Pain Physician 2008; 11:161-186 • ISSN 1533-3159

Invited Review

Evidence-Based Medicine, Systematic Reviews, and Guidelines in Interventional Pain Management, Part I: Introduction and General Considerations

Laxmaiah Manchikanti, MD

From: Pain Management Center of Paducah, Paducah, KY.

Dr. Manchikanti is Medical Director of the Pain Management Center of Paducah, Paducah, KY, and Associate Clinical Professor of Anesthesiology and Perioperative Medicine, University of Louisville, Louisville, KY.

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

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

Manuscript received: 02/04/2008
Accepted for publication: 02/28/2008

Free full manuscript:
www.painphysicianjournal.com

Evidence-based medicine, systematic reviews, and guidelines are part of modern interventional pain management. As in other specialties in the United States, evidence-based medicine appears to motivate the search for answers to numerous questions related to costs and quality of health care as well as access to care. Scientific, relevant evidence is essential in clinical care, policy-making, dispute resolution, and law. Consequently, evidence based practice brings together pertinent, trustworthy information by systematically acquiring, analyzing, and transferring research findings into clinical, management, and policy arenas. In the United States, researchers, clinicians, professional organizations, and government are looking for a sensible approach to health care with practical evidence-based medicine. All modes of evidence-based practice, either in the form of evidence-based medicine, systematic reviews, meta-analysis, or guidelines, evolve through a methodological, rational accumulation, analysis, and understanding of the evidentiary knowledge that can be applied in clinical settings.

Historically, evidence-based medicine is traceable to the 1700s, even though it was not explicitly defined and advanced until the late 1970s and early 1980s. Evidence-based medicine was initially called "critical appraisal" to describe the application of basic rules of evidence as they evolve into application in daily practices. Evidence-based medicine is defined as a conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. Evidence-based practice is defined based on 4 basic and important contingencies, which include recognition of the patient's problem and construction of a structured clinical question, thorough search of medical literature to retrieve the best available evidence to answer the question, critical appraisal of all available evidence, and integration of the evidence with all aspects and contexts of the clinical circumstances.

Systematic reviews provide the application of scientific strategies that limit bias by the systematic assembly, critical appraisal, and synthesis of all relevant studies on a specific topic. While systematic reviews are close to meta-analysis, they are vastly different from narrative reviews and health technology assessments.

Clinical practice guidelines are systematically developed statements that aim to help physicians and patients reach the best health care decisions. Appropriately developed guidelines incorporate validity, reliability, reproducibility, clinical applicability and flexibility, clarity, development through a multidisciplinary process, scheduled reviews, and documentation. Thus, evidence-based clinical practice guidelines represent statements developed to improve the quality of care, patient access, treatment outcomes, appropriateness of care, efficiency and effectiveness and achieve cost containment by improving the cost benefit ratio. Part 1 of this series in evidence-based medicine, systematic reviews, and guidelines in interventional pain management provides an introduction and general considerations of these 3 aspects in interventional pain management.

Key words: Evidence-based medicine, systematic reviews, clinical guidelines, narrative reviews, health technology assessments, grading of evidence, recommendations, grading systems, strength of evidence.

Pain Physician 2008; 11:161-186

The pace of innovation in health care has never been greater, and this innovation is constantly adding to a broad and complex area of health care interventions and systems. Thus, the need for careful scientific evaluation of clinical practice became a prominent focus during the second half of the twentieth century (1). The demonstration of pervasive and persistent unexplained variability in clinical practice (2) and high rates of inappropriate care (3), combined with increased expenditures (4-6), have fueled a steadily increasing demand for evidence of clinical effectiveness (7-36). Consequently, a body of evidence regarding safety, effectiveness, appropriate indications, cost-effectiveness, and other attributes of medical care are demanded. This demand is partly based on the limited amount of high-quality evidence, geographic variation, inappropriate care, and the limited success of quality improvement efforts (7,8). Failure to understand which services work best, under what circumstances, and for which types of patients contributes to the increasing cost of care, threats to patient safety, and avoidable loss of life (9). Landmark reports of the Institute of Medicine, including To Err Is Human: Building a Safer Health System (10) and Crossing the Quality Chasm: A New Health System for the 21st Century (11), have drawn national attention to shortcomings in quality and patient safety. As a result, increased attention is being directed to the development of methods that can provide valid and reliable information about what works best in health care. However, achieving many of the opportunities to improve health care based on evidence depends on the ability of clinicians, patients, and policy makers to interpret and apply this body of evidence. In the United States, health care competes for consumers with other items in the marketplace (37). Thus, there is a need for high quality evidence, which is scientific, understandable, evidence-based, and practical – not anecdotal, consensus only, opinionated, nihilistic, or simply economic-based (12-20,30-32,35,38-47).

Researchers, clinicians, professional organizations, and government in the United States, along with other countries, are looking for a sensible approach to health care with practical evidence-based medicine (EBM). However, each segment has their own interpretation and agenda, which is not based on science and best care for the patient, but rather on 3 important aspects – economics, economics, economics (30-32,35,38-47). The trend to develop and implement research in support of evidence-based practice has been the convention of medicine for the past decade (12). Consequently, this emphasis has been fostered, at least partly, by a perceived need to improve patient care through applied clinical decision-making in diagnosis and treatment, even though this emphasis has not been well defined. Evidence-based practice, either in the form of EBM, systematic reviews, meta-analysis, or guidelines, evolves through a methodological, rational accumulation, analysis, and understanding of the evidentiary knowledge that can be applied in the clinical setting(s) (48). While it is important to remember that the actual value of the evidence is related to the application in which it will be used, and the circumstances in, and agents for whom such evidence may or may not have relevance, it is also essential to remember that the value of evidence is only as good as the type of evidence reviewed, methodology utilized, knowledge, and experience of the reviewers and many other factors, including bias, self-interest, economics.

Osler wrote, "Medicine is the art of probability" (49). Currently medicine is defined as the art and science of diagnosis, treatment, and prevention of disease and the maintenance of good health. Based on the current working definition of medicine, most clinical decisions applied in medicine are based upon the knowledge that health is a stochastic process, that outcomes are probabilistic, and that it is difficult to predict where a patient will fall in a bell-shaped curve (50). This results in the art of probability aspects of medicine with health care being dependent on probabilities and decisions that are based on population-based information.

Similar to medicine, evidence can be defined as any ground or reason for knowledge or certitude in knowledge; proof, whether immediate or derived from inference; a fact or body of facts on which a proof, belief, or judgment is based (51). However, the nature of belief and the foundation upon which the evidence rests provides the utility of evidence (52,53). In simplistic terms, for medical purposes, any data or information, whether solid or weak, obtained through experience or research can be considered as evidence (54).

Evidence-Based Medicine

Evidence-based medicine is not new or a twentieth century phenomenon. Historically it is traceable to the 1700s, even though it was not explicitly defined and advocated until the late 1970s and early 1980s (21). Initially, EBM was called "critical appraisal" to de-

scribe the application of basic rules of evidence. This evidence was presented by a group of clinical epidemiologists at McMaster University. Since then, epidemiologists, rather than clinicians, have been heavily involved in the development of evidence-based practice. The term "evidence-based medicine" was used as part of an informational document at McMaster University in 1990, which was later published in the ACP Journal Club in 1991 (22). These initial efforts evolved into collaboration between evidence-based medical educators at McMaster University and a group of academic physicians, primarily from the United States, resulting in the first international EBM working group. The work of this group was published in a 25-part series, "Users Guides to the Medical Literature" in JAMA between 1993 and 2000, which ultimately resulted in a textbook (23-28). Around the same time, multiple other organizations also devoted themselves to advancing EBM, including Archie Cochrane who developed the Cochrane Collaboration, which started in 1993 (29). The Agency for Healthcare Research and Quality (AHRQ), a US agency, also was chartered in 1999 (55), which evolved from the Agency for Healthcare Policy and Research (AHCPR), with its mission to fill the gap of lack of appropriate information, by contributing to the health care knowledge base itself and identifying priority areas for assembling, interpreting, and translating to users findings from this knowledge base. AHRQ has been recognized as a well-founded organization and source to provide evidence in health care, specifically through its evidence-based practice centers program (9). The US Department of Health and Human Services has been working to develop evidence to reduce healthcare expenditures since the mid 1980s through multiple organizations – the US Public Health Service, the National Center for Health Services Research and Health Care Technology Assessment, AHCPR and now AHRQ.

An operational definition of EBM is a conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients (21). Thus, EBM is about solving clinical problems (22). Consequently, EBM acknowledges that intuition, unsystematic clinical experience, and pathophysiological rationale are insufficient grounds for clinical decision making, and stresses the examination of evidence from clinical research (56), in contrast to the traditional paradigm of medical practice. EBM suggests that a formal set of rules must compliment medical training and common sense for clinicians to interpret the results of clinical research effectively. Further, EBM places a lower value on authority than does the traditional medical paradigm. EBM is focused on the use of the right (types and extent of) knowledge to guide the right and good intentions and actions of medical practice. Consequently, this process is fundamental to potential clinical decision-making (12,57-59). In essence, EBM requires the prudent, specific contextual application of knowledge gained by integration of individual clinical expertise and experience, in concert with the best available external evidence gained from systematic research (12,18-20,57-61). As a result, clinicians should be extremely cognizant of the systematic evidence and make practical and ethical decisions that affect patient care with due weight given to all valid, relevant information (12,53,62). This leads to the question of whether or not the evidence should be only a certain type, for example, derived from randomized clinical trials, conducted by a certain group of individuals, conducted in ceratin countries, or published in certain journals. To the contrary, the evidence should include not only the evidence derived from randomized controlled trials from academic medical centers and published in 2 or 3 perceived top journals, but from all types of evidence, in conjunction with both patient preferences to accept or refuse a particular treatment, and patient access to available, affordable resources. This contradicts the argument that only one form of evidence should necessarily be the determining factor in decision-making. Thus, EBM explicitly mandates the necessity for an active search for all information that is valid and relevant, and an ongoing assessment to ascertain both the accuracy of information and the applicability of evidence to the decision in question. Thus, evidence-based practice emphasizes an integration of the best research evidence with patient's circumstances and values (24,63).

An ethical and practical approach to EBM involves 2 fundamental principles. First, delineating that scientific evidence alone is never sufficient to make a clinical decision and decision makers must always consider the patient's values when evaluating the benefits, risks, and burdens associated with any/all treatment strategies (22). The second principle is that, while EBM describes a hierarchy of informational value to guide clinical decision-making (22,56), this hierarchy is never absolute, and must reflect how different types and levels of evidence can be relative to, and inform the calculus of, circumstances, agents, and the consequences of decisions and actions (53,61-64).

Evidence-based practice is defined based on 4 basic and important contingencies (65):

- Recognition of the patient's problem and construction of a structured clinical question.
- Thorough search of medical literature to retrieve the best available evidence to answer the question.
- Critical appraisal of all available evidence.
- Integration of the evidence with all aspects and contexts of the clinical circumstances to facilitate the decisional process that determines the best clinical care of each patient.

In spite of clear definitions, confusion surrounds the definitions and understanding of evidence-based medicine or practice. Evidence-based medicine or practice is understood differently by academicians, practitioners, managed care executives, professional organizations representing the industry, organizations with an economic focus, for-profit evidence-based organizations, organizations with a relationship to industry, industry, researchers, attorneys, policy makers, and patients (30-39,66). Considering the multiple forces working against each other, the application of the principles of EBM in practice is not a simple issue and it is no wonder that most clinicians' practices do not reflect the principles of EBM but rather are based upon tradition, their most recent experience, what was learned years ago in medical school, or anecdotal information acquired from colleagues, unless EBM or such principles are mandated in the form of Local Coverage Determinations (LCDs) or other instruments to affect the economic return.

Bogduk et al (67-70) introduced EBM as, ". . . the medical practice that uses techniques with proven reliability, validity, and efficacy, while shunning those that partly lack reliability, validity, or efficacy." While this approach does not include literature search, literature assessment, methodological review, or evidence grading, it is less conciliatory than other definitions such as that provided by Sackett et al (21). In yet another terminology, a supermarket approach to the evidence-informed management of chronic low back pain was introduced in 2008 (33,34). These authors proposed that in the modern era, it is possible to fill an entire shopping cart with treatments that one can try simultaneously or serially under the guise of "multidisciplinary care," which essentially eliminates a patient browsing through multiple aisles and temptation by more than one product (33). Window shopping for those unwilling (or unable) to purchase one of the numerous products displayed in the aisles, may take the form of coping and acceptance, activity modification, self education, patient initiated comfort methods passed on from their parents, grandparents, or friends, and what has become known as watchful waiting and reassurance (33). Thus, a simplified, partial inventory of treatment options available to a person with chronic low back includes over 200 different medications, therapies, injections, products, or procedures (33). They conceded that systematic review methodology confined to high-quality randomized controlled trials would likely find only limited evidence for many of the interventions used in chronic low back pain. Hence, given the wealth of clinical experience among invited authors, it was concluded that an evidence informed approach would be more appropriate than strictly evidence-based recommendations.

Some have considered EBM as cookbook medicine or cost saving medicine. However, EBM is neither. In essence, EBM may even increase the cost. Thus, EBM, based on numerous propositions and requirements, may not be based on evidence and values to sustain that evidence.

While numerous self reinforcing interests have been criticized as driving EBM, some have felt that the researcher, also known as the messenger, is under attack due to intimidation by special-interest groups (71). A special interest group focusing on intimidation of researchers by special-interest groups described the attacks on them, as similar to the attacks on health researchers, such as Pierre Lewis, who was vilified nearly 2 centuries ago for suggesting that bloodletting was an ineffectual therapy (72). While this is an overdramatic statement, another researcher (73) argued that litigation, fear, bias, and greed interfere with scientific effort to answer questions of importance to public health and an antiscientific social attitude encourages premature or ill-informed political and legal solutions to medical questions. Deyo et al (71) compare research issues and differences of opinion as similar to "hot-button" policy issues, such as chemical exposure, firearm injuries, and breast implants. A story from these authors tells about AHCPR guidelines and their demise due to non-recommendation of fusion as a treatment modality (35,55,74-76). However, they had also ignored the implications of AHCPR guidelines on all aspects of management (35,55). Still, they believe that they have been unfairly attacked through marketing, professional, media, legal, administrative, or political channels for publishing legitimate scientific results that

ran counter to financial interests and strong beliefs. However, at least in the case of low back pain guidelines (35), evidence shows otherwise (55). Further, as described by these authors, the arguments between various special interest groups may not be analogous to the strategy embodied in so-called slapsuits (strategic lawsuits against public participation) (40,77). Generally, slapsuits are brought by private financial interest against activists who have opposing points of view and engage in such activities as circulating petitions or testifying at public hearings, turning them into political and judicial forums while these investigators in the multiple examples quoted in this manuscript (71) were considered to be trained in a method of scientific discourse that is generally cautious in interpreting data, acknowledges faults and limitations, and places findings in the context of scientific knowledge. It is contended by opponents that many may lack clinical knowledge required, thus, leading to misinterpretation of the data (42,43,78-81).

Systematic Reviews

A systematic review is defined as, "the application of scientific strategies that limit bias by the systematic assembly, critical appraisal and synthesis of all relevant studies on a specific topic" (82,83). Systematic reviews are labor intensive and require expertise in both the subject matter and review methods. Thus, expertise in one or the other area is not enough and may lead to inaccurate conclusions in turn leading to inappropriate application of the results (84-86). While expertise in the subject matter is crucial, expertise in review methods is also particularly important.

Historically, expert opinion has been presented in narrative reviews which are not evidence-based, and, consequently have significant limitations (87,88). Unsystematic narrative reviews are more likely to include only research selected by the authors, thereby introducing bias; hence, they frequently lag behind and contradict available evidence (89,90). Characteristically, 2 types of approaches are utilized – narrative reviews, also known as focused reviews, and health technology assessments, apart from systematic reviews (91,92). A narrative review, while similar to a systematic review, does not employ the methodological safeguards to control bias. Thus, the major difference between a systematic review and narrative review is that a systematic review attempts to minimize bias by the comprehensiveness and reproducibility of the search and selection of articles for review and provides assessment of the methodological quality of the studies (88,93-98). Systematic reviews are generated to answer specific, often narrow, clinical questions in depth. However, in a systematic review, while systematic searching, selecting, appraising, interpreting, and summarizing of data from original studies is essential, these original studies may be observational studies or randomized trials (91,99). Further, the study summaries may be qualitative or quantitative. Systematic review articles are one type of integrative publication compared to practice guidelines, economic evaluations, and clinical decision analysis which are separate and different. However, other types of integrative articles often incorporate the results of systematic reviews, thus, evidence-based practice guidelines are based on systematic reviews of the literature, appropriately adapted to local circumstances and values. In the past 20 years, the publication of systematic reviews and meta-analysis has grown exponentially as shown in Fig. 1.(12-20,100).

A third type of review is the health technology assessment (HTA), a multidisciplinary approach that studies the medical, social, ethical and economic implications of the development, use, and diffusion of health technologies. HTAs have been described as, "the bridge between the world of research and the world of decision making" (92) and are being used with increasing frequency to influence both practice and policy meaning these guidelines are used by not only the insurers and industry, but also AHRQ in assessing the evidence. However, HTAs must not only be scientifically accurate, but must also be optimally timed so as to affect the sensitivity of the political decision makers, to effectively influence the policy.

Table 1 shows the differences between narrative and systematic reviews, whereas, Table 2 illustrates differences between systematic reviews and health technology assessments. Since the foundation of EBM practices in the use of information gained from systematic reviews or more importantly in the synthesis of evidence from systematic reviews, it is vital to consider that the strength of this foundation reflects the quality of the systematic reviews, and it is therefore necessary to evaluate the evidence summaries and synthesis themselves before evidence-based decisional processes can be built upon them. Thus, Evidence-Based Practice Centers Partners Guide (9) from AHRQ describes that systematic reviews are only as complete and useful as

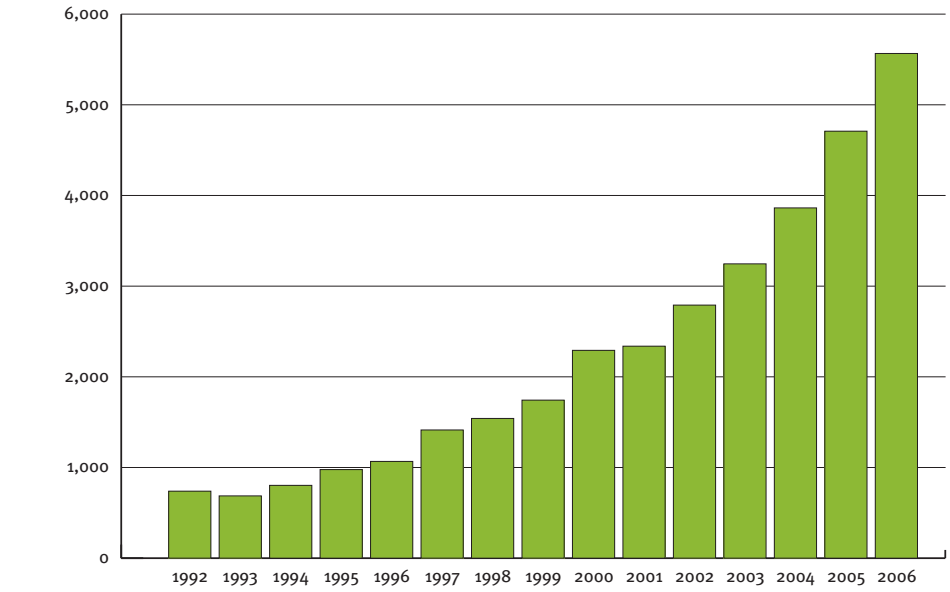

Fig. 1. *Explosion of systematic reviews/meta-analyses.*

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 from Ref. 191

Table 2. *Differences between systematic reviews and health technology assessments (HTAs)*

	Systematic Reviews	HTAs
Methodological standards	Only include studies with the best methodological evidence	Include studies of topics of interest to policy-makers, even if evidence is suboptimal
Repeating previous studies	No need to repeat if previous studies were high quality, and no new high-quality evidence	The need to defend the report's conclusions often necessitates repetition
Breadth versus depth	Only include topics for which there is good evidence; topics driven by scientists' interests	Include topics most relevant to policy-makers; exclude those not of relevance even if there is good quality evidence
Inclusion of content experts and policy-makers	Content experts, but not policy-makers usually included	Can be concerns that content experts and policy-makers are biased
Performance of economic evaluations	Usually not done	Economic evaluations are an important component of HTAs, but lack of good evidence about effectiveness/ diagnostic accuracy limit their impact
Making policy recommendations	Almost never done	Sometimes done, but with caution
Active dissemination	Rarely done	Sometimes done

Adapted from Ref. 105

the evidence that exists on a particular topic or the scope and nature of the evidence questions that guide the review.

It is important to acknowledge the types of evidence other than randomized controlled trials o interventions can be systematically reviewed. Even thought, the review methods for such primary articles are the same, the methods of pooling results will differ. Certain questions must be considered in assessing the quality of a systematic review as described in Table 3 (101). Further, in assessing the value of the review, it is important to consider if the results can be applied to patients in a particular setting and will the results help me care for patients. And, one shoulder consider if the benefits are worth the harm and costs.

Guidance has been provided for writing (82), as well as reading and interpreting (87, 102-108) systematic reviews. Oxman et al (101,105,108) provided guidance for critical appraisal of the evidence. Table 3 illustrates questions that should be considered in determining if the results of a systematic review are valid (101). Oxman (108) noted the need for check lists, as analogous to flying an airplane due to the complex practice of medicine. The most dangerous errors in reviews are systematic ones (bias) rather than ones that

Table 3. *Questions that should be considered in determining if the results of systematic review are valid*

1) **Did the overview address a focused clinical question?**
2) **Were the criteria used to select articles for inclusion both defined and appropriate?**
3) **What is the likelihood that relevant studies were missed?**
4) **Was the validity of the included studies assessed?**
5) **Were the assessments reproducible?**
6) **Were the study-to-study results congruent?**
7) **How precise were the results of the overview?**
8) **Were all clinically important outcomes considered?**

Adapted from Crowther al (Ref. 99)

occur by chance alone (random errors). Therefore, the most important thing for doers and users of the review to check is its "validity," the extent to which its design and conduct are likely to have been protected against bias. Bias, as well as random errors, are deadly. If a review is done systematically and quantitative results are presented, the confidence interval around the results provides a good indication of "precision," the extent to which the results are likely to differ from the "truth" because of a chance alone (108,109). However, a confidence interval does not provide any indication of the likelihood of bias. Thus, the questions listed in Table 3 must be asked for every systematic review. Oxman et al (105) provided guidance for presentation of evaluation of synthesis. They (105) described a systematic review of 2 instruments critically appraising systematic reviews (110,111) and studies how to present the results of systematic review to policy makers (112), the general public (113), and users of Cochrane reviews (114). West et al (110) reviewed different instruments for critically appraising systematic reviews and found 20 systems concerned with the appraisal of systematic or meta-analysis, including one scale, 10 checklists, and 9 guidance documents, and identified 7 domains that they considered important to appraise: study question, search strategy, inclusion and exclusion criteria, data abstraction, study quality, data synthesis and analysis, and funding or ownership as listed in Table 4 (115-119). The authors concluded that based on coverage of the 7 domains that they considered key, these 5 systems represented "best practice" for appraising the systematic reviews.

In contrast, another review used a detailed process to evaluate and select a system and expanded the work by AHRQ up until the year 2005 (111). They identified approximately 240 quality assessment instruments for systematic reviews, randomized controlled trials, and observational studies, as well as nearly 50 evidence grading systems. Following critical and extensive review, the Amstar 2005 was selected as the best instrument for appraising systematic reviews as illustrated in Table 5 (111).

Assessment by the National Institute for Health and Clinical Excellence (NICE) (120) assessed 20 technology assessment reports and found that a more selective approach to database searching would suffice in most cases and would save resources, whereas, searching other sources, including contact with experts and checking reference lists, appeared to be a more productive way of identifying further studies (120).

It is essential to evaluate the search strategy and assessment of study quality. One of the most powerful arguments used by the supporters of systematic reviews is that they overcome most of the limitations of narrative reviews by being the product of a scientific process to reduce bias and impression and by providing detailed information to allow replication by others (103,121,122). Ideally, a systematic review includes all the relevant trials available (123). Consequently, identifying all the relevant trials for systematic review has been recognized as a most fundamental challenge (121). Identification of all relevant trials for a systematic review is an excruciating process but by all means possible (124). However, considering that searching the literature can be an onerous, resource consuming task, reviewers with limited resources have to set priorities regarding what sources to use to identify trials in the most cost-effective way.

Table 4. *Evaluation of systems to grade the quality of systematic reviews.*

Instrument	Critical domains in the evaluation criteria						
	Study question	Search strategy	Inclusion/ exclusion	Data extraction	Study quality	Data synthesis/ analysis	Funding
Irwig et al (115)	•	•	•	•	•	•	°
Sacks et al (116)	•	•	•	•	•	•	•
Auperin et al (117)	=	•	•	•	=	•	•
Barnes and Bero (119)	•	=	•	°	•	•	•
Khan et al (118)	•	•	•	•	•	•	°

Legend: • = yes; = partial; ° = not met or no information

Source: Lohr KN (Ref. 106)

Table 5. *A measurement tool to assess reviews (AMSTAR), 2005.*

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

Source: Oxman et al (Ref. 105)

Two of the most effective mechanisms for a systematic review to reduce bias and impression are including the maximum possible number of relevant individual trials and providing a detailed description of their strengths and limitations. Publication bias has been described as publication of positive trails and non-inclusion of unpublished studies (125). However, in recent years, it appears that only manuscripts published in society journals are positive trials of their own specialty and negative trials of other specialties. Consequently, systematic reviews that fail to identify and include unpublished trials and those who fail to identify all journals and databases beyond their own specialty, are at the risk of overestimating the effect

of interventions they are interested in and they are practicing and underestimating the interventions they oppose. Readers, however, must take into account the reviewer bias, which includes non-identification (or omission) of unpublished trials and the articles from other journals based on the preference of the reviewers. While it has been extremely difficult without compulsory registration of trials at inception to know how many unpublished trials exist, the modern rules of clinical registry makes it easier. Further, many journals refuse to publish reviews that include unpublished data (126). On a pragmatic basis, admittedly without empirical evidence supporting this, a systematic review in interventional pain management at minimum must have a comprehensive review using at least 3 sources and provide a description of efforts to identify all databases, and journals, if not unpublished trials. An effective combination of comprehensive search includes a minimum of 3 bibliographic databases (Medline, EMBASE, Cochrane library), a hand search of references of eligible trials, and direct contact with the corresponding authors of eligible trials asking for additional published or unpublished trials.

The next issue is with regards to the quality of trials included. While almost every systematic review has supporters and detractors, both groups agree on the relevance of the dictum, "garbage in, garbage out" (103). Essentially, this tells us that evidence is in the eyes of the reviewer and shows that the extent to which a systematic review could guide health care decisions depends on the quality of the trials available. Thus, it is always argued that if the trial quality was assessed appropriately, if it was assessed at all, the expertise of various authors of reviews vary widely with some considering the quality assessment as an important strategy to identify and reduce bias, and others who see assessment as a source of bias or as completely uninformative, whereas, yet some others criticize the criteria utilized on a multitude of personal biases (127,128). In a perfect world, it would be best if only ideal trials are included in the reviews, which would have answered all the questions. Among other things, those trials should include the following:

1) To answer clear and relevant clinical questions
2) To be designed, conducted, and reported by researchers who did not have conflicts of interest
3) To follow strict ethical principles
4) To include all patients available
5) To evaluate all possible interventions for all possible variations of the conditions of interest, in all possible types of patients, in all settings, and using all relevant outcome measures
6) To include strategies to eliminate bias during the administration of the interventions, during the evaluation of the outcomes, and during reporting of the results, thus reflecting the true effect of the intervention
7) To include perfect statistical analysis
8) To be described in clear and unambiguous language, including an exact account of all the events that occurred during the design and conduct of the trial, individual patient data, and an accurate description of the patients who were included, excluded, and withdrawn and who dropped out.

Once reviewers have assessed the trial quality, one should look at the nature and type of the quality assessment, including the definition and assessment tools employed. Further, it is important to recognize the incorporation of quality assessments into systematic reviews (97,130,131).

Finally, the crucial issue and interpretation of the systematic review is how the authors synthesized the evidence. Thus, understanding and replicating the evidence synthesis and even the conclusions would be appropriate.

In contrast to a systematic review, meta-analysis is a statistical procedure that integrates the results of several independent studies considered to be combinable (84-88,132,133). Well conducted analyses allow a more objective appraisal of the evidence than traditional narrative reviews, provide a more precise estimate of treatment effect, and may explain heterogeneity between the results of individual studies (134). However, an ill conducted meta-analysis, on the other hand, may be biased owing to exclusion of relevant studies or inclusion of inadequate or irrelevant studies (135). Thus, to understand meta-analysis, meta-analyses should follow a few basic principles:

1) Meta-analysis should be as carefully planned as any other research project with a detailed written protocol being prepared in advance.
2) The priori definition of eligibility criteria for studies to be included in a comprehensive search for such studies are central to high quality meta-analysis.
3) The graphical display of results from individual studies on a common scale is an important intermediate step, which allows a visual examination of the degree of heterogeneity between studies.
4) Different statistical methods exist for combining the data, but there is no single "correct" method.

5) A thorough sensitivity analysis is essential to assess the robustness of combined estimates to different assumptions and inclusion criteria.

It has been described that bias in meta-analysis may be detected by a simple, graphical test (135). Based on the findings of some meta-analyses contradicted by large trials, bias was found in 38% of the meta-analyses published in leading general medicine journals and in 13% of the reviews from the Cochrane database of systematic reviews. As a result,, it is strongly recommended that all systematic reviews and meta-analyses must be checked for biases. Publication bias (136), and other factors affecting comparability of meta-analyses and the largest trials results was also described (137). Similar to systematic reviews, in conducting a meta-analysis, investigators should make strenuous efforts to find all published studies, search for unpublished work and carefully assess the quality of component studies (138,139). Stroup et al (139) developed a checklist for authors, editors, and reviewers of meta-analyses of observational studies as illustrated in Table 6. Empiric research on the quality of systematic reviews has shown that not all systematic reviews are truly systematic (140,141), that the quality of systematic reviews is highly variable (87,88), and that the Cochrane reviews, on average, may be more rigorous and better reported than journal reviews (141,142). However recent studies have shown that even Cochrane reviews have methodological problems (143,144). In addition, in evaluation of thequality of primary studies which sets apart systematic reviews from traditional reviews, empiric research shows that not all systematic reviews assess study quality. It has been shown that among evaluation of 240 systematic

Table 6. *A proposed reporting checklist for authors, editors, and reviewers of meta-analyses of observational studies*

Reporting of background should include:
Problem definition
Hypothesis statement
Description of study outcome(s)
Type of exposure or intervention used
Type of study designs used
Study population

Reporting of search strategy should include:
Qualifications of searchers (e.g., librarians and investigators)
Search strategy, including time period included in the synthesis and keywords
Effort to include all available studies, including contact with authors
Databases and registries searched
Search software used, name and version, including special features used (e.g., explosion)
Use of hand searching (e.g., reference lists of obtained articles)
List of citations located and those excluded, including justification
Method of addressing articles published in languages other than English
Method of handling abstracts and unpublished studies
Description of any contact with authors

Reporting of methods should include:
Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested
Rationale for the selection and coding of data (e.g., sound clinical principles or convenience)
Documentation of how data were classified and coded (e.g., multiple raters, blinding, and interrater reliability)
Assessment of confounding (e.g., comparability of cases and controls in studies where appropriate)
Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results
Assessment of heterogeneity
Description of statistical methods (e.g., complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient detail to be replicated
Provision of appropriate tables and graphics

Reporting of results should include:
Graphic summarizing individual study estimates and overall estimate
Table giving descriptive information for each study included
Results of sensitivity testing (e.g., subgroup analysis)
Indication of statistical uncertainty of findings

Reporting of discussion should include:
Quantitative assessment of bias (e.g., publication bias)
Justification for exclusion (e.g., exclusion of non-English-language citations)
Assessment of quality of included studies

Reporting of conclusions should include:
Consideration of alternative explanations for observed results
Generalization of the conclusions (i.e., appropriate for the data presented and within the domain of the literature review)
Guidelines for future research
Disclosure of funding source

Source: Stroup et al (Ref. 139)

reviews from journals only 48% assessed quality (145), in the evaluation of 480 systematic reviews in DARE only 52% assessed quality (140), and in the evaluation of 50 systematic reviews on asthma only 28% reported validity assessment criteria (142). Further, among meta-analyses heterogenicity is a common finding (146). However, empiric work has shown that evaluation of heterogenicity is not universally done, and that only approximately 45% to 68% of reviews tested for heterogenicity (140,142,147). Consequently, due to inability of quality of systematic reviews to be taken for granted, the reader has the responsibility of critically appraising them.

Guidelines

Clinical practice guidelines are systematically developed statements that aim to help physicians and patients reach the best health care decisions (148). Appropriately developed guidelines incorporate validity, reliability, reproducibility, clinical applicability and flexibility, clarity, development through a multidisciplinary process, scheduled reviews, and documentation (149). Thus, evidence-based clinical practice guidelines are statements developed to improve thequality of care, patient access, treatment outcomes, appropriateness of care, efficiency and effectiveness and achieve cost containment by improving the cost benefit ratio. Guidelines are sponsored by various organizations, most commonly by specialty societies. The National Guideline Clearinghouse and AHRQ (150) lists over 1,856 guidelines on their website. Guidelines are developed based on the evidence and opinion, thus, they are neither infallible nor a substitute for clinical judgment (148). In contrast to systematic reviews, guidelines recommend what should and should not be done in specific clinical circumstances.

Guidelines are highly variable in length. While some are widely respected and standardize the care with diminution of variations and improved health outcomes, others are developed with economic goals in mind and are controversial. Guidelines could be extremely controversial even when developed by governmental agencies, even though the best guidelines are considered the ones from the US Preventive Services Task Force, the Advisory Committee on Immunization Practices, the National Academies, Centers for Disease Control and Prevention, World Health Organization, and the National Institute for Clinical Excellent (NICE) in the United Kingdom. Guidelines may be controversial for numerous reasons including the type of recommendations and the restrictions on practice patterns. A prime example is the demise of the AHCPR in 1995 following the development of acute low back pain guidelines (35), which issued 19 guidelines between 1992 and 1996 at a cost of $750 million for 15 guidelines, at a cost of $50 million per guideline (151).

Aside from the federal government initiatives, practice guidelines have been developed by national medical speciality societies in a number of formats. As early as 1938, the American Academy of Pediatrics began publishing its guidelines for the treatment of infectious diseases (152). In fact, the American Medical Association (AMA) supported guidelines development as an alternative to expenditure targets and established its own organizational structure for the development of clinically sound and relevant guidelines through the forum on practice parameters in 1989 (153).

A comparison of Cochrane reviews with articles published in paper-based journals also showed that Cochrane reviews appear to have greater methodological rigor and aremore frequently updated than systematic reviews or meta-analyses published in paper-based journals (141). However, reviews found in Medline included more trials, and more patients than Cochrane reviews. Guidelines have been questioned on various fronts based on pharmaceutical and medical device company sponsorship, when members of guideline committee have substantial financial associations with industry, and relationship of the developing organization to the industry when there is no relevant relationship or expertise in developing the guidelines except for the sole purpose of financial gain (30-32,44,45). It has been argued that public disclosure of sponsorship and of the financial associations of committee members, along with other adequate safeguards and rules to prevent sponsors from influencing the selection of panel members and the content of guidelines be mandatory. It is maintained that practice recommendations will invariably be viewed with skepticism unless corporate sponsorship and financial relationship as well as experts with financial ties and also experts promoting their own specialty while making decisions on other specialties are completely avoided. At present, the financial ties between guidelines panels and industry appear to be extensive. A survey of 685 disclosure statements by authors of guidelines concerning medications found that 35% declared a potential financial conflict of interest (157). The guidelines developed by American College

of Occupational and Environmental Medicine (30) and the ODG guidelines (Official Disability Guidelines) have been criticized for their relationship with industry and numerous ties of the guidelines panel. Hayes Guidelines are commercially available for over $15,000 each making it practically impossible for anyone to review them (32).

Thus, conflict management in the guideline profession is an essential ingredient. A conflict of interest exists when an individual's secondary interest (e.g., personal, financial) interferes with or influences judgments regarding the individual's primary interests (patient welfare, e.g., education, research integrity) (155). Further, there is evidence demonstrating the association of financial ties with a breakdown in research integrity. It has been shown that industry funding for research is associated with favorable outcomes for the sponsor (156-160) and financial ties of investigators with their sponsors such as Stark ownership laws, consulting income, etc., are also associated with favorable research outcomes for the sponsor (160). As expected, this evidence has been accentuated by lay media stories documenting how financial conflicts of interests have led to biased and even dangerous research (161,162). Biased research may be intentional or unintentional (163) and may result from loss of objectivity at multiple stages in the research process, including conceptualization of the question, design, or conduct of the research, interpretation of the results, and publication or lack thereof of the research (164,165). Consequently, regardless of its source, the bias associated with financial and other conflicts of interest may damage both the public's and other researcher's trust in science (166), whereas, the type of conflict most likely to affect the public trust is financial conflict where the scientist tends to gain financially from a particular research outcome (166-171). However, other competing interests, such as professional advancement, are also extremely important. Thus, conflict of interest policies are designed to protect the integrity of research and decision-making processes through disclosure and transparency.

The NIH Consensus Development Program (172), started in 1977, sponsors evidence-based assessments of important medical issues. These assessments include a systematic review of literature through the AHRQ, public presentation of the research, and a consensus statement that is disseminated widely. In this forum, panel members can have neither financial nor other potential conflicts, and panels are independent of both NIH and the Department of Health and Human Services. Therefore, the consensus statements reflect the conclusions of the panels, not those of institutes. However, the conference speakers may have industry ties, but if they do, those ties are disclosed, thus the process despite it rigor has limitations. This process is also expensive with costs of $500,000, and time consuming taking 18 months from conception to completion. In addition, AHRQ sponsors about 20 to 25 systematic reviews each year, providing these public and private organizations, including the Centers for Medicare and Medicaid Services (CMS), NIH, and the various specialty societies. However, the researchers can have no financial associations related to the subject.

In contrast, specialty societies are highly variable with diverse policies for corporate sponsorship of the guidelines and the financial associations of committee members. While all the criticism has been focused on direct sponsorship by the industry, very little has been said and there has been very little investigation with regards to organizations developing guidelines beyond their scope for political and financial purposes (44,45). By the same token, the best clinical practice guidelines are developed with the necessary financial and methodological support to ensure their quality.

Guidelines must be based on the practice of EBM, which is based on 4 basic contingencies originally defined by evidence-based practice (65). These 4 contingencies have been described in the EBM section, which include recognition of the patient's problem and the construction of a structured clinical question, effective and extensive search of the medical literature to obtain the best available evidence, critical appraisal of the evidence, and, finally, integration of the evidence in patient decision making to determine the best clinical care of the patient. The National Health and Medical Research Council (173) described 9 basic principles in development of the guidelines:

- Outcomes (survival rates to quality-of-life attributes)
- Best available evidence (according to its quality, relevance and strength)
- Appropriate systems to synthesize the available evidence (judgment, experience and good sense)
- Multidisciplinary process of development
- Flexibility and adaptability
- Cost-effectiveness of treatments
- Appropriate dissemination
- Evaluation of implementation and impact of guidelines

- Appropriate revision of the guidelines on a regular basis

Shaneyfelt et al (174) recommended the following criteria for appropriate guideline development and adherence:

1. Purpose of the guideline is specified
2. Rationale and importance of the guideline are explained
3. The participants in the guideline development process and their areas of expertise are specified
4. Targeted health problem or technology is clearly defined
5. Targeted patient population is specified
6. Intended audience or users of the guideline are specified
7. The principal preventive, diagnostic, or therapeutic options available to clinicians and patients are specified
8. The health outcomes are specified
9. The method by which the guideline underwent external review is specified
10. An expiration date or date of scheduled review is specified
11. Method of identifying scientific evidence is specified
12. Time period from which evidence is reviewed is specified
13. The evidence used is identified by citation and referenced
14. Method of data extraction is specified
15. Method for grading or classifying the scientific evidence is specified
16. Formal methods of combining evidence or expert opinion are used and described
17. Benefits and harms of specific health practices are specified
18. Benefits and harms are quantified
19. The effect on health care costs from specific health practices is specified
20. Costs are quantified
21. The role of value judgments used by the guideline developers in making recommendations is discussed
22. The role of patient preferences is discussed
23. Recommendations are specific and apply to the stated goals of the guideline
24. Recommendations are graded according to the strength of the evidence
25. Flexibility in the recommendations is specified

They evaluated 279 guidelines covering a wide range of topics (174). Overall, the mean (SD) number of standards satisfied out of 25, was 10.77 (3.71), or 43.1% with a range of 2 to 24. However, they noted that guidelines showed significant improvement from 1985 to 1997, but still only 50.4% of the standards were met, on average, for each guideline in 1997. Overall adherence to methodological standards on guideline development and format was only fair in 51%. While guidelines are developed to improve health outcomes, outcomes of interest were specified only in 40%. In addition, fewer than 50% described the patient population to which the guideline applied, while slightly greater than 50% described the intended audience of the guideline. While mean adherence to standards by each guideline improved from 36.9% in 1985 to 50.4% in 1997, there was little improvement over time and adherence to standards on identification and summary of evidence from 34.6% prior to 1990 to 36.1% after 1995. There was no difference in the mean number of standards satisfied by guidelines produced by subspecialty medical societies, general medical societies or government agencies.

The AGREE (Appraisal of Guidelines Research and Evaluation) also describes the critical assessment of clinical data of guideline development (175-177). Utilizing the AGREE evaluation, the Occupational Medicine Practice Guidelines developed by the American College of Occupational and Environmental Medicine (ACOEM) showed average domain scores in rigor of development of 26.59, application of 31.48, editorial independence of 29.17, and stakeholder involvement of 46.06, whereas they were 86.81 for clarity and presentation (38).

Consequently, the guideline development process should be precise and rigorous to ensure that the results are reproducible and not vague (178-180). In reference to multiple guidelines existent in today's medicine, specifically in interventional pain management, they impact the manner in which patient care is assessed by peer review and often serve as basis for payor decision making regarding the delivery of interventional pain management techniques to patients. In essence, the ACOEM guidelines have been adapted by several compensation systems as a standard for evaluation and management of work injuries (181) and Hayes Guidelines have been accepted by many private insurers (32). It is generally accepted in the medical and research community that peer review and testing

of practice guidelines should be performed prior to their acceptance as being valid and their subsequent utilization in wide arena of clinical practice (182,183). Consequently, the guidelines must be developed by systematically acquiring, analyzing, and transferring research findings into clinical, management, and policy arenas (106). The process involves:

- Developing the question in a way that it can be answered by a systematic review with specification of the populations, settings, problems, interventions, and outcomes of interest.
- Establishment of eligibility criteria.
- Extensive and appropriate search of the literature to capture any and all of the evidence.
- Review of abstracts or publications to determine eligibility of studies.
- Reviewing the retained studies to determine final eligibility.
- Abstracting data on these studies into evidence tables.
- Determining the quality of studies and the overall strength of evidence.
- Synthesizing and combining data from evidence tables.
- Writing a draft review, to be subjected to peer review, editing and revising, and producing the final review.

Guidelines for Guidelines (184) developed in a WHO series identified 19 components as follows:

1) Priority Setting
2) Group composition
3) Declaration and avoidance of conflicts of interest
4) Group processes
5) Identification of important outcomes including cost
6) Explicit definition of the question and eligibility criteria
7) Type of study designs for different types of questions
8) Identification of evidence
9) Synthesis and presentation of evidence
10) Specification and integration of values
11) Making judgments about desirable and undesirable effects
12) Taking account of equity
13) Grading evidence and recommendations
14) Taking account of costs
15) Applicability, transferability and adaptation of guidelines
16) Structure of reports
17) Methods of peer review
18) Planned methods of dissemination and implementation
19) Evaluation of the impact of the guideline

Grading the strength of recommendations and quality of evidence in clinical guidelines has been changing (185). While grading the strength of recommendations and the quality of underlying evidence enhances the usefulness of clinical guidelines, the profusion of guideline grading systems undermines the value of the grading exercise (185,186). The GRADE working group, an acronym for Grading of Recommendations, Assessment, Development and Evaluation, recommended grading quality and strength of evidence. The steps in this approach were to make sequential judgments about the quality of evidence across studies for each important outcome, which outcomes were critical to a decision, the overall quality of evidence across those critical outcomes, the balance between benefits and harms, and the strength of recommendations.

Table 7 illustrates the sequential process for developing guidelines, Table 8 shows criteria for assigning grade of evidence, whereas Table 9 shows definitions of grades of evidence. The GRADE system enables more consistent judgments, and communication of such judgments can support better informed choices in health care. Table 10 summarizes differences between GRADE and other systems, whereas Table 11 illustrates summary points to apply GRADE system in evidence synthesis for guidelines.

Guyatt et al (185) also established grading strength of recommendations and quality of evidence in clinical guidelines. This working group examined currently available systems and ultimately modified an approach formulated by the International Grade Group. The Task Force from the American College of Chest Physicians (ACCP) developed criteria that defined an optimal grading system. Based on the philosophy that guidelines panels should make recommendations to administer, or not administer, an intervention on the basis of tradeoffs between benefits on the one hand, and risks, burdens, and potentially, costs on the other, the Task Force chose to classify recommendations into 2 levels, strong and weak, as illustrated in Table 12. Consequently, if guideline panels are very certain that benefits do, or do not, outweigh the risks and burdens, they will make a strong recommendation, Grade I. At the same time, if the panels think that the benefits and the risks and burdens are finely balanced or, if ap-

Table 7. *Sequential process for developing guidelines.*

First steps

1. *Establishing the process*—For example, prioritizing problems, selecting a panel, declaring conflicts of interest, and agreeing on group processes.

Preparatory steps

2. *Systematic review*—The first step is to identify and critically appraise or prepare systematic reviews of the best available evidence for all important outcomes.
3. *Prepare evidence profile for important outcomes*—Profiles are needed for each subpopulation or risk group, based on the results of systematic reviews, and should include a quality assessment and a summary of findings.

Grading quality of evidence and strength of recommendations

4. *Quality of evidence for each outcome*—Judged on information summarized in the evidence profile and based on the criteria in Table 2.
5. *Relative importance of outcomes*—Only important outcomes should be included in evidence profiles. The included outcomes should be classified as critical or important (but not critical) to a decision.
6. *Overall quality of evidence*—The overall quality of evidence should be judged across outcomes based on the lowest quality of evidence for any of the critical outcomes.
7. *Balance of benefits and harms*—The balance of benefits and harms should be classified as net benefits, trade-offs, uncertain trade-offs, or no net benefits based on the important health benefits and harms.
8. *Balance of net benefits and costs*—Are incremental health benefits worth the costs? Because resources are always limited, it is important to consider costs (resource utilization) when making a recommendation.
9. *Strength of recommendation*—Recommendations should be formulated to reflect their strength—that is, the extent to which one can be confident that adherence will do more good than harm.

Subsequent steps

10. *Implementation and evaluation*—For example, using effective implementation strategies that address barriers to change, evaluation of implementation, and keeping up to date.

Source: Atkins et al (Ref. 186)

Table 8. *Criteria for assigning grade of evidence*

Type of evidence
Randomized trial = high
Observational study = low
Any other evidence = very low

Decrease grade if:
- Serious (– 1) or very serious (– 2) limitation to study quality
- Important inconsistency (– 1)
- Some (– 1) or major (– 2) uncertainty about directness
- Imprecise or sparse data (– 1)
- High probability of reporting bias (– 1)

Increase grade if:
- Strong evidence of association—significant relative risk of > 2 (< 0.5) based on consistent evidence from two or more observational studies, with no plausible confounders (+1)46
- Very strong evidence of association—significant relative risk of > 5 (< 0.2) based on direct evidence with no major threats to validity (+2)46
- Evidence of a dose response gradient (+1)
- All plausible confounders would have reduced the effect (+1)

Source: Atkins et al (Ref. 186)

Table 9. *Imprecise or sparse data.*

There is not an empirical basis for defining imprecise or sparse data. Two possible definitions are:

- Data are sparse if the results include just a few events or observations and they are uninformative
- Data are imprecise if the confidence intervals are sufficiently wide that an estimate is consistent with either important harms or important benefits.

These different definitions can result in different judgments. Although it may not be possible to reconcile these differences, we offer the following guidance when considering whether to downgrade the quality of evidence due to imprecise or sparse data:

- The threshold for considering data imprecise or sparse should be lower when there is only one study. A single study with a small sample size (or few events) yielding wide confidence intervals spanning both the potential for harm and benefit should be considered as imprecise or sparse data
- Confidence intervals that are sufficiently wide that, irrespective of other outcomes, the estimate is consistent with conflicting recommendations should be considered as imprecise or sparse data

Source: Atkins et al (Ref. 186)

Table 10. *Comparison of GRADE and other systems.*

Factor	Other systems	GRADE	Advantages of GRADE system*
Definitions	Implicit definitions of quality (level) of evidence and strength of recommendation	Explicit definitions	Makes clear what grades indicate and what should be considered in making these judgments
Judgments	Implicit judgments regarding which outcomes are important, quality of evidence for each important outcome, overall quality of evidence, balance between benefits and harms, and value of incremental benefits	Sequential, explicit judgments	Clarifies each of these judgments and reduces risks of introducing errors or bias that can arise when they are made implicitly
Key components of quality of evidence	Not considered for each important outcome. Judgments about quality of evidence are often based on study design alone	Systematic and explicit consideration of study design, study quality, consistency, and directness of evidence in judgments about quality of evidence	Ensures these factors are considered appropriately
Other factors that can affect quality of evidence	Not explicitly taken into account	Explicit consideration of imprecise or sparse data, reporting bias, strength of association, evidence of a dose-response gradient, and plausible confounding	Ensures consideration of other factors
Overall quality of evidence	Implicitly based on the quality of evidence for benefits	Based on the lowest quality of evidence for any of the outcomes that are critical to making a decision	Reduces likelihood of mislabelling overall quality of evidence when evidence for a critical outcome is lacking
Relative importance of outcomes	Considered implicitly	Explicit judgments about which outcomes are critical, which ones are important but not critical, and which ones are unimportant and can be ignored	Ensures appropriate consideration of each outcome when grading overall quality of evidence and strength of recommendations
Balance between health benefits and harms	Not explicitly considered	Explicit consideration of trade-offs between important benefits and harms, the quality of evidence for these, translation of evidence into specific circumstances, and certainty of baseline risks	Clarifies and improves transparency of judgments on harms and benefits
Whether incremental health benefits are worth the costs	Not explicitly considered	Explicit consideration after first considering whether there are net health benefits	Ensures that judgments about value of net health benefits are transparent
Summaries of evidence and findings	Inconsistent presentation	Consistent GRADE evidence profiles, including quality assessment and summary of findings	Ensures that all panel members base their judgments on same information and that this information is available to others
Extent of use	Seldom used by more than one organization and little, if any empirical evaluation	International collaboration across wide range of organizations in development and evaluation	Builds on previous experience to achieve a system that is more sensible, reliable, and widely applicable

Source: Atkins et al (Ref. 186)

Table 11. *Criteria for an Optimal Grading System*

Criteria	Description
1	Separation of grades of recommendations from quality of evidence
2	Simplicity and transparency for clinician consumer
3	Sufficient (but not too many) categories
4	Explicitness of methodology for guideline developers
5	Simplicity for guideline developers
6	Consistent with general trends in grading systems
7	Explicit approach to different levels of evidence for different outcomes

Source: Guyatt et al (Ref. 185)

Table 12. *Grading recommendations*.

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

Source: Guyatt et al (Ref. 185)

preciable, uncertainties exist above the magnitude of the benefits and risks, they must offer a weak, Grade 2, recommendation. However, guideline panels must consider a number of factors in grading recommendations as illustrated in Table 13.

For over 25 years, a growing number of organizations have employed various systems to grade the quality of evidence and the strength of recommendation. This has led to confusion and criticisms of each other. Recommendations are based on the judgment about the quality of evidence and the balance of benefits and risks. Frequently these judgments are made implicitly rather than explicitly and judgments about the quality of evidence are confused

Table 13. *Factors panels should consider in deciding on a strong or weak recommendation.*

Issue	Example
Methodological quality of the evidence supporting estimates of likely benefit, and likely risk, inconvenience, and costs	Many high-quality randomized trials have demonstrated the benefit of therapy with inhaled steroids in patients with asthma, while only case series have examined the utility of pleurodesis in patients with pneumothorax
Importance of the outcome that treatment prevents	Preventing postphlebitic syndrome with thrombolytic therapy in DVT patients in contrast to preventing death from PE
Magnitude of treatment effect	Clopidogrel vs aspirin leads to a smaller stroke reduction in patients with TIAs (RRR,198.7%) than anticoagulation vs placebo in patients with AF (RRR, 68%)
Precision of estimate of treatment effect	ASA therapy vs placebo in AF patients has a wider confidence interval than ASA therapy for stroke prevention in patients with TIA
Risks associated with therapy	ASA and clopidogrel for anticoagulation therapy in patients with acute coronary syndromes has a higher risk for bleeding than ASA alone
Burdens of therapy	Therapy with adjusted-dose warfarin is associated with a higher burden than that with aspirin; warfarin requires monitoring the intensity of anticoagulation and a relatively constant dietary vitamin K intake
Risk of target event	Some surgical patients are at very low risk of post-operative DVT and PE while other surgical patients have considerably higher rates of DVT and PE
Costs	Clopidogrel has a much higher cost in patients with TIA than does aspirin
Varying values	Most young, healthy people will put a high value on prolonging their lives (and thus incur suffering to do so); the elderly and infirm are likely to vary in the value they place on prolonging their lives (and may vary in the suffering they are ready to experience to do so)

*DVT = deep vein thrombosis; PE = pulmonary embolism; TIA = transient ischemic attack; AF = atrial fibrillation; ASA = aspirin.

Source: Guyatt et al (Ref. 185)

with judgments about the balance of benefits and risks (187). Further, numerous systems that are used to grade the quality of evidence and the strength of recommendations confuse these judgments by equating the strength of recommendation with the quality of evidence, with some recommendations of strong based on high quality evidence, without explicitly considering the balance of benefits and risks. However, quality of evidence, while essential, is not sufficient for making judgments about the strength of a recommendation (188). The relationship between the quality of evidence and strength of recommendation are complex issues and require careful consideration of numerous factors. In a series of 16 international meetings and correspondence over 5 years, the GRADE working group has derived a set of criteria to assess the quality of evidence (Table 14) and strength of recommendations (Table 15) (185,186,189-192).

The World Health Organization (WHO), like many other organizations around the world, has recognized the need to use more vigorous processes to ensure that health care recommendations are informed by the best available research evidence. Consequently, a series of 16 reviews that have been prepared as background for advice from the WHO Advisory Committee on health research to WHO on how to achieve this have been published (105,155,156,184,187,193-203). These articles titled "Improving the Use of Research Evidence in Guideline Development" are as follows:

1) Guidelines for Guidelines (184)
2) Priority setting (193)
3) Group composition and consultation process (194)
4) Managing conflicts of interest (155)
5) Group processes (195)
6) Determining which outcomes are important (196)
7) Deciding what evidence to include (197)
8) Synthesis and presentation of evidence (105)
9) Grading evidence and recommendations (187)
10) Integrating values and consumer involvement (198)
11) Incorporating considerations of cost-effectiveness, affordability and resource implications (199)
12) Incorporating considerations of equity (200)
13) Applicability, transferability and adaptation (201)
14) Reporting guidelines (202)

Table 14. *GRADE quality assessment criteria.*

Quality of evidence	Study design	Lower if *	Higher if *
High	Randomised trial	**Study quality**: -1 Serious limitations -2 Very serious limitations -1 Important **inconsistency** **Directness**: -1 Some uncertainty -2 Major uncertainty -1 **Sparse data** -1 High probability of **Reporting bias**	**Strong association**: +1 Strong, no plausible confounders, consistent and direct evidence** +2 Very strong, no major threats to validity and direct evidence*** +1 Evidence of a **Dose response** gradient +1 All **plausible confounders** would have reduced the effect
Moderate			
Low	Observational study		
Very low			

* 1 = move up or down one grade (for example from high to intermediate) 2 = move up or down two grades (for example from high to low)

** A statistically significant relative risk of >2 (< 0.5), based on consistent evidence from two or more observational studies, with no plausible confounders

*** A statistically significant relative risk of > 5 (< 0.2) based on direct evidence with no major threats to validity

Source: Schünemann et al (Ref. 187)

Table 15. *Decisions about the strength of a recommendation.*

Factors that can weaken the strength of a recommendation	Explanation
Lower quality evidence	Will create greater uncertainty about the size of the (relative) effects (benefits and harms)
Uncertainty about the balance of benefits versus harms and burdens	Uncertainty about the baseline risk, prevalence of a problem or health status, which could affect the size of the (absolute) effects
Uncertainty or differences in values	Uncertainty about the relative importance of the benefits and downsides to those affected, or differences in how important they are to different people, which could affect the balance between the benefits versus harms and burden
Marginal net benefits or downsides	The anticipated net benefits or downsides are small (and uncertain)
Uncertainty about whether the net benefits are worth the costs	Uncertainty related to lack of information about the cost or whether the resource expenditure is justified by the anticipated benefit

Source: Schünemann et al (Ref. 187)

15) Disseminating and implementing guidelines (203)
16) Evaluation (156)

Conclusion

Evidence-based medicine, systematic reviews, and guidelines are an inevitable part of the practice of interventional pain management which motivates the search for answers to numerous questions related to the costs and quality of health care and access to care. The modern physician realizes that scientific and relevant evidence is essential in clinical care, policy-making, dispute resolution, and law. Thus, evidence-based practice brings pertinent, trustworthy information by systematically acquiring, analyzing and transferring research findings into clinical, management, and policy arenas. Appropriately derived evidence-based medicine, systematic reviews and guidelines are an essential part of modern interventional pain management.

Acknowledgements

The authors wish to thank Tonie M. Hatton and Diane E. Neihoff, transcriptionists, for their assistance in the preparation of this manuscript.

References

1. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: Increasing the value of clinical research for decision-making in clinical and health policy. *JAMA* 2003; 290:1624-1632.
2. Wennberg J, Gittelsohn A. Small area variation in health care delivery. *Sci Am* 1973; 182:1102-1108.
3. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? *Milbank Q* 1998; 76:517-563.
4. Manchikanti L, Giordano J. Physician payment 2008 for interventionalists: Current state of health care policy. *Pain Physician* 2007; 10:607-626.
5. Manchikanti L, Boswell MV. Interventional techniques in ambulatory surgical centers: A look at the new payment system. *Pain Physician* 2007; 10:627-650.
6. Manchikanti L. Health care reform in the United States: Radical surgery needed now more than ever. *Pain Physician* 2008; 11:13-42.
7. Eddy DM, Billings J. The quality of medical evidence: Implications for quality of care. *Health Aff (Millwood)* 1988; 7:19-32.
8. McNeil BJ. Shattuck lecture—hidden barriers to improvement in the quality of care. *N Engl J Med* 2001; 345:1612-1620.
9. *Evidence-based Practice Centers Partner's Guide*. Prepared for Agency for Healthcare Research and Quality by the EPC Coordinating Center. January 2005. www.ahrq.gov/Clinic/epcpartner/epcpartner.pdf
10. Kohn LT, Corrigan JM, Donaldson MS, editors. To Err Is Human: Building a Safer Health System. A Report of the Committee on Quality of Health Care in America. Institute of Medicine. National Academy Press, Washington, DC, 2000.
11. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press, 2001.
12. Manchikanti L, Boswell MV, Giordano J. Evidence-based interventional pain management: Principles, problems, potential and applications. *Pain Physician* 2007; 10:329-356.
13. Manchikanti L, Singh V, Bakhit CE, Fellows B. Interventional techniques in the management of chronic pain. Part 1.0. *Pain Physician* 2000; 3:7-42.
14. Manchikanti L, Singh V, Kloth DS, Slipman CW, Jasper JF, Trescot AM, Varley KG, Atluri SL, Giron C, Curran MJ, Rivera JJ, Baha A, Bakhit CE, Reuter M. Interventional techniques in the management of chronic pain: Part 2.0. *Pain Physician* 2001; 4:24-96.
15. Manchikanti L, Staats PS, Singh V, Schultz DM, Vilims BD, Jasper JF, Kloth DS, Trescot AM, Hansen HC, Falasca TD, Racz GB, Deer T, Burton AW, Helm S, Lou L, Bakhit CE, Dunbar EE, Atluri SL, Calodney AK, Hassenbusch S, Feler CA. Evidence-based practice guidelines for interventional techniques in the management of chronic spinal pain. *Pain Physician* 2003; 6:3-80.
16. Boswell MV, Shah RV, Everett CR, Sehgal N, Mckenzie-Brown AM, Abdi S, Bowman RC, Deer TR, Datta S, Colson JD, Spillane WF, Smith HS, Lucas-Levin LF, Burton AW, Chopra P, Staats PS, Wasserman RA, Manchikanti L. Interventional techniques in the management of chronic spinal pain: Evidence-based practice guidelines. *Pain Physician* 2005; 8:1-47.
17. Boswell MV, Trescot AM, Datta S, Schultz DM, Hansen HC, Abdi S, Sehgal N, Shah RV, Singh V, Benyamin RM, Patel VB, Buenaventura RM, Colson JD, Cordner HJ, Epter RS, Jasper JF, Dunbar EE, Atluri SL, Bowman RC, Deer TR, Hansen HC, Staats PS, Smith HS, Burton AW, Kloth DS, Giordano J, Manchikanti L. Interventional techniques: Evidence-based practice guidelines in the management of chronic spinal pain. *Pain Physician* 2007; 10:7-111.
18. Manchikanti L. Evidence-based interventional pain medicine: Is there any evidence? *Pain Physician* 2002; 5:1-7.
19. Manchikanti L, Abdi S, Lucas LF. Evidence synthesis and development of guidelines in interventional pain management. *Pain Physician* 2005; 8:73-86.
20. Manchikanti L, Heavner J, Racz GB, Mekhail NA, Schultz DM, Hansen HC, Singh V. Methods for evidence synthesis in interventional pain management. *Pain Physician* 2003; 6:89-111.
21. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ* 1996; 312:71-72.
22. Haynes RB, Sackett RB, Gray JMA, Cook DC, Guyatt GH. Transferring evidence from research into practice, 1: The role of clinical care research evidence in clinical decisions. *ACP J Club* 1996; 125: A14-A15.
23. Sackett D, Richardson WS, Rosenberg W, Haynes B. *Evidence-based Medicine*. Churchill Livingstone, London, 1996.
24. Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of EBM and patient choice. *ACP J Club.* 2002;136: A11-A14.
25. Guyatt G, Drummond R. Part 1. The basics: Using the medical literature. 1A. Introduction: The philosophy of evidence-based medicine. In. *Users' Guides to the Medical Literature. A Manual for Evidence-based Clinical Practice*. EBM Working Group. AMA Press, 2002, pp 3-12.
26. Guyatt G, Drummond R. Preface. In *Users' Guides to the Medical Literature. A Manual for Evidence-based Clinical Practice*. EBM Working Group. AMA Press, 2002, pp xiii-xvi.
27. Guyatt GH, Rennie D. Users' Guides to the Medical Literature [editorial]. *JAMA* 1993; 270:2096-2097.
28. Guyatt G, Drummond R. *Users' Guides to the Medical Literature. A Manual for Evidence-based Clinical Practice*. EBM Working Group. AMA Press, 2002.
29. The Cochrane Collaboration. *Evidence Health Care*. Available at: www.cochrane.org/docs/ebm.htm.
30. American College of Occupational and Environmental Medicine, Glass LS. *Occupational Medicine Practice Guidelines: Evaluation and Management Of Common Health Problems And Functional Recovery Of Workers, 2nd ed*. OEM Press, Beverly Farms, 2004.
31. Dennison PL. *Official Disability Guidelines*. 13th Ed. Work Loss Data Institute, 2008.
32. HAYES, Inc. Independent Health Technology Assessment Company. www.hayesinc.com
33. Haldeman S, Dagensai S. A supermarket approach to the evidence-informed management of chronic low back pain. *Spine J* 2008; 8:1-7.
34. Haldeman S, Dagenais S. What have we learned about the evidence-informed management f chronic low back pain? *Spine J* 2008; 8:266-277.
35. Bigos SJ, Boyer OR, Braen GR, Brown K, Deyo R. *Acute Low Back Problems in Adults*. Clinical Practice Guideline Number 4. AHCPR Publication No. 95-0642. Agency for Health Care Policy and Re-

search, Public Health Service, US Department of Health and Human Services, Rockville, December 1994.

36. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H , Zanoli G. Chapter 4: European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006; 15:S192-S300.
37. Budetti PP. Market justice and US health care. *JAMA* 2008; 299:92-94.
38. Cates JR, Young DN, Bowerman DS, Porter RC. An independent AGREE evaluation of the *Occupational Medicine Practice Guidelines. Spine J* 2006; 6:72-77.
39. Helm S. California workers' compensation system: Are occupational medicine practice guidelines sufficient for the treatment of chronic spinal pain or do they require supplementation by guidelines for interventional techniques? *Pain Physician* 2004; 229-238.
40. Canan P, Kretzmann M, Hennessy M, Pring GW. Using law ideologically: The conflict between economic and political liberty. *J Law Polit* 1992; 8:539-58.
41. Gatchel RJ, McGeary D. Cochrane collaboration-based reviews of healthcare interventions: Are they unequivocal and valid scientifically, or simply nihilistic? *Spine J* 2002; 2:315-319.
42. Mowatt G, Shirran L, Grimshaw JM, Rennie D, Flanagin A, Yank V, MacLennan G, Gotzsche PC, Bero LA. Prevalence of honorary and ghost authorship in Cochrane Reviews. *JAMA* 2002; 287:2769-2771.
43. Errico TJ, Gatchel RJ, Schofferman J, Benzel EC, Faciszewski T, Eskay-Auerbach M, Wang JC. A fair and balanced view of spine fusion surgery. *Spine J* 2004; 4:S129-S138.
44. LaDou J, Teitelbaum DT, Egilman DS, Frank AL, Kramer SN, Huff J. American College of Occupational and Environmental Medicine (ACOEM): A professional association in service to industry. *Int J Occup Environ Health* 2007; 13:404-426.
45. Greenberg M. Commentary on effects of exposure to industry influence on ACOEM. *Int J Occup Environ Health* 2007; 13:427.
46. Niemistö L, Kalso E, Malmivaara A, Seitsalo S, Hurri H. Radiofrequency denervation for neck and back pain: A systematic review within the framework of the Cochrane Collaboration Back Review Group. *Spine* 2003; 28:1877-1888.
47. Manchikanti L, Kloth DS, Singh V. The role of guidelines in interventional pain medicine: Let us separate apples and oranges. *Pain Physician* 2001; 4:13-23.
48. Jenicek M. How do we see medicine, health and disease? A basic set of rules and fundamental paradigms. In *Foundations of Evidence-Based Medicine*. The Parthenon Publishing Group, Inc., New York, 2005, pp, 3-13.
49. Osler W, McCrae T. *The Principles and Practice of Medicine, Designed for the Use of Practitioners and Students of Medicine*. 10th ed., Appleton, New York, 1925.
50. Eisenberg JM, What does evidence mean? Can the law and medicine be reconciled? *J Health Polit Policy Law* 2001; 26:369-381.
51. *New Illustrated Webster's Dictionary of the English Language*. PAMCO Publishing Co., New York, 1992.
52. Giordano J. Pain as disease and illness: Part 2 – Structure and function of the ethics of pain medicine. *Prac Pain Management* 2006; 6:65-68.
53. Jonas WB. The evidence house: How to build an inclusive base for complementary medicine. *West J Med.* 2001; 175:79–80.
54. Jenicek M. The work of physicians with individuals and communities. Epidemiology and other partners in evidence-based medicine. In *Foundations of Evidence-Based Medicine*. The Parthenon Publishing Group, Inc., New York, 2005, pp, 15-40.
55. Gonzalez EG, Materson (eds). *The Nonsurgical Management of Acute Low Back Pain*. Demos Vermane, New York 1997.
56. Napodano RJ. *Values in Medical Practice*. Human Sciences Press, New York, 1986.
57. Giordano J. Pain, the patient and the physician: Philosophy and virtue ethics in pain medicine. In Schatman M (ed.) *Ethics of Chronic Pain Management*. Taylor-Francis, NY, 2006 pp 1-18.
58. Giordano J. Moral agency in pain medicine: Philosophy, practice and virtue. *Pain Physician* 2006; 9:71-76.
59. Giordano J. On knowing: Domains of knowledge and intellectual virtue in practical pain management. *Prac Pain Management* 2006; 6:65-67.
60. Maricich Y, Giordano J. Pain, suffering and the ethics of pain medicine: Is a deontic foundation sufficient? *Am J Pain Manage* 2007; 17:130-138.
61. Giordano J. Pain research: Can paradigmatic revision bridge the needs of medicine, scientific philosophy and ethics? *Pain Physician* 2004; 7:459-463.
62. Jonas WB. Building an evidence house: Challenges and solutions to research in complementary and alternative medicine. *Forsch Komplementärmed Klass Naturheilkd* 2005; 12:159-167.
63. Giordano J. Good as gold? The random randomized controlled trial: Pragmatic and ethical issues in pain research. *Am J Pain Manage* 2006; 16;68-71.
64. Giordano J, Engebretson J, Garcia MK. Challenges to CAM research: Focal issues influencing integration into a cancer care model. *J. Integrative Cancer Ther* 2005; 4:210-218.
65. Hatala R, Guyatt G. Evaluating the teaching of evidence-based medicine. *JAMA* 2002; 288:1110-1112.
66. Keckley PH. EBM in managed care: A survey of current and emerging strategies. *Mud Gen Med* 2004; 6:56.
67. Bogduk N, Govind J. *Medical Management of Acute Lumbar Radicular Pain: An Evidence-Based Approach*. Cambridge Press, Newcastle, 1999.
68. Bogduk N, McGuirk B. *Medical Management of Acute and Chronic Low Back Pain. An Evidence-Based Approach: Pain Research and Clinical Management*. Elsevier Science BV, Amsterdam:2002.
69. Bogduk N, McGuirk B. *Management of Acute and Chronic Neck Pain. An Evidence-Based Approach*. Elsevier, 2006.
70. Bogduk N. *Medical Management of Acute Cervical Radicular Pain: An Evidence-based Approach*, 1st edition. University of Newcastle, Newcastle Bone and Joint Institute, Australia, 1999.
71. Deyo RA, Psaty BM, Simon G, Wagner EH, Omenn GS. The messenger under attack – intimidation of researchers by special-interest groups. *N Engl J Med* 1997; 336:1176-1180.
72. Feinstein AR. *Clinical Judgment*.. Robert E. Krieger, Huntington, NY 1967:220-222.
73. Angell M. Evaluating the health risks of breast implants: The interplay of medical science, the law, and public opinion. *N Engl J Med* 1996; 334:1513-1518.
74. Deyo RA, Cherkin DC, Loeser JD, Bigos SJ, Ciol MA. Morbidity and mortality in association with operations on the lumbar spine: The influence of age,

diagnosis, and procedure. *J Bone Joint Surg Am* 1992; 74:536-543.

75. Deyo RA, Ciol MA, Cherkin DC, Loeser JD, Bigos SJ. Lumbar spinal fusion: A cohort study of complications, reoperations, and resource use in the Medicare population. *Spine* 1993; 18:1463-1470.
76. Turner JA, Ersek M, Herron L, Haselkorn J, Kent D, Ciol MA, Devo R.Patient outcomes after lumbar spinal fusions. *JAMA* 1992; 268:907-911.
77. Pring GW, Canan P. SLAPPs: An introduction for bench, bar and bystanders. *Bridgeport Law Rev* 1992; 12:937-62.
78. 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.
79. Furlan AD, Clarke J, Esmail R, Sinclair S, Irvin E, Bombardier C. A critical review of reviews on the treatment of chronic low back pain. *Spine* 2001; 26:E155-E162.
80. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, Lau J. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002; 287:2973-2982.
81. Hopayian K. The need for caution in interpreting high quality systematic reviews. *Br Med J* 2001; 323:681-684.
82. Wright RW, Brand RA, Dunn W, Spindler KP. How to write a systematic review. *Clin Orthop Relat Res* 2007; 455:23-29.
83. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine*. 2nd ed. : Churchill Livingstone, Edinburgh, UK2000.
84. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol* 1995; 48:167-171.
85. Mulrow C, Langhorne P, Grimshaw J. Integrating heterogeneous pieces of evidence in systematic reviews. *Ann Intern Med* 1997; 127:989-995.
86. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10:31-56.
87. McAlister FA, Clark HD, van Walraven C, Straus SE, Lawson FM, Moher D, Mulrow CD. The medical review article revisited: Has the science improved? *Ann Intern Med* 1999; 131:947-951.
88. Mulrow CD. The medical review article: State of the science. *Ann Intern Med* 1987; 106:485–488.
89. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *JAMA* 1992; 268:240–248.
90. Montori VM, Swiontkowski MF, Cook DJ. Methodologic issues in systematic reviews and meta-analyses. *Clin Orthop Relat Res* 2003; 413:43–54.
91. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: Synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997; 126:376-380.
92. Rotstein D, Laupacis A. Differences between systematic reviews and health technology assessments: A trade-off between the ideals of scientific rigor and the realities of policy making. *Int J Technol Assess Health Care* 2004; 20:177-183.
93. Lohr KN, Carey TS. Assessing "best evidence": Issues in grading the quality of studies for systematic reviews. *Jt Comm J Qual Improv* 1999; 25:470-479.
94. Clark HD, Wells GA, Huet C, McAlister FA, Salmi LR, Fergusson D, Laupacis A. Assessing the quality of randomized trials: Reliability of the Jadad scale. *Control Clin Trials* 1999; 20:448-452.
95. Chalmers TC, Smith H Jr., Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981; 2:31-49.
96. 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.
97. 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.
98. Coulter ID. Evidence summaries and synthesis: Necessary but insufficient approach for determining clinical practice of integrated medicine? *Integr Cancer Ther* 2006; 5:282-286.
99. Crowther MA, Cook DJ. Trials and tribulations of systematic reviews and meta-analyses. *Hematology Am Soc Hematol Educ Program* 2007; 2007:493-497.
100. Mulrow CD, Cook DJ, Davidoff F. Systematic reviews: critical links in the great chain of evidence. *Ann Intern Med* 1997; 126:389-391.
101. Oxman AD, Cook DJ, Guyatt GH. Users' guide to the medical literature, VI: How to use an overview. Evidence-based Medicine Working Group. *JAMA* 1994; 272:1367-1371.
102. Klassen TP, Jadad AR, Moher D. Guides for reading and interpreting systematic reviews: I. Getting started. *Arch Pediatr Adolesc Med* 1998; 152:700-704.
103. Jadad AR, Moher D, Klassen TP. Guides for reading and interpreting systematic reviews: II. How do the authors find the studies and assess their quality? *Arch Pediatr Adolesc Med* 1998; 152:812-817.
104. Moher D, Jadad AR, Klassen TP. Guides for reading and interpreting systematic reviews: III. How did the authors synthesize the data and make their conclusions? *Arch Pediatr Adolesc Med* 1998; 152:915-920.
105. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. *Health Res Policy Syst* 2006; 4:20.
106. Lohr KN. Rating the evidence of scientific evidence: Relevance for quality improvement programs. *Int J Qual Health Care* 2004; 16:9-18.
107. Lohr KN, Eleazar K, Mauskopf J. Health policy issues and applications for evidence-based medicine and clinical practice guidelines. *Health Policy* 1998; 46:1-19.
108. Oxman AD. Systematic reviews: Checklists for review articles. *BMJ* 1994; 309:648-651.
109. Altman DG. Confidence intervals in research evaluation. *Ann Interen Med* 1991; 116:A28.
110. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L. *Systems to Rate the Strength of Scientific Evidence*, Evidence Report, Technology Assessment No. 47. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality, 2002. www.thecre.com/pdf/ahrq-system-strength.pdf
111. Proposed Evaluation Tools for COMPUS [www.ccohta.ca/compus/compus_pdfCOMPUS_Evaluation_Methodology_draft_e.pdf]. Ottawa: Canadian Coordinating Office for Health Technology As-

sessment November 29, 2005.

112. Lavis JN, Davies HTO, Oxman AD, Denis JL, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform healthcare management and policy making. *Journal of Health Services Research and Policy* 2005, 10:35-48.
113. Glenton C, Underland V, Kho M, Pennick V, Oxman AD. Summaries of findings, descriptions of interventions, and information about adverse effects would make reviews more informative. *J Clin Epidemiol* 2006; 59:770-778.
114. Glasziou P, Oxman AD, Higgins J. Summary of Findings Tables within Cochrane Reviews: Draft Specification for Rev Man 5.0. December 2004. Obtaining a consensus on the content and methods of a Summary of Findings table for Cochrane Reviews. Report to the Cochrane Collaboration Steering Group 2005.
115. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmer TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994, 120:667-676.
116. Sacks HS, Reitman D, Pagano D, Kupelnick B. Metaanalysis: An update. *Mt Sinai J Med* 1996; 63:216-224.
117. 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.
118. Khan KS, Ter Riet G, Glanville J, Sowden AJ, Kleijnen J. *Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews*: University of York, NHS Centre for Reviews and Dissemination; York, England: 2000.
119. Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. *J Am Med Associ* 1998; 279:1566-1570.
120. Royle P, Waugh N. Literature searching for clinical and cost effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. *Health Technology Assessment* 2003, 7(34): [www.ncchta.org/execsumm/summ734.htm].
121. Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *BMJ* 1992; 305:786-788.
122. Haynes RB. Clinical review articles: should be as scientific as the articles they review. *BMJ* 1992; 304:330-331.
123. Peto R. Why do we need systematic overviews of randomized trials? *Stat Med* 1987;6:233-240.
124. Jadad AR, McQuay HJ. Searching the literature: Be systematic in your searching. *BMJ* 1993;307:66.
125. Dickersin K. The existence of publication bias and risk factors of its occurrence. *JAMA* 1990; 263:1385-1389.
126. Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, McIlroy W, Oxman AD. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993; 269:2749-2753.
127. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. *Int J Technol Assess Health Care*1996; 12:195-208.
128. Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol*1994; 140:300-301.
129. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. In (eds) Chalmers I, Altman DG. *Systematic Reviews*. BMJ Publishing Group; London, 1995:17-36.
130. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992; 45:255-265.
131. Rennie D, Bero L. The Cochrane Collaboration: Preparing, maintaining and disseminating systematic reviews of the effects of health care. *JAMA* 1995; 274:1935-1938.
132. Egger M, Smith GD, Phillips AN. Meta-analysis: Principles and procedures. *BMJ* 1997; 315:1533-1537.
133. Huque MF. Experiences with meta-analysis in NDA submissions. Proceedings of the Biopharmaceutical Section of the American Statistical Association 1988; 2:28-33.
134. Egger M, Davey Smith G. Meta-analysis: Potentials and promise. *BMJ* 1997; 315:1371-1374.
135. Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315:629-634.
136. Sterne JAC, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; 323:101-105.
137. Villar J, Piaggio G, Carroli G, Donna A. Factors affecting the comparability of meta-analyses and the largest trials results in perinatology. *J Clin Epidemiol* 1997; 50:997-1002.
138. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323:42-46.
139. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; 283:2008-2012.
140. Petticrew M, Song F, Wilson P, Wright K. Quality-assessed review of health care interventions and the Database of Abstracts of Reviews of Effectiveness (DARE). *Int J Technol Assess Health Care* 1999; 15:671-678.
141. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: A comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998; 280:278-280.
142. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, Stevens R. Systematic reviews and meta-analyses of treatment of asthma: critical evaluation. *BMJ* 2000; 320:537-540.
143. Olsen O, Middleton P, Ezzo J, Gøtzsche PC, Hadhazy V, Herxheimer A, Kleijnen J, McIntosh H. Quality of Cochrane reviews: Assessment of sample from 1998. *BMJ* 2001; 323:829-831.
144. Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper journals. *Eval Health Prof* 2002; 25:116-129.
145. Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, Pham B, Klassen TP. Assessing the quality of reports of randomized trials: Implications for the conduct of meta-analyses. *Health Technol Assess* 1999; 3:1-98.
146. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significant in meta-analysis: An empirical study of 125 meta-analyses. *Stat Med* 2000; 19:1707-1728.
147. Petitti DB. Approaches to heterogeneity in meta-analysis. *Stat Med* 2001; 20:3625-3633.
148. Steinbrook R. Guidance for guidelines. *N Engl J Med* 2007; 356:331-333.
149. Field MJ, Lohr KN (eds). *Clinical Practice Guidelines: Directions for a New Pro-*

gram. National Academy Press, Washington, DC, 1990.

150. www.guideline.gov/resources/guideline_resources.aspx
151. Gonzalez EG, Materson RS. The guidelines, the controversy, the book. In Gonzalez ER, Materson RS (eds). *The Nonsurgical Management of Acute Low Back Pain*. Demos Vermande, New York, 1997, pp 1-4.
152. American Academy of Pediatrics. *Report of the Committee on Immunization Procedures of the American Academy of Pediatrics*. American Academy of Pediatrics, Evanston, 1938.
153. Turk MA. Practice guidelines. In Gonzalez ER, Materson RS (eds). *The Nonsurgical Management of Acute Low Back Pain*. Demos Vermande, New York, 1997, pp 5-13.
154. Hasenfeld R, Shekelle PG. Is the methodological quality of guidelines declining in the US? Comparison of the quality of US Agency for Health Care Policy and Research (AHCPR) guidelines with those published subsequently. *Qual Saf Health Care* 2003; 12:428-434.
155. Boyd EA, Bero LA. Improving the use of research evidence in guideline development: 4. Managing conflicts of interest. *Health Res Policy Syst* 2006; 4:16.
156. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 16. Evaluation. *Health Res Policy Syst* 2006; 4:28.
157. Barnes DE, Bero LA. Scientific quality of original research articles on environmental tobacco smoke. *Tob Control* 1997; 6:19-26.
158. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *JAMA* 2003; 289:454-465.
159. Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Intern Med* 1996; 124:485-489.
160. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003; 326:1167-1170.
161. Wilson D, Heath D. He saw the tests as a violation of 'trusting desperate human beings'. In *The Seattle Times* Seattle; 2001.
162. Wilson D, Heath D. Uninformed consent: they call the place Mother Hutch. In *The Seattle Times* Seattle; 2001.
163. Dana J. Harm avoidance and financial conflict of interest. *J Med Ethics* 2003; Online Electronic Version: 1-18.
164. Bero LA. Tobacco industry manipulation of research. *Public Health Rep* 2005; 120:200-208.
165. Montori VM, Jaeschke R, Schunemann HJ, Bhandari M, Brozek JL, Devereaux PJ, Guyatt GH. Users' guide to detecting misleading claims in clinical research reports. *BMJ* 2004; 329:1093-1096.
166. Friedman PJ. The impact of conflict of interest on trust in science. *Sci Eng Ethics* 2002; 8:413-420.
167. Drazen JM, Koski G. To protect those who serve [editorial]. *N Engl J Med* 2000, 343:1643-1645.
168. Baltimore D. The worsening climate for biological research: Is the public losing faith in science? *Technol Review* 1989; 92:22.
169. Haerlin B, Parr D. How to restore public trust in science [commentary]. *Nature* 1999; 400:499.
170. Cohen JJ. Trust us to make a difference: Ensuring public confidence in the integrity of clinical research. *Acad Med* 2001; 76:209-214.
171. DeAngelis CD. Conflict of interest and the public interest. *JAMA* 2000; 284:2237-2238.
172. www.consensus.nih.gov/
173. 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.
174. 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.
175. Cluzeau FA, Burgers JS, Brouwers M, Grol R, Mäkelä M, Littlejohns P, Grimshaw J, Hunt C, Asua J, Bataillard A, Browman G, Burnand B, Durieux P, Fervers B, Grilli R, Hanna S, Have P, Jovell A, Klazinga N, Kristensen F, Madsen PB, Miller J, Ollenschläger G, Qureshi S, Rico-Iturrioz R, Vader J-P, Zaat J. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: The AGREE project. *Qual Saf Health Care* 2003; 12:18–23.
176. Thomason M, Cluzeau F, Littlejohns P, Ollenschlaeger G, Grilli R, Rico-Iturrioz R, Grol R, Burgers J, Kristensen FB, Palmhoj-Nielsen C, Fervers B, Jovell A, Moran S, Grimshaw J, Burnand B, Klazinga N, Ten Have P, Durieux P, Miller J, Feder G, Makela M. Guideline development in Europe: An international comparison. *Int J Technol Assess Health Care* 2000; 16:1039-1049.
177. AGREE Collaboration (Appraisal of Guidelines, Research, and Evaluation in Europe [AGREE] Collaborative Group). Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. www.agreecollaboration.org
178. Harris JS. Development, use, and evaluation of clinical practice guidelines. *J Occupa Environ Med/Am College Occupa Environ Med* 1997; 39:23-34.
179. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. *Int J Technol Assess Health Care* 2003; 19:148-157.
180. Vergnenegre A. [Clinical practice guidelines: A reader's guide]. *Revue Des Maladies Respiratoires* 2003; 20:920–927.
181. California Workers' Compensation Institute 2004. AB749: An Analysis of Medical Utilization Trends. Available at www.cwci.org/icis/ViewReport.CFM?&ReportIndex519.
182. Hayward R, Wilson M, Tunis S, Bass E, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. *JAMA* 1995; 274:570–574.
183. Cates JR,Young DN, Guerriero DJ, Jahn WT, Armine JP, Korbett AB, Bowerman DS, Porter RC, Sandman TD, King RA.. Evaluating the quality of clinical practice guidelines. *J Manip Physiol Therap* 2001; 24:170–176.
184. Schünemann HJ, Fretheim A, Oxman AD, WHO Advisory Committee on Health Research. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. *Health Res Policy Syst* 2006; 4:13.
185. Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, Raskob G, Lewis SZ, Schünemann H. Grading strength of recommendations and quality of evidence in clinical guidelines. Report from an American College of Chest Physicians task force. *Chest* 2006; 129:174-181.
186. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magri-

ni N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr, Zaza S; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328:1490.

187. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. *Health Res Policy Syst* 2006; 4:21.
188. Buller HR, Agnelli G, Hull RD, Hyers TM, Prins MH, Raskob GE. Antithrombotic therapy for venous thromboembolic disease: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest* 2004; 126:401S-428S.
189. Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, Schunemann H. An emerging consensus on grading recommendations? (Editorial). *ACP J Club* 2006; 144:A8-9.
190. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, Hill S, Jaeschke R, Liberati A, Magrini N, Mason J, O'Connell D, Oxman AD, Phillips B, Schunemann H, Edejer TT, Vist GE, Williams JW Jr. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005; 5:25.
191. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schunemann H, Edejer TT, Vist GE, Williams JW Jr. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004; 4:38.
192. Schünemann HJ, Best D, Vist G, Oxman AD. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003, 169:677-680.
193. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 2. Priority setting. *Health Res Policy Syst* 2006; 4:14.
194. Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. *Health Res Policy Syst* 2006; 4:15.
195. Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 5. Group processes. *Health Res Policy Syst* 2006; 4:17.
196. Schünemann HJ, Oxman AD, Fretheim A. Improving the use of research evidence in guideline development: 6. Determining which outcomes are important. *Health Res Policy Syst* 2006; 4:18.
197. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 7. Deciding what evidence to include. *Health Res Policy Syst* 2006; 4:19.
198. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement. *Health Res Policy Syst* 2006; 4:22.
199. Edejer TT. Improving the use of research evidence in guideline development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications. *Health Res Policy Syst* 2006; 4:23.
200. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 12. Incorporating considerations of equity. *Health Res Policy Syst* 2006; 4:24.
201. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 13. Applicability, transferability and adaptation. *Health Res Policy Syst* 2006; 4:25.
202. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 14. Reporting guidelines. *Health Res Policy Syst* 2006; 4:26.
203. Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 15. Disseminating and implementing guidelines. *Health Res Policy Syst* 2006; 4:27.