

METHODS FOR EVIDENCE SYNTHESIS IN INTERVENTIONAL PAIN MANAGEMENT

Laxmaiah Manchikanti, MD, James E. Heavner, DVM, PhD, Gabor B. Racz, MD, Nagy Mekhail, MD, David M. Schultz, MD, Hans C. Hansen, MD and Vijay Singh, MD

Healthcare decisions are increasingly being made on research-based evidence, rather than on expert opinion or clinical experience alone. Consequently, the process by which the strength of scientific evidence is evaluated and developed by means of evidence-based medicine recommendations and guidelines has become crucial resulting in the past decade in unprecedented interest in evidence-based medicine and clinical practice guidelines.

Systematic reviews, also known as evidence-based technology assessments, attempt to minimize bias by the comprehensiveness and reproducibility of the search for and selection of articles for review. Evidence-based medicine is defined as the conscientious, explicit and judicious use of the current best evidence in making decisions about the care of individual patients. Thus, the practice of evidence-based medicine requires the integration of individual clinical expertise with the best available external evidence from systematic research. To arrive at evidence-based medicine decisions all valid and relevant evidence should be considered alongside randomized controlled trials, patient preferences and resources. However, many systematic reviews in interventional pain management fail to follow evidence-based medicine principles.

Clinical practice guidelines are systematically developed statements that assist clinicians, consumers and policy makers to make appropriate healthcare decisions. The complex processes of guideline development depend on integration of a number of activities, from collection and processing of scientific literature to evaluation of the evidence, development of evidence-based recommendations or guidelines and implementation and dissemination of the guidelines to relevant professionals and consumers. Guidelines are being designed to improve the quality of healthcare and decrease the use of unnecessary, ineffective or harmful interventions.

This review describes various aspects of evidence-based medicine, systematic reviews in interventional pain management, evaluation of the strength of scientific evidence, differences between systematic and narrative reviews, rating the quality of individual articles, grading the strength of the body of evidence and appropriate methods for searching for the evidence.

Keywords: Evidence-based medicine, interventional pain management, systematic reviews, narrative reviews, randomized controlled trials, observational trials, levels of evidence, quality of evidence

The past decade has been marked by unprecedented interest in evidence-based medicine and clinical practice guidelines. Healthcare decisions are increasingly being made on research-based evidence rather than on expert opinion or clinical experience alone. At the core of the evidence-based approach to clinical or public health issues is, inevitably, the evidence itself which needs to be carefully gathered and collated from a systematic literature review of the particular issues. Consequently, the processes by which the strength of scientific evidence is evaluated in the development of evidence-based medicine recommendations and guidelines is crucial. Re-

search-based evidence and evidence-based medicine are not synonymous. Evidence-based medicine recognizes that expert consensus is one level of evidence. Evidence-based medicine recognizes that the patients must be treated even if the highest level of evidence is not available to support any treatment option. However, quite often there is a tendency to inappropriately apply evidence-based medicine principles to the development of guidelines, especially when developing guidelines for interventional techniques.

Systematic reviews represent a rigorous method of compiling scientific evidence to answer questions regarding healthcare issues of treatment, diagnosis, or preventive services. Traditional opinion-based narrative reviews and systematic reviews differ in several ways. Systematic reviews also known as evidence-based technology assessments attempt to minimize bias by the comprehensiveness and reproducibility of the search for and selection of

articles for review. In contrast, narrative reviews are often broad in scope without all the safeguards to control against bias, even though they are similar to a systematic review. In addition, systematic reviews assess the methodological quality of the included studies, including the study design, methodology, and analysis, with an evaluation of the overall strength of that body of evidence. Thus, systematic reviews and technology assessments increasingly form the basis for evidence-based medicine, and the development of guidelines, and consequently for making individual and policy-level healthcare decisions.

This review will describe the role of evidence-based medicine in interventional pain management, various systems to rate the strength of scientific evidence and the importance of the application of appropriate systematic reviews and evidence-based principles in developing practice guidelines.

From Pain Management Center of Paducah, Paducah, KY; Texas Tech University, Lubbock TX; The Cleveland Clinic, Cleveland, OH; Medical Advanced Pain Specialists Pain Clinic Minneapolis, MN; The Pain Relief Centers, Conover, NC; and Pain Diagnostics Associates, Niagara, WI. Address Correspondence: Laxmaiah Manchikanti, MD, 2831 Lone Oak Road, Paducah, Kentucky 42003. E-mail: drm@apex.net

EVIDENCE-BASED MEDICINE

Since Hippocrates, clinicians have been primarily interested in making accurate diagnosis and selecting optimal treatments for the patients in their practice. However, they must also avoid harmful exposures and offer patients prognostic information. Thus, evidence-based medicine, also known as EBM is about solving clinical problems (1). The prolific growth of evidence-based medicine dates back to the late 1970s when a group of clinical epidemiologists at McMaster University, led by David Sackett, planned a series of articles advising clinicians on how to read clinical journals. This series was subsequently published in the *Canadian Medical Association Journal* beginning in 1981. The group proposed the term *critical appraisal* to describe the application of the basic rules of evidence presented in that series. These authors, with their extensive experience of teaching critical appraisal for a number of years, became increasingly aware of both the necessity and challenges of motivating clinicians to go beyond merely browsing the literature and, rather, to actually use the information in solving patient problems. The term suggested by David Sackett was *bringing critical appraisal to the bedside* to describe the process of the practical application of evidence from the medical literature to the patient care. This concept of bringing critical appraisal to the bedside had evolved into a philosophy of medical practice based on knowledge and understanding of the medical literature supporting each clinical decision at McMaster University. It was believed that this represented a fundamentally different style of practice warranting a formal term that would capture the difference. In 1990, Guyatt (2) suggested a new approach and coined the term *scientific medicine*. This bothered some implying that they were practicing unscientific medicine. Hence, the name was changed and, evidence-based medicine, was born.

The term *evidence-based medicine* first appeared in 1990 in an informational document intended for residents entering or considering application to the residency program at McMaster University. This term also subsequently appeared in print in the ACP Journal Club in 1991 (3). Based on the conclusions by the innovators at McMaster University that the concept of a new approach to medical practice would prove useful for the larger commu-

nity of medical educators, evidence-based medicine evolved. This process further developed by collaboration with a larger group of academic physicians, primarily from the United States, to form the first international evidence-based medicine working group. This working group expanded on the then existing description of evidence-based medicine describing this phenomenon as a paradigm shift (4). The journal of the American Medical Association (JAMA), subsequently published a 25-part series called "The Users Guide to the Medical Literature" between 1993 and 2000, which was published as a manual (2). From evidence-based medicine, other terms have developed to fit the various needs of the healthcare profession, some of which include: evidence-based healthcare, evidence-based practice, and evidence-based interventional pain management (5).

Evidence-based medicine means many things to many people. Currently, evidence-based medicine is defined as the conscientious, explicit and judicious use of the current best evidence in making decisions about the care of individual patients. Thus, evidence-based medicine is essentially what most clinicians have been trying to practice all their working lives. The practice of evidence-based medicine requires the integration of individual clinical expertise with the best available external evidence from systematic research. Decisions that affect the care of patients should be made with *due weight* accorded to *all* valid, relevant information. There are many other factors in addition to the results of randomized controlled trials, which may weigh heavily in both clinical and policy decisions, for example, patient preferences and resources. Valid, relevant evidence should be considered alongside these other factors in the decision-making process. Thus, no one sort of evidence should necessarily be the determining factor in decision-making. *All* implies that there should be an active search for all that is valid, relevant information and that an assessment should be made of the accuracy of information and the applicability of the evidence to the decision in question (6). Four basic contingencies originally defined evidence-based practice (7). First, recognition of the patient's problem and construction of a structured clinical question. Second, the ability to efficiently and effectively search the medical literature to retrieve the best available evidence to an-

swer the clinical question. Third, critical appraisal of the evidence. Fourth, integration of the evidence with all aspects of individual patient decision making to determine the best clinical care of the patient. Thus, evidence-based medicine is a loose term which has been used based not only on the necessity to present a particular view, but also based on personal philosophy, bias and conjecture. This has led to a multitude of questions as to whether evidence-based medicine is truly based on evidence.

In the 1990s, numerous guidelines were published in various countries around the world (8). Many professional organizations produced consensus guidelines, and the Cochrane collaboration of systematic reviews, which started in 1993, now has more than 3,000 collaborations worldwide (8, 9). In pain management, the first so-called evidence-based guidelines were produced by the Agency for Health Care Policy Research (AHCPR) in 1994 (10). AHCPR produced 15 guidelines at a cost of \$750,000,000, each at varying costs (11). The agency was eventually replaced with a small portion of its original budget and without the mandate to develop practice guidelines. AHCPR was renamed as The Agency for Healthcare Research and Quality (AHRQ). However, guideline development continued experiencing an explosive growth with numerous publications appearing in the form of consensus statements, clinical guidelines, and books (6, 8, 9, 12-30).

REVIEWS IN INTERVENTIONAL PAIN MANAGEMENT

Systematic reviews are a necessary part of evidence-based guidelines, but they are not intended to be clinical guidelines (31). A comprehensive text entitled *An Evidence-Based Resource for Pain Relief* edited by McQuay and Moore (6), with analysis of many techniques applied in the management of pain, has been largely ignored. There have been numerous systematic reviews, guidelines, policies, and practice parameters describing pain management, and in particular, interventional pain management (8, 12-30).

In one such review, Nelemans et al (19), reviewed multiple methods of treatment, reaching an inaccurate singular conclusion. They concluded that, "convincing evidence is lacking regarding the effects of injection therapy on low back pain." They vaguely described interven-

tional techniques such as *local injection therapy* stating that *local injection therapy* is a badly defined term. It is well understood and well known that no one in interventional pain management uses the term *local injection therapy*. They also reported that there was no evidence for lumbar facet joint syndrome. Manchikanti et al (32) contended that the validity of facet joint injections has been strongly documented by properly designed studies in the diagnosis of facet joint pain (13). Nelemans et al (19) described only one study of facet joint injections which also used placebo injections and which has been criticized extensively (33). Nelemans et al (19) also combined epidural injections with other studies of interventional techniques including disc injections, trigger point injections and facet joint injections. Thus, the review consisted of flawed criteria. The use of intradiscal injections, other than those for provocative discography, is not a common practice. In addition, the combination of all types of epidural injections into one category is also a major drawback. Epidural injections are administered by multiple routes, which include caudal, interlaminar, and transforaminal. They failed to understand the significant differences in techniques and outcomes among these approaches. The literature thus far has demonstrated that there is strong evidence for the efficacy of caudal injections and moderate evidence for transforaminal epidural injections when they are analyzed separately (8). Further analysis of the review by Nelemans et al (19) by Manchikanti et al (32) showed that four of the five studies involving caudal epidural steroid injections produced positive results, whereas five of seven studies on interlaminar lumbar epidural steroid injections produced negative results.

A second review pertains to a systematic review of randomized clinical trials evaluating the efficacy of radiofrequency procedures for the treatment of spinal pain (20). This review was similar to the review by Nelemans et al (19). Geurts et al (34) also have reached inaccurate conclusions. In addition, these false conclusions were supported in an editorial by Carr (32). Guerts et al (20) reviewed 6 total studies, two of which were dorsal root ganglion radiofrequency studies, and a third study was intraarticular facet denervation. Therefore, out of six, only three studies were relevant. They also failed to

include a meticulously performed study by Dreyfuss et al (35) in the analysis and review, because this study had no control group. Radiofrequency neurotomy of dorsal root ganglion is not a common procedure and has not been proven to be an effective modality for facet joint pain, whereas it is used, for segmentally radiating pain. Further, intraarticular radiofrequency, which is not an acceptable technique and has no physiologic or scientific basis for denervation as it should be performed on the medial branches, rather than the joint itself was also included inaccurately in this review. Apart from all the confusion with regards to the identification of the best evidence hypothesis, three of the three studies were positive for radiofrequency management of facet joint pain with neurolysis. Thus, this should have yielded moderate to strong evidence rather than their strongest conclusion as Carr (34) noted, "insufficient evidence supporting the effectiveness of most radiofrequency treatments for spinal pain." Bogduk (36) responding to the systematic review by Geurts et al (20) defended radiofrequency neurotomy and identified numerous deficiencies in the systematic review of Geurts et al (20) and also elaborated on the practical difficulties with randomized trials. Bogduk (36) described that the tenure of the review was unfortunately nihilistic, and he defended radiofrequency lest the articles be abused by organizations intent upon discrediting radiofrequency neurotomy. He described numerous mishaps of this review with evaluation of a study based on total number of patients, evaluation of only randomized controlled trials, and described the difficulties of obtaining a grant, either by academicians or practitioners in private practices.

Manchikanti et al (37) also evaluated the medial branch neurotomy in the management of chronic spinal pain. This review utilized inclusion/exclusion criteria, search strategy and followed key domains in rating quality of systematic reviews as described by The Agency for Healthcare Research and Quality (AHRQ). Based on the stringent criteria, after identifying seven randomized trials of radiofrequency neurotomy for spinal pain, they identified only four related to medial branch neurotomy. They included only two randomized trials for evidence synthesis and excluded two trials due to various deficiencies. They also considered multiple

observational studies and included four prospective evaluations and three retrospective evaluations in evidence synthesis. Based on two randomized evaluations, four prospective evaluations, and three retrospective evaluations, Manchikanti et al (37) concluded that combined evidence of radiofrequency neurotomy of medial branches provided strong evidence of short-term relief and moderate evidence of long-term relief of chronic spinal pain of facet joint origin. This evidence synthesis appeared to have been more in line with the practice patterns and based on actual evidence-based medicine.

A third review, consisting of multiple reviews in the form of a well recognized book by Natchemson and Jonsson (38), also inaccurately reached negative conclusions about diagnostic, as well as therapeutic interventional techniques, due to a lack of proper review of interventional techniques.

Fourthly, there have been multiple systematic reviews of the effectiveness of epidural steroid injections published. The first review by Kepes and Duncalf in 1985 (21) concluded that the rationale for epidural systemic steroids was not proven. However, Benzon in 1986 (22) utilizing the same studies, concluded that mechanical causes of low back pain, especially those accompanied by signs of nerve root irritation, may respond to epidural steroid injections. The difference in the conclusion of Kepes and Duncalf (21) and Benzon (22) may have been due to the fact that the earlier study included studies on systemic steroids, whereas, the later analysis was limited to studies on epidural steroid injections. The Australian National Health and Medical Research Council Advisory Committee on epidural steroid injections, by Bogduk et al (23) extensively studied caudal, interlaminar, and transforaminal epidural injections, including all the literature available at the time. They concluded that the balance of the published evidence supported the therapeutic use of caudal epidurals. However, they also concluded that the results of lumbar interlaminar epidural steroids strongly refuted the utility of epidural steroids in acute sciatica. Bogduk (24) in updated recommendations in 1999, opined against epidural steroids by the lumbar route as requiring too high a number necessary for treatment, but supported the potential usefulness of transforaminal steroids for disc prolapse. Koes et al (25) in 1995

reviewed 12 trials of lumbar and caudal epidural steroid injections and reported positive results from only 6 studies. This analysis showed that there were 5 studies for caudal epidural steroid injections and 7 studies for lumbar epidural steroid injections available at the time which were randomized. Four of the five studies involving caudal epidural steroid injections were positive, whereas 5 of 7 studies were negative for lumbar epidural steroid injections. Koes et al (25), based on this flawed analysis, concluded that the efficacy of epidural steroid injection has not yet been established and the benefits of epidural steroid injections, if any, seemed to be of short duration only. Koes et al (26) updated their review of epidural steroid injections for low back pain and sciatica, including three more studies with a total of 15 trials, which met the inclusion criteria. In this study, again, they concluded that of the 15 trials, 8 reported positive results of epidural steroid injections. Thus, their basic conclusions remained the same. However, the same flaws as in 1995 applied to this evaluation also. In both the studies, when caudal epidural steroid injections were separated from interlaminar epidural steroid injections, there was significant proof of effectiveness of epidural steroids. Watts and Silagy (27) in 1995 performed a meta-analysis of the available data and defined efficacy in terms of pain relief (at least 75% improvement) in the short-term (60 days) and in the long-term (1 year). They concluded that epidural steroid injections increased the odds ratio of pain relief to 2.61 in the short-term and to 1.87 in the long-term suggesting that epidural steroids were effective. van Tulder et al (39), in analyzing numerous treatments based on scientific evidence in conservative treatment of chronic low back pain, also analyzed seven studies of epidural steroid injections. Similar to the previous studies, they also concluded that there was conflicting evidence with inconsistent findings with regards to the effectiveness of epidural steroid injections. They also utilized the criteria by Koes et al (25, 26). McQuay and Moore (28) in 1998 reviewed the literature and concluded that epidural corticosteroid injections were effective for back pain and sciatica. They also emphasized the fact that even though epidural steroid injections can optimize conservative therapy and provide substantial pain relief for up to 12 weeks in patients with acute or subacute sciati-

ca, a few patients with chronic pain report complete relief. Consequently, the majority must return for repeat epidural injections. Bernstein (40) reviewed injections in surgical therapy in chronic spinal pain and concluded that there was limited evidence of effectiveness of interlaminar or caudal epidural steroid injections for sciatica with low back pain. In a pragmatic review of data provided by available systematic reviews and seminal controlled studies pertaining to the treatment of regional musculoskeletal problems, Curatolo and Bogduk (41) concluded that epidural steroids may offer limited, short-term benefit for sciatica. Vroomen et al (42) reviewed conservative treatment of sciatica with 19 randomized controlled trials, including epidural steroid injections, and concluded that epidural steroids may be beneficial for subgroups for nerve root compression. Rozenberg et al (43) in a systematic review of 13 trials of epidural steroid therapy concluded that five trials demonstrated greater pain relief within the first month in the steroid group as compared to the control group, whereas, eight trials found no measurable benefits. BenDebba et al (44) in a large multicenter study from various departments of neurosurgery and orthopedic surgeries, failed to include any physicians from interventional pain management departments.

Finally, Cepada et al (45) in a narrative and systematic review of therapeutic role of local anesthetic sympathetic blockade in complex regional pain syndrome included both randomized and observational studies. However, the authors failed to utilize appropriate search criteria, inclusion and exclusion criteria, and appropriate quality evaluation forms. This review showed full response in 29% of patients, partial response in 41% and absent response in 32% with sympathetic blockade in regional pain syndrome.

COCHRANE COLLABORATION-BASED REVIEWS

The Cochrane collaboration was founded in 1993 in order to provide "systematic, up-to-date reviews of all relevant randomized controlled trials (RCTs) of health care" (9). Cochrane reviews of the effects of health care interventions and procedures have increased, and they are published electronically in the Cochrane database of systematic reviews. Some of them are also published in peer-reviewed literature. Many of these reviews are not

just reviews of the relevant RCTs but are reviews of other reviews of the relevant literature (46). Along with other reviews, the Cochrane reviews are used as gold standards for determining the treatment effectiveness, specifically to deny care. Gatchel and McGeary (46) in a critical editorial about Cochrane collaboration-based reviews of health care interventions referred to various deficiencies and questioned the scientific validity of these reviews. Recently, multiple critical responses were elicited in response to some of the publications (47-52). Numerous Cochrane reviews were also published pertaining to pain management and spine care. Gatchel and McGeary (46) highlight some of the issues related to Cochrane reviews in spine care (54-58). As described earlier in this review, similar to most publish reviews, the Cochrane reviews also must be viewed with caution regarding the issues of impartiality of the conclusions. The most important troubling feature is that the primary authors of most of the reviews of interventional pain management techniques are either non-physicians or physicians without expertise in interventional techniques. Their expertise and credibility in this field is questioned. Further, as Mowatt et al (55) found, a considerable proportion of Cochrane reviews had strong evidence of either honorary or ghost authorship with very lax disclosure policy in conjunction with potential conflicts of interest. Thus, these reviews no doubt represent some major shortcomings, with potentially harmful health care implications for patients in the United States. This was elegantly pointed out by Gatchel and McGeary (46) in their editorial questioning the scientific validity of Cochrane reviews, which were described as simply nihilistic. They organized their critique around four primary points of concern, namely systematic reviews, meta-analyses, randomized controlled trials, and arbitrary rating criteria.

Furlan et al (59) described the quality of individual reviews of Cochrane collaboration reviews varied considerably. Cochrane reviews included often a number of low-quality trials and combined them with better quality trials, which resulted in a heterogeneity of trials consequently resulting in inconclusive judgments. Such flaws in the basic methodology can lead to invalid, prejudiced and biased conclusions with potentially serious implications for the quality of patient

care. In some cases, the opposite is true as critical evidence is omitted, citing various reasons such as low quality or non-randomization. In addition, Cochrane reviews never included observational studies in interventional pain management, thus, not fitting into the concept of the true evidence-based medicine. Furlan et al (56) also reported that when they evaluated systematic reviews of the effectiveness of common interventions for chronic nonspecific low back pain, they found that interventions for which there was evidence in the form of multiple reviews, often yielded conflicting conclusions. In addition, similar systematic reviews of analgesic interventions conducted by others yielded discordance among reviews with similar quality ratings (59). Further, it was also shown that quality measures used in systematic reviews are not reliable in measuring treatment effect (57). It was also shown that use of a checklist in evaluating studies through a systematic review may actually damage the validity of the review by not considering certain aspects of the studies being analyzed (58). Hopayian (58) elegantly pointed out examples of systematic reviews in which a study with fatal flaws was not only included, but was also given a relatively high rating in three different systematic reviews of the same phenomenon. He cautioned that although the quality and rigor of systematic reviews continues to evolve, studies being evaluated should be considered from both a reviewer and a clinician viewpoint to ensure the validity of the analysis. Once again, this reinforces the importance of involving interventionalists with clinical background in evaluation of strength of studies in evidence synthesis.

In contrast to systematic reviews, meta-analyses involves both quantitative and qualitative interpretive features. This is required because it will take into account various factors that differ among the studies, such as sample sizes, strength of experimental methodology used and various issues of internal and external validity (60). Thus, meta-analysis requires subjective decisions by the implementers, resulting in the possibility that disparate researchers may interpret the same data differently (6, 46, 50). This phenomenon has been observed in interventional pain management quite frequently. For example, there have been several criticisms of Cochrane's review of Nelemans et al (19) and the review by Geurts

et al (20). Further, many concerns about the use of meta-analysis to draw conclusions about health care treatment without careful scrutiny through peer review have been raised (50). Questions also have been raised about the tools utilized by Cochrane collaboration (61). Senn (61) suggested that the authors may be unfairly biased against certain kinds of trials that do not fit what they are accustomed to analyzing. He noted that the Cochrane collaboration's favorite tool, RevMan, is able to analyze only single-center parallel-group trials with no covariates; therefore, their investigators are inclined to see any other sort of trial as problematic. Further, earlier versions of RevMan also could accommodate only dichotomous variables. This may be especially problematic in chronic spinal pain disorders research because of the complexities of biopsychosocial etiologies, assessment and treatments (46, 62, 63). RevMan is a free computer program readily available for preparing and submitting a Cochrane review.

The third issue is with regards to randomized controlled trials. One of the criteria that had to be met for a review to be included in the Cochrane review is the presence of at least one randomized or clinical controlled trial. This also has been blindly followed by other systematic reviewers. Sometimes, the entire reviews are of only randomized trials. Thus, there is no consistency at all in the literature, specifically, in the emerging field of interventional pain management. However, the rationale for this is puzzling. This phenomenon of randomized trial no longer exists in light of the fact that Concato et al (64) have shown that the popular belief that only randomized controlled trials will unequivocally produce trustworthy results, and that all observational studies may be misleading, is no longer true. Further, Concato et al (64) highlighted the fact that the results of a well-designed observational study does not systematically overestimate the magnitude of the effects of treatment, relative to those in RCTs on the same topic. Feinstein (65) also outlined several advantages of observational studies over randomized controlled trials, including lower costs, greater timeliness and a broader range of patients. Benson and Hartz (66) also concluded that there was little evidence that estimates of treatment effects in observational studies differ from those in randomized controlled trials.

Concato et al (64) searched 122 citations of meta-analyses, including 6 that examined both randomized, controlled trials and observational studies of the same clinical topic. They challenged the current consensus about hierarchy of study designs in clinical research. Contrary to the prevailing beliefs, the "average results" from well-designed observational studies did not systematically overestimate the magnitude of the associations between exposure and outcome as compared with the results of randomized, controlled trials of the same topic. Rather, the summary results of randomized, controlled trials and observational studies were remarkably similar for each clinical topic they examined. Viewed individually, the observational studies had less variability in point estimate with less heterogeneity of results than randomized, controlled trials on the same topic. In fact, only among randomized, controlled trials did some studies report results in a direction opposite that of the pulled point estimate, representing a paradoxical finding (64). Concato et al (64) concluded that even though their data were a challenge to accepted beliefs, the findings were consistent with three other types of available evidence. First, the previous investigations have shown that observational cohort studies can produce results similar to those of randomized, controlled trials when similar criteria were used to select study subjects. Second, data from non-medical research do not support a hierarchy of research designs. Finally, the findings that there is substantial variation in the results of randomized, controlled trials is consistent with prior evidence of contradictory results among randomized, controlled trials. Further, Concato et al (64) also stated that, there is evidence that observational studies can be designed with rigorous methods that mimic those of clinical trials. An analysis by McKee et al (67) of 18 randomized and observational studies in health-services research found that treatment effects may differ according to research design, but that "one method does not give a consistently greater effect than the other." McKee et al (67) also stated that the treatment effects were most similar when the exclusion criteria were similar and when the prognostic factors were accounted for in observational studies. Horwitz et al (68) described a specific method used to strengthen observational studies adapting

principles of the design of randomized, controlled trials to the design of an observational study. These principles include an identification of a “zero time” for determining a patient’s eligibility and baseline features, use of inclusion and exclusion criteria similar to those of clinical trials, adjustment for differences in baseline susceptibility to the outcome, and use of statistical methods (eg, intention-to-treat analysis) similar to those of randomized controlled trials. When these procedures were used in a cohort study evaluating the benefit of beta blockers after recovery from myocardial infarction, the use of restricted cohort produced results consistent with corresponding findings from beta blocker heart attack trial (68, 69).

Finally, Concato et al (64) also described that data in the literature of other scientific disciplines supporting their contention that research design should not be considered a rigid hierarchy. They described that a comprehensive research on various psychological, educational, and behavioral treatments (70) identified a 302 meta-analyses and examined the reports on the basis of several features, including research design. Results were presented from the 74 meta-analyses that included studies with randomized and observational designs. To allow for comparisons among various topics with different outcome variables, effect size was used as a unit-free measure of the effect of the intervention. The observational designs did not consistently overestimate or underestimate the effect of treatment. Manchikanti and Pampati (71) evaluated the concept of randomization and research designs in interventional pain management. They examined the concept if randomization does provide the protective statistical shield that some think it provides in an interventional pain management population. In this study, they compared randomized and non-randomized samples. Randomization was accomplished by the use of random number tables and random sampling, whereas, non-randomization was achieved by allocation into various groups by two different means. The results of this evaluation showed that there was only one significant difference when patients were allocated by means of non-randomization among the groups or compared to the total sample. In contrast, randomization showed significant differences in seven parameters. They concluded that based on the results of this study

in interventional pain management settings, non-randomized sampling is valid. Based on these results, it appears that in interventional pain management settings, non-randomized type of evaluation may not only be valid, but also may be superior as far as sampling is concerned. Horwitz (72), in a review of 200 randomized controlled trials on 36 clinical topics, found numerous examples of conflicting results.

Benson and Hartz (66) reviewed 136 reports about 19 diverse treatments, such as calcium-channel-block therapy for coronary artery disease, appendectomy, and interventions for subfertility. In most cases, the estimates of the treatment effects from observational studies and randomized, controlled trials were similar. In only 2 of the 19 analyses of treatment effects did the combined magnitude of the effect in the observational studies lie outside the 95% confidence interval for the combined magnitude in the randomized, controlled trials.

Multiple randomized trials have shown diverse results recently (73-77). However, even though observational studies may generally give valid results, there are known limitations. Thus, it is of paramount importance to evaluate methodology used in both randomized controlled trials, as well as observational studies to integrate the outcome of specific study in the evidence synthesis (78). While randomization is considered as the gold standard of clinical studies to remove confounding variables that might otherwise weaken the usefulness of a particular study, this can also be achieved by blinding the clinician and subject by removing the problems of preconceived notions of subjects or investigations from systematically introducing bias into the outcomes. The difficulties in conducting a scientifically valid and clinically useful RCT include ethical concerns, difficulty in randomization, blinding issues, and patient recruitment issues in the United States (79-87). The use of placebos as controlled treatment groups has been common and was believed to facilitate the attainment of new clinical knowledge. With the revision in 2000 of the declaration of Helsinki by the World Medical Association (84), placebo groups have fallen into disfavor among some countries, specifically in the United States. This revision makes it unethical to use placebos in research if there is a known treatment or intervention for that particular disease or illness being

studied. If there is no known treatment, placebos are acceptable groups in randomized controlled trials. Thus, the issues of placebos remains unsettled in the United States (85). Patients in the United States if given a placebo or a knowingly inactive treatment may feel anger that time was wasted at their expense, and some conditions will actually deteriorate, resulting in potential harm to the subject. This seriously violates the ethical concern of the right to treatment. Freedman (82) and Levine (88) have reviewed significant bioethical concerns associated with placebo-control groups. In addition, because of ethical concerns, it has been stated that it is difficult for randomized controlled trials conducted elsewhere to be replicated in the United States, thereby eliminating the external validity of any conclusions derived from these many European studies (46). Gatchel and McGeary (46) concluded that Cochrane’s studies done in countries with different social and medical systems would be expected to bias the attitudes and methodological approaches of investigators. Consequently, an important question is whether studies conducted by non-US investigators would be relevant in the US health care system. Fortunately, there are a host of other experimental designs that may be appropriately employed to yield important scientific data to help in delineating cause-effect relationships as are meant to be used for true evidence-based medicine approaches. As Manchikanti and Pampati (71) have shown, these designs may even be superior to randomized designs and provide more homogeneity in the patient population.

Finally, Gatchel and McGeary (46) described arbitrary rating criteria used in the Cochrane reviews. Apparently many reviewers have shown that arbitrary rating criteria is not useful by any means and lacks validity and reliability. Gatchel and McGeary (46) concluded that it is a disservice to the health care community (provider and patient community) if one reviews a series of reviews/articles that are quite heterogeneous in quality and concludes that the overall evidence for effectiveness is “inconclusive.” This is simply providing a misguided conclusion. In today’s age of financially motivated attempts by managed care administrators and so-called independent clinicians without a need for accountability to deny quality care to patients, such faulty

reviews provide “ammunition” to these companies by enabling them to cite that “the Cochrane review” justifies the denial of potentially important services (46). These authors (46) recommend to:

- ◆ Carefully scrutinize Cochrane reports as any scientific literature
- ◆ Carefully assess authors’ credentials and disclosures to identify whether their interpretations are unbiased
- ◆ Carefully evaluate their prior track record, explicit or implicit agenda, honorary or ghost authorship
- ◆ Carefully evaluate potential limitations in external validity when comparing studies conducted in different countries that have different social and medical/health care systems
- ◆ Not to accept the argument that an RCT is the only research methodology available to produce scientifically acceptable outcome result
- ◆ Remember that interpretation of results from any study, regardless of research methodology employed, is only inferential process
- ◆ Finally, remember, the statement “unequivocal results or conclusions” can rarely be made in the scientific literature of clinical outcomes research.

These recommendations may also be applied to other reviews.

CLINICAL PRACTICE GUIDELINES

Clinical practice guidelines are systematically developed statements that assist clinicians, consumers and policy makers to make appropriate healthcare decisions. Such guidelines present statements of best “practice” based on a thorough evaluation of the evidence from published studies on the outcomes of treatment. The methods used for collecting and evaluating evidence and developing guidelines can be applied to a wide range of clinical interventions and disciplines, including interventional procedures, both diagnostic and therapeutic, pharmaceuticals and others.

The development of the guideline processes are complex. However, these mostly depend on integration of a number of activities, from collection and processing of scientific literature to evaluation of the evidence, development of evidence-based recommendations or guidelines, and implementation and dissemination of the guidelines to relevant professionals and consumers.

Guidelines are being designed to improve the quality of healthcare and decrease the use of unnecessary, ineffective or harmful interventions. In an era of evidence-based medicine, guidelines are becoming one of the critical links between the best available evidence and good clinical practice. Guidelines constitute one element of a systems approach to quality healthcare. Clinical practice guidelines are one component of good medical decision-making, which takes into account patient’s preferences and values, clinician’s values and experience, and the available resources. The guidelines’ main purpose is to achieve better health outcomes by improving the practice of health professionals and providing consumers with better information about treatment options. Guidelines can inform consumers about risk factors and how to avert them; they can be used to broaden the education of practitioners in the community, thus, contributing to quality assurance processes; and they can assist in the resolution of legal disputes and ethical dilemmas. Research has shown that clinical practice guidelines can be an effective means of changing the process of healthcare and improving health outcomes (89-92). Traditionally, guidelines have been based on consensus among experts. However, now it has been acknowledged that guideline recommendations should be based on systematic identification and synthesis of the best available scientific evidence (92). Considering the extensive research activity, the lack of a single source to identify the appropriate literature, significant bias in the systematic evaluations, and substantial reports outside the published and peer reviewed literature; identification and synthesis of the available evidence; and publication of this evidence in the form of guidelines can be a major undertaking. The National Health and Medical Research Council of Australia published: *A Guide to the Development, Implementation and the Evaluation of Clinical Practice Guidelines* in 1999 (92). This comprehensive document includes the principles, the development, the dissemination and implementation of guidelines.

Nine basic principles described for development guidelines are as follows (92):

1. Focus on outcomes. Outcome measures can range from survival rates to quality-of-life attributes.

2. Best available evidence. Evidence is graded according to its quality, relevance and strength.
3. Appropriate systems to synthesize the available evidence. Turning the evidence into a clinically useful recommendation depends on the judgment, experience and good sense of the authors of guidelines. **The fact of having evidence from a high-level study does not automatically result in a good clinical recommendation.**
4. Multidisciplinary process of development.
5. Flexibility and adaptability
6. Cost effectiveness of treatments
7. Appropriate dissemination
8. Evaluation of implementation and impact of guidelines
9. Appropriate revision of the guidelines on a regular basis

Legal considerations and potential liability of practitioners is an important aspect of guidelines. Many practitioners are concerned about their potential legal liability if a patient does not receive treatment as specified in clinical practice guidelines. It is possible that guidelines could be produced as evidence of what constitutes reasonable conduct by an interventional pain management practitioner. It is generally believed that following the guidelines provides a measure of protection. However, physicians should provide all appropriate information about all types of treatments, along with associated risks of any treatment, especially risks that may influence the patient’s decision. Patients should be provided with as much information as they seek, and in a form that is appropriate to their culture and level of education. Finally, all the patients should be encouraged to make their own decisions. The potential for any guidelines to be used as evidence in a court of law depends on the process used to develop them, the extent to which they are evidence-based, the degree of consensus about them, and whether they are up to date (92). However, guideline developers are unlikely to be held liable for any negative consequences of the implementation of guidelines. In general, guidelines should be summaries of the evidence, should have an expiration date, should not be unduly prescriptive, and should acknowledge areas where there is disagreement.

EVALUATION OF THE STRENGTH OF SCIENTIFIC EVIDENCE

Shaneyfelt et al (93) reviewed the methodological quality of clinical guidelines in the peer-reviewed medical literature, with evaluation of 279 guidelines developed by 69 different organizations and published from 1985 to 1997. They showed that mean overall adherence to standards by each guideline was 43.1%. They concluded that guidelines on the peer-reviewed medical literature during the past decade did not adhere well to published methodological standards. They also added that while all areas of guideline development need improvement, the greatest improvement is needed in the identification, evaluation, and synthesis of the scientific evidence. Almost all systematic evaluations in interventional pain management included only randomized controlled trials. This is in contrast to the definition of evidence-based medicine, which explicitly states that no one sort of evidence should necessarily be the determining factor in decision-making. Further, evidence-based medicine also emphasizes *all* implies that there should be an active search for all that is valid, relevant information and that assessment should be made of the accuracy of information and the applicability of the evidence to the decision in question. Recent systematic analyses have increasingly utilized observational studies, as well as other types of evidence (94, 95) even though it has not been applied to interventional pain management. It is also recognized that, meta-analysis restricted to randomized clinical trials is usually preferred to meta-analysis of observational studies (96-98). The number of published meta-analysis of observational studies in healthcare has increased substantially with 678 in 1955 to 1992, to more than 400 in 1996 alone (95). However, this has not been demonstrated in interventional pain management. In many situations, randomized designs are not feasible, and only data from observational studies are available (99).

The acme of clinical research is the randomized, double blind, controlled trials, but such trials must be undertaken responsibly and are extremely difficult to conduct in interventional pain management. Randomized controlled trials were introduced into clinical medicine when streptomycin was evaluated in the treat-

ment of tuberculosis (100). Since then, randomized controlled trials have become the gold standard for assessing the effectiveness of therapeutic agents (101-103). Sacks et al (104) compared published randomized controlled studies with those that used observational designs. In this landmark evaluation, they showed that the agent being tested was considered effective in 44 of 56 trials (79%) in observational studies utilizing historical controls, whereas the agent was considered positive only in 10 of 50 (20%) randomized controlled trials. Thus, they concluded that bias in patient selection may irretrievably weigh the outcome of historically controlled trials in favor of new therapies in observational studies. However, a recent evaluation by Kjaergard and Als-Nielsen (105) evaluating the association between competing interests and authors' conclusions; epidemiological study of randomized clinical trials published in the *British Medical Journal* concluded that randomized clinical trials, significantly favored experimental interventions, if financially competing interests were declared. These conclusions were based on review of 159 trials from 12 medical specialties. They also concluded that other competing interests were not significantly associated with authors' conclusions. Similar conclusions were drawn in a study of trials of multiple myeloma, the authors' conclusion- that is, the authors' reported interpretation of the overall trial results - were more positive towards the benefit of experimental interventions in those trials that were funded by the pharmaceutical industry compared with trials that were funded by non-profit organizations (106). Many stumbling blocks, including the issues of ethics, feasibility, cost and reliability, insurmountable challenges to randomized, double-blind trials in interventional pain management have been discussed (58-86, 107-113).

The ability to assign subjects randomly to either experimental or controlled status is considered to be science that is unsurpassed. However, random assignment does not confer an absolute protection against bias. It simply reduces the likelihood that such bias has occurred. Because randomized controlled trials are complicated and difficult to conduct, they are usually restricted to very tightly targeted groups of patients. Often, the investigators are not actively concerned about how subjects are obtained and rely on random

allocation to distribute any differences equally across the two groups. As a result, randomized trials often trade internal validity (tightness of comparisons) for external validity (generalizability) (114). Hence, randomization does not provide the protective shield that some think. Further, many patients refuse to participate in the process with the belief that randomization always puts them in the control groups. Thus, it does not seem feasible to rely exclusively on randomized controlled trials for all, or even most, of the needed empirical data, linking outcome to the process of care (64). Generally, a difference in outcome between a treatment and a control group can be due to chance, confounding, or bias due to differences between the groups, differences in handling the groups; and the true effect of intervention. Confounding and bias are avoided in the design of a trial by randomization, single-blinding or double-blinding. Thus, randomization is considered as a cornerstone to avoid bias and to maintain similarity between treatment and control groups, influencing the eventual outcome. Randomization by the tossing of a coin (or any equivalent method) ensures that the physician running the trial is not consciously or unconsciously allocating the certain patients to a particular group. Without randomization, trials of surgical versus medical techniques are wide open to selection bias. It is assumed that low-risk cases are much more likely to be assigned to the operative group, leaving high-risk patients to be managed by the physicians. Assigning volunteers to the treatment group and those who do not volunteer to the control group is also likely to result in a biased comparison - volunteers will be quite different, in many respects, from patients who do not volunteer (115). The criticism also has been advanced against allocation and treatment or control groups based on alternate days, alternate numbers or another assigned preformed methodology. Even though, it is believed that randomization does ensure that the two groups will differ only by chance, it does not guarantee that in practice, the balance will be actually achieved through the randomization. In fact, Manchikanti et al (71), in evaluating the influence of randomization over other types of allocation research designs in interventional pain management, showed that there was only one significant difference when patients were allocated by

means of non-randomization among the groups or compared to the total sample, whereas, randomization showed significant differences in seven parameters evaluated. Multiple ethical issues applicable to randomization and placebo treatment in the United States are discussed in earlier sections (82-88).

Throughout the 1990s and into the 21st century, the Agency for Health Care Research and Quality (AHRQ) has been the foremost federal agency providing research support and policy guidance in health services research in the United States (116). AHRQ published Evidence Report/Technology Assessment; No. 47 entitled *Systems to Rate the Strength of Scientific Evidence* in 1999 (116). This comprehensive document includes methodology and results to systems for rating the quality of individual articles, as well as systems for grading the strength of a body of evidence. AHRQ (116) commissioned this document with the overarching goals of this project to describe systems to rate the strength of scientific evidence, including evaluating the quality of individual articles that make up a body of evidence on a specific scientific question in healthcare, and to provide some guidance as to “best practices” in this field today. “Methodological quality” has been defined as “the extent to which all aspects of a study’s design and conduct can be shown to protect against systematic bias, non-systematic bias, and inferential error (117). AHRQ (116) acknowledged that quality varied depending on the instrument used for its measurement.

The National Health and Medical Research Council of Australia considered scientific evidence to be at the core of evidence-based approach to clinical or public health issues (117). They emphasized that evidence needs to be carefully gathered and collated from a systematic literature review of a particular issue in question. They published a comprehensive document entitled *How to Use the Evidence; Assessment and Application of Scientific Evidence* in 2000. They conceded that the interpretation of this evidence and its use to frame appropriate guidelines or recommendations has been a major challenge for expert committees compiling clinical practice guidelines over the last few years as an evidence-based approach has been developed and trailed.

The National Coordinating Center for Health Technology Assessment of

the United Kingdom also published systematic reviews of trials and other studies edited by Sutton et al (118). The objectives of this review were to identify applications of systematic review and meta-analytical methods in health technology assessment; to promote further, appropriate use of such approaches in these areas of application; and to begin to identify priorities for further methodological developments in this field. Sutton et al (118) recommended that, for the most part, standard and widely agreed upon approaches should be followed. They also noted that it may be appropriate to provide greater latitude in the nature of studies potentially eligible for review, including non-randomized studies and the results of audit exercises. Sutton et al (118) also described extensively the methodology for meta-analysis, searching the literature and identifying primary studies, evaluating the study quality, applications of meta-analysis in other context and using other data types, extensions of meta-analytic methods, and recommendations for further research.

SYSTEMATIC VS NARRATIVE REVIEWS

Cook et al (119) in 1997 described a systematic review as a type of scientific investigation of the literature on a given topic in which the “subjects” are the articles being evaluated. In a systematic review, before a research team conducts a review, it develops a well-designed protocol that lists: 1) a focused study question, 2) a specific search strategy, including the databases to be searched, and how studies will be identified and selected for the re-

view according to inclusion and exclusion criteria, 3) the types of data to be abstracted from each article and 4) how the data will be synthesized, either as a text summary or as some type of quantitative aggregation or meta-analysis. In a sense, these steps are taken to protect the work against various forms of unintended bias in the identification, selection and use of published work in these reviews. In contrast, a narrative review is similar to a systematic review but without all the safeguards to control against bias (Table 1). The major difference between these two approaches to synthesizing the clinical or scientific literature is that a systematic review attempts to minimize bias by the comprehensiveness and reproducibility of the search for and selection of articles for review. Selection bias can arise in systematic reviews by the poor choice of articles that are reviewed, if the literature search is not broad enough or the reasons for inclusion and exclusion of articles are not clearly specified. The concern about study quality of systematic reviews has been described since early 1980s, with numerous studies providing the evidence that study quality is important when producing systematic reviews (120-124).

RATING THE QUALITY OF INDIVIDUAL ARTICLES

Multiple types of studies used for assessing clinical and public health interventions are described in Table 2, which include systematic reviews, experimental studies, randomized trials, observational studies, and diagnostic test studies

Table 1. Differences between narrative and systematic reviews

Core Feature	Narrative Review	Systematic Review
Study Question	Often broad in scope.	Often a focused clinical question.
Data sources and search strategy	Specifications of database searched and search strategy are not typically provided.	Comprehensive search of many databases as well as the so-called gray literature. Explicit search strategy provided.
Selection of articles for study	Not usually specified. If specified, potentially biased.	Criterion-based selection, uniformly applied.
Article review or appraisal	Variable, depending on who is conducting the review.	Rigorous critical appraisal, typically using a data extraction form.
Study quality	Usually not assessed. If assessed, may not use formal quality assessment.	Some assessment of quality is almost always included as part of the data extraction process.
Synthesis	Often a qualitative summary.	Quantitative or qualitative summary.
Inferences	Occasionally evidence-based.	Usually evidence-based.

Adapted and modified from Cook et al (119)

(125). AHRQ described important evaluation domains and elements for evaluating systems related to rating the quality of individual articles, including systematic reviews, randomized clinical trials, observational studies and diagnostic test studies (116). Table 3 shows the important domains and elements for systems to rate quality of individual articles (116).

Systematic Reviews

Authors of AHRQ document reviewed 20 systems concerned with systematic reviews or meta-analyses. These authors characterized one as a scale (126) and 10 as checklists (127-137). They considered several others as guidance documents (94, 125, 126). To arrive at a set of high-performing scales or checklists pertaining to systematic reviews, the authors of AHRQ (116) took account of 7 key domains as shown in Table 3: study question, search strategy, inclusion and exclusion criteria, data extraction, and funding or sponsorship. Only one checklist (130) addressed all 7 domains, whereas a second checklist, while addressing all 7 domains, merited only a “partial” score for a study question and study quality (131). In addition, two checklists (129, 135) and one scale (94) also addressed 6 of the 7 domains.

Table 4 shows domains and elements for systematic reviews with further descriptions of various domains and elements. Table 5 shows systematic review quality evaluation form.

Randomized Clinical Trials

In systematic reviews, randomized clinical trials are the most commonly utilized instruments. However, numerous variations have been described in these instruments. AHRQ in evaluating systems concerned

Table 2. Types of studies used for assessing clinical and public health interventions

Study design	Protocol
Systematic reviews	Systematic location, appraisal and synthesis of evidence from scientific studies
Experimental studies	
Randomized controlled trial	Subjects are randomly allocated to groups for either the intervention/treatment being studied or control/placebo (using a random mechanism, such as coin toss, random number table, or computer-generated random numbers) and the outcomes are compared.
Pseudorandomized controlled trial	Subjects are allocated to groups for intervention/treatment or control/placebo using a nonrandom method (such as alternate allocation, allocation by days of the week, or odd-even study numbers) and the outcomes are compared.
Clustered randomized trial	Groups of subjects are randomized to intervention or control groups (eg, community intervention trials).
Comparative (nonrandomized and observational) studies	
Concurrent control or cohort	Outcomes are compared for a group receiving the treatment/intervention being studied, concurrently with control subjects receiving the comparison treatment/intervention (eg, usual or no care).
Case-control	Subjects with the outcome or disease and an appropriate group of controls without the outcome or disease are selected and information is obtained about the previous exposure to the treatment/intervention or other factor being studied.
Historical control	Outcomes for a prospectively collected group of subjects exposed to the new treatment/intervention are compared with either a previously published series or previously treated subjects at the same institutions.
Interrupted time series	Trends in the outcome or disease are compared over multiple time points before and after the introduction of the treatment/intervention or other factor being studied.
Other observational studies	
Case series	A single group of subjects are exposed to the treatment/intervention.
-- post-test	Only outcomes after the intervention are recorded in the case series, so no comparisons can be made.
-- pretest/post-test	Outcomes are measured in subjects before and after exposure to the treatment/intervention for comparison (also called a ‘before-and-after’ study).

Adapted and modified from NHMRC (125)

Table 3. Important domains and elements for systems to rate quality of individual articles

Systematic Reviews	Randomized Clinical Trials	Observational Studies	Diagnostic Test Studies
<i>1. Study question</i>	Study question	<i>Study question</i>	<i>Study population</i>
<i>2. Search strategy</i>	Study population	<i>Study population</i>	<i>Adequate description of test</i>
<i>3. Inclusion and exclusion criteria</i>	Randomization	<i>Comparability of subjects</i>	<i>Appropriate reference standard</i>
<i>4. Interventions</i>	Blinding	<i>Exposure or intervention</i>	<i>Blinded comparison of test and reference</i>
<i>5. Outcomes</i>	Interventions	<i>Outcome measurement</i>	<i>Avoidance of verification bias</i>
<i>6. Data extraction</i>	Outcomes	<i>Statistical analysis</i>	
<i>7. Study quality and validity</i>	Statistical analysis	<i>Results</i>	
<i>8. Data synthesis and analysis</i>	Results	<i>Discussion</i>	
<i>9. Results</i>	Discussion	<i>Funding or sponsorship</i>	
<i>10. Discussion</i>	Funding or sponsorship		
<i>11. Funding or sponsorship</i>			

* Key domains in italics

Adapted from AHRQ (116)

Table 4. Domains and elements for systematic reviews

Domain	Elements*
Study Question	• Question clearly specified and appropriate
Search Strategy	• <i>Sufficiently comprehensive and rigorous with attention to possible publication biases</i>
	• <i>Search restrictions justified (eg, language or country of origin)</i>
	• 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 reviews
	• 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 AHRQ (116)

Table 5. Systematic Review Quality Evaluation Form

Study	Author(s): _____ Journal: _____ Other: _____ _____	
1. Study Question	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Clearly focused and appropriate Question		
2. Search Strategy	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Sufficiently comprehensive and rigorous with attention to possible publication biases		
- Search restrictions justified (eg, language and country of origin)		
- Sufficiently detailed to reproduce study		
3. Inclusion and Exclusion Criteria	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Selection methods specified and appropriate, with a priori criteria specified if possible		
4. Interventions	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Intervention(s) clearly detailed for all study groups		
5. Outcomes	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- All potentially important harms and benefits considered		
6. Data Extraction	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Rigor and consistency of process		
- Number and types		
- Blinding of reviewers		
- Measure of agreement or reproducibility		
- Extraction of clearly defined interventions/exposures and outcomes for all relevant subjects and subgroups		
7. Study Quality/Validity	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Assessment method specified and appropriate		
- Method of incorporation specified and appropriate		
8. Data Synthesis and Analysis	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Appropriate use of qualitative and/or quantitative synthesis, with consideration of robustness of results and heterogeneity issues		
- Presentation of key primary study elements sufficient for critical appraisal and replication		
9. Results	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Narrative summary and/or quantitative summary statistic and measure of precision, as appropriate		
10. Discussion	<input type="checkbox"/> Yes	<input type="checkbox"/> No
- Conclusions supported by results with possible biases and limitations taken into consideration		
11. Funding/Sponsorship	<input type="checkbox"/> Yes	<input type="checkbox"/> No
i. Government	_____	
ii. Agency	_____	
iii. Manufacturer	_____	
iv. Insurer(s)	_____	
- Type and level of support		

Adapted from AHRQ (116)

with randomized clinical trials reviewed 20 scales (122, 138-156), 11 checklists (117, 125-135, 137, 157-164), one component evaluation and 7 guidance documents (117, 134, 165-170). The authors of AHRQ designated a set of high-performing scales or checklists pertaining to randomized clinical trials by assessing their coverage of the following 7 domains as shown in Table 3; study pop-

ulation, randomization, blinding, interventions, outcomes, statistical analysis and funding or sponsorship. Following this extensive evaluation, authors of the AHRQ document concluded that only 8 systems for randomized clinical trials represented acceptable approaches that could be used today without major modifications (122, 137, 138, 140, 149, 151, 153, 159). Of the extensive review and

multiple systems evaluated, only 2 systems fully addressed all 7 domains (122, 159) and 6 of them addressed multiple domains except the funding domain (122, 137, 140, 149, 151, 153). Among these, two were described as rigorously developed (151, 153). However, the significance of this description of rigorous development is not yet known. Table 6 gives the description of various domains and

Table 6. Domains and elements for randomized controlled trials

Domain	Elements*
Study Question	<ul style="list-style-type: none"> • Clearly focused and appropriate question
Study Population	<ul style="list-style-type: none"> • Description of study population • Specific inclusion and exclusion criteria • Sample size justification
Randomization	<ul style="list-style-type: none"> • <i>Adequate approach to sequence generation</i> • Adequate concealment method used • <i>Similarity of groups at baseline</i>
Blinding	<ul style="list-style-type: none"> • Double-blinding (eg, of investigators, caregivers, subjects, assessors, and other key study personnel as appropriate) to treatment allocation
Interventions	<ul style="list-style-type: none"> • Intervention(s) clearly detailed for all study groups (eg, dose, route, timing for drugs, and details sufficient for assessment and reproducibility for other types of interventions) • Compliance with intervention • Equal treatment of groups except for intervention
Outcomes	<ul style="list-style-type: none"> • Primary and secondary outcome measures specified • Assessment method standard, valid, and reliable
Statistical Analysis	<ul style="list-style-type: none"> • Appropriate analytic techniques that address study withdrawals, loss to follow-up, missing data, and intention to treat • Power calculation • Assessment of confounding • Assessment of heterogeneity, if applicable
Results	<ul style="list-style-type: none"> • Measure of effect for outcomes and appropriate measure of precision • Proportion of eligible subjects recruited into study and followed up at each assessment
Discussion	<ul style="list-style-type: none"> • Conclusions supported by results with possible biases and limitations taken into consideration
Funding or Sponsorship	<ul style="list-style-type: none"> • 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 AHRQ (116)

Table 7. Randomized Control Trial Quality Evaluation Form

Study Author(s): _____ Journal: _____ Other: _____ _____	
1. Study Question <input type="checkbox"/> Yes <input type="checkbox"/> No - Clearly focused and appropriate Question - Medial branch neurotomy-pain relief and/or functional improvement at least at 2 months	
2. Study Population <input type="checkbox"/> Yes <input type="checkbox"/> No - Description of study population - Specific inclusion and exclusion criteria - Appropriate diagnostic criteria - Sample size justification	
3. Randomization <input type="checkbox"/> Yes <input type="checkbox"/> No - Adequate approach to sequence generation - Adequate concealment method used - Similarity of groups at baseline	
4. Binding <input type="checkbox"/> Yes <input type="checkbox"/> No - Double-blinding to treatment allocation	
5. Interventions <input type="checkbox"/> Yes <input type="checkbox"/> No - Intervention(s) clearly detailed for all study groups - Compliance with intervention - Equal treatment of groups except for intervention	
6. Outcomes <input type="checkbox"/> Yes <input type="checkbox"/> No - Primary/secondary outcome measures specified - Assessment method standard, valid and reliable	
7. Statistical Analysis <input type="checkbox"/> Yes <input type="checkbox"/> No - Appropriate analytic techniques that address study withdrawals, loss to follow-up, missing data, and intention to treat - Power calculation - Assessment of confounding - Method of handling withdrawals, losses to follow up and missing data - Assessment of heterogeneity, if applicable	
8. Results <input type="checkbox"/> Yes <input type="checkbox"/> No - Measure of effect for outcomes and appropriate measure of precision - Proportion of eligible subjects recruited into study and followed up at each assessment	
9. Discussion <input type="checkbox"/> Yes <input type="checkbox"/> No - Conclusions supported by results with possible biases and limitations taken into consideration	
10. Funding/Sponsorship <input type="checkbox"/> Yes <input type="checkbox"/> No i. Government _____ ii. Agency _____ iii. Manufacturer _____ iv. Insurer(s) _____ - Type and level of support	

Adapted from AHRQ (116)

elements for randomized controlled trials whereas Table 7 describes randomized controlled trial quality evaluation form.

Observational Studies

In the true spirit of evidence-based medicine, AHRQ recognizes the importance of observational studies. Thus, they considered numerous systems concerning

the evaluation of observational studies. The authors utilized multiple described scales (145, 146, 153, 171), 8 checklists (125, 135-137, 159, 161, 163, 171), and multiple guidance documents (96-99, 116). Authors of AHRQ considered several key domains and arrived at a set of 5 high-performing scales or checklists pertaining to observational studies as described in Table 3: comparabili-

ty of subjects, exposure or intervention, outcome measurement, statistical analysis and funding or sponsorship. Apparently these systems that cover these domains represent acceptable approaches for assessing the quality of observational studies. Table 8 describes domains and elements for observational studies in detail. Table 9 shows observational trial quality evaluation form.

Table 8. Domains and elements for observational studies

Domain	Elements*
Study Question	• Clearly focused and appropriate question
Study Population	• Description of study populations • Sample size justification
Comparability of Subjects†	For all observational studies: • Specific inclusion/exclusion criteria for all groups • Criteria applied equally to all groups • Comparability of groups at baseline with regard to disease status and prognostic factors • Study groups comparable to non-participants with regard to confounding factors • <i>Use of concurrent controls</i> • Comparability of follow-up among groups at each assessment Additional criteria for case-control studies: • Explicit case definition • Case ascertainment not influenced by exposure status • Controls similar to cases except without condition of interest and with equal opportunity for exposure
Exposure or Intervention	• Clear definition of exposure • Measurement method standard, valid and reliable • Exposure measured equally in all study groups
Outcome Measurement	• Primary/secondary outcomes clearly defined • Outcomes assessed blind to exposure or intervention status • Method of outcome assessment standard, valid and reliable • Length of follow-up adequate for question
Statistical Analysis	• Statistical tests appropriate • Multiple comparisons taken into consideration • Modeling and multivariate techniques appropriate • Power calculation provided • Assessment of confounding • Dose-response assessment, if appropriate
Results	• Measure of effect for outcomes and appropriate measure of precision • Adequacy of follow-up for each study group
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 AHRQ (116)

Table 9. Observational Study Quality Evaluation Form

Study	
Author(s):	_____
Journal:	_____
Other:	_____
1. Study Question <input type="checkbox"/> Yes <input type="checkbox"/> No - Clearly focused and appropriate question - Medial branch neurectomy- pain relief and/or functional improvement at least at 2 months	
2. Study Population <input type="checkbox"/> Yes <input type="checkbox"/> No - Description of study population - Appropriate diagnostic criteria - Sample size justification - Specific inclusion/exclusion criteria	
3. Comparability of Subjects <input type="checkbox"/> Yes <input type="checkbox"/> No	
4. For All Observational Studies <input type="checkbox"/> Yes <input type="checkbox"/> No - Criteria applied equally to all groups - Comparability of groups at baseline with regard to disease status and prognostic factors - Study groups comparable to non-participants with regard to confounding factors - <i>Use of concurrent controls</i> - Comparability of follow-up among groups of each assessment	
5. Additional criteria for case-control studies <input type="checkbox"/> Yes <input type="checkbox"/> No - Explicit case definition - Case ascertainment not influenced by exposure status - Controls similar to cases except without condition of interest and with equal opportunity for exposure	
6. Exposure/Intervention <input type="checkbox"/> Yes <input type="checkbox"/> No - Clear definition of exposure - Measurement method standard, valid and reliable - Exposure measured equally in all study groups	
7. Outcome Measurement <input type="checkbox"/> Yes <input type="checkbox"/> No - Primary/secondary outcomes clearly defined - Outcomes assessed blind to exposure or intervention status - Method of outcome assessment standard, valid and reliable - Length of follow-up adequate for question	
8. Statistical Analysis <input type="checkbox"/> Yes <input type="checkbox"/> No - Statistical tests appropriate - Multiple comparisons taken into consideration - Modeling and multivariate techniques appropriate - Power calculation provided - Assessment of confounding - Dose-response assessment, if appropriate	
9. Results <input type="checkbox"/> Yes <input type="checkbox"/> No - Measure of effect for outcomes and appropriate measure of precision - Adequacy of follow-up for each study group	
10. Discussion <input type="checkbox"/> Yes <input type="checkbox"/> No - Conclusions supported by results with possible biases and limitations taken into consideration	
11. Funding/Sponsorship <input type="checkbox"/> Yes <input type="checkbox"/> No i. Government _____ ii. Agency _____ iii. Manufacturer _____ iv. Insurer(s) _____ - Type and level of support	

Adapted from AHRQ (116)

Studies of Diagnostic Tests

Multiple precision diagnostic blocks utilized in interventional pain management have never been reviewed systematically. However, the value and valid-

ity of multiple diagnostic interventions with precision diagnostic blocks has been described extensively and also has been questioned repeatedly (8, 13, 173-215). AHRQ Assessment identified 6 checklists

to evaluate the quality of diagnostic studies (125, 135, 137, 213, 214, 216, 217). Following this extensive review, the authors identified 5 key domains for making judgments about the quality of diagnostic test

Table 10. Domains and elements for diagnostic studies

Domain	Elements*
Study Population	<ul style="list-style-type: none"> • <i>Subjects similar to populations in which the test would be used and with a similar spectrum of disease</i>
Adequate Description of Test	<ul style="list-style-type: none"> • <i>Details of test and its administration sufficient to allow for replication of study</i>
Appropriate Reference Standard	<ul style="list-style-type: none"> • <i>Appropriate reference standard ("gold standard") used for comparison</i>
	<ul style="list-style-type: none"> • Reference standard reproducible
Blinded Comparison of Test and Reference	<ul style="list-style-type: none"> • Evaluation of test without knowledge of disease status, if possible
	<ul style="list-style-type: none"> • <i>Independent, blind interpretation of test and reference</i>
Avoidance of Verification Bias	<ul style="list-style-type: none"> • <i>Decision to perform reference standard not dependent on results of test under study</i>

*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.
Adapted from AHRQ (116)

Table 11. Diagnostic Study Quality Evaluation Form

1. Study	
Author(s):	_____
Journal:	_____
Other:	_____
2. Study Population <input type="checkbox"/> Yes <input type="checkbox"/> No	
- <i>Subjects similar to populations in which the test would be used and with a similar spectrum of disease</i>	
3. Adequate Description of Test <input type="checkbox"/> Yes <input type="checkbox"/> No	
- <i>Details of test and its administration sufficient to allow for replication of study</i>	
4. Appropriate Reference Standard <input type="checkbox"/> Yes <input type="checkbox"/> No	
- <i>Appropriate reference standard ("gold standard") used for comparison</i>	
- Reference standard reproducible	
5. Blinded Comparison of Test <input type="checkbox"/> Yes <input type="checkbox"/> No	
- Evaluation of test without knowledge of disease status, if possible	
- <i>Independent, blind interpretation of test and reference</i>	
6. Avoidance of Verification Bias <input type="checkbox"/> Yes <input type="checkbox"/> No	
- <i>Decision to perform reference standard not dependent on results of test under study</i>	

Adapted from AHRQ (116)

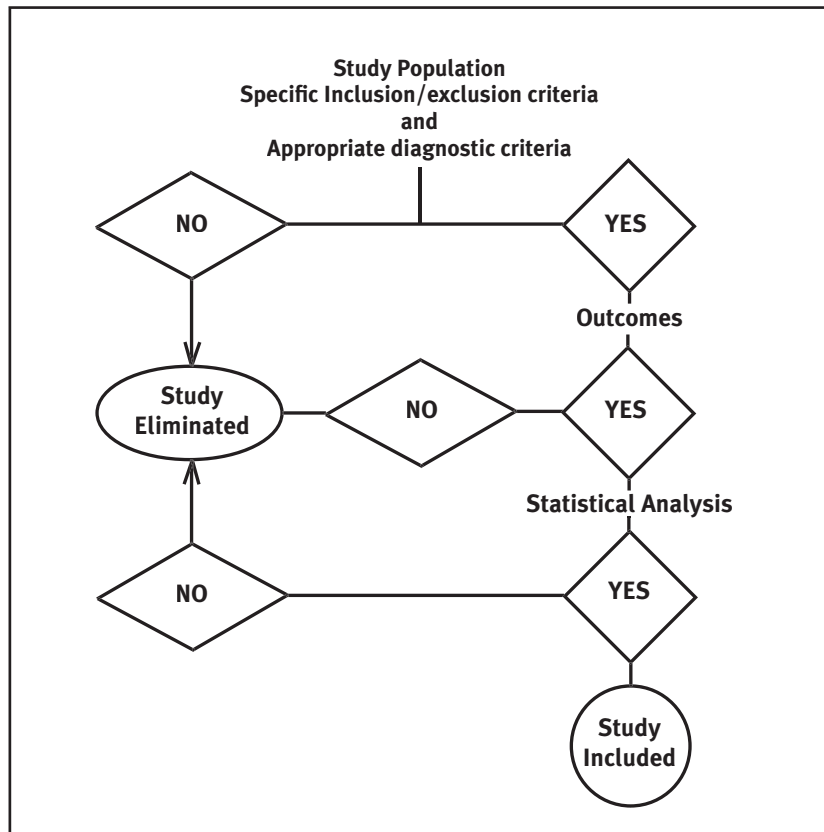
reports as described in Table 3: study population, adequate description of the test, appropriate reference standard, blinded comparison of test and reference and avoidance of verification bias. They described that 3 of the 6 checklists met all 5 criteria (125, 216, 217). Table 10 illustrates multiple domains and elements for diagnostic studies, whereas Table 11 illustrates the diagnostic study quality evaluation form.

Based on these extensive evaluations and guidance in evidence synthesis, a comprehensive study evaluation was developed as illustrated in Table 12.

Evaluation in Interventional Pain Management

Numerous types of systematic evaluations have been utilized in interventional pain management. Koes et al (25) in reviewing the efficacy of epidural steroid injections for low back pain and sciatica utilized the criteria list for the methodological assessment of randomized clinical trials of epidural steroid injection therapy for low back pain as shown in Table 13. However, in their selection of studies, they utilized on a MEDLINE literature search from 1966 to 1993 with inclusion of randomized clinical trials.

Table 12. Study evaluation (inclusion/exclusion) algorithm



Nelemans et al (19) performed a systematic evaluation of injection therapy for subacute and chronic benign low back pain utilizing the methodological quality criteria as described in Table 14 based on

Table 13. Criteria list for the methodological assessment of randomized clinical trials of epidural steroid injection therapy for low back pain

Criterion	Weight
<i>Study population:</i>	
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 to follow-up	2
<10% loss to follow-up	2
F >50 subjects in the smallest group	8
>100 subjects in the smallest group	9
<i>Interventions:</i>	
G Interventions included in protocol and described	10
H Pragmatic study	5
I Co-interventions avoided	5
J Placebo-controlled	5
<i>Effect:</i>	
K Patients blinded	5
L Outcome measures relevant	10
M Blinded outcome assessments	10
N Follow-up period adequate	5
<i>Data-presentation and analysis:</i>	
O Intention-to-treat analysis	5
P Frequencies of most important outcomes presented for each treatment group	5

Adapted from Koes et al (25)

Table 14. Methodological quality criteria

<ul style="list-style-type: none"> • A - selection and restriction (4 points) • B - treatment allocation (15 points) • C - study size (12 points) • D - prognostic comparability (10 points) • E - dropouts (12 points) • F - loss to follow-up assessment (10 points) • G - description of intervention (5 points) • H - extra treatments (2 points) • I - blinding of patients (4 points) • J - blinding of physician (4 points) • K - blinding of observer (4 points) • L - outcome measures (5 points) • M - timing of outcome measurements (6 points) • N - side effects (2 points) • O - analysis and presentation of data (5 points)

Adapted from Nelemans et al (19) and Riet et al (218)

a criteria list of Riet et al (218) in 1990. Ezo et al (219) in evaluating the effectiveness of acupuncture for the treatment of chronic pain utilized criteria developed by Jadad et al (156).

Table 15. Criteria list for assessment of the quality of randomized controlled trials in acupuncture and chronic pain

Item ^a	Score
A. Was the study described as randomized?	0/1
B. Was the randomization scheme described and appropriate?	0/1
C. Was the study described as double blind?	0/1
D. Was the method of double blinding appropriate?	0/1
D1. Were patients reported as blinded?	0/1
D2. Was the outcomes assessor reported as blinded?	0/1
E. Was there a description of dropouts and withdrawals?	0/1
F. Were cointerventions avoided or controlled for?	0/1
G. Was compliance satisfactory?	0/1
H. Was the study population adequately homogenous?	0/1
I. Was the therapeutic time equivalent between groups?	0/1

^a Scoring for Jadad scale A + B + C + D + E = 5 possible points; 0-2, low quality; 3-5 high quality. Items D1, D2, F, G, H, I are included for sensitivity analysis. Coding 1 = yes, 0 = no
Adapted from Jadad et al (156) and Ezo et al (219)

GRADING THE STRENGTH OF BODY OF EVIDENCE

Systems for grading the strength of a body of evidence are much less uniform and inconsistent than those for rating study quality (116). As with the quality rating systems, selecting among the evidence grading systems will depend on the reason for measuring evidence strength, the type of studies that are being summarized, and the structure of the review panel. Some systems are extremely cumbersome to use, requiring substantial resources, whereas others are incomplete and incomprehensive. Multiple systems have been utilized in preparation of guidelines. Table 16 shows evidence dimensions-definitions. Tables 17-20 illustrate multiple levels of evidence utilized at the present time. Table 21 shows the designation of levels of evidence from level I through V considered in interventional pain management with guideline preparation. It was developed by with modification of various publications.

The National Health and Medical Research Council (NHMRC) described five key points for considering levels of evidence as follows (125):

- ◆ Resolution of differences in the conclusions reached about effectiveness from studies at differing levels of evidence or within a given level of evidence
- ◆ Resolution of the discrepancies is an important task in the compilation of an evidence summary.
- ◆ Inclusion of biostatistical and epidemiological advice on how to search for possible explanation for the disagreements before data are rejected as being unsuit-

Table 16. Evidence dimensions - definitions

Type of evidence (dimension)	Definition
Strength of evidence	
Level	The study design used, as an indicator of the degree to which bias has been eliminated by design (see Table 20).
Quality	The methods used by investigators to minimize bias within a study design.
Statistical precision	The P-value or, alternatively, the precision of the estimate of the effect (as indicated by the confidence interval). It reflects the degree of certainty about the existence of a true effect.
Size of effect	The distance of the study estimate from the 'null' value and the inclusion of only clinically important effects in the confidence interval.
Relevance of evidence	The usefulness of the evidence in clinical practice, particularly the appropriateness of the outcome measures used.

Adapted from NHMRC (125)

Table 17. *Classifying the relevance of evidence*

Ranking	Relevance of the Evidence
1	Evidence of an effect on patient-relevant clinical outcomes, including benefits and harms, and quality of life and survival
2	Evidence of an effect on a surrogate outcome that has been shown to be predictive of patient-relevant outcomes for the same intervention
3	Evidence of an effect on proven surrogate outcomes but for a different intervention
4	Evidence of an effect on proven surrogate outcomes but for a different intervention and population
5	Evidence confined to unproven surrogate outcomes

Adapted from NHMRC(125)

Table 18. *An example of a hierarchy of evidence*

I	Well-designed randomized controlled trials
<i>Other types of trial:</i>	
II-1a	Well-designed controlled trial with pseudo-randomization
II-1b	Well-designed controlled trials with no randomization
<i>Cohort studies:</i>	
II-2a	Well-designed cohort (prospective study) with concurrent controls
II-2b	Well-designed cohort (prospective study) with historical controls
II-2c	Well-designed cohort (retrospective study) with concurrent controls
II-3	Well-designed case-control (retrospective) study
III	Large differences from comparisons between times and/or places with and without intervention (in some circumstances these may be equivalent to level II or I)
IV	Opinions of respected authorities based on clinical experience; descriptive studies and reports of expert committees

Source: Sutton et al (118)

Table 19. *Designation of levels of evidence*

I	Evidence obtained from a systematic review of all relevant randomized controlled trials
II	Evidence obtained from at least one properly designed randomized controlled trial
III-1	Evidence obtained from well-designed pseudo-randomized controlled trials (alternate allocation or some other method)
III-2	Evidence obtained from comparative studies with concurrent controls and allocation not randomized (cohort studies), case-control studies, or interrupted time series with a control group
III-3	Evidence obtained from comparative studies with historical control, two or more single-arm studies, or interrupted time series without a parallel control group
IV	Evidence obtained from case series, either post-test or pre-test and post-test

Adapted from NHMRC (125)

Table 20. *Designation of levels of evidence*

Levels of evidence	Study design
I	Evidence obtained from a systematic review of all relevant randomized controlled trials
II	Evidence obtained from at least one properly designed randomized controlled trial
III-1	Evidence obtained from well-designed pseudo-randomized controlled trials (alternate allocation or some other method)
III-2	Evidence obtained from comparative studies (including systematic reviews of such studies) with concurrent controls and allocation not randomized (cohort studies), case-control studies, or interrupted time series with a control group
III-3	Evidence obtained from comparative studies with historical control, two or more single-arm studies, or interrupted time series without a parallel control group
IV	Evidence obtained from case series, either post-test or pre-test/post-test

Source: NHMRC (125)

able basis on which to make recommendations.

- ◆ Recognition of the fact that it may not be feasible to undertake randomized controlled trials in all situations. Guidelines should be used on the best available evidence.
- ◆ Recognition of the fact that it may be necessary to use evidence from different

study designs for different aspects of the treatment effect.

SEARCHING FOR THE EVIDENCE

To achieve balance in evidence-based interventional pain management and also to include all types of evidence, *all* types of evidence must be literally included. These include not only systematic

reviews and randomized clinical trials but also all published literature of observational studies and diagnostic test studies. Cook et al (220) presented a list of possible sources of literature that could be included in a systematic review. These are listed in Table 22. Thus, a search strategy should include all sources easily available to obtain the literature: 1) Index Medi-

Table 21. Designation of levels of evidence

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 of appropriate size (with at least 60 patients in smallest group); or research-based evidence from multiple properly designed studies of smaller size; or at least one randomized trial, supplemented by predominantly positive prospective and/or retrospective evidence.
Level III	Moderate: Evidence from a well-designed small randomized trial or evidence from well-designed trials without randomization, or quasi-randomized studies, single group, pre-post cohort, time series, or matched case-controlled studies or positive evidence from at least one meta-analysis.
Level IV	Limited: Evidence from well-designed nonexperimental studies from more than one center or research group.
Level V	Indeterminate: Opinions of respected authorities, based on clinical evidence, descriptive studies, or reports of expert committees.

cus, 2) EMBASE, 3) all peer-reviewed but non-indexed journals, 4) scientific meeting proceedings, 5) scientific newsletters, 6) cross-references from articles and 7) cross-references from reviews. Other sources as described in Table 22 may be utilized if feasible. The reviewer(s) should establish inclusion and exclusion criteria for each article. It has been shown that using only MEDLINE, 30% to 80% of all known published randomized controlled trials are identifiable, depending on the area or specific question (118, 221). Sutton et al (118) in systematic reviews of trials and other studies, described that non-English language references are under-represented in MEDLINE and only published articles are included (222). Thus, there is the potential for publication bias and language bias (221-224). Further, it was shown that depending on the country of origin, there is also potential for geographical biases (225). Another problem with databases is that even though many

Table 22. Possible sources of primary studies for inclusion in a systematic review

• Trial (research) registries
• Computerized bibliographic databases of published and unpublished research
• Review articles
• Published and unpublished research
• Abstracts
• Conference/symposia proceedings
• Dissertations
• Books
• Expert informants
• Granting agencies
• Industry
• Journal handsearching

Adopted and modified from Cook et al (220)

of the studies may be in a database such as MEDLINE, it may not be easy to identify all those which are relevant (226). Dickersin et al (227) investigated the problem of MEDLINE and showed that MEDLINE failed to find 44% of known trials. Possible reasons for poor retrieval are as follows: the search used was too narrow, the indexing of studies in MEDLINE is inadequate and the original reports may have been too vague. The same issues are applicable to EMBASE. In general, MEDLINE provides wide coverage of many English language journals. In contrast, EMBASE can be used to increase coverage of articles in the European languages. The overlap between the MEDLINE and EMBASE is approximately 34% (228), even though it can vary between 10% and 75% for specific topics (222). Thus, one cannot rely on searching a single database. Further, dependence on databases also may miss many non-indexed journals, proceedings of the scientific meetings, and peer-reviewed articles from scientific newsletters. Search of the reference lists of articles found through databases may also identify further studies for consideration (225). In fact, the Cochrane handbook advises that reviewers should check the references of all relevant articles that are obtained. Thus, additional potentially relevant, articles that are identified should be retrieved and assessed for possible inclusion in the review. The potential for reference bias or a tendency to potentially cite studies supporting one's own views, however, should be kept in mind when doing this type of search. This bias can be guarded against by using a multitude of search strategies. The idea of reference bias was originally described by Sackett (229). He found evidence of reference bias and also commented on many multiple publications of the same trials, another potential source of bias to be aware of when carrying out

a review, sometimes described as "salami science."

Keyword notes can be hand searched to check if the search has missed anything using the alternate method (118). Missing can occur due to poor indexing in electronic databases. Thus, hand searching carefully selected journals, may reveal a high percentage of relevant studies. Further, results may have been published in reports, book chapters, conference proceedings, technical reports, discussion papers, or other formats, which are not indexed on the main databases (229-231). Sutton et al (118) termed this literature as "grey literature." However, identifying such literature is not easy, even though some databases do exist, such as SIGLE (system for information on grey literature), NTIS (National Technical Information Service, DHSS-data, and the British reports, translations, and thesis, which is received by the BLDS (British Library Document Supplies Center). Further, one should also be cognizant of the fact that even if these materials are identified, obtaining them may be problem.

Conference proceedings may be easier to obtain than the other types of grey literature. These may be obtained directly from the sources or in England from various databases. Dickersin et al (221) compared the state of the art (hand search and MEDLINE) with only MEDLINE searches. They concluded that using MEDLINE only, omitted half of the relevant studies. Clarke (232) also highlighted that search of MEDLINE was insufficient. Adams et al (233) also summarized further investigations into searching using MEDLINE, and concluded that between 20% to 60% of the randomized controlled trials are missed by skilled MEDLINE searches when compared to hand searching or using trial registers. Jadad and McQuay (234) in investigating the pain literature, reported that MEDLINE was the

most time efficient, as it identified 87% of known trials with 52% precision. They also noted that MEDLINE search took only one tenth of the time, that of hand searching. Kleinjen and Knipschild (235) investigated the effectiveness of computed database searches using MEDLINE and EMBASE exploring 3 subjects. Their conclusion was that the number of articles found with computer searches depended very much on the subject at hand, and that the better methodological studies were found on the whole in the electronic databases.

The main point in the searching strategy is that there is no one single search strategy that would provide adequate results. Further, in performing reviews, researchers should maintain a healthy degree of skepticism about any or all of their searches.

CONCLUSION

At the core of evidence-based approach to clinical or public health issues is inevitably the evidence itself, which needs to be carefully gathered and collated from a systematic literature review of the particular issues. Systematic reviews and clinical practice guidelines are inter-related. A systematic review is a type of scientific investigation of the literature on a given topic in which the "subjects" are the articles being evaluated. In contrast, clinical practice guidelines are systematically developed statements that assist clinicians, consumers and policy makers to make appropriate healthcare decisions. The practice of evidence-based medicine requires the integration of individual clinical expertise with the best available external evidence from systematic research. All types of valid, relevant evidence should be considered alongside a multitude of other factors in the decision-making process.

The Agency for Health Care Research and Quality (AHRQ) along with a multitude of agencies around the world have developed systems relating to systematic reviews, meta-analysis, randomized trials, and non-randomized evaluations. We have discussed various aspects of evidence-based medicine, rating the quality of individual articles, grading the strength of body of evidence and searching for the evidence.

Author Affiliation:

Laxmaiah Manchikanti, MD, Medical Director, Pain Management Center of Paducah, 2831 Lone Oak Road, Paducah, KY 42003, or e-mail drm@apex.net.

James E. Heavner, DVM, PhD, Professor of Anesthesiology and Physiology, Texas Tech University Health Sciences Center, Lubbock, TX 79430-0001, or e-mail anejeh@ttuhsc.edu

Gabor B. Racz, MD, Grover Murray Professor, Professor and Chair Emeritus, Director Pain Services, Texas Tech University Health Sciences Center, Lubbock, TX 79430-0001, or e-mail aneplib@ttuhsc.edu

Nagy Mekhail, MD, Chairman, Department of Pain Management The Cleveland Clinic Foundation Cleveland, The Cleveland Clinic, 17350 Hawksview Lane, Chagrin Falls, OH 44023 or e-mail elsayeh@ccs.org.

David M. Schultz, MD, Medical Director, Medical Advanced Pain Specialists, Assistant Professor of Anesthesiology, University of Minnesota, 2104 Northdale Blvd NW, Suite 220, Minneapolis, MN 55433 or e-mail dschultz@painphysicians.com

Hans C. Hansen, MD, Medical Director, The Pain Relief Centers, PA, 3451 Greystone Place SW, Conover, NC 28613 or e-mail hans@hippocrates.org

REFERENCES

- Haynes RB, Sackett RB, Gray JMA et al. Transferring evidence from research into practice, 1: the role of clinical care research evidence in clinical decisions. *ACP J Club*, 1996; 125A:14-15.
- Guyatt GH, Rennie D. *User's Guides to the Medical Literature*. A manual for evidence-based clinical practice. AMA Press, Chicago, 2002.
- Guyatt GH. Evidence-based medicine. *ACP J Club* 1991; 114:A-16.
- Evidence-based medicine working group. Evidence-based medicine: a new approach to the teaching of medicine. *JAMA* 1992; 268:2420-2425.
- Manchikanti L. Evidence-based interventional pain medicine: Is there any evidence? *Pain Physician* 2002; 5:1-7.
- McQuay H, Moore A (eds). *An evidence-based resource for pain relief*. Oxford University Press, New York, 1998.
- Hatala R, Guyatt G. Evaluating the teaching of evidence-based medicine. *JAMA* 2002; 288:1110-1112.
- Manchikanti L, Singh V, Kloth D et al. Interventional techniques in the management of chronic pain: Part 2.0. *Pain Physician* 2001; 4:24-96.
- www.Cochrane.org.
- Bigos SJ, Boyer OR, Braen GR et al. Acute Low Back Problems in Adults. *Clinical Practice Guideline Number 4*. AHCPR Publication No. 95-0642. Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, Rockville, December 1994.
- Gonzalez EG, Materson (eds). *The nonsurgical management of acute low back pain*. Demos Vermane, New York 1997.
- Spitzer WO, Leblanc FE, Dupuis M (eds). Quebec Task Force on Spinal Disorders. Scientific approach to the assessment and management of activity-related spinal disorders: A monograph for clinicians. *Spine* 1987; (suppl) 12:1-59.
- Bogduk N. International Spinal Injection Society guidelines for the performance of spinal injection procedures. Part 1: Zygapophyseal joint blocks. *Clin J Pain* 1997; 13:285-302.
- Campbell JK, Penzien DB, Wall EM. Evidence-based guidelines for migraine headache: *Behavioral and physical treatments*. 2000. Available at: http://www.aan.com/public/practiceguidelines/headache_gl.htm.
- American Pain Society. Guideline for the management of acute and chronic pain in sickle-cell disease. American Pain Society, Glenview, 1999.
- American Geriatrics Society. The management of chronic pain in older persons: New guidelines from the American Geriatrics Society. *J Am Geriatr Soc* 1998; 46: 128-150.
- Sanders SH, Rucker KS, Anderson KO et al. Clinical practice guidelines for chronic non-malignant pain syndrome patients. *J Back Musc Rehabil* 1995; 5:115-120.
- Sanders SH, Harden RN, Benson SE et al. Clinical practice guidelines for chronic non-malignant pain syndrome patients II: An evidence-based approach. *J Back Musc Rehabil* 1999; 13:47-58.
- Nelemans PJ, Debie RA, DeVet HC et al. Injection therapy for subacute and chronic benign low back pain. *Spine* 2001; 26:501-515.
- Geurts JW, van Wijk RM, Stolker RJ et al. 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.
- Kepes ER, Duncalf D. Treatment of backache with spinal injections of local anesthetics, spinal and systemic steroids. *Pain* 1985; 22:33-47.
- Benzon HT. Epidural steroid injections for low back pain and lumbosacral radiculopathy. *Pain* 1986; 24:277.

23. Bogduk N, Christophidis N, Cherry D et al. Epidural use of steroids in the management of back pain. *Report of working party on epidural use of steroids in the management of back pain*. National Health and Medical Research Council. Canberra, Commonwealth of Australia, 1994, pp 1-76.
24. Bogduk N. Epidural steroids for low back pain and sciatica. *Pain Digest* 1999; 9: 226-227.
25. Koes BW, Scholten RJPM, Mens JMA 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.
26. Koes BW, Scholten R, Mens JMA et al. Epidural steroid injections for low back pain and sciatica. An updated systematic review of randomized clinical trials. *Pain Digest* 1999; 9:241-247.
27. 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.
28. McQuay HJ, Moore RA. Epidural corticosteroids for sciatica. In *An Evidence-Based Resource for Pain Relief*. Oxford University Press, Oxford, 1998; pp 216-218.
29. Spitzer WO, Skovron ML, Salmi LR et al. Scientific monograph of the Quebec Task Force on whiplash-associated disorders: Redefining "whiplash" and its management. *Spine* 1995; (Suppl) 20:1-73.
30. Practice guidelines for chronic pain management. A report by the American Society of Anesthesiologists Task Force on Pain Management, Chronic Pain Section. *Anesthesiology* 1997; 86:995-1004.
31. van Tulder M, Malmivarira A, Esmail R et al. Letter to the editor: In reply. *Spine* 2001; 26:840-844.
32. Manchikanti L, Jasper J, Singh V. Cochrane review is mixing apples and oranges (ltr). *Spine* 2001; 26:2641-2643.
33. Carrette S, Marcoux S, Truchon R et al. A controlled trial of corticosteroid injections into facet joints for chronic low back pain. *N Engl J Med* 1991; 325:1002-1007.
34. Carr DB, Goudas LC. Evidence-based pain medicine: The good, the bad, and the ugly. *Reg Anesth Pain Med* 2001; 26:389-393.
35. Dreyfuss P, Halbrook B, Pauza K et al. Efficacy and validity of radiofrequency neurotomy for chronic lumbar zygapophysial joint pain. *Spine* 2000; 25:1270-1277.
36. Bogduk N. In defense of radiofrequency neurotomy. *Reg Anesth Pain Med* 2002; 27:439-447.
37. Manchikanti L, Singh V, Vilims B et al. Medial branch neurotomy in management of chronic spinal pain: Systematic review of the evidence. *Pain Physician* 2002; 5:405-418.
38. Nachemson AL, Jonsson E (eds). *Neck and back pain. The scientific evidence of causes, diagnosis, and treatment*. Lippincott Williams & Wilkins, Philadelphia, 2000.
39. van Tulder MW, Goossens M, Waddell G et al. Conservative treatment of chronic low back pain. In Nachemson AL, Jonsson E (eds). *Neck and Back Pain*. Lippincott Williams & Wilkins, Philadelphia, 2000, pp 271-304.
40. Bernstein RM. Injections and surgical therapy in chronic pain. *Clin J Pain* 2001; 17:S94-S104.
41. Curatolo M, Bogduk N. Pharmacologic pain treatment of musculoskeletal disorders: Current perspectives and future prospects. *Clin J Pain* 2001; 17:25-32.
42. Vroomen PC, De Krom MC, Slofstra PD et al. Conservative treatment of sciatica: A systematic review. *J Spin Disord* 2000; 13: 463-469.
43. Rozenberg S, Dubourg G, Khalifa P et al. Efficacy of epidural steroids in low back pain and sciatica: A critical appraisal by a French task force of randomized trials. *Revue du Rhumatisme* 1999; 66:79-85.
44. BenDebba M, Torgerson W, Boyd R et al. Persistent low back pain and sciatica in the United States: Treatment outcomes. *J Spinal Disord Tech* 2002; 15:2-15.
45. Cepeda MS, Lau J, Carr DB. Defining the therapeutic role of local anesthetic sympathetic blockade in complex regional pain syndrome: A narrative and systematic review. *Clin J Pain* 2002; 18:216-233.
46. Gatchel RJ, McGeary D. Cochrane collaboration-based reviews of health-care interventions: Are they unequivocal and valid scientifically, or simply nihilistic? *The Spine Journal* 2002; 2:315-319.
47. Olsen O, Gøtzsche PC. Cochrane Review on screening for breast cancer with mammography. *Lancet* 2001; 358:1340-1342.
48. Olsen O, Gøtzsche PC. Cochrane Review on screening for breast cancer with mammography. Cochrane Library, issue 4. Update Software, Oxford, 1991.
49. Horton R. Screening mammography: An overview revisited. *Lancet* 2001; 358: 1284-1285.
50. Lee JH, Zuckerman D. Correspondence: Screening for breast cancer with mammography. *Lancet* 2001; 358:2164-2165.
51. Miller AB. Correspondence: Screening for breast cancer with mammography. *Lancet* 2001; 358:2164.
52. Thornton H. Correspondence: Screening for breast cancer with mammography. *Lancet* 2001; 358:2165.
53. Duffy SW, Tabar L, Smith RA. Correspondence: Screening for breast cancer with mammography. *Lancet* 2001; 358:2166.
54. Hoving JL, Gross AR, Gasner D. A critical appraisal of review articles on the effectiveness of conservative treatment for neck pain. *Spine* 2001; 26:196-205.
55. Mowatt G, Shirran L, Grimshaw JM et al. Prevalence of honorary and ghost authorship in Cochrane Reviews. *JAMA* 2002; 287:2769-2771.
56. Furlan AD, Clarke J, Esmail R et al. A critical review of reviews on the treatment of chronic low back pain. *Spine* 2001; 26: E155-E162.
57. Balk EM, Bonis PA, Moskowitz H et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002; 287:2973-2982.
58. Hopayian K. The need for caution in interpreting high quality systematic reviews. *Br Med J* 2001; 323:681-684.
59. Jadad A, McQuay HS. Meta-analyses to evaluate analgesic interventions: A systematic qualitative review of their methodology. *J Clin Epidemiol* 1996; 49:235-243.
60. Gatchel RJ, Matt-Maddrey A. Experimental design issues in clinical research of musculoskeletal pain disabilities. *Crit Rev Phys Rehab Med* 2000; 12:9-101.
61. Senn S. Correspondence: Screening for breast cancer with mammography. *Lancet* 2001; 358:2164.
62. Gatchel RJ, Bell G. The biopsychosocial approach to spine care and research. *Spine* 2000; 25:2561-2571.
63. Turk DC, Gatchel RJ. Psychosocial factors and pain: Revolution and evolution. In Gatchel RJ, Turk DC (eds). *Psychological Factors in Pain: Critical Perspectives*. Guilford, New York, 1999:481-494.
64. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; 342:1887-1892.
65. Feinstein AR. Epidemiologic analyses of causation: the unlearned scientific lessons of randomized trials. *J Clin Epidemiol* 1989; 42:481-489.
66. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; 342: 1878-1886.
67. McKee M, Britton A, Black N et al. Methods in health services research: interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999; 319:312-315.
68. Horwitz RI, Viscoli CM, Clemens JD et al. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; 89:630-638.
69. A randomized trial of propranolol in patients with acute myocardial infarction. I. Mortality results. *JAMA* 1982; 247:1707-1714.
70. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *Am Psychol* 1993; 48:1181-1209.
71. Manchikanti L, Pampati V. Research designs in interventional pain management: Is randomization superior, desirable or essential? *Pain Physician* 2002; 5:275-284.
72. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med* 1987; 82:498-510.
73. Horn KD. Evolving strategies in the treat-

- ment of sepsis and systemic inflammatory response syndrome (SIRS). *QJM* 1998; 91: 265-277.
74. Angus DC, Birmingham MC, Balk RA et al. E5 murine monoclonal antiendotoxin antibody in gram-negative sepsis: A randomized controlled trial. *JAMA* 2000; 283: 1723-1730.
 75. LeLorier J, Grégoire G, Benhaddad A et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997; 337:536-542.
 76. Lefering R, Neugebauer E. Problems of Randomized Controlled Trials in Surgery. Nonrandomized Comparative Clinical Studies. In Able U, Koch A (eds). In *Proceedings of the International Conference on Nonrandomized Comparative Clinical Studies in the International conference on Nonrandomized Comparative Clinical Studies in Heidelberg*, Symposium Publishing, April 10-11, 1997. <http://www.symposium.com/nrccs/lefering.htm>.
 77. Bigby M. Challenges of the hierarchy of evidence: Does the emperor have no clothes? *Arch Dermatol* 2001; 137:345-346.
 78. Guyatt GH, Haynes RB, Jaeschke RZ et al. Users' guides to the medical literature, XXV. Evidence-based medicine: Principles for applying the users' guides to patient care. *JAMA* 2000; 284:1290-1296.
 79. Fung EK, Lore JM. Randomized controlled trials for evaluating surgical questions. *Arch Otolaryngol Head Neck Surg* 2002; 128:631-634.
 80. Schulz KF, Chalmers I, Haynes RJ et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412.
 81. Chalmers I. Unbiased, relevant and reliable assessments in health care. *BMJ* 1998; 317:1167-1168.
 82. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987; 317: 141-145.
 83. Burton v. Brooklyn Doctors Hosp., 452 N.Y.S. 2d. 875 (N.Y. app. Div. 1982).
 84. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. World Medical Association, Edinburgh, Scotland, October 2000.
 85. Vastag B. Helsinki discord? A controversial declaration. *JAMA* 2000; 284:2983-2985.
 86. Emanuel EJ, Miller FG. The ethics of placebo-controlled trials – A middle ground. *N Engl J Med* 2001; 345:915-919.
 87. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomized controlled trials. *Lancet* 1998; 351: 47-52.
 88. Levine RJ. The need to revise declaration of Helsinki. *N Engl J Med* 1999; 341:531-534.
 89. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: A systematic review of rigorous evaluation. *Lancet* 1993; 342:1317-1322.
 90. Effective Health Care Bulletin (EHCB) Implementing clinical practice guidelines. Can guidelines be used to improve practice? *Effective Health Care Bulletin*, No. 6, University of Leeds, UK 1994.
 91. Manchikanti L, Rivera J, Pampati V et al. Effectiveness of clinical guidelines in interventional pain management. *Pain Physician* 2002; 5:127-132.
 92. A guide to the development, implementation and evaluation of clinical practice guidelines. National Health and Medical Research Council, Canberra, Commonwealth of Australia, 1998, pp 1-79.
 93. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; 281:1900-1905.
 94. Stroup DF, Berlin JA, Morton SC et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA* 2000; 283:2008-2012.
 95. Stroup DF, Thacker SB, Olson CM et al. Characteristics of meta-analyses submitted to a medical journal. From: International Congress on Biomedical Peer Review and Global Communications; September 17-21, 1997; Prague, Czech Republic.
 96. Greenland S. Invited commentary. *Am J Epidemiol* 1994; 140:290-296.
 97. Lau J, Ioannidis JP, Schmid CH. Summing up evidence. *Lancet* 1998; 351:123-127.
 98. Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994; 140:771-778.
 99. Berlin JA. Invited commentary. *Am J Epidemiol* 1995; 142:383-387.
 100. Streptomycin treatment of pulmonary tuberculosis: A Medical Research Council investigation. *BMJ* 1948; 2:769-782.
 101. Byar DP, Simon RM, Friedewald WT et al. Randomized clinical trials: Perspectives on some recent ideas. *N Engl J Med* 1976; 295:74-80.
 102. Feinstein AR. Current problems and future challenges in randomized clinical trials. *Circulation* 1984; 70:767-774.
 103. Abel U, Koch A. The role of randomization in clinical studies: Myths and beliefs. *J Clin Epidemiol* 1999; 52:487-497.
 104. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72:233-240.
 105. Kjaergard LK, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. *BMJ* 2002; 325:1-4.
 106. Djulbegovic B, Lacey M, Cantor A et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000; 356: 635-638.
 107. Hopwood MB, Manning DC. Lumbar epidural steroid injections: Is a clinical trial necessary or appropriate? *Reg Anesth Pain Med* 1999; 24:5-7.
 108. Turner JA, Loeser JD, Bell KG. Spinal cord stimulation for chronic low back pain. A systematic literature synthesis. *Neurosurgery* 1995; 37:1088-1098.
 109. Weinstein JN. The tortoise and the hare. Is there a place in spine surgery for randomized trials? *Spine* 1999; 24:2548-2549.
 110. Winter RB. The prospective, randomized, controlled clinical trial in spine surgery. Fact or fiction? *Spine* 1999; 24:2550-2552.
 111. Fairbank J. Randomized controlled trials in the surgical management of spinal problems. *Spine* 1999; 24:2556-2561.
 112. Tosteson TD. Point of view. *Spine* 1999; 24:2562-2563.
 113. Carey TS. Randomized controlled trials in surgery. An essential component of scientific progress. *Spine* 1999; 23:2553-2555.
 114. Kane RL. Approaching the outcomes question. In Kane RL (ed). *Understanding Health Care Outcomes Research*. Aspen Publication, Gaithersburg, 1997, pp 1-15.
 115. Daly LE, Bourke GJ. Epidemiological and clinical research methods. In Daly LE, Bourke GJ (eds). *Interpretation and Uses of Medical Statistics*. Blackwell Science, 2000, Oxford, pp 143-201.
 116. Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment No. 47 University of North Carolina: Agency for Healthcare Research and Quality. AHRQ Publication No. 02-E016; April 2002.
 117. Lohr KN, Carey TS. Assessing 'best evidence': Issues in grading the quality of studies for systematic reviews. *Joint Commission J Qual Improvement* 1999; 25:470-479.
 118. Sutton AJ, Abrams KR, Jones DR et al. Systematic reviews of trials and other studies. *Health Technol Assess* 1998; 2: 1-276.
 119. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: Synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997; 126:376-380.
 120. Mulrow CD. The medical review article: State of the science. *Ann Intern Med* 1987; 106:485-488.
 121. Clark HD, Wells GA, Huet C et al. Assessing the quality of randomized trials: Reliability of the Jadad scale. *Control Clin Trials* 1999; 20:448-452.
 122. Chalmers TC, Smith H Jr., Blackburn B et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; 2:31-49.
 123. Hemminki E. Quality of reports of clinical trials submitted by the drug industry to the Finnish and Swedish control authorities. *Eur J Clin Pharmacol* 1981; 19:157-165.
 124. 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.
125. 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.
 126. Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. *JAMA* 1998; 279:1566-1570.
 127. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44:1271-1278.
 128. Oxman AD, Guyatt GH, Singer J et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; 44:91-98.
 129. Irwig L, Tosteson AN, Gatsonis C et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994; 120:667-676.
 130. Sacks HS, Reitman D, Pagano D et al. Meta-analysis: An update. *Mt Sinai J Med* 1996; 63:216-224.
 131. 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.
 132. Beck CT. Use of meta-analysis as a teaching strategy in nursing research courses. *J Nurs Educ* 1997; 36:87-90.
 133. Smith AF. An analysis of review articles published in four anesthesia journals. *Can J Anaesth* 1997; 44:405-409.
 134. Clarke M, Oxman AD. Cochrane Reviewer's Handbook 4.0. The Cochrane Collaboration; 1999.
 135. Khan KS, ter Riet G, Glanville J et al. Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews: York, England: University of York, NHS Centre for Reviews and Dissemination; 2000.
 136. New Zealand Guidelines Group. Tools for Guideline Development & Evaluation.
 137. Harbour R, Miller J. A new system [Scottish Intercollegiate Guidelines Network (SIGN)] for grading recommendations in evidence based guidelines. *BMJ* 2001; 323:334-336.
 138. de Vet HCW, de Bie RA, van der Heijden GJMG et al. Systematic review on the basis of methodological criteria. *Physiotherapy* 1997; 83:284-289.
 139. Evans M, Pollock AV. A score system for evaluating random control clinical trials of prophylaxis of abdominal surgical wound infection. *Br J Surg* 1985; 72:256-260.
 140. 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.
 141. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: *Medical Stat Med* 1989; 8:441-454.
 142. Gotsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal anti-inflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10:31-56.
 143. Kleijnen J, Knipschild P, ter Riet G. Clinical trials of homeopathy. *BMJ* 1991; 302:316-323.
 144. Detsky AS, Naylor CD, O'Rourke K et al. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992; 45:255-265.
 145. Cho MK, Bero LA. Instruments for assessing the quality of drug studies published in the medical literature. *JAMA* 1994; 272:101-104.
 146. Goodman SN, Berlin J, Fletcher SW et al. Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Ann Intern Med* 1994; 121:11-21.
 147. Fahey T, Hyde C, Milne R et al. The type and quality of randomized controlled trials (RCTs) published in UK public health journals. *J Public Health Med* 1995; 17:469-474.
 148. Khan KS, Daya S, Collins JA et al. 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.
 149. van der Heijden GJ, van der Windt DA, Kleijnen J et al. Steroid injections for shoulder disorders: A systematic review of randomized clinical trials. *Brit J Gen Pract* 1996; 46:309-316.
 150. Bender JS, Halpern SH, Thangaroopan M et al. Quality and retrieval of obstetrical anesthesia randomized controlled trials. *Can J Anaesth* 1997; 44:14-18.
 151. 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.
 152. van Tulder MW, 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.
 153. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. *J Epidemiol Community Health* 1998; 52:377-384.
 154. Moher D, Pham B, Jones A et al. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352:609-613.
 155. Turlik MA, Kushner D. Levels of evidence of articles in podiatric medical journals. *J Am Podiatr Med Assoc* 2000; 90:300-302.
 156. Jadad AR, Moore RA, Carroll D et al. Assessing the quality of randomized clinical trials: Is blinding necessary? *Controlled Clin Trials* 1996; 17:1-12.
 157. DerSimonian R, Charette LJ, McPeck B et al. Reporting on methods in clinical trials. *N Engl J Med* 1982; 306:1332-1337.
 158. Poynard T, Naveau S, Chaput J. Methodological quality of randomized clinical trials in treatment of portal hypertension. In *Methodology and Reviews of Clinical Trials in Portal Hypertension*. *Excerpta Medica* 1987:306-311.
 159. Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989; 84:815-827.
 160. Imperiale TF, McCullough AJ. Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials. *Ann Intern Med* 1990; 113:299-307.
 161. Spitzer WO, Lawrence V, Dales R et al. Links between passive smoking and disease: A best-evidence synthesis. A report of the Working Group on Passive Smoking. *Clin Invest Med* 1990; 13:17-46.
 162. Verhagen AP, de Vet HC, de Bier RA et al. The Delphi list: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; 51:1235-1241.
 163. Zaza S, Wright-De Augero LK et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000; 18:44-74.
 164. Schulz KF, Chalmers I, Hayes RJ et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412.
 165. Prendiville W, Elbourne D, Chalmers I. The effects of routine oxytocin administration in the management of the third stage of labour: An overview of the evidence from controlled trials. *Br J Obstet Gynaecol* 1988; 95:3-16.
 166. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA* 1994; 271:59-63.
 167. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1993; 270:2598-2601.
 168. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272:1926-1931.
 169. The Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. Checklist of information for inclusion in reports of clin-

- ical trials. *Ann Intern Med* 1996; 124:741-743.
170. Mohor D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; 285: 1987-1991.
 171. Corrao G, Bagnardi V, Zamboni A et al. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: A meta-analysis. *Addiction* 1999; 94:1551-1573.
 172. Hildebrandt J. Relevance of nerve blocks in treating and diagnosing low back pain – is the quality decisive? *Schmerz* 2001; 15:474-483.
 173. Bogduk N. Low back pain. In *Clinical Anatomy of the Lumbar Spine and Sacrum*, ed 3. Churchill Livingstone, New York, 1997; pp 187-214.
 174. Lord SM, Barnsley L, Wallis BJ et al. Chronic cervical zygapophysial joint pain with whiplash: A placebo-controlled prevalence study. *Spine* 1996; 21:1737-1745.
 175. Schwarzer AC, Aprill CN, Derby R et al. Clinical features of patients with pain stemming from the lumbar zygapophysial joints. Is the lumbar facet syndrome a clinical entity? *Spine* 1994; 19:1132-1137.
 176. Schwarzer AC, Wang S, Bogduk N et al. Prevalence and clinical features of lumbar zygapophysial joint pain: A study in an Australian population with chronic low back pain. *Am Rheum Dis* 1995; 54:100-106.
 177. Schwarzer AC, Aprill CN, Derby R et al. The relative contributions of the disc and zygapophysial joint in chronic low back pain. *Spine* 1994; 19:801-806.
 178. Manchikanti L, Pampati VS, Fellows B et al. Prevalence of lumbar facet joint pain in chronic low back pain. *Pain Physician* 1999; 2:59-64.
 179. Manchikanti L, Pampati V, Fellows B et al. The inability of the clinical picture to characterize pain from facet joints. *Pain Physician* 2000; 3:158-166.
 180. Manchikanti L, Singh V, Pampati VS et al. Evaluation of the relative contributions of various structures in chronic low back pain. *Pain Physician* 2001; 4:308-316.
 181. Manchikanti L, Pampati VS, Fellows B et al. The diagnostic validity and therapeutic value of medial branch blocks with or without adjuvants. *Curr Rev Pain* 2000 4: 337-344.
 182. Barnsley L, Lord SM, Wallis BJ et al. The prevalence of chronic cervical zygapophysial joint pain after whiplash. *Spine* 1995; 20:20-26.
 183. Manchikanti L, Singh V, Rivera J et al. Prevalence of cervical facet joint pain in chronic neck pain. *Pain Physician* 2002; 5:243-249.
 183. Manchikanti L, Singh V, Pampati VS et al. Evaluation of the prevalence of facet joint pain in chronic thoracic pain. *Pain Physician* 2002; 5: 354-359.
 184. Manchikanti L, Singh V, Pampati VS et al. Is there correlation of facet joint pain in lumbar and cervical spine? An evaluation of prevalence in combined chronic low back and neck pain. *Pain Physician* 2002; 5:365-371.
 185. Manchikanti L, Pampati VS, Fellows B et al. Influence of psychological factors on the ability to diagnose chronic low back pain of facet joint origin. *Pain Physician* 2001; 4:349-357.
 186. Schwarzer AC, Derby R, Aprill CN et al. The value of the provocation response in lumbar zygapophysial joint injections. *Clin J Pain* 1994; 10:309-313.
 187. Schwarzer AC, Aprill CN, Derby R et al. The false-positive rate of uncontrolled diagnostic blocks of the lumbar zygapophysial joints. *Pain* 1994; 58:195-200.
 188. Barnsley L, Lord S, Wallis B et al. False-positive rates of cervical zygapophysial joint blocks. *Clin J Pain* 1993; 9:124-130.
 189. Barnsley L, Bogduk N. Medial branch blocks are specific for the diagnosis of cervical zygapophysial joint pain. *Reg Anesth* 1993 18:343-350.
 190. Kaplan M, Dreyfuss P, Halbrook B et al. The ability of lumbar medial branch blocks to anesthetize the zygapophysial joint. *Spine* 1998; 23:1847-1852.
 191. Dreyfuss P, Schwarzer A, Lau P et al. Specificity of lumbar medial branch and L5 dorsal ramus blocks. *Spine* 1997; 22:895-902.
 192. Bogduk N. International Spinal Injection Society guidelines for the performance of spinal injection procedures. Part 1: Zygapophysial joint blocks. *Clin J Pain* 1997; 13:285-302.
 193. Manchikanti L, Singh V, Fellows B. Structural basis of chronic low back pain. In Manchikanti L, Slipman CW, Fellows B (eds), *Interventional Pain Management: Low Back Pain – Diagnosis and Treatment*. ASIPP Publishing, Paducah, KY 2002; 77-96.
 194. Schwarzer AC, Aprill CN, Bogduk N. The sacroiliac joint in chronic low back pain. *Spine* 1995; 20:31-37.
 195. Maigne J, Aivaliklis A, Pfefer F. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. *Spine* 1996; 21: 1889-1892.
 196. Fortin JD. Precision diagnostic disc injections. *Pain Physician* 2000; 3:271-288.
 197. Saal JS. General principles of diagnostic testing as related to painful lumbar spine disorders. *Spine* 2002; 27:2538-2545.
 198. Schellhas KP, Smith MD, Gundry CR et al. Cervical discogenic pain. Prospective correlation of magnetic resonance imaging and discography in asymptomatic subjects and pain sufferers. *Spine* 1996; 21: 300-312.
 199. Wood KB, Schellhas KP, Garvey TA et al. Thoracic discography in health individuals. A controlled prospective study of magnetic resonance imaging and discography in asymptomatic and symptomatic individuals. *Spine* 1999; 24:1548-1555.
 200. Walsh TR, Weinstein JN, Spratt KP et al. Lumbar discography in normal subjects. *J Bone Joint Surg* 1990; 72A:1081-1088.
 201. Holt EP Jr. The question of lumbar discography. *J Bone Joint Surg (Am)* 1968; 50: 720-726.
 202. Holt EP Jr. Fallacy of cervical discography. Report of 50 cases in normal subjects. *JAMA* 1964; 188:799-801.
 203. Carragee EJ. Is lumbar discography a determinate of discogenic low back pain: Provocative discography reconsidered. *Curr Rev Pain* 2000; 4:301-308.
 204. Carragee EJ, Chen Y, Tanner CM et al. Provocative discography in patients after limited lumbar discectomy. A randomized study of pain response in symptomatic and asymptomatic subjects. *Spine* 2000; 25:3065-3071.
 205. Carragee EJ, Tanner CM, Khurana S et al. The rates of false-positive lumbar discography in select patients without low back symptoms. *Spine* 2000; 25:1373-1381.
 206. Carragee EJ, Paragioudakis SJ, Khurana S. 2000 Volvo award winner in clinical studies. Lumbar high-intensity zone and discography in subjects without low back problems. *Spine* 2000; 25:2987-2992.
 207. Wetzel FT. Rates of false-positive lumbar discography: Point of view. *Spine* 2000; 25:1381.
 208. Block A, Vanharanta H, Ohnmeiss D et al. Discographic pain report: Influence of psychological factors. *Spine* 1996; 1:334-338.
 209. Ohnmeiss DD, Vanharanta H, Guyer RD. The association between pain drawings and computed tomographic/discographic pain responses. *Spine* 1995; 20:729-733.
 210. Heggeness MH, Watters WC, Gray PM. Discography of lumbar discs after surgical treatment for disc herniation. *Spine* 1997; 22:1606-1609.
 211. Manchikanti L, Singh V, Pampati VS et al. Provocative discography in low back pain patients with or without somatization disorder: A randomized, prospective evaluation. *Pain Physician* 2001; 4:227-239.
 212. Hildebrandt J. Relevance of nerve blocks in treating and diagnosing low back pain-- is the quality decisive? *Schmerz* 2001; 15: 474-483.
 213. Sheps SB, Schechter MT. The assessment of diagnostic tests. A survey of current medical research. *JAMA* 1984; 252:2418-2422.
 214. Arroll B, Schechter MT, Sheps SB. The assessment of diagnostic tests: A comparison of medical literature in 1982 and 1985. *J Gen Intern Med* 1988; 3:443-447.
 215. Carragee EJ, Alamin TF, Miller J et al. Provocative discography in volunteer subjects with mild persistent low back pain. *The Spine Journal* 2002; 2:25-34.

216. Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests. Recommended Methods; 1996.
217. Lijmer JG, Mol BW, Heisterkamp S et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061-1066.
218. ter Riet G, Kleijnen J, Knipschild P. Acupuncture and chronic pain: A criteria-based meta-analysis. *J Clin Epidemiol* 1990; 43:1191-1199.
219. Ezzo J, Berman B, Hadhazy V et al. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000; 86:217-225.
220. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized controlled trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol* 1995; 48:167-171.
221. Dickersin K, Scherer R, Lefebvre C. Systematic reviews – identifying relevant studies for systematic review. *BMJ* 1994; 309:1286-1291.
222. Oxman AD. The Cochrane Collaboration Handbook: Preparing and maintaining systematic reviews. Second Edition. Cochrane Collaboration, Oxford, 1992.
223. Begg CB, Berlin JA. Publication bias: A problem in interpreting medical data (with discussion). *J R Statist Soc A* 1988; 151:419-463.
224. Dickersin K, Chan S, Chalmers TC et al. Publication bias and clinical trials. *Controlled Clin Trials* 1987; 8:343-353.
225. Deeks J, Glanville J, Sheldon T. Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews. Centre for Reviews and Dissemination, Report #4, York Publishing Services Ltd, York, 1996.
226. Dickersin K, Cooper H, Hedges LV. Research register. In *The Handbook of Research Synthesis*. Russell Sage Foundation, New York, 1994, pp 71-84.
227. Dickersin K, Hewitt P, Mutch L et al. Pursuing the literature: Comparison of MEDLINE searching with a perinatal trials database. *Controlled Clin Trials* 1985; 6:306-317.
228. Smith BJ, Darzins PJ, Quinn M et al. Modern methods of searching the medical literature. *Med J Aust* 1992; 157:603-611.
229. Sackett DL. Bias in analytic research. *J Chron Dis* 1979; 32:51-63.
230. Chalmers TC, Levin H, Sacks HS. Meta-analysis of clinical trials as a scientific discipline. I: control of bias and comparison with large co-operative trials. *Stat Med* 1987; 6:315-325.
231. Reed JG, Baxter PM, Cooper H et al. Using reference databases. In *The Handbook of Research Synthesis*. Russell Sage Foundation, New York, 1994; pp 57-70.
232. Clarke M. Searching MEDLINE for randomized trials. *BMJ* 1993; 307:565.
233. Adams CE, Power A, Frederick K et al. An investigation of the adequacy of MEDLINE searches for randomized controlled trials (RCTs) of the effects of mental health care. *Psychol Med* 1994; 24:741-748.
234. Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. *Online J Curr Clin Trials* 1993; No. 33.
235. Kleijnen J, Knipschild P. The comprehensiveness of MEDLINE and EMBASE computer searches. *Pharm Weekbl [Sci]* 1992; 14:316-320.

