reason for high quality systematic synthesis of evidence and for its application in clinical practice soon after it is available (110). Antman et al (63) summarized their findings that some experts have not yet mentioned effective therapies, while others continue to recommend those that are ineffective or possibly harmful, concluded that meta-analyses that pool the findings of high-quality trials could help opinion leaders and regulatory bodies to synthesize the burgeoning literature and help improve informed choices for approximated therapies.

7.0 Methodologic Quality Assessment of Systematic Reviews

Often, systematic reviewers seem to ignore the basic principles of EBM and the very different hierarchies necessary for issues of diagnosis, prognosis, and therapy. The (EPCs) Partner's Guide (111) from 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. Even though there is an explosion of systematic reviews and meta-analyses, empiric research on the quality of systematic reviews has shown that not all systematic reviews are truly systematic (16,46). Further, the quality of systematic reviews is highly variable (44,47); thus, methodologic quality assessment of systematic reviews is not only essential, but mandatory.

Due to the complex practice of medicine, Oxman (112) noted the need for checklists analogous to flying an airplane. The most dangerous errors in reviews are systematic ones (bias) rather than ones that occur by chance alone (random errors). Therefore, most important for doers and users of the review is to check its "validity," the extent its design and conduct are likely to have been protected against bias. Random errors and biases are considered to be deadly. In a properly performed systematic review with quantitative results, the confidence intervals (CIs) around the results should provide a good indication of "precision," the extent to which the results are likely to differ from "truth" because of chance alone (47,58,112,113). Oxman (112) provided guidance for the presentation of evaluation synthesis with a description of a systematic review of 2 instruments critically appraising systematic reviews (114,115) and studies how to present the results of a systematic review to policy-makers (116), the general public (117), and users of Cochrane reviews (118).

West et al (114) reviewed different instruments for critically appraising systematic reviews and found 20 systems concerned with the appraisal of systematic reviews or meta-analyses, 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 as shown in Table 2 (39,119-123).

Another review used a detailed process to evaluate and select a system and expanded the work by AHRQ up until the year 2005 (115). In this review, approximately 240 quality assessment instruments were identified for systematic reviews, RCTs, and observational studies, as well as nearly 50 evidence grading systems. Following this critical and extensive review, the AMSTAR 2005 was selected as the best instrument for appraising systematic reviews as illustrated in Table 3 (112).

Further, assessment by the National Institute for Health and Clinical Excellence (NICE) (124) assessed 20 technology assessment reports and found that a more selective approach to database searching would suffice in most cases and would save resources, whereas, searching other sources, including contact with experts and checking reference lists appeared to be a more productive way of identifying further studies.

Coulter (97) has proposed 3 criteria to assess the quality of systematic reviews.

7.1 Who Did the Review?

Reviews are performed by a variety of researchers and institutions. These vary considerably in both expertise and in the resource available to conduct the review. The effect of funding on results has been noted in the literature and strongly consistent evidence shows that industry-sponsored research tends to draw pro-industrial conclusions (125). The most important issue is whether or not there were sufficient resources available to ensure that the review was comprehensive with adequate literature search analysis and expertise, and that the use of these resources did not incur bias.

7.2 What Was the Objective of the Review?

Most objectives involve effectiveness and/or complications of a medical technique. In general, randomized trials tend not to report complications and safety in detail, and these tend to be better reported in observational studies. As well, most interventional pain medicine techniques have not been studied using well-performed randomized, controlled trials. Much of the available literature reflects interventions performed as much as 10 to 15 years earlier, with inadequate or dated methodology.

7.3 How Was the Review Done?

Namely, how was the database searched; were appropriate search terms used; if inclusion and exclusion criteria were utilized, how was the evidence evaluated, what synthesis was possible, and how was safety evaluated? Many systematic reviews in interventional

Table 2. *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 (114).

pain management have had problems with the aforementioned issues.

While there have been a significant number of appropriately performed systematic reviews, which may often be overlooked, a multitude of reviews performed, specifically in the interventional pain management literature, may be poorly performed, misleading, or inappropriate (126-140).

Table 3. *A measurement tool to assess 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	Not 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 (90).

8.0 How to Conduct Systematic Reviews or Meta-analyses

Guidance has been provided for writing (6), as well as reading and interpreting systematic reviews/meta-analyses (47,59,98,112,141-144). Oxman et al (58,112) provided guidance for critical appraisal of the evidence.

Multiple documents describe steps for a systematic review or meta-analysis (11,14,16,40,43,63,145). The central objective of a systematic review is to summarize the evidence on a specific clinical question (12,14,27,146). Secondary objectives are to critically evaluate the quality of the primary studies, check for and identify sources of heterogeneity in results across studies, and, if necessary and possible, determine sources of heterogeneity. Systematic reviews are also helpful in identifying a new research question. Ideally, every research study should begin with a systematic review and build upon the existing evidence base. The Centre for Reviews and Dissemination (CRD) guidance for systematic reviews (14) provides this to be performed in 3 stages and 9 phases as illustrated in Table 4.

In contrast, Cochrane methodology (145) recommends the formulation of the problem, location and selection of studies, assessment of study quality, collection of data, analysis and presentation of results, interpretation of results, and finally, improvement and update of the reviews. The QUORUM statement (43) provides quality reporting of a meta-analyses. In a document presented by Australian and U.S. researchers, 6 key steps were described (101,147), 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.

Planning the review is the first of 3 stages in producing a high quality systematic review and starts with establishing the need for undertaking a review (14). Having established a clear need for a new review, reviewers should undertake a preliminary assessment of the extent of the potentially eligible component studies that are available, and the degree to which it can be used to answer the review questions. Further, all the participants must understand the objective of the review and the methodology to address the objectives, along with assessment of appropriateness and feasibility of the objectives and methodology. The scientific and administrative aspects of the review should be documented in a protocol, which should be discussed before commencing the review itself. Along with the protocol and methodology, a timetable should be arranged to guide the progress of the review work.

Table 4. *Steps in conducting systematic reviews.*

Stage 1
Planning the review

Phase 0
Identification of the need for a review

Phase 1
Preparation of a proposal for a review

Phase 2
Development of a review protocol

Stage II
Conducting a review

Phase 3
Identification of research

Phase 4
Selection of studies

Phase 5
Study quality assessment

Phase 6
Data extraction and monitoring program

Phase 7
Data synthesis

Stage III
Reporting and dissemination

Phase 8
The report and recommendations

Phase 9
Getting evidence into practice

Adapted and modified from: CRD Centre for Reviews and Dissemination, University of York, York, UK. Undertaking Systematic Reviews of Research on Effectiveness. CRDs Guidance for Carrying Out or Commissioning Reviews. CRD Report Number 4 (2nd), March 2001. www.york.ac.uk/inst/crd/report4.htm (14)

8.1 Formulating a Question

As with any research, the first and most important decision in preparing a review is to determine its focus (147). Clearly framed questions are essential for determining the structure of a systematic review or meta-analysis (148-150). In essence, the 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. A properly formulated question also provides relevance for the initial assessment in the review process.

8.1.1 Key Components of a Question

A well formulated question consists of several key components which provide criteria for selecting studies (151,152). Thus, a clearly defined 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. Equal precision in addressing each component is not necessary.

8.1.1.1 Types of Participants

Inclusion criteria for types of participants must be clear. First, define the disease or conditions that are of interest, such as facet joint pain, discogenic pain, or radicular pain. Second, the population of interest must be identified which involves deciding whether one is interested in a special population group determined on the basis of factors such as age, sex, race, educational status, or the presence of a particular condition such as low back pain or radiculitis. Third, setting may also be important such as a community setting, ambulatory surgery setting, hospital outpatient setting, office setting, or inpatient setting.

Any restrictions with respect to specific population characteristics or settings should be based on sound evidence (147). For example, focusing a review on the effectiveness of caudal epidural injections in a Medicare population can be justified based on controversy with coverage policies and previously published systematic reviews. Focusing a review on a particular subgroup of people based on some irrelevant factor based on personal interest or biases when there is no underlying biological or sociological justification for doing so should be avoided.

8.1.1.2 Types of Interventions

It is crucial to define the interventions in formulating a question, along with the specification of the interventions that are of interest. Further, interventions should be clearly described as there are many types of randomized trials with control groups and blinding (2). Studies can be placebo-controlled or pragmatic. Multiple types of controls include placebo, active treatment, no treatment, different dose or regimen of the study treatment, and external or historical control (153). Usefulness of specific control types in various situations has been illustrated as shown in Table 5.

Explanatory trials test whether an intervention is efficacious; that is whether it can have a beneficial effect in an ideal situation. Pragmatic trials measure effectiveness; that is they measure the degree of beneficial effect in real clinical practice. Thus, the explanatory trial seeks to maximize the internal validity by issuing rigorous control of all variables other than the intervention, and the pragmatic trial seeks to maximize external validity to ensure that the results can be generalized. There are advantages and limitations for both types of trials. In modern medicine, specifically in interventional pain management, pragmatic or practical clinical trials measuring effectiveness are considered more appropriate than explanatory trials measuring efficacy (2,154-167). Further, reviewers should realize that explanatory trials are most commonly conducted in academic settings measuring the efficacy, whereas pragmatic or practical trials are best designed to provide the results of benefit of the treatment produced in routine clinical practice (1,2,154,155,163). A further advantage is that with practical clinical trials, information about risks, benefits, and costs of an intervention may be obtained as they occur in routine clinical practice better than in an explanatory trial in an academic setting (2). The issue of lack of a placebo group is addressed in pragmatic trials with a treatment response accounting for the total difference between 2 treatments, including both treatment and associated placebo effect. Consequently, the treatment response in a pragmatic trial is a combination of the treatment effect and placebo effect, as this will best reflect the likely clinical response in actual clinical practice.

Table 5. *Usefulness of specific control types in various situations.*

Trial Objective	Type of Control						
	Placebo Control	Active Control	Dose Response (D/R)	Placebo + Active	Placebo + D/R	Active + D/R	Placebo + Active + D/R
Measure absolute effect size	Y	N	N	Y	Y	N	Y
Show existence of effect	Y	Y	Y	Y	Y	Y	Y
Show dose-response relationship	N	N	Y	N	Y	Y	Y
Compare therapies	N	Y	N	Y	N	P	Y

Y =Yes, N = No, P = Possible, depending on whether there is historical evidence of sensitivity to drug effects. Adapted and modified from: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials E10. July 20, 2000 (153).

8.1.1.3 Types of Outcomes

The third key component of a well-formulated question is the delineation of particular outcomes that are of interest. While it is important to utilize primary outcomes (pain relief) and secondary outcomes such as functional status, opioid intake, or employment, trivial outcomes should not be included as they only overwhelm and confuse the readers by including data that is of little or no importance alongside the data that is important. However, it is also crucial that important data should not be left out. Consequently, explicit criteria for establishing the presence of appropriate outcomes and if necessary, their combinations must be specified. For example, outcomes may be only pain relief or a combination of pain relief with increased function, return to work, or patient satisfaction.

8.1.1.4 Types of Study Designs

The differences between various types of control designs must be understood (Table 5). Thus, studies may not be excluded based merely on personal philosophy or that the study has no placebo control. The essence of an RCT is randomization but not blinding or placebo control. The critical question remains the control design and appropriate performance of the study.

8.1.2 Usefulness of Key Components of a Question

Properly focused questions should determine the initial searching strategies. This is related to the condition being studied, intervention being assessed, and the population being studied. Further, details relevant to key components of the questions are what the authors will be collecting from individual studies. The questions that the review addresses may be broad or narrow in scope, both associated with certain advantages and disadvantages. Finally, the questions may be refined based on the data which is available during the review. However, it is essential to guard against bias in modifying questions, as post-hoc questions are more susceptible to bias than those asked a priori, and data-driven questions can generate false conclusions based on spurious results. Further, any changes to the protocol that results from revising the question for the review should be documented clearly.

8.2 Finding Relevant Studies

Finding the relevant studies is a complex and time-consuming process. The aim 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 (168-172). Identification of relevant randomized trials by a thorough, unbiased search strategy is crucial. In essence, the comprehensiveness of the search used to capture the relevant trials, determines the validity of the systematic review. Further, the level of precision in the effect estimate that can be generated by a systematic review depends on the volume of information included in the review. In essence, a comprehensive search for relevant RCTs which seeks to minimize bias is one of the essential steps in doing a systematic review and one of the factors that distinguishes a systematic review from narrative or focused review (172).

Recent analysis of Cochrane reviews of interventional pain management and other extensively quoted reviews (135,173-176) showed a lack of appropriate criteria and absence of many key manuscripts. The same was true with highly outspoken critics of interventional pain management in reviews (177,178).

8.2.1 Searching for Studies

A "quick and dirty" search of, for example, MEDLINE, is generally not considered adequate. Studies have shown that only 30% to 80% of all known published RCTs were identifiable using MEDLINE (179). 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 the report (180-182). Even though there is evidence 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 (183-186), 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 (184,187). It has been shown that there is significant value to adding EMBASE to MEDLINE in the search strategy (180). The overlap of EMBASE and MEDLINE has been estimated to be 10% to 87% depending in the topic under investigation (188-192). Researchers comparing databases have concluded that relevant studies would be missed if only MEDLINE were searched for studies in pharmacology (193), toxicology (194,195), psychiatry (181), alternative medicine (196), and other specialties (115,197-203). Lefebvre et al (180) identified that the cumulative sensitivity of the search of 80,000 reports of trials on MEDLINE, EMBASE, and Cochrane library ranged from 0.1% to 60% and cumulative precision ranged from 8% to 61%. Further, the results of many studies are never published, and most of these probably remain unknown. It is believed that studies showing an intervention to be effective are more likely to be published, thus any summary of only the published reports may result in an overestimate of effectiveness due to a publication bias (172,204). However, in more recent years, due to the explosion of many journals and the bias exerted by them, it appears that negative trials are published more frequently than the positive trials. Consequently, it appears that only manuscripts published in society journals are positive trials of their own specialty and negative trials of other specialties. While it has been extremely difficult without compulsory registration of trials at inception to know how many unpublished trials exist, the modern regulations of clinical registry make it easier. Further, many journals refuse to publish reviews that include unpublished data. On a pragmatic basis, admittedly without empirical evidence supporting this, a systematic review in interventional pain management at a minimum must have a comprehensive review using at least 3 sources and provide a description of efforts to identify all databases and journals, if not, unpublished trials. An effective combination of a comprehensive search includes a minimum of 3 bibliographic databases (MEDLINE, EMBASE, Cochrane library), a hand search of references of eligible trials, and direct contact with the corresponding authors of eligible trials asking for additional published or unpublished trial information (205).

A search strategy must be developed and documented appropriately. During this process it is always necessary to strike a balance between comprehensiveness and precision. Increasing the comprehensiveness of a search entails reducing its precision and retrieving more non-relevant articles. An electronic search strategy generally includes 3 sets of terms: terms to search for the health condition of interest, terms to search for interventions evaluated, and terms to search for the types of study design.

8.3 Study Selection

Once the search for potentially relevant studies is completed, the studies should be retrieved and 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 the risk of errors of judgment (206-208). Quality assessment of primary studies is used at various stages in the review process, from study selection to generation of recommendations for practice and research (207,208).

An explicit and standardized method for selecting studies from among all of those identified and then assessing the selections is a key part of a systematic review. Such a method serves the dual purpose of choosing the highest-quality studies and also demonstrates that the selection and assessments have been as free from bias as possible (75,209-213). It is essential that decisions about inclusion or exclusion of studies are made according to predetermined written criteria as stated in the protocol.

8.3.1 Inclusion Criteria

Both inclusion and exclusion criteria should follow logically from the review question and they should be defined in terms of the population, their interventions, the outcomes, and the study designs of interest. Thus, only studies that meet all of the inclusion criteria and none of the exclusion criteria should be included in a review. The inclusion criterion specifying the type of study design stems from the desire to base reviews on the highest quality evidence (214). There are many conditions and interventions in interventional pain management which have not been evaluated with methodologically sound studies. Thus, studies of methodologically lower quality may have to be included.

8.3.2 Study Selection Process

Study selection is a multi-stage process. Initially, the selection criteria is applied liberally to the citation generated from computed database searching. Unless they can be definitely excluded, the titles and abstracts identified as being potentially relevant from searches should be provisionally included for consideration on the basis of full text articles (206).

The final inclusion and exclusion decision should be made after retrieving the full texts of all potentially relevant citations. Reviewers should assess the information contained in these reports to see whether the criteria have been met or not. Most of the citations initially included may be excluded at later stages.

Further, a list of excluded studies may be made, detailing the reason for each exclusion. A final report of the review may also include a flow chart or a table detailing the studies included and excluded from the review (43), which are described in the text.

8.4 Study Quality Assessment

Quality is a construct about which there are different views (208,215,216). These include study quality, the degree to which a study employs measures to minimize biases; focusing on internal validity or methodological 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, namely routine practice. The information gained from quality assessment is crucial in determining the strength of inferences and in assigning grades to recommendations generated within a review. Quality assessment can be used at various stages in a review, starting with the study selection to data synthesis and interpretation.

While almost every systematic review has supporters and detractors, both groups agree on the relevance of the dictum, "garbage in, garbage out" (141). Essentially, one can say that evidence is in the eyes of the reviewer, which illustrates that the extent to which a systematic review could guide health care decisions depends on the quality of the trials available. It is always argued that if the trial quality was assessed appropriately (if it was assessed at all), the expertise of various authors of reviews vary widely with some considering the quality assessment as an important strategy to identify and reduce bias, and others who see assessment as a source of bias or as completely uninformative, whereas, yet some others criticize the criteria utilized on a multitude of personal biases (217,218).

Once reviewers have assessed the trial quality, they should also look at the nature and type of the quality assessment, including the definition and assessment tools employed. Further, it is important to recognize the incorporation of quality assessments into systematic reviews (96,219-221).

8.4.1 Validity

In the context of a systematic review, 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 a measure of the likelihood of chance effect leading to random errors. It is reflected in the CI around the estimate of effect from each study and the weight given to the results of each study when an overall estimate of effect or weighted average is derived. However, more precise results are given more weight.

Variation and validity can explain variation in the results studies included in a systematic review. More rigorous studies may more likely yield results that are closer to the truth. Quantitative analysis of results from studies of variable validity can result in false-positive conclusions (erroneously concluding an intervention is effective) if the less rigorous studies are biased toward overestimating an intervention's effectiveness. They

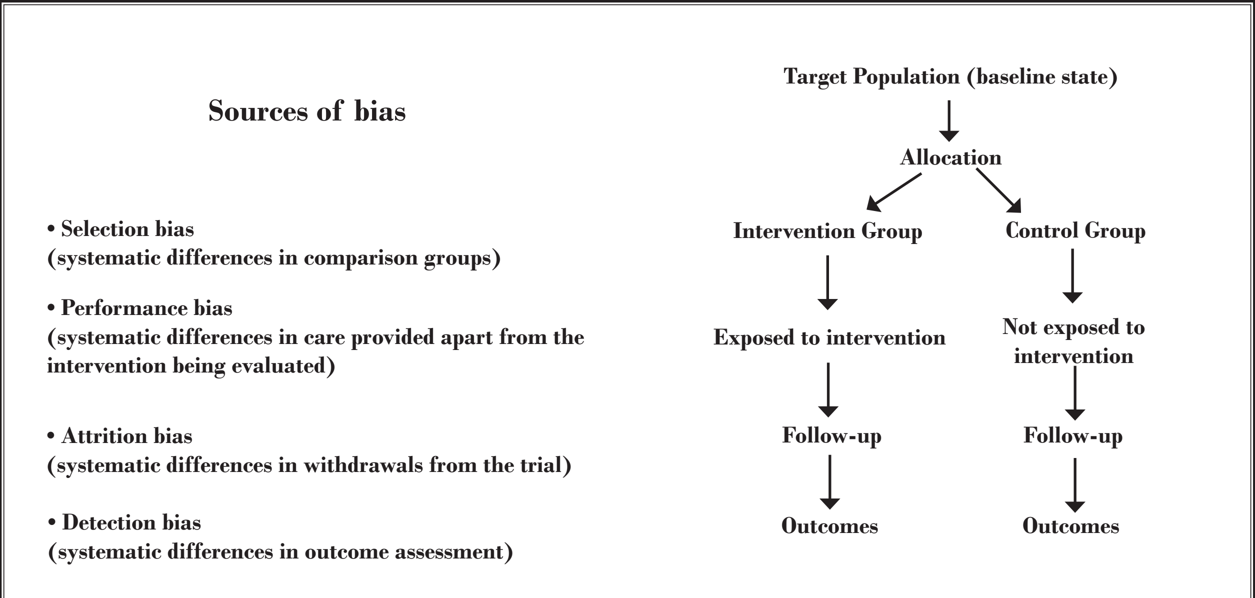

Fig. 2. *Sources of bias in trials of health care interventions. Adapted and modified from Glasziou P et al.* Systematic Reviews in Health Care. A Practical Guide. *University Press, Cambridge, 2001 (146).*

might also come to false-negative conclusions (erroneously concluding no effect) if the less rigorous studies are biased towards underestimating an intervention's effect (199). Thus, it is important to systematically complete critical appraisal of all studies in a review even if there is no variability in either the validity or results of the included studies. In a hypothetical situation, the results may be consistent among studies, but all the studies may be flawed, providing conclusions which are flawed and the conclusions would not be as strong as if a series of rigorous studies yielded consistent results about an intervention's effect.

8.4.2 Assessment of Bias

There are 4 sources of systematic bias in trials of the effects of health care, which include selection bias, performance bias, attrition bias, and detection bias. Fig. 2 illustrates these biases.

However, there is no strong empirical evidence of a relationship between trial outcomes and specific criteria or sets of criteria used to assess the risks of these biases (215,217). A logical basis for suspecting such relationships and a good reason to consider these 4 potential biases in assessing studies for review has been developed.

First, selection bias is one of the most important factors that may lead to bias and to start treatment comparisons, which can result from the way the comparison groups are assembled (221). Using an appropriate method for preventing foreknowledge of treatment assignment is crucially important in the trial design (208). Ideally, the selection process of patients should be impervious to any influence by the individuals making the allocation. This is most certainly achieved by using appropriate randomization by someone who is not responsible for recruiting subjects, such as someone based in a central trial office or pharmacy (2,222,223). Thus, it is essential to evaluate randomization procedures and allocation concealment.

Second, performance bias refers to systematic differences in the care provided to the participants in the comparison groups other than the intervention under investigation. To protect against unintended differences in care and placebo effects, those providing and receiving care can be "blinded" so that they do not know the group to which the recipients of care have been allocated. Some research suggests that such blinding is important in protecting against bias (105,224,225). Studies have shown that contamination (provision of the intervention to the control group) and co-intervention (provision of unintended additional care to either comparison group) can affect the study results (226,227). In addition, there is also evidence that participants who are aware of their assignment status report more symptoms, leading to biased results (224).

Third, attrition bias refers to systematic differences between comparison groups in the loss of participants from the study, also called exclusion bias. However, it may be confused with pre-allocation exclusion and inclusion criteria for enrolling participants. Thus, attrition bias is different from exclusion bias. Because of inadequacies in reporting how losses of participants such as withdrawals, dropouts, protocol deviations, etc., are handled, authors should be cautious about implicit accounts of follow-up (208).

Fourth, detection bias refers to systematic differences between the comparison groups in outcome assessment (208). Trials that blind the people who will assess outcomes to the intervention allocation should logically be less likely to be biased than trials that do not. Blinding is likely to be particularly important in research with subjective outcome measures such as pain (105,224,225). However, at least 2 empirical studies have failed to demonstrate a relationship between blinding of outcome assessment and study results, which may have been due to inadequacies in the reporting of studies (228). In addition, bias due to the selective reporting of results is somewhat different from bias in outcome assessment (208). This source of bias may be important in areas where multiple outcome measures are used, such as evaluations of treatment for arthritis or chronic pain (10). Thus, it is essential to look for specification of predefined primary outcomes and analysis by the investigators as indicators of validity. Alternatively, selective reporting of particular outcomes could be taken to suggest the need for better reporting and efforts by authors to obtain the same missing data.

8.5 Quality Assessment Instruments

Classifying studies according to the level of methodological rigor will help to identify those studies which are of better quality. Both pooled design of studies and lack of rigor in execution of a study may result in biased estimates of effects. Differences in study quality may provide an explanation for heterogeneity in results. When heterogeneity exists, reviewers should weigh the better quality studies in generating inferences (207). A meta-analysis may be conducted where the study results are weighed in proportion to quality (96,106,215,219,229). Alternatively, studies may be pooled cumulatively from high to low quality. Quality assessment instruments are usually based on individual aspects or components of study design, conduct, and analysis. These items can be assembled into a checklist, which can be used to systematically evaluate each study (207). Assigning numerical values to checklist items creates a scale. Checklists and scales offer an overall qualitative and quantitative index of study quality, which cannot be captured by single items alone. There are many generic checklists and scales available in assessing the methodologic quality of studies of randomized trials. Quality assessment instruments also may have multiple disadvantages based on their development, specifically, they have not been developed using rigorous criteria (114,215,230,231). Further, as different checklists and scales emphasize different dimension of quality, variation in these tools may produce differing assessments for the same studies. In addition, variation in quality scales may also produce differing summary estimates in a meta-analysis (104). This observation can be explained, at least in part, by variations in the purpose, scope, and degree of the development of different checklists and scales, and the use of an arbitrary dichotomy (low or high quality) in classifying studies (232). Scales, in particular, have been criticized for ignoring the direction of bias in their schema (218).

Multiple systems to rate the strength of scientific evidence of randomized trials have been published (114,131-133,135,173,176,215,230,231). Table 6 illustrates the Cochrane criteria with weighted scoring (132). AHRQ developed systems to rate the strength of scientific evidence, a comprehensive document (114), which evaluated numerous systems concerned with RCTs including 20 scales, 11 checklists, one component evaluation, and 7 guidance documents, along with the review of 10 rating systems used by AHRQ's EPCs. Subsequently, they designed a set of high-performing scales or checklists pertaining to RCTs by assessing their coverage of the 7 domains, which included study question, study population, randomization, blinding, interventions, outcomes, statistical analysis, results, discussion, and funding or sponsorship. They concluded that 8 systems for RCTs present acceptable approaches that could be used today without major modifications (94,230,233-237). Rating systems used by AHRQ's EPCs are also considered critically developed and reviewed (238-249). However, quite often the researchers tend to use modified systems to meet their needs, and biases, or use outdated systems (31,130,131,133,175,177,178,250-255).

For simplicity purposes, interventional systematic reviews should use methodologic quality assessment criteria as per Koes et al (132). These have been applied in multiple systematic reviews (256-261). Nele-

Table 6. *Modified weighted Cochrane methodologic quality assessment criteria as described by Koes et al (132).*

CRITERION		Weighted Score (points)
1. Study population		**35**
A	Homogeneity	2
B	Comparability of relevant baseline characteristics	5
C	Randomization procedure adequate	4
D	Drop-outs described for each study group separately	3
E	< 20% loss for follow-up	2
	< 10% loss for follow-up	2
F	> 50 subject in the smallest group	8
	> 100 subjects in the smallest group	9
2. Interventions		**25**
G	Interventions included in protocol and described	10
H	Pragmatic study	5
I	Co-interventions avoided or similar	5
J	Placebo-controlled	5
3. Effect		**30**
K	Patients blinded	5
L	Outcome measures relevant	10
M	Blinded outcome assessments	10
N	Follow-up period adequate	5
4. Data-presentation and analysis		**10**
O	Intention-to-treat analysis	5
P	Frequencies of most important outcomes presented for each treatment group	5
	TOTAL SCORE	100

Adapted from Koes BW et al. Efficacy of epidural steroid injections for low-back pain and sciatica: A systematic review of randomized clinical trials. Pain 1995; 63:279-288 (132).

man's et al (135) also utilized modified Cochrane criteria as shown in Table 7. Even then, reviewers of interventional techniques should make several basic decisions regarding the assessment of studies, similar to those made regarding the process of selecting studies. A prime consideration is the number of authors. It is recommended that multiple authors should be involved with an explicit procedure or decision rule identified a priori for assessment, and for identifying and resolving disagreements. Some have suggested blinding the names of the authors to assessment. Empirical evidence suggests that blind assessment of reports might produce lower and more consistent scores than open assessments (231), but there is also contrary evidence with very little or no benefit from blind assessments (262), which are very time consuming and difficult. Even though this aspect continues to be evaluated (208), it may be impossible to blind the reviewers if they are knowledgeable about the subject and practicing clinicians.

Methodologic quality assessments may be used in several ways in a review including as a threshold for inclusion of studies as possible explanations for differences in results between studies, in sensitivity analysis, and as weights in statistical analysis of the study results. If reviewers decide on a methodologic cut-point for including studies, there will be less variation in validity among the included reports. Thus, assessment of validity would characterize studies by the risk of bias within the range above the inclusion cut-point. With a significantly high cut-point, any variation in validity among included studies may be too small to be important.

Table 7. *Modified weighted Cochrane methodologic quality assessment criteria (Neleman et al 135).*

Criterion		Weighted Score (points)
1.	Internal validity	63
A	Selection and restriction	4
B	Treatment allocation (randomization process and concealment should be provided in detail)	15
C	Prognostic comparability. The distribution of baseline characteristics is similar and clearly presented for intervention groups.	10
D	Blinding of patients	4
E	Blinding of physician	4
F	Blinding of observer	4
G	Dropouts	12
H	Loss to follow-up assessment	10
2.	Relevance	20
I	Extra treatments or co-interventions similar.	2
J	Intervention Detailed description of interventions	5
K	Outcome measures	5
L	Timing of outcome measurements	6
M	Side effects	2
3.	Statistical approaches	17
N	Analysis and presentation of data	5
O	Study size	12
	TOTAL	100

Adapted and modified from Nelemans PJ et al. Injection therapy for subacute and chronic benign low back pain. Spine 2001; 26:501-515 (135).

8.6 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. Collection of data, either electronically or on a paper format serves 3 important functions. First, the data collection (263) is directly linked to the formulated review question and planned assessment of included studies, and, therefore, provides a visual representation of these. Second, the data collection format is the historical record of the multitude of decisions and changes to decisions that occur throughout the review process. And thirdly, the data collection format is the data repository from which the analysis will emerge. Data management software is available. Whether it is paper or electronic, key components of a data collection form should include essential information and also methodologic quality assessment criteria as shown in Tables 6 and 7.

8.7 Summarizing and Synthesizing Relevant Study Results

One of the chief goals of systematic reviews of the evidence is to summarize the findings of the best studies available (264,265). A concise written summary of each of the relevant studies is usually provided, 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 a treatment's effects are routinely given more weight in the meta-analysis calculations. One of the most reliable forms of a 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-analyses.

Randomized trials comparing health care interventions use the outcomes of participants to compare the effects of different interventions. While in primary studies, the investigators select and collect data from individual patients, in systematic reviews, the investigators select and collect data from primary studies (264). Meta-analysis focuses on pair-wise comparisons of interventions, such as experimental intervention versus a controlled intervention or the comparison of 2 experimental interventions.

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 (265). In addition to generating a summary of the effects of interventions, it is an integral part of the data synthesis to investigate whether the effects are consistent across the included studies, and if not, to investigate the reasons for the differences.

8.7.1 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 about the characteristics (population, interventions, outcomes, and study quality) and results of the studies included in the review are summarized in a meaningful way. This is best done by tabulation, which allows readers to look at the evidence, its methodological rigor, and the differences between the studies. The descriptive overview is an essential part of the data on which an understanding of the data, planning the quantitative data synthesis, and preventing errors in its interpretation are dependent.

The process of carrying out the descriptive part of data synthesis should be explicit and rigorous (210,266). In general, the effectiveness of a health care intervention is dependent on a large number of factors, some known and others unknown, relating to who receives it, who delivers it and how, and in what context. The key elements in the descriptive approach to data synthesis may include multiple characteristics such as population; interventions; settings where the technology was applied; environmental, social, 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 measure of the study's importance. Each study's weight usually relates to its sample size and the resulting precision of the estimate of effect. Statistical methods of a meta-analysis are explicit numerical formulations of this process, and should be used wherever possible.

8.7.2 Quantitative Synthesis (Meta-Analysis)

An assessment of the tabular summaries helps in planning the quantitative synthesis by highlighting the comparisons that could be made, the outcomes that could be combined (meta-analysis), and the study characteristics that should be considered when investigating variation in effects, also known as heterogeneity (265). First, it should be determined whether quantitative synthesis is at all possible and if so, whether it could be appropriate. Meta-analysis is not 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 heterogeneous to be sensibly combined.

The meta-analysis is performed 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 (264).

Once it is established that a meta-analysis is possible and appropriate, reviewers have to make 3 choices before beginning (265). First, which comparison should be made? Second, which outcome measures should be used in the synthesis? Third, which effect measure (a measure of association quantifying the effect of intervention) should be used to describe effectiveness?

The choice of an effect measure is also essential as there are 4 issues of importance in selecting an effect measure (265). These are what type of data is the outcome measure? Is the measure interpretable by those who will use the review? Is the measure likely to be consistent across the studies and transferrable? Does the measure have the mathematical properties required to give a valid answer?

There are 3 types of data commonly encountered in systematic reviews which include dichotomies of binary data where each individual must be in one of the 2 states, such as dead or alive, which can be summarized using odds ratios, risk ratios, or risk differences; or second, continuous data or outcomes that are summarized as means, arising through measurements or for use of assessment scales, and are summarized in systematic reviews as differences in means, or standardized differences in means (effect sizes). Third, some outcomes measures do not fit the above classification which may be short ordinal scales, such as pain scales, for which it is not sensible to calculate a mean, or are event accounts such as the number of painful attacks per month. In such cases, although there are specific methods dealing with such data, often, the measures are dichotomized and treated as binary data.

In a measure of the effect of an intervention generated by comparing outcomes in an interventional group with those in the control group, the objective is to determine the extent to which outcomes are better or worse in the intervention group compared to the control group. Depending on the measurement scale of the outcome, an effect measure can be generated as a change in an event rate or as a change on a continuous scale. For event data, these comparisons could be generated in terms of relative differences (odds ratio and relative risk) or absolute differences (absolute risk reduction and number needed to treat) between the groups. For continuous data, the effect measures are based on differences in means or standardized mean differences (d-statistics, z scores, or effect sizes). There are multiple methods of meta-analyses.

8.7.3 Interpreting the Review Findings

Interpretation of the results is considered as the final step in the systematic review which essentially returns to the original question that was formulated in explaining how well results have answered it. Even though it is argued that results of a systematic review should stand on their own, many faced with the decision look to the decision and authors' conclusions for help interpreting the results (267,268). Discussion and conclusions about the strength of evidence, the applicability of results, other information, such as consideration of costs and current practice, that might be relevant to someone making a decision, and clarification of any important trade-offs between the expected benefits, harms, and costs of interventions can help to make decisions.

8.7.4 Strength of Evidence

One of the major goals of interpretation is to try to explain the strength of the evidence from the different studies that the review summarized. In other words, for a clinical question for an intervention, the user of the review needs to know whether the best available evidence comes from study designs at a high level in the hierarchy of evidence. A good starting point for the discussion section of the review is to address any important methodological limitations of the included trials and the methods used in the review that might affect the practical decisions about health care or future research. Conclusions regarding the strength of inferences about the effectiveness of an intervention are essentially causal inferences.

Along with interpreting the strength of evidence, the systematic review will attempt to assess the quality of the key studies being reviewed, whatever their level of evidence. RCTs can be poorly run or well run. Providing information about both the level and quality of evidence is a key role of the systematic review. One tool to help assess the quality of RCTs is the revised CONSORT statement (222), the extension of the CONSORT statement of reporting of non-inferiority and equivalence randomized trials (223), and improving the reporting in pragmatic trials (156), developed to try and improve the design, reporting, and analysis of randomized trials (269).

8.7.5 Level of Evidence

Throughout the 1990s and into the twenty-first century, AHRQ has been the foremost federal agency providing research support and policy guidance in health services research in the United States. Its ongoing work includes systems to rate the quality of individual articles, as well as systems for grading the strength of a body of evidence (114,270). The National Health and Medical Research Council (NHMRC) of Australia considers 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 (271). Multiple organizations have described instruments to assess the level of evidence of clinical studies (114,270-272). Grading the quality of individual studies and rating the strength of a body of evidence are both crucial elements. Specific sets of guidelines have been formulated from synthesized sets of evidence providing clear instructions on how systematic reviews (Quality of Reporting of Meta-analysis or QUOROM) (43) and RCTs (CONSORT) (156,222,223,269) may be reported. In addition, AHRQ (114), Cochrane reviews (53), and other reports evaluating evidence-based studies have been published (273-275).

Strength of evidence has a range of definitions, all taking into account 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 study quality (114). 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 8.

However, not all systems are viable or facile; some are extremely cumbersome to use — requiring substantial resources — whereas others are incomplete and are non-comprehensive. Multiple systems have been utilized in the preparation of systematic reviews evaluating the level of evidence. West et al (114) reviewed 40 systems that addressed grading the strength of a body of evidence: 34 from sources other than AHRQ EPCs and 6 from EPCs. The evaluation criteria involved 3 domains — quality, quantity, and consistency that are well established variables for characterizing how confidently one can conclude that a body of knowledge provides information on which clinicians or policy-makers can act. The 34 non-EPC systems incorporated quality, quantity, and consistency to varying degrees. Seven systems fully addressed the quality, quantity, and consistency domains (276-282). Nine others incorporated the 3 domains at least in part (7,28,121,131,283-286). Of these quality of evidence criteria systems routinely used by multiple organizations, there is one from AHRQ (then known as AHCPR) that is now outdated. An example of its use is illustrated in Table 9.

The American Society of Interventional Pain Physicians (ASIPP) guidelines utilized a graded strength of evidence over the years as shown in Table 10.

Recently, the quality of evidence developed by USPSTF is utilized more commonly (256-261,287,288) (Table 11).

8.7.6 Grading Recommendations

Guyatt et al (289) developed an optimal grading system based on the philosophy that guideline panels should make recommendations to administer or not administer an intervention on the basis of a trade-off between benefits on the one hand and risks, burdens, and potential costs on the other. They provided recommendations at 2 levels: strong and weak as illustrated in Table 12. A Grade 1 recommendation (strong) is if guideline panels are very certain that benefits do or do not outweigh the risks and burdens. A Grade 2 (weak) recommendation is if panels think that the benefits and the risks and burdens are finely balanced or applicable and uncertainties exist above the magnitude of the benefits and risks. However, guideline panels must consider a number of factors in grading recommenda-

Table 8. *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 (271).

Table 9. *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 (22). AHCPR was extinguished by Congress in 1995, changing AHCPR to AHRQ. Acute Low Back Pain Guidelines (22) provide a disclaimer "not for patient care."

tions 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.

8.7.7 Applicability

Applicability or generalizability of the results of a systematic review is extremely important. Decisions about applicability depend on knowledge of particular circumstances in which decisions about health care are being made. In addressing the applicability of the results of a review, authors should be cautious not to assume that their own circumstances, or the circumstances reflected in the included studies, are necessarily the same as those of others. Authors can, however, help people to make decisions about applicability by drawing attention to the spectrum of circumstances to which the evidence is likely applicable (268).

8.7.8 Limitations

The interpretation may also discuss tradeoffs between benefits and harms, and, less often, costs. Discussion of costs may be called cost-effective analysis, cost-benefit analysis, economic evaluation, or pharmacoeconomics when applied to drugs. However, the question of whether a treatment or policy intervention is integrated is not yet included in many systematic reviews.

8.8 Improving and Updating Reviews

Updating and improving access to the reviews is considered so important enough by many scientists that it is regarded as the final step in the review process. The Cochrane Collaboration requires that reviewers consider updating each synthesis every 2 years in some cases. ASIPP also requires updating every 2 – 4 years. The emergence of important new evidence from a fresh study can mean that updating is needed even sooner. Sometimes the results of a new trial will

Table 10. *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 11. *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 evidence
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) (287).

Table 12. *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 (289).

mean that the updated review will include a new variable, like quality of life or mortality and morbidity measures. Further, the updated review may introduce a new method of statistical analysis. It has been shown that some reviews expire sooner than others. Consequently, a 2 year limit is optimal; however, 2 to 4 years may be acceptable.

Shojania et al (290) attempted to estimate the average time of changes in evidence that are sufficiently important to warrant updating systematic reviews. This study evaluated survival analysis of 120 quantitative systematic reviews. Quantitative signals for updating were changes in statistical significance or relative changes in effect magnitude of at least 50% involving one of the primary outcomes of the original systematic review or any mortality outcome. Qualitative signals included substantial differences in characterization of effectiveness, new information about harm, and caveats about the previously reported findings that could affect clinical decision-making. The results showed that a qualitative or quantitative signal for updating occurred for 57% of reviews (95% CI, 47% to 67%). Median duration of survival free of signal for updating was 5.5 years. However, a signal occurred within 2 years for 23% of reviews and within one year for 15%. In 7%, a signal had already occurred at the time of the publication. Only 4% of the reviews had a signal within one year of the end of the reported search period; 11% had a signal within 2 years of the search. This study showed shorter survival was associated with cardiovascular topics. Survival for interventional pain management topics may also correlate with cardiovascular topics.

Shea et al (291) compared methodological and reporting quality of original versus updated Cochrane systematic reviews. They concluded that overall quality of Cochrane systematic reviews was fair-to-good. There was reporting quality improved on certain individual items; however, there was no overall improvement seen with updating and methodologic quality remaining unchanged with other items. They concluded that there was room for improvement of methodologic quality and authors updating reviews should address identified methodological or reporting weaknesses. These aspects apply for interventional pain management reviews, either Cochrane or non-Cochrane.

9.0. Reporting of Systematic Reviews

The QUOROM statement (43) has been prepared to improve the quality of reports of meta-analysis of RCTs. Similarly, the CONSORT statements have been developed (156,222,223) to improve the reporting of RCTs. The QUOROM statement provides a checklist and a flow diagram. The checklist describes the preferred way to present the abstract, introduction, methods, results, and discussion sections of a report of a meta-analysis. It is organized into 21 headings and subheadings regarding searches, selection, validity assessment, data extraction, study characteristics, and quantitative data synthesis, and in the results with "trial flow," study characteristics, and quantitative data synthesis; with research documentation being identified for 8 of the 18 items. The flow diagram provides information about both the numbers of RCTs identified, included, and excluded and the reasons for the exclusion of trials. Following the development of the CONSORT statement (222), the authors organized the QUOROM conference to address these issues as they relate to meta-analysis of RCTs (43). Table 13 illustrates QUOROM statement (185,231,262,292-305).

In recent years, with the increasing emphasis on shortening the duration of clinical training in the United Kingdom and continental Europe, many surgical training programs are re-evaluating the role of instruction and clinical research methodology (306). In the USA, the National Institute of Health's (NIH) road map initiative is seeking to expand, enhance, and empower the clinical research workforce by offering clinical research training programs (307). Mahid et al (306) published a clinical research premier for the surgeon by focusing on the methodology of conducting a systematic literature review and meta-analysis.

Even then, a publication evaluating the role of acetylcysteine in the prevention of contrast associated nephropathy with compliance with QUOROM and quality of reporting of overlapping meta-analyses concluded that multiple systematic reviews on the same clinical topic varied in quality of reporting and recommendations (308). They also concluded that longer manuscripts and not-for-profit funding manuscripts were associated with higher quality of reporting.

9.1 Title

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

9.2 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 (222,223,269,292,304) and they also allow readers to find information more easily (292,301). These headings include 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 characteristics; quantitative data synthesis in sufficient detail to permit replication; results; characteristics of the RCTs included and excluded; quantitative and qualitative findings, and subgroup analysis if available; and the conclusion with the main results.

9.3 Introduction

The introduction includes the scientific background and an explanation of rationale. Typically, it includes 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. Further, the introduction 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 (302,303). The authors should also explicitly explain if the systematic review is limited to the review itself or if the meta-analysis is planned.

9.4 Methods

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

9.4.1 Searching

The information sources must be described in detail with databases, registers, personal files, expert informants, agencies, hand searching, and any restrictions such as years considered, publication status, or language of publication (294-296).

Description of the literature search is essential. The authors should also describe if a professional librarian has been used or any other assistance obtained from professionals. The authors should also describe the search terminology utilized in the systematic review.

Table 13. *Quality of reporting of meta-analysis.*

Heading	Subheading	Descriptor	Reported? (Y/N)	Page number
Title		Identify the report as a meta-analysis [or systematic review] of RCTs		
Abstract		Use a structured format (292)		
		Describe		
	Objectives	The clinical question explicitly		
	Data sources	The databases (e.g., list) and other information sources		
	Review methods	The selection criteria (e.g., population, intervention, outcome, and study design); methods for validity assessment, data abstraction, and study characteristics, and quantitative data synthesis in sufficient detail to permit replication		
	Results	Characteristics of the RCTs included and excluded; qualitative and quantitative findings (e.g., point estimates and confidence intervals); and subgroup analyses		
	Conclusion	The main results		
Introduction		The explicit clinical problem, biological rationale for the intervention, and rationale for review		
Methods	Searching	The information sources, in detail (293) (e.g., databases, registers, personal files, expert informants, agencies, hand-searching), and any restrictions (years considered, publication status, (301), language of publication (295,296)		
	Selection	The inclusion and exclusion criteria (defining population, intervention, principal outcomes, and study design (297)		
	Validity assessment	The criteria and process used (e.g., masked conditions, quality assessment, and their findings) (106,231,262,298)		
	Data abstraction	The process or processes used (e.g., completed independently, in duplicate) (185,262)		
	Study characteristics	The type of study design, participants' characteristics, details of intervention, outcome definitions, etc., (122), and how clinical heterogeneity was assessed		
	Quantitative data synthesis	The principal measures of effect (e.g., relative risk), method of combining results (statistical testing and confidence intervals), handling of missing data; how statistical heterogeneity was assessed; (299) a rationale for any a-priori sensitivity and subgroup analyses; and any assessment of publication bias (300)		
Results	Trial flow	Provide a meta-analysis profile summarizing trial flow (see Fig. 3)		
	Study characteristics	Present descriptive data for each trial (e.g., age, sample size, intervention, dose, duration, follow-up period)		
	Quantitative data synthesis	Report agreement on the selection and validity assessment; present simple summary results (for each treatment group in each trial, for each primary outcome); present data needed to calculate effect sizes and confidence intervals in intention-to-treat analyses (e.g., 2 x 2 tables of counts, means and SDs, proportions)		
Discussion		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 (e.g., publication bias); and suggest a future research agenda		

Source: Moher et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Quality of reporting of met-analyses. *Lancet* 1999; 354:1896-1900 (43).

9.4.2 Selection

The authors should clearly describe the inclusion and exclusion criteria with definition of the population, intervention, principle outcomes, and study design (297).

Precise details of the population, setting and locations, interventions, outcomes, and objectives must be clearly described.

9.4.3 Validity Assessment

The multiple criteria and processes used to assess the validity must be described. These may include appropriate randomization, allocation concealment, quality assessment, the instruments utilized, and the results (106,147,262,298).

9.4.4 Data Extraction

Data extraction should be described clearly whether it was completed independently or in duplicate (147,262).

9.4.5 Study Characteristics

Under this section, the type of study design, participants' characteristics, details of intervention, outcome definitions, and the assessment of clinical heterogeneity (122) must be described.

9.5 Quantitative Data Synthesis

The principle measure of effect (relative risk), method of combining results (statistical testing and CIs), handling of missing data, how statistical heterogeneity was assessed (299), a rationale for any prior sensitivity and subgroup analysis, and any assessment of publication bias (300) should be clearly documented and reported.

9.6. Results

The results section includes trial flow, study characteristics, and quantitative data synthesis.

9.6.1 Trial Flow

A trial flow figure should be inserted which shows how the literature was searched and inclusion/exclusion criteria were met. This is illustrated in Fig. 3.

9.6.2 Study Characteristics

The authors should present descriptive data for each trial, along with sample size, intervention, dose, duration, and follow-up periods, etc.

9.6.3 Quantitative Data Synthesis

Results should show the principle measures of effect (method of combining results), statistical testing, and CIs; handling of missing data; the 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 CIs; and intention-to-treat analysis with tables of counts, means, and standard deviations (SDs) or proportions.

9.7 Level of Evidence

Level of evidence may be presented based on the conditions. However, this is not a QUOROM requirement.

9.8 Recommendations

Grading of recommendations may be provided again which is not a recommendation of the QUOROM statement. Cost-effective analysis may be provided which is not a QUOROM requirement.

9.9 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. Table 13 illustrates the QUOROM statement of quality of reporting of meta-analysis.

Some journals have encouraged a structure to the authors' discussion of the results (309-311). The Annals of Internal Medicine (310) 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 weaknesses and limitations of the study (222,223,312,313). It is also essential to describe the difference between statistical significance and clinical importance.

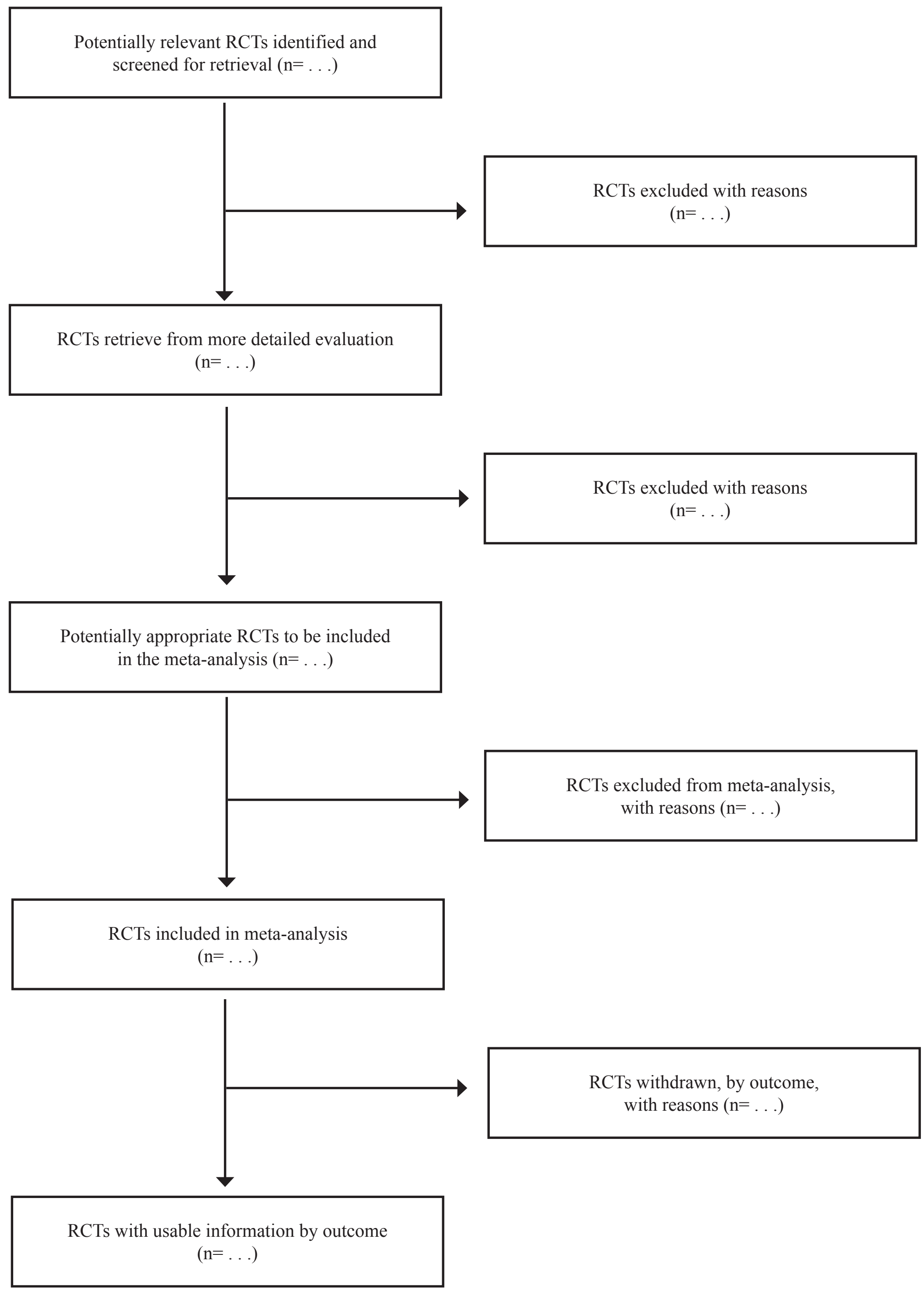

Fig. 3. *Progress through the stages of a meta-analysis for RCTs.*

10.0 Discussion

Assessment of healthcare interventions can be misleading unless investigators ensure unbiased comparisons. Even though in the hierarchy of evidence, systematic evaluations of randomized trials take the number 2 place, one can consider it to be number one in providing the most internally valued evidence for medical decision-making. 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 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, a high index of suspicion for unanticipated difficulties in bias, potentially unnoticed problems, and methodological deficiencies; and skills to report the findings appropriately with close attention to minimizing bias. Sound reporting encompasses adequate reporting and conduct of the review which rests on the footing of sound science, which may not subject readers to speculation. Interventional pain specialists must understand the differences between multiple types of reviews — systematic, meta-analysis, narrative, focused, health technology assessment (HTA), 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 preparation of this manuscript.

References

1. 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.
2. 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.
3. 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.
4. Petrie A, Bulman JS, Osborn JF. Further statistics in dentistry Part 8: Systematic reviews and meta-analyses. *Br Dent J* 2003; 194:73-78.
5. Crowther MA, Cook DJ. Trials and tribulations of systematic reviews and meta-analyses. *Hematology Am Soc Hematol Educ Program* 2007; 2007:493-497.
6. Wright RW, Brand RA, Dunn W, Spindler KP. How to write a systematic review. *Clin Orthop Relat Res* 2007; 455:23-29.
7. 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.
8. 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.
9. Mulrow C, Langhorne P, Grimshaw J. Integrating heterogeneous pieces of evidence in systematic reviews. *Ann Intern Med* 1997; 127:989-995.
10. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10:31-56.
11. 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.
12. Egger M, Smith GD, Altman DG (eds). *Systematic Reviews in Health Care. Meta-Analysis in Context*. BMJ Publishing Group, London, 2001.
13. Clarke M, Oxman AD (eds). *Cochrane Reviewers Handbook 4.2.0* [updated March 2003]. In: *The Cochrane Library*, Issue 2, 2003. Oxford, Update Software.
14. *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
15. Guyatt GH, Rennie D (eds). *Users' Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice*. AMA Press, Chicago, 2002.
16. 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.
17. 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.
18. 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 Village, Epublished August 14, 2008.
19. Denniston PL. *Official Disability Guidelines*, 13th ed. Work Loss Data Institute, 2008.
20. Field MJ, Lohr KN (eds). Clinical Practice

Guidelines: Directions for a New Program. Washington, DC. National Academy Press, 1990.

21. www.guideline.gov/resources/guideline_resources.aspx
22. 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.
23. Gonzalez EG, Materson RS. The guidelines, the controversy, the book. In: Gonzalez ER, Materson RS (eds). *The Nonsurgical Management of Acute Low Back Pain*. Demos Vermande, New York, 1997, pp 1-4.
24. HAYES, Inc. Independent Health Technology Assessment Company. http://www.hayesinc.com
25. Deeks J, Glanville J, Sheldon T. Undertaking Systematic Reviews of Research on Effectiveness. CRD Report NO. 4. York: University of York. NHS Centre for Reviews and Dissemination, 1996.
26. Huston P. The Cochrane Collaboration helping unravel tangled web woven by international research. *Can Med Assoc J* 1996; 154:1389-1392.
27. 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.
28. Guyatt GH, Sackett DL, Sinclair J, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 1995; 274:1800-1804.
29. 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.
30. 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.
31. 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.
32. Manchikanti L, Singh V, Bakhit CE, Fellows B. Interventional techniques in the management of chronic pain. Part 1.0. *Pain Physician* 2000; 3:7-42.
33. Manchikanti L, Singh V, Kloth DS, Slipman CW, Jasper JF, Trescot AM, Varley KG, Atluri SL, Giron C, Curran MJ, Rivera JJ, Baha A, Bakhit CE, Reuter M. Interventional techniques in the management of chronic pain: Part 2.0. *Pain Physician* 2001; 4:24-96.
34. Manchikanti L, Staats PS, Singh V, Schultz DM, Vilims BD, Jasper JF, Kloth DS, Trescot AM, Hansen HC, Falasca TD, Racz GB, Deer T, Burton AW, Helm S, Lou L, Bakhit CE, Dunbar EE, Atluri SL, Calodney AK, Hassenbusch S, Feler CA. Evidence-based practice guidelines for interventional techniques in the management of chronic spinal pain. *Pain Physician* 2003; 6:3-80.
35. Trescot AM, Helm S, Hansen H, Benyamin R, Glaser SE, Adlaka R, Patel S, Manchikanti L. Opioids in the management of chronic non-cancer pain: An update of American Society of the Interventional Pain Physicians' (ASIPP) guidelines. *Pain Physician* 2008; 11:S5-S62.
36. Boswell MV, Shah RV, Everett CR, Sehgal N, McKenzie-Brown AM, Abdi S, Bowman RC, Deer TR, Datta S, Colson JD, Spillane WF, Smith HS, Lucas-Levin LF, Burton AW, Chopra P, Staats PS, Wasserman RA, Manchikanti L. Interventional techniques in the management of chronic spinal pain: Evidence-based practice guidelines. *Pain Physician* 2005; 8:1-47.
37. 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.
38. 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.
39. Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: An update. *Mt Sinai J Med* 1996; 63:216-224.
40. 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.
41. LaDou J, Teitelbaum DT, Egilman DS, Frank AL, Kramer SN, Huff J. American College of Occupational and Environmental Medicine (ACOEM): A professional association in service to industry. *Int J Occup Environ Health* 2007; 13:404-426.
42. Greenberg M. Commentary on effects of exposure to industry influence on ACOEM. *Int J Occup Environ Health* 2007; 13:427.
43. 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 met-analyses. *Lancet* 1999; 354:1896-1900.
44. Mulrow CD. The medical review article: State of the science. *Ann Intern Med* 1987; 106:485-488.
45. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: Synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997; 126:376-380.
46. 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.
47. 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.
48. 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.
49. 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.
50. Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic re-

views published in paper journals. *Eval Health Prof* 2002; 25:116-129.

51. Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, Pham B, Klassen TP. Assessing the quality of reports of randomized trials: Implications for the conduct of meta-analyses. *Health Technol Assess* 1999; 3:1-98.
52. Petitti DB. Approaches to heterogeneity in meta-analysis. *Stat Med* 2001; 20:3625-3633.
53. van Tulder M, Assendelft W, Koes B, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for spinal disorders. *Spine* 1997; 22:2323-2330.
54. Vivekananthan DP, Penn MS, Sapp SK, Hsu A, Topol EJ. Use of antioxidant vitamins for the prevention of cardiovascular disease: Meta-analysis of randomised trials. *Lancet* 2003; 361:2017-2023.
55. 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.
56. 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.
57. 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.
58. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA* 1994; 272:1367-1371.
59. Klassen TP, Jadad AR, Moher D. Guides for reading and interpreting systematic reviews: 1. Getting started. *Arch Pediatr Adolesc Med* 1998; 152:700-704.
60. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987; 107:224-233.
61. Olkin I. A critical look at some popular meta-analytic methods. *Am J Epidemiol* 1984; 140:287-288.
62. Olkin I. Statistical and theoretical considerations in meta-analysis. *J Clin Epidemiol* 1995; 48:133-146.
63. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992; 268:240-248.
64. Allen IE, Olkin I. Estimating time to conduct a meta-analysis from number of citations retrieved. *JAMA* 1999; 282:634-635.
65. Moynihan R. Evaluating health services: A reporter covers the science of research synthesis. Millbank Memorial Fund. March 2004. www.milbank.org/reports/2004Moynihan/040330Moynihan.html
66. 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. http://pc-78-120.udac.se:8001/WWW/Nautica/Medicine/Lind(1753).html (accessed Nov. 24, 2003).
67. www.jameslindlibrary.org/
68. Pearson K. Report on certain enteric fever inoculation statistics. *Br Med J* 1904; 3:1243-1246.
69. Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res* 1976; 5:3-8.
70. Cochrane AL. *Effectiveness and Efficiency. Random Reflections on Health Services*. Royal Society of Medicine Press Limited, London, 1972.
71. Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Eval Health Prof* 2002; 25:12-37.
72. 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.
73. Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research*. Harvard University Press, Cambridge, 1984.
74. Mulrow C. The medical review article: State of the science. *Annals of Intern Med* 1987; 106:485-488.
75. Oxman A, Guyatt G. The science of reviewing research. *Ann NY Acad Sci* 1993; 703:125-134.
76. Hill S, Henry D, Stevens A. The use of evidence in drug selection: The Australian Pharmaceutical Scheme. In *Informing Judgment: Case Studies of Health Policy and Research in Six Countries*, Milbank Memorial Fund and Cochrane Collaboration, New York, 2001.
77. www.ahrq.gov/about/profile.htm
78. www.cms.hhs.gov/coverage/
79. Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? *Int J Epidemiol* 2002; 31:1-5.
80. Plackett RL. Studies in the history of probability and statistics: VII. The principle of the arithmetic mean. *Biometrika* 1958; 45:130-135.
81. Cochrane AL. 1931-1971: A critical review, with particular reference to the medical profession. In: *Medicine for the Year 2000*. Office of Health Economics, London, 1979.
82. Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995; 274:1935-1938.
83. Lewis SJ, Egger M, Sylvester PA, Thomas S. Early enteral feeding versus "nil by mouth" after gastrointestinal surgery: Systematic review and meta-analysis of controlled trials. *BMJ* 2001; 323:773-776.
84. Clarke MJ, Stewart LA. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in Context*. BMJ Publishing Group, London, 2001, pp 228-247.
85. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: An overview of the randomised trials. *Lancet* 1998; 351:1451-1467.
86. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007; 4:e78.
87. Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB, Doig CJ. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit Care* 2005; 9:R575-R582.
88. McElvenny DM, Armstrong BG, Järup L, Higgins JP. Meta-analysis in occupational epidemiology: A review of practice. *Occup Med (Lond)* 2004; 54:336-344.
89. 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.
90. 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.

91. Egger M, Smith GD. Misleading meta-analysis. *Br Med J* 1995; 310:753-754.
92. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significant in meta-analysis: An empirical study of 125 meta-analyses. *Stat Med* 2000; 19:1707-1728.
93. Montori VM, Swiontkowski MF, Cook DJ. Methodologic issues in systematic reviews and meta-analyses. *Clin Orthop Relat Res* 2003; 413:43-54.
94. Rotstein D, Laupacis A. Differences between systematic reviews and health technology assessments: A trade-off between the ideals of scientific rigor and the realities of policy making. *Int J Technol Assess Health Care* 2004; 20:177-183.
95. Clark HD, Wells GA, Huet C, McAlister FA, Salmi LR, Fergusson D, Laupacis A. Assessing the quality of randomized trials: Reliability of the Jadad scale. *Control Clin Trials* 1999; 20:448-452.
96. Khan KS, Daya S, Jadad A. The importance of quality of primary studies in producing unbiased systematic reviews. *Arch Intern Med* 1996; 156:661-666.
97. Coulter ID. Evidence summaries and synthesis: Necessary but insufficient approach for determining clinical practice of integrated medicine? *Integr Cancer Ther* 2006; 5:282-286.
98. Lyman GH, Kuderer N. The strengths and limitations of meta-analyses based on aggregate data. *BMC Med Res Methodol* 2005; 5:14.
99. Higgins JPT, Green S. Introduction. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 (updated September 2006), Section 1. In *The Cochrane Library*, Issue 4, 2006. John Wiley & Sons, Ltd., Chichester, UK.
100. Mulrow CD. Rationale for systematic reviews. *BMJ* 1994; 309:597-599.
101. Mulrow C, Cook D (eds). *Systematic Reviews: Synthesis of Best Evidence for Health Care Decisions*. ACP Press, Philadelphia, 1998.
102. Mulrow C. Rationale for systematic reviews. In Chalmers I, Altman D (eds). *Systematic Reviews*, BMJ Books, London, 1995, pp 1-9.
103. 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.
104. 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.
105. 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.
106. Moher D, Pham B, Jones A, Cook D, Jadad A, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998 352:609-613.
107. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007; (2):MR000012.
108. Moore T. *Deadly Medicine*. Simon and Schuster, New York, 1995.
109. Furberg CD. Effect of antiarrhythmic drugs on mortality after myocardial infarction. *Am J Cardiol* 1983; 52:32C-36C.
110. Chalmers I. Using systematic reviews and registers of ongoing trials for scientific and ethical trial design, monitoring, and reporting. In Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in Context*. BMJ Publishing Group, London, 2001, pp 429-443.
111. Agency for Healthcare Research and Quality. Evidence-based Practice Centers Partner's Guide. January 2005. www.ahrq.gov/clinic/epcpartner/
112. Oxman AD. Systematic reviews: Checklists for review articles. *BMJ* 1994; 309:648-651.
113. Altman DG. Confidence intervals in research evaluation. *Ann Intern Med* 1991; 116:A28.
114. 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 Healthcare Research and Quality, 2002. www.thecre.com/pdf/ahrq-system-strength.pdf
115. 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.
116. Lavis JN, Davies HTO, Oxman AD, Denis JL, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform healthcare management and policy making. *Journal of Health Services Research and Policy* 2005; 10:35-48.
117. Glenton C, Underland V, Kho M, Pennick V, Oxman AD. Summaries of findings, descriptions of interventions, and information about adverse effects would make reviews more informative. *J Clin Epidemiol* 2006; 59:770-778.
118. Glasziou P, Oxman AD, Higgins J. Summary of Findings Tables within Cochrane Reviews: Draft Specification for Rev Man 5.0. December 2004. *Obtaining a Consensus on the Content and Methods of a Summary of Findings Table for Cochrane Reviews. Report to the Cochrane Collaboration Steering Group* 2005.
119. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmer TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994, 120:667-676.
120. Auperin A, Pignon JP, Poynard T. Review article: Critical review of meta-analyses of randomized clinical trials in hepatogastroenterology. *Alimentary Pharmacol Ther* 1997, 11:215-225.
121. Khan KS, Ter Riet G, Glanville J, Sowden AJ, Kleijnen J. *Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews*. CRD Centre for Reviews and Dissemination, University of York, York, UK, 2000.
122. Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. *JAMA* 1998; 279:1566-1570.
123. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, Streiner DL. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; 44:91-98.
124. 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.
125. Krimsky S. The funding effect in science and its implications for the judiciary. *J Law Policy* 2005; 13:43, 59.
126. Boswell MV, Colson JD, Sehgal N, Dunbar EE, Epter R. A systematic review of therapeutic facet joint interventions in chronic spinal pain. *Pain Physician* 2007; 10:229-253.
127. 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.

128. 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.
129. 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.
130. Geurts JW, van Wijk RM, Stolker RJ, Groen GJ. Efficacy of radiofrequency procedures for the treatment of spinal pain: A systematic review of randomized clinical trials. *Reg Anesth Pain Med* 2001; 26:394-400.
131. Niemisto L, Kalso E, Malmivaara A, Seitsalo S, Hurri H. Cochrane Collaboration Back Review Group. Radiofrequency denervation for neck and back pain: A systematic review within the framework of the Cochrane Collaboration back review group. *Spine* 2003, 28:1877-1888.
132. Koes BW, Scholten RJ, Mens JM, Bouter LM. Efficacy of epidural steroid injections for low-back pain and sciatica: A systematic review of randomized clinical trials. *Pain* 1995; 63:279-288.
133. van Tulder MWV, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 1997; 22:2128-2156.
134. Buenaventura RM, Shah RV, Patel V, Benyamin R, Singh V. Systematic review of discography as a diagnostic test for spinal pain: An update. *Pain Physician* 2007; 10: 147-164.
135. Nelemans PJ, DeBie RA, DeVet HCW, Sturmans F. Injection therapy for subacute and chronic benign low back pain. *Spine* 2001; 26:501-515.
136. Kepes ER, Duncalf D. Treatment of backache with spinal injections of local anesthetics, spinal and systemic steroids. *Pain* 1985; 22:33-47.
137. Benzon HT. Epidural steroid injections for low back pain and lumbosacral radiculography. *Pain* 1986; 24:277.
138. Watts RW, Silagy CA. A meta-analysis on the efficacy of epidural corticosteroids in the treatment of sciatica. *Anaesth Intens Care* 1995; 23:564-569.
139. McQuay HJ, Moore RA. Epidural corticosteroids for sciatica. In *An Evidence-Based Resource for Pain Relief*. Oxford University Press, Oxford, 1998, pp 216-218.
140. Bernstein RM. Injections and surgical therapy in chronic pain. *Clin J Pain* 2001; 17:S94-S104.
141. 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.
142. 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.
143. Lohr KN. Rating the evidence of scientific evidence: Relevance for quality improvement programs. *Int J Qual Health Care* 2004; 16:9-18.
144. 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.
145. 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.
146. Glasziou P, Irwig L, Bain C, Colditz G. *Systematic Reviews in Health Care. A Practical Guide.* University Press, Cambridge, 2001.
147. 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.
148. Jackson GB. Methods for integrative reviews. *Rev Educ Res* 1980; 50:438-460.
149. 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.
150. Hedges LV. Statistical considerations. In: Cooper H, Hedges LV (eds). *The Handbook of Research Synthesis*. Russell Sage Foundation, New York, 1994; 29-38.
151. 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.
152. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997; 127:380-387.
153. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials E10. July 20, 2000.
154. Hotopf M. The pragmatic randomized controlled trial. *Adv Psychiatr Treat* 2002; 8:326-333.
155. Hotopf M, Churchill R, Lewis G. Pragmatic randomized controlled trials in psychiatry. *Br J Psychiatry* 1999; 175:217-223.
156. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D; CONSORT group; Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ* 2008; 337:1223-1226.
157. Riew KD, Park JB, Cho YS, Gilula L, Patel A, Lenke LG, Bridwell KH. Nerve root blocks in the treatment of lumbar radicular pain. A minimum five-year follow-up. *J Bone Joint Surg Am* 2006; 88:1722-1725.
158. Manchikanti L, Singh V, Falco FJE, Cash KA, Pampati V. Lumbar facet joint nerve blocks in managing chronic facet joint pain: One-year follow-up of a randomized, double-blind controlled trial: Clinical Trial NCT00355914. *Pain Physician* 2008; 11:121-132.
159. Manchikanti L, Singh V, Falco FJ, Cash KM, Fellows B. Cervical medial branch blocks for chronic cervical facet joint pain: A randomized, double-blind, controlled trial with 1-year follow-up. *Spine* 2008; 33:1813-1820.
160. Manchikanti L, Singh V, Falco FJ, Cash KM, Pampati V. Effectiveness of thoracic medial branch blocks in managing chronic pain: A preliminary report of a randomized, double-blind controlled trial: Clinical Trial NCT00355706. *Pain Physician* 2008; 11:491-504.
161. Manchikanti L, Rivera JJ, Pampati V, Damron KS, McManus CD, Brandon DE, Wilson SR. One day lumbar epidural adhesiolysis and hypertonic saline neurolysis in treatment of chronic low back pain: A randomized, double-blind trial. *Pain Physician* 2004; 7:177-186.
162. Manchikanti L, Boswell MV, Rivera JJ, Pampati V, Damron KS, McManus CD,

Brandon DE, Wilson SR. A randomized, controlled trial of spinal endoscopic adhesiolysis in chronic refractory low back and lower extremity pain. *BMC Anesthesiol* 2005; 5:10.

163. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials. Increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003; 290:1624-1632.
164. Manchikanti L, Cash KA, McManus CD, Pampati V, Smith HS. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 1. Discogenic pain without disc herniation or radiculitis. *Pain Physician* 2008; 11:785-800.
165. Manchikanti L, Singh V, Cash KA, Pampati V, Damron KS, Boswell MV. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 2. Disc herniation and radiculitis. *Pain Physician* 2008; 11:801-815.
166. Manchikanti L, Singh V, Cash KA, Pampati V, Datta S. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 3. Post surgery syndrome. *Pain Physician* 2008; 11:817-831.
167. Manchikanti L, Cash KA, McManus CD, Pampati V, Abdi S. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 4. Spinal stenosis. *Pain Physician* 2008; 11:833-848.
168. 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
169. 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.
170. Section 5. Locating and selecting studies. In Clarke M, Oxman A (eds). *Cochrane Reviewers' Handbook*. 4.1 ed. Cochrane Collaboration, Oxford, 2000.
171. 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.
172. Higgins JPT, Green S (eds). 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, 2006. John Wiley & Sons, Ltd., Chichester, UK.
173. Staal JB, de Bie R, de Vet HC, Hildebrandt J, Nelemans P. Injection therapy for subacute and chronic low-back pain. *Cochrane Database Syst Rev* 2008; 3: CD001824.
174. 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.
175. 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.
176. Nelemans P, deBie R, deVet H, Sturmans F. Injection therapy for subacute and chronic benign low-back pain. *Spine* 2001; 26:2641-2643.
177. 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* 2008; 33:S153-S169.
178. Côté P, van der Velde G, Cassidy JD, Carroll LJ, Hogg-Johnson S, Holm LW, Carragee EJ, Haldeman S, Nordin M, Hurwitz EL, Guzman J, Peloso PM, Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. The burden and determinants of neck pain in workers. Results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* 2008; 33:S60-S74.
179. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309:1286-1291.
180. Lefebvre C, Eisinga A, McDonald S, Paula N. Enhancing access to reports of randomized trials published world-wide – the contribution of EMBASE records to the Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library. Emerg Themes Epidemiol 2008; 5:13.
181. 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.
182. 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.
183. 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.
184. Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technol Assess* 2003; 7:1-90.
185. 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.
186. 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.
187. 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.
188. 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.
189. Smith BJ, Darzins PJ, Quinn M, Heller RF. Modern methods of searching the medical literature. *Med J Aust* 1992; 157:603-611.
190. Rovers JP, Janosik JE, Souney PF. Cross-

over comparison of drug information on-line database vendors: Dialog and MEDLARS. *Ann Pharmacother* 1993; 27:634-639.

191. 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.
192. Royle P, Bain L, Waugh N. Systematic reviews of epidemiology in diabetes: Finding the evidence. *BMC Med Res Methodol* 2005; 5:2.
193. Woods D, Trewheellar K: MEDLINE and EMBASE complement each other in literature searches. *BMJ* 1998; 316:1166.
194. Biarez O, Sarrut B, Doreau CG, Etienne J. Comparison and evaluation of nine bibliographic databases concerning adverse drug reactions. *Drug Intell Clin Pharm* 1991; 25:1062-1065.
195. Barillot MJ, Sarrut B, Doreau CG. Evaluation of drug interaction citation in nine on-line bibliographic databases. *Ann Pharmacother* 1997; 31:45-49.
196. Kleijnen J, Knipschild P. The comprehensiveness of MEDLINE and EMBASE computer searches: Searches for controlled trials of homeopathy, ascorbic acid for common cold and ginkgo biloba for cerebral insufficiency and intermittent claudication. *Pharm Weekbl Sci* 1992; 14:316-320.
197. Wolf FM, Grum CM, Bara A, Milan S, Jones PW. Comparison of MEDLINE and EMBASE retrieval of RCTs of the effects of educational interventions on asthma-related outcomes [abstract]. In *Third International Cochrane Colloquium: 4–8 October 1995; Oslo* the Nordic Cochrane Centre and the Health Services Research Unit, National Institute of Public Health, Oslo. Oslo: The Cochrane Collaboration 1995; V-15.
198. Brazier H, Murphy AW, Lynch C, Bury G. Searching for the evidence in pre-hospital care: A review of randomised controlled trials. *J Accid Emerg Med* 1999; 16:18-23.
199. Topfer LA, Parada A, Menon D, Noorani H, Perras C, Serra-Prat M. Comparison of literature searches on quality and costs for health technology assessment using the MEDLINE and EMBASE databases. *Int J Technol Assess Health Care* 1999; 15:297-303.
200. Minozzi S, Pistotti V, Forni M. Searching for rehabilitation articles on MEDLINE and EMBASE. An example with cross-over design. *Arch Phys Med Rehabil* 2000; 81:720-722.
201. Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE is not enough. *Control Clin Trials* 2000; 21:476-487.
202. Mitchell R, McDonald S, Craig J. How useful is searching Biological Abstracts (BIOSIS) for reports of randomized trials? A comparison with MEDLINE and EMBASE in renal disease [abstract]. In *Ninth International Cochrane Colloquium: 9–13 October 2001; Lyon* the French Cochrane Centre. Lyon: The Cochrane Collaboration 2001; 22.
203. Royle PL, Bain L, Waugh NR. Sources of evidence for systematic reviews of interventions in diabetes. *Diabet Med* 2005; 22:1386-1393.
204. Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, McIlroy W, Oxman AD. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993; 269:2749-2753.
205. Hopewell S, Clark M, Lefebvre C, Scherer R. Handsearching still a valuable element of the systematic review. *Evid Based Dent* 2008; 9:85.
206. Khan KS, Kleijnen J. Stage II – Conducting the review, Phase 4 – Selection of studies. In *Undertaking Systematic Rev iews 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
207. 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
208. Higgins JPT, Green S (eds). 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, 2006. John Wiley & Sons, Ltd., Chichester, UK.
209. 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.
210. Slavin B. Best evidence synthesis: An intelligent alternative to meta-analysis. *J Clin Epidemiol* 1995; 48:9-18.
211. Goodman C. Step 2: Specify inclusion criteria for studies. Swedish Council on Technology Assessment in Health Care, 1993.
212. Clarke M, Oxman A. Section 5. Locating and selecting studies. In: Clarke M, Oxman A (eds). *Cochrane Reviewers' Handbook. 4.1 ed*. Oxford: Cochrane Collaboration 2000.
213. Cooper H, Ribble RG. Influences on the outcome of literature searches for integrative research reviews. *Knowledge* 1989; 10:179-201.
214. Eysenck HJ. Meta-analysis and its problems. *BMJ* 1994; 309:789-792.
215. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Control Clin Trials* 1995; 16:62-73.
216. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, Knipschild PG. The Delphi list: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; 51:1235-1241.
217. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: Current issues and future directions. *Int J Technol Assess Health Care* 1996; 12:195-208.
218. Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994; 140:300-301.
219. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992; 45:255-265.
220. Rennie D, Bero L. The Cochrane Collaboration: Preparing, maintaining and disseminating systematic reviews of the effects of health care. *JAMA* 1995; 274:1935-1938.
221. Kunz R, Oxman AD. The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998; 317:1185-1190.
222. Moher D, Schulz KF, Altman D, for the CONSORT Group. The CONSORT state-

ment: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; 285:1987-1991.

223. 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.
224. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JI. Ascorbic acid for the common cold: A prophylactic and therapeutic trial. *JAMA* 1975; 231:1038-1042.
225. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. *Stat Med* 1989; 8:441-454.
226. The Canadian Cooperative Study Group. The Canadian trial of aspirin and sulfinpyrazone in threatened stroke. *N Engl J Med* 1978; 299:53-59.
227. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; 32:51-63.
228. Reitman D, Chalmers TC, Nagalingam R, Sacks H. Can efficacy of blinding be documented by meta-analysis? Presented to the Society for Clinical Trials, San Diego, 23-26 May, 1988.
229. Liberati A. Meta-analysis: Statistical alchemy for the 21st century. Discussion: A plea for a more balanced view of meta-analysis and systematic overviews of the effect of health care interventions. *J Clin Epidemiol* 1995; 48:81-86.
230. Downs S, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998; 52:377-384.
231. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Control Clin Trials* 1996; 17:1-12.
232. ter Riet G, Leffers P, Zeegers M. Scoring the quality of clinical trials. *JAMA* 2000; 283:1421.
233. de Vet HCW, de Bie RA, van der Heijden GJMG, Verhagen AP, Sijpkes P, Kipschild PG. Systematic reviews on the basis of methodological criteria. *Physiotherapy* 1997; 83:284-289.
234. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986; 4:942-951.
235. van der Heijden GJ, van der Windt DA, Kleijnen J, Koes BW, Bouter LM. Steroid injections for shoulder disorders: A systematic review of randomized clinical trials. *Brit J Gen Pract* 1996; 46:309-316.
236. Sindhu F, Carpenter L, Seers K. Development of a tool to rate the quality assessment of randomized controlled trials using a Delphi technique. *J Adv Nurs* 1997; 25:1262-1268.
237. Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989; 84:815-827.
238. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323:334-336.
239. Seidenfeld J, Samson DJ, Aronson N, Albertson PC, Bayoumi AM, Bennett C, Brown A, Garber A, Gere M, Hasselblad V, Wilt T, Ziegler K. Relative Effectiveness and Cost-Effectiveness of Methods of Androgen Suppression in the Treatment of Advanced Prostate Cancer. Evidence Report/Technology Assessment No. 4. Rockville, MD. Agency for Healthcare Policy and Research. AHCPR Publication No.99-E0012; 1999.
240. Lau J, Ioannidis J, Balk E, Milch C, Chew P, Terrin N, Lang TA, Salem D, Wong JB. Evaluating Ischemia in Emergency Departments: Evidence Report/Technology Assessment: No. 26. Rockville, MD. Agency for Healthcare Research and Quality. AHRQ Publication No. 01-E006 (Contract 290-97-0019 to the New England Medical Center); 2000.
241. Chestnut RM, Carney N, Maynard H, Patterson P, Mann NC, Helfand M. Rehabilitation for Traumatic Brain Injury. Evidence Report/Technology Assessment No. 2. Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No. 99-E006; 1999.
242. Jadad AR, Boyle M, Cunningham C, Kim M, Schachar R. Treatment of Attention-Deficit/Hyperactivity Disorder. Evidence Report/Technology Assessment No. 11. Rockville, MD. Agency for Health care Research and Quality. AHRQ Publication No. 00-E005; 1999.
243. Heidenreich PA, McDonald KM, Hastie T, Fadel B, Hagan V, Lee BK, Hlatky MA. An Evaluation of Beta-Blockers, Calcium Antagonists, Nitrates, and Alternative Therapies for Stable Angina. Rockville, MD. Agency for Healthcare Research and Quality. AHRQ Publication No. 00-E003; 1999.
244. Mulrow CD, Williams JW, Trivedi M, Chiquette E, Aguilar C, Cornell JE. Treatment of Depression: Newer Pharmacotherapies. Evidence Report/Technology Assessment No. 7. Rockville, MD. Agency for Health Care Policy and Research. AHRQ Publication No. 00-E003; 1999.
245. Vickrey BG, Shekelle P, Morton S, Clark K, Pathak M, Kamberg C. Prevention and Management of Urinary Tract Infections in Paralyzed Persons. Evidence Report/Technology Assessment No. 6. Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No. 99-E008; 1999.
246. West SL, Garbutt JC, Carey TS, Lux LJ, Jackman AM, Tolleson-Rinehart S, Lohr KN, Crews FT. Pharmacotherapy for Alcohol Dependence. Evidence Report/Technology Assessment No. 5; Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No. 99-E004; 1999.
247. McNamara RL, Miller MR, Segal JB, Goodman SN, Kim NL, Robinson KA, Powe NR. Management of New Onset Atrial Fibrillation. Evidence Report/Technology Assessment No. 12. Rockville, MD. Agency for Health Care Policy and Research; AHRQ Publication No. 01-E026; 2001.
248. Ross S, Eston R, Chopra S, French J. Management of Newly Diagnosed Patients With Epilepsy: A Systematic Review of the Literature. Evidence Report/Technology Assessment No. 39; Rockville, MD. Agency for Healthcare Research and Quality. AHRQ Publication No. 01-E-029; 2001.
249. Goudas L, Carr DB, Bloch R, Balk E, Ioannidis JPA, Terrin N, Gialeli-Goudas M, Chew P, Lau J. Management of Cancer Pain. Evidence Report/Technology Assessment. No. 35 (Contract 290-97-0019 to the New England Medical Center). Rockville, MD. Agency for Health Care Policy and Research. AHCPR Publication No. 99-E004; 2000.
250. Chou R. Using evidence in pain practice: Part I: Assessing quality of systematic reviews and clinical practice guidelines. *Pain Med* 2008; 9:518-530.
251. Chou R. Using evidence in pain practice: Part II: Interpreting and applying systematic reviews and clinical practice guidelines. *Pain Med* 2008; 9:531-541.
252. Sanders SH, Harden RN, Benson SE, Vicente PJ. Clinical practice guidelines for chronic non-malignant pain syndrome patients II: An evidence-based approach. *J Back Musc Rehabil* 1999; 13:47-58.

253. Resnick DK, Choudhri TF, Dailey AT, Groff MW, Khoo L, Matz PG, Mummaneni P, Watters WC 3rd, Wang J, Walters BC, Hadley MN; American Association of Neurological Surgeons/Congress of Neurological Surgeons. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 13: Injection therapies, low-back pain, and lumbar fusion. *J Neurosurg Spine* 2005; 2:707-715.
254. Levin JH. Prospective, double-blind, randomized placebo-controlled trials in interventional spine: What the highest quality literature tells us. *Spine J* 2008; Sep 11 [Epub ahead of print].
255. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G. Chapter 4: European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006; 15: S192-S300.
256. Atluri S, Datta S, Falco FJ, Lee M. Systematic review of diagnostic utility and therapeutic effectiveness of thoracic facet joint interventions. *Pain Physician* 2008; 11:611-629.
257. 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.
258. 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.
259. Benyamin R., 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.
260. Helm S, Hayek S, Benyamin R, Manchikanti L. Systematic review of effectiveness of thermal annular procedures in treating discogenic low back pain. *Pain Physician* 2009; 12:207-232.
261. Buenaventura R, Datta S, Abdi S, Smith HS. Systematic review of therapeutic lumbar transforaminal epidural steroid injections. *Pain Physician* 2009; 12:233-251.
262. Berlin JA.Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. *Lancet* 1997; 350:185-186.
263. Higgins JPT, Green S (eds). Collecting data. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 (updated September 2006); Section 7. In *The Cochrane Library*, Issue 4, 2006. John Wiley & Sons, Ltd.Chichester, UK.
264. 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
265. 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
266. 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.
267. 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
268. 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.
269. 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.
270. Manchikanti L, Boswell MV, Giordano J. Evidence-based interventional pain management: Principles, problems, potential and applications. *Pain Physician* 2007; 10:329-356.
271. 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.
272. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technol Assess* 1998; 2: pp 1-276.
273. 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.
274. 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. www.clinchem.org/cgi/content/full/49/1/1
275. Bruns DE, Huth EJ, Magid E, Young DS. Toward a checklist for reporting of studies of diagnostic accuracy of medical tests. *Clin Chem* 2000; 46:893-895.
276. Clarke M, Oxman AD (eds). *Cochrane Reviewer's Handbook 4.0*. The Cochrane Collaboration, 1999.
277. Greer N, Mosser G, Logan G, Halaas GW. A practical approach to evidence grading. *Jt Comm J Qual Improv* 2000; 26:700-712.
278. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, Wilson MC, Richardson WS. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: Principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. *JAMA* 2000; 284:1290-1296.
279. NHS Centre for Reviews and Dissemination Centre of Evidence-Based Medicine. Levels of Evidence. www.cebm.net/index.aspx?o=1025
280. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D, Methods Work Group, Third US Preventive Services Task Force. Current methods of the U.S. Preventive Services Task Force: A review of the process. *Am J Prev Med* 2001; 20:21-35.
281. Gyorkos TW, Tannenbaum TN, Abrahamowicz M, Oxman AD, Scott EA, Millson ME, Rasooly I, Frank JW, Riben PD, Mathias RG, Best A. An approach to the development of practice guidelines for community health interventions. *Can J Public Health* 1994; 85:S8-13.
282. Briss PA, Zaza S, Pappaioanou M, Fielding J, Wright-De Agüero L, Truman BI, Hopkins DP, Mullen PD, Thompson RS, Woolf SH, Carande-Kulis VG, Anderson

L, Hinman AR, McQueen DV, Teutsch SM, Harris JR. Developing an evidence-based Guide to Community Preventive Services — methods. The Task Force on Community Preventive Services. *Am J Prev Med* 2000; 18:35-43.

283. Prady SL, Richmond SJ, Morton VM, MacPherson H. A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS ONE* 2008; 3:e1577.
284. Ariens GA, van Mechelen W, Bongers PM, Bouter LM, van der Wal G. Physical risk factors for neck pain. *Scand J Work Environ Health* 2000; 26:7-19.
285. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre (1981d). How to read clinical journals IV. To determine etiology or causation. *Can Med Assoc J* 1981; 124:985-990.
286. Guyatt GH, Cook DJ, Sackett DL, Eckman M, Pauker S. Grades of recommendation for antithrombotic agents. *Chest* 1998; 114:441S-444S.
287. Berg AO, Allan JD. Introducing the third U.S. Preventive Services Task Force. *Am J Prev Med* 2001; 20:S3-S4.
288. 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.
289. 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.
290. 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.
291. 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.
292. 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.
293. Tramér M, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997; 315:635-640.
294. McAuley L, Moher D, Tugwell P. The influence of grey literature on meta-analysis. MSc Thesis: University of Ottawa, 1999.
295. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A; International Cochrane Colloquium. Does the language of publication of reports of randomized trials influence the estimates of intervention effectiveness reported in meta-analyses? [conference presentation]. In: 6th Cochrane Colloquium; 1998; Baltimore. Providence: New England Cochrane Center Providence Office.
296. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997; 350: 326-329.
297. 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.
298. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412.
299. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994; 309:1351-1355.
300. Simes RJ. Publication bias: The case for an international registry of clinical trials. *J Clin Oncol* 1986; 4:1529-1541.
301. Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: Are structured abstracts more efficient? *Journal of Information Science* 1996; 22:349-356.
302. World Medical Association declaration of Helsinki. Recommendations guiding physicians in biomedical research involving human subjects. *JAMA* 1997; 277:925-926.
303. Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996; 313:1390-1393.
304. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990; 113:69-76.
305. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication. International Committee of Medical Journal Editors (ICMJE). Updated October 2008. www.icmje.org
306. Mahid SS, Hornung CA, Minor KS, Turina M, Galandiuk S. Systematic reviews and meta-analysis for the surgeon scientist. *Br J Surg* 2006; 93:1315-1324.
307. National Institutes of Health. NIH Roadmap for Medical Research. http://nihroadmap.nih.gov
308. Biondi-Zoccai GG, Lotrionte M, Abbate A, Testa L, Remigi E, Burzotta F, Valgimigli M, Romagnoli E, Crea F, Agostoni P. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: Case study. *BMJ* 2006; 332:202-209.
309. Horton R. The rhetoric of research. *BMJ* 1995; 310:985-987.
310. *Annals of Internal Medicine.* Information for authors. Available at www.annals.org.
311. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999; 318:1224-1225.
312. 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.
313. 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.