Pain Physician 2021; 24:S1-S26 • ISSN 2150-1149

Health Services Research

Methodology for Evidence Synthesis and Development of Comprehensive Evidence-Based Guidelines for Interventional Techniques in Chronic Spinal Pain

Laxmaiah Manchikanti, MD, Sairam Atluri, MD, Mark V. Boswell, MD, PhD, Aaron K. Calodney, MD, Sudhir Diwan, MD, Sanjeeva Gupta, MD, Alan D. Kaye, MD, PhD, Nebojsa Nick Knezevic, MD, PhD, Kenneth D. Candido, MD, Alaa Abd-Elsayed, MD, Miguel A. Pappolla, MD, PhD, Gabor B. Racz, MD, Harsh Sachdeva, MD, Mahendra R. Sanapati, MD, Shalini Shah, MD, Vanila Singh, MD, Amol Soin, MD, and Joshua A. Hirsch, MD

From: Author affiliations can be found on pp. S204-S205

Address Correspondence:
Laxmaiah Manchikanti, MD
Pain Management Centers of Amercia
67 Lakeview Drive
Paducah, Kentucky 42001
E-mail: drlm@thepainmd.com

Disclaimer: There was no external funding in the preparation of this manuscript.

Conflict of interest: Dr. Calodney is a consultant for Medtronic, Stryker, Nevro, and Boston Scientific; Dr Gupta has organized meetings, which were supported by Pharmaceutical companies and companies that manufacture equipment used in pain medicine; Dr. Kaye has been a speaker for Merck in the past; Dr. Abd-Elsayed is a consultant of Medtronic and Avanos; Dr. Racz is a Consultant for and has family ownership of Epimed International, is a Consultant to Cosman RF Company, and has Medtronic patent issues; Dr. Shah is a consultant for Masimo Corporation and a speaking consultant for Allergan Corporation; Dr. Singh is Independent Director Bio Delivery Sciences International (BDSI) since November 2019, Independent Director Lucid Lane since April 2020, and Consultant to SPR Therapeutics since March 2020; Dr. Soin is the founder and CEO of Soin Neuroscience, which is developing a spinal cord stimulator to treat spinal pain; Dr. Hirsch is a consultant for Medtronic and Senior Affiliate Research Fellow at the Neiman Policy Institute.

Accepted for publication: 12/01/2020

Free full manuscript:
www.painphysicianjournal.com

Background: The re-engineered definition of clinical guidelines in 2011 from the IOM (Institute of Medicine) states, "clinical practice guidelines are statements that include recommendations intended to optimize patient care that is informed by a systematic review of evidence and an assessment of the benefit and harms of alternative care options." The revised definition distinguishes between the term "clinical practice guideline" and other forms of clinical guidance derived from widely disparate development processes, such as consensus statements, expert advice, and appropriate use criteria.

Objective: To assess the literature and develop methodology for evidence synthesis and development of comprehensive evidence-based guidelines for interventional techniques in chronic spinal pain.

Methods: A systematic review of the literature including methodology of guideline development encompassing GRADE approach for guidance on evidence synthesis with recommendations.

Results: Some of the many factors described in 2011 continue as of 2020 and impede the development of clinical practice guidelines. These impediments include biases due to a variety of conflicts and confluence of interest, inappropriate and poor methodological quality, poor writing and ambiguous presentation, projecting a view that these are not applicable to individual patients or too restrictive with the elimination of clinician autonomy, and overzealous and inappropriate recommendations, either positive, negative, or non-committal. Thus, ideally, a knowledgeable, multidisciplinary panel of experts with true lack of bias and confluence of interest must develop guidelines based on a systematic review of the existing evidence.

This manuscript describes evidence synthesis from observational studies, various types of randomized controlled trials (RCTs), and, finally, methodological and reporting quality of systematic reviews. The manuscript also describes various methods utilized in the assessment of the quality of observational studies, diagnostic accuracy studies, RCTs, and systematic reviews.

Limitations: Paucity of publications with appropriate evidence synthesis methodology in reference to interventional techniques.

Conclusion: This review described comprehensive evidence synthesis derived from systematic reviews, including methodologic quality and bias measurement. The manuscript described various methods utilized in the assessment of the quality of the systematic reviews, RCTs, diagnostic accuracy studies, and observational studies.

Key words: Evidence-based medicine (EBM), interventional pain management, evidence synthesis, methodological quality assessment, conflict of interest, confluence of interest, comparative effectiveness research (CER), clinical practice guidelines, systematic reviews, meta-analysis

Pain Physician 2021: 24:S1-S26

Health care research, practice, and policy focus on improving the organization, delivery, quality, cost, outcomes of care, and accountability (1-9). Critical to achieving these objectives is the need for guidance based on currently available knowledge generated through research, combined with professional experience and consideration of each individual patient (1-5,7,10). Thus the emphasis on evidence synthesis and development of guidelines continues to grow, despite numerous developments in health care policy and regulation. In 2011, the Institute of Medicine (IOM) re-engineered its definition of clinical guidelines (10) as, "clinical practice guidelines are statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options." This definition departed from a 1990 IOM report, which defined guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" (11).

The 2011, IOM definition was hailed for providing a clear distinction between the term "clinical practice guideline" and other forms of clinical guidance derived from widely disparate development processes, such as consensus statements, expert advice, and appropriate use criteria. In addition, this definition also emphasized importance of systematic review and both benefits and harms assessment as essential components of clinical practice guidelines. However, the IOM committee recognized other forms of clinical guidance, which may have value. The concerns of the IOM committee of lack of high-quality evidence for many clinical domains continue, riddled with the confluence of interest from high-level, biased assessments (1-5,7,10,12-32). Consequently, the recommendation of the IOM to be able to produce trustworthy clinical practice guidelines, if their development reflects the committee standards as described by the IOM (10) despite such constraints, continues to fail to achieve optimum value.

Any group of individuals can designate itself as evidence-based medicine (EBM), comparative effectiveness research (CER), or guideline group. Different groups have reviewed the same procedure or problem in interventional pain management and reached vastly different conclusions (1-5,7,12-29). Consequently, it is clear that the process of preparation of EBM or CER manuscripts and guidelines is inadequately monitored, and replication, which is the distinguishing characteristic of scientific knowledge and an essential test of the validity of any scientific statement, is essentially impossible. Multiple factors influencing guideline development include the nature of the newly recommended practice or technology itself; characteristics of health care providers; organizational capacity to collect, adapt, share, and apply evidence; system-level environmental factors; and policies dictated by governmental agencies and the insurance community (10,27,30-35). These factors, however, are considered to be manifestations of the downstream of guideline development. Consequently, the application of single and combined interventions in assessment has been recommended to address these barriers and improve compliance with guideline recommendations, even though their impact can be variable and inconsistent (33-39). Other factors including bias related to various conflicts of interest, variable methodological quality, inappropriate or poor writing, and ambiguous presentation, projecting a view that these are not applicable to individual patients or too restrictive, with reduction or elimination of clinician autonomy and inappropriate overzealous recommendations, are intrinsic to guideline development and can be addressed during the development process. The volume of guidelines currently available may be overwhelming, particularly given that recommendations for the same clinical indication may be inconsistent across different guidelines related to individual biases and conflicts of interests (1-5,7,10-32). The IOM provided guidance for trustworthy guidelines (10), which was developed based on the following standards:

- Based on a systematic review of the existing evidence
- Developed by knowledgeable, multidisciplinary panel of experts and representatives from key affected groups
- Considerate of important patient subgroups and patient preferences, as appropriate
- Based on an explicit and transparent process that minimizes distortions, biases, and conflicts of interest
- Clear in their explanation of the logical relationships between alternative care options and health outcomes, and provide ratings of both the quality of evidence and the strength of recommendations
- Reconsidered and revised as appropriate when important new evidence warrants modifications of recommendations.

Multiple frameworks have been developed to improve the ability to implement clinical guidelines by developing national and international standards (10,33-56). Even then, conflicting opinions about whether guidelines are a solution to rationing or politics disguised as science continue (32,57). According to Saarni and Gylling (57), EBM is often seen as a scientific tool for quality improvement, even though its application requires consideration of scientific facts along with value judgments and the cost of different treatments. Thus guideline development depends on whether we approach the problem from the perspective of patients, doctors, or public health administrators. The EBM exerts a fundamental influence on certain key aspects of medical professionalism. Thus each segment has its own interpretation and agenda, often seemingly based on factors other than science and best care for the patient. The actual value of evidence is related to the application and circumstances in which and for whom it will be used. It is also essential to remember that the value of evidence is only as good as the type of evidence reviewed, the methodology utilized, the reviewers' knowledge and experience, and many other factors, including bias, self-interest, and financial factors.

EBM

EBM begins with the assertion that it is a shift in medical paradigms and is about solving clinical problems (58). For clinicians to interpret the results of clinical research effectively, a formal set of rules must complement medical training and common sense. Thus knowing the tools of evidence-based practice is necessary, but not sufficient, for delivering the highest quality of patient care. It therefore continues to be a challenge for EBM, CER, and interventional pain management to better integrate new scientific innovations with the time-honored craft of caring for the sick. However, EBM also has been characterized as a stick by which policy-makers and academicians beat clinicians (32). There is an extensive role for EBM, CER, and clinical guidelines based on EBM in interventional pain management. EBM is commonly defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (58). The term is loosely used and can include conducting a statistical meta-analysis of accumulated research, promoting randomized controlled trials (RCTs), supporting uniform reporting styles for research, or having a personal orientation toward critical self-evaluations (59).

In contrast, CER is defined as "the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care" (60).

The notion that EBM promises to create better informed patients and clinicians by offering collectively agreed on and publicly available information about treatment options is contradicted by a significant proportion of physician providers. In practice, EBM clinical practice guidelines are created by a small group of interested parties. Even so, there has been an explosion in the development of clinical practice guidelines, as well as the literature focusing on EBM and CER, all of which are unregulated and unchecked.

The utilization of interventional techniques has been a significant concern over the years (61-65). In addition, health care costs in the United States continue to increase in the years following enactment of the Affordable Care Act, with low back and neck pain and other musculoskeletal disorders contributing to $134.5 and $129.8 billion, respectively, in costs in 2016 (66). Further, overall health care expenditures have been rapidly increasing (67).

Development of Clinical Guidelines

Appropriately developed guidelines must incorporate validity, reliability, reproducibility, clinical applicability and flexibility, clarity, development through a multidisciplinary process, scheduled reviews, and documentation (1,10,36,37,41-49). When appropriately applied, rigorously developed guidelines have the potential to reduce undesirable practice variation, reduce the use of services that are of minimal or questionable value, increase utilization of services that are effective but underused, and target services to those populations most likely to benefit (1,10,47-60).

The IOM committee on clinical practice guidelines (10) described 8 standards for developing trustworthy clinical practice guidelines, including transparency, conflict of interest, group composition, clinical practice guidelines–systematic review interception, articulation of recommendations, external review, and updating. Furthermore, the committee has focused increased attention on aspects of conflicts of interest, such as details of guideline development group exclusions; aspects of guideline group composition, including training of patient and consumer representatives in evidence appraisal; the specific nature of working relationships between systematic review teams and clinical

pain guideline developers; critical steps in establishing evidence foundation for clinical recommendations and rating recommendations strength; external review of clinical practice guidelines, including specifying mechanisms for ensuring public stakeholder comment; and elements essential to clinical practice guideline updating, including ongoing monitoring and review of the clinical guideline-relevant scientific literature and factors indicating the need for updates. Unlike many development methodologies, which are specific to particular guideline development, entity, and clinical problem, the 8 standards described by the IOM provide sufficient flexibility to be applicable to all guideline development groups, whether the evidence in a particular clinical area is lacking or is abundant.

Several manuscripts have described the development of clinical practice guidelines, along with the development of international standards and the updating of clinical practice guidelines (1,10,33,36-38,47-49,67-71). Woolf et al (67) described that clinical practice guidelines are one of the foundations of efforts to improve health care. The context for guideline development has changed with the emergence of guideline clearinghouses and large-scale guideline production organizations, such as the Agency for Healthcare Research and Quality (AHRQ) (68) and the National Institute for Health and Care Excellence (NICE) (69). AHRQ described adherence to IOM standards, and launched the National Guideline Clearinghouse Extent Adherence to Trustworthy Standards (NEATS) to be followed in the preparation of guidelines to be posted on the AHRQ website (70). These recommendations are similar to the statements made by the IOM (10).

Apart from the IOM, AHRQ, and NICE, the most commonly utilized including in the other systems is the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach, utilized as a system of rating and quality of evidence for guidelines development that reportedly offers a transparent and structured process for carrying out the steps involved in developing recommendations (36,37,44,47-49). The guidance was developed by the GRADE working group, describing a systematic approach to make decisions about the quality of evidence and strength of recommendations. More specifically, in this approach, methodological rigorousness of the studies included in the guidelines development process, consistency of results across the studies, and generalized ability of the results to a wider patient base are evaluated (36,37,44,47-49,72). The GRADE approach for evidence-based guideline development has been utilized by many national and international organizations. The GRADE approach provides a useful framework for grading both the quality of the evidence behind a recommendation and consideration of how strong the recommendation should be. Consequently, the GRADE approach may overcome the limitations of various other guideline developments, as it is evidenced by the adaptation of GRADE by 70 organizations, including the earlier-described guidance. The GRADE approach has been applied in multiple fields, including interventional pain management, and for both diagnostic and therapeutic guidance.

Methodology of Guideline Development

High quality, relevant, clinical practice guidelines are developed to provide evidence-based advice on diagnosis and treatment to clinicians at the point of care. To be successful, these guidelines must be trustworthy. During the last 2 decades, major advances have been made in the development, dissemination, and implementation of guidelines to improve health care outcomes. Extensive guidance on how to develop guidelines have been published, and standards are suggested for the development of trustworthy guidelines. Although the process of developing guidelines is complex, time-consuming, with the involvement of large multidisciplinary teams, despite the progress in preparing guidelines, numerous challenges remain in interventional pain management with lack of trustworthiness; ineffective guideline authoring and adoption; inefficient guideline dissemination to clinicians at the point of care; and suboptimal presentation formats of guideline content (35). In a survey of pain physicians, the awareness of national guidelines published by the Faculty of Pain Medicine of the Royal College of Anaesthetists and British Pain Society varied between 38% and 90%, which led to publication of a guidance and implementation document to improve uptake (73,74). In addition, newer problems include inconsistent and underdeveloped systems for integration of trustworthy guidelines in electronic medical records, and limited support for shared decision-making at point of care (35).

Similar to guidelines, there have been substantial increases in the number of treatment alternatives available to providers and patients, the proportions of patients receiving interventional pain management services, the volume of studies describing the effectiveness (or ineffectiveness) of those options, guidelines, and systematic reviews (1-5,12-28).

Further, the body of evidence available continues to increase in its complexity, with numerous conflicts of interest, and to be acceptable by the majority of the providers. However, these guidelines are a critical tool for summarizing the available literature and organizing it in a format that is accessible to interventional pain physicians.

GRADE Approach for Guidance

GRADE approach provides a rating of a body of evidence, also described as the quality of evidence or confidence in evidence in the context of systematic reviews, developing health care recommendations, and supporting decisions. GRADE's approach to rating certainty of the evidence is based on a 4-level system: high, moderate, low, and very low. GRADE system also has described the overall rating of the evidence, utilization of evidence to move to recommendations, and decisions on multiple diagnostic and therapeutic issues. The GRADE evidence approach is utilized in the United States and internationally by 70 organizations, including the Centers for Disease Control and Prevention (CDC), NICE, and Cochrane Collaboration. Consequently, GRADE has been utilized in multiple studies (36,37,47,71,72).

The GRADE approach is most often used to assess the quality of evidence for specific outcomes, most commonly to a meta-analysis in the context of a systematic review. However, it can also be applied to individual studies or nonquantitative synthesis when meta-analyses are not available. Evidence from RCTs begins as high-quality evidence but can be downgraded according to various factors, including risk of bias, inconsistency, indirectness, imprecision, and publication bias. In contrast, the inclusion of nonrandomized studies provides low-quality evidence. However, rating from nonrandomized studies may be upgraded, provided there are no other limitations identified according to the 5 factors.

Three primary factors, namely large magnitude of effect, evidence of a dose-response effect, and all plausible findings considered and upgrading occurs. After the process of downgrading or upgrading, the quality of the evidence for each outcome are separated as high, moderate, low, or very low. In fact, Meader et al (48) published a checklist designed to aid consistency and reproducibility of GRADE assessments. In this pilot validation study, authors examined measures of agreement and found that for most of the items designed to assess the risk of bias, an agreement was found to be either almost perfect or substantial. However, for one item, designed to measure attrition bias, the agreement was moderate. For other items concerning no other bias and selective reporting, the level of agreement was relatively low.

Multiple organizations and authors have concluded that GRADE approach to grading the quality of evidence and strength of recommendations for diagnostic phase, therapeutic interventions, and strength of recommendations provides a comprehensive and transparent approach (36,72). However, owing to lack of extensive publications in interventional pain management, utilizing GRADE approach with systematic utilization of the recommendations, further understanding and discussions are essential to reach a common consensus to be applied to guidelines and recommendations in interventional pain management.

The robustness, systematic, and comprehensive approach for guidelines and recommendation development have been described extensively (36). With GRADE approach, assessment, and evidence synthesis is a comprehensive and structured way to rate the quality of evidence in systematic reviews or other synthesized evidence. GRADE approach includes multiple steps, which are interconnected and not necessarily sequential.

Evidence Synthesis

The GRADE approach starts by defining the question in terms of PICO (Population, Intervention, Comparison, and Outcomes) criteria, and proceeds with a systematic search to identify all the available evidence on the subject matter.

Quality of Evidence

The quality of the evidence is rated based on 5 factors that may downgrade and 3 factors that may upgrade the quality of evidence. The factors that may upgrade the evidence include large magnitude effect, dose-response, and effect of plausible confounding factors. In contrast, the 5 factors that may downgrade the evidence include risk of bias, inconsistency, indirectness, imprecision, and publication bias.

Balance of Consequences

Balance of consequences is determined by the quality of evidence, trade-off between benefits and risks, values and preferences, feasibility, equity and acceptability, and resource views. Thus after considering all the information from evidence synthesis, a

decision is made about the importance and criticality of the outcomes based on the recommendations being formulated, and overall quality of evidence is assigned based on the assessment.

Recommendations

Recommendations are formulated with a decision for or against in strength, strong or weak of the recommendations. The strength of recommendations is based on overall evidence and balance of consequences. Table 1 also shows various steps in the application of GRADE approach in guideline preparation.

Application and Intervention

Based on the earlier-described discussion and available literature in various specialties, GRADE approach may be easily applied for interventional techniques, as shown in Fig. 1.

IOM Guidance

The IOM explained that to be trustworthy, guidelines must be (10):

- Based on a systematic review of existing evidence
- Developed by a knowledgeable, multidisciplinary panel of experts and representatives from key affected groups
- Considerate of important patient subgroups and patient preferences, as appropriate
- Based on an explicit and transparent process that minimizes distortion, biases, and conflicts of interest

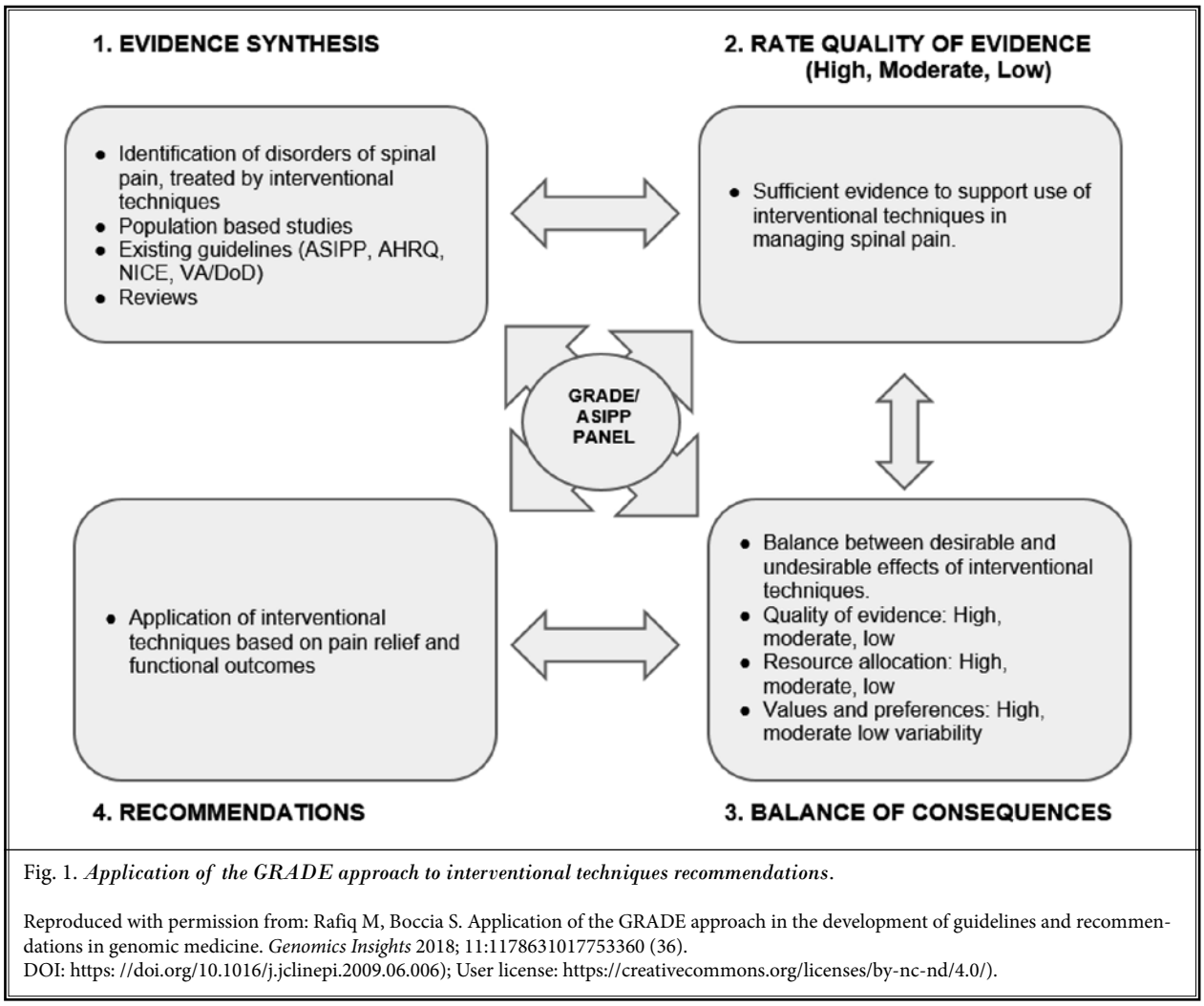

Fig. 1. *Application of the GRADE approach to interventional techniques recommendations.*

Reproduced with permission from: Rafiq M, Boccia S. Application of the GRADE approach in the development of guidelines and recommendations in genomic medicine. *Genomics Insights* 2018; 11:1178631017753360 (36).
DOI: https: //doi.org/10.1016/j.jclinepi.2009.06.006); User license: https://creativecommons.org/licenses/by-nc-nd/4.0/).

Table 1. *Agreement for all checklist items.*

Item	Kappa (95% CI)	Magnitude of Agreement
Risk of bias		
Was random sequence generation used (i.e.,, no potential for selection bias)?	0.89 (0.69 to 1)	Almost perfect
Was allocation concealment used (i.e.,, no potential for selection bias)?	0.69 (0.29 to 1)	Substantial
Was there blinding of participants and personnel (i.e.,, no potential for performance bias)?	0.71 (0.41 to 1)	Substantial
Was there blinding of outcome assessment (i.e.,, no potential for detection bias)?	0.98 (0.67 to 1)	Almost perfect
Was an objective outcome used?	1	Almost perfect
Were more than (80%)a of participants enrolled in trials included in the analysis? (i.e., no potential attrition bias)	0.44 (0.07 to 0.81)	Moderate
Were data reported consistently for the outcome of interest (i.e., no potential selective reporting)? (no potential reporting bias)	0.25 (0 to 0.61)	Fair
No other biases reported? (no potential of other bias)	0.20 (0 to 0.62)	Slight
Did the trials end as scheduled (i.e., not stopped early)?	1	Almost perfect
Inconsistency		
Point estimates did not vary widely? (i.e., no clinical meaningful inconsistency)	0.65 (0.37 to 0.93)	Substantial
To what extent do confidence intervals overlap?	0.50 (0.17 to 0.77)	Moderate
Was the direction of effect consistent?	1	Almost perfect
What was the magnitude of statistical heterogeneity (as measured by I2)?	1	Almost perfect
Was the test for heterogeneity statistically significant (p<0.1)?	1	Almost perfect
Indirectness		
Were the populations in included studies applicable to the target population?	Below chance	Poor
Were the interventions in included studies applicable to target intervention?	Below chance	Poor
Was the included outcome not a surrogate outcome?	1	Almost perfect
Was the outcome timeframe sufficient?	0.47 (0 to 1)	Moderate
Were the conclusions based on direct comparisons?	1	Almost perfect
Imprecision		
Was the confidence interval for the pooled estimate not consistent with benefit and harm?	1	Almost perfect
What was the magnitude of the median sample size?	1	Almost perfect
What was the magnitude of the number of included studies?	1	Almost perfect
Was the outcome a common event? (e.g., occurs more than 1/100)a	1	Almost perfect
Was there no evidence of serious harm associated with treatment?	0.89 (0.67 to 1)	Almost perfect
Publication bias		
Did the authors conduct a comprehensive search?	0.65 (0 to 1)	Substantial
Did the authors search for grey literature?	0.26 (0 to 0.67)	Fair
Authors did not apply restrictions to study selection on the basis of language?	0.74 (0.45 to 1)	Substantial
There was no industry influence on studies included in the review?	0.71 (0.45 to 0.98)	Substantial
There was no evidence of funnel plot asymmetry?	0.62 (0.35 to 0.89)	Substantial
There was no discrepancy in findings between published and unpublished trials?	1	Almost perfect

[a]These thresholds can be replaced with different ones based on the context of the particular review.
Source: Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: Development and pilot validation. *Syst Rev* 2014; 3:82 (48).

- Clear in their explanation of logical relationships between alternative care options and health outcomes, providing ratings of both the quality of evidence and the strength of recommendations
- Reconsidered and revised as appropriate when important new evidence warrants modifications and recommendations.

The IOM also described the multiple factors commonly undermining the quality and trustworthiness of clinical practice guidelines, including:

- Variable quality of individual scientific studies
- Limitations of systematic reviews on which clinical guidelines are based
- Lack of transparency of development groups' methodologies, particularly with respect to evidence quality and strength of recommendation appraisals
- Failure to convene multistakeholder and multidisciplinary guideline development groups, and resulting nonreconciliation of conflicting guidelines
- Unmanaged conflicts of interest
- Overall failure to use rigorous methodologies during development.

In addition, the IOM committee noted that evidence supporting clinical decision-making and clinical practice guideline development relevant to subpopulations, such as patients with comorbidities, the socially and economically disadvantaged, and those with rare conditions, is usually absent. Overall, the committee concluded that the quality of clinical practice guideline development processes and guideline developer adherence to quality standards have remained unsatisfactory and unreliable for decades. Nonstandardized development results in significant variation in clinical recommendations. Even though the IOM once again depended on unreliable tools and evidence, they have formulated a new definition and also developed standards for trustworthy clinical practice guidelines. The committee's 8 proposed standards are reproduced herewith in Table 2 (10).

Table 2. *IOM standards for developing trust worthy guidance.*

STANDARD 1:	Establishing transparency
STANDARD 2:	Conflict of Interest 2.1 Management of conflict of interest (COI) 2.2 Disclosure of COIs within GDG 2.3 Divestment of Final Investments 2.4 Exclusions
STANDARD 3:	Guideline development group composition
STANDARD 4:	Clinical practice guideline–systematic review intersection
STANDARD 5:	Establishing evidence foundations for and rating strength of recommendations
STANDARD 6:	Articulation of recommendations
STANDARD 7:	External review
STANDARD 8:	Updating

Systematic Reviews

Guidelines are developed from the evidence from systematic reviews, RCTs, and observational studies, but preferably from relevant high-quality systematic reviews with or without meta-analysis.

Systematic reviews are considered the highest level of evidence in medicine, particularly if meta-analysis is possible and applicable (75). Systematic reviews and guidelines are also used in making decisions about health care resource allocation in interventional pain management through economic evaluations (1-5,10,12-29). As expected with dynamic changes in medicine, the number of published systematic reviews continues to grow exponentially. The fundamental importance of systematic reviews is the strength of available evidence and if there is sufficient high-quality evidence to support the majority of recommendations in health care. However, systematic reviews may vary in quality and rigor and may not guarantee high methodological quality. In fact, Riado Minguez et al (54) assessed the methodological and reporting quality of systematic reviews published in the highest-ranking journals in the field of pain.

Risk of bias and methodological and reporting quality assessment for systematic reviews is generally performed utilizing 3 tools: A Measurement Tool to Assess Systematic Reviews (AMSTAR) (56), Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (55), and Scottish Intercollegiate Guidelines Network (SIGN) (76).

PRISMA

PRISMA was developed by a multidisciplinary team of experts in 2005 as an extension of QUOROM (Quality of Reporting of Meta-Analyses) and was published in 2009 (77). The PRISMA statement consists of a 27-item checklist and a 4-phase flow diagram. The checklist includes items deemed essential for transparent reporting for a systematic review. Liberati et al (77) published an explanation and elaboration in 2009. Although PRISMA focuses on ways in which authors can ensure the transparent and complete reporting of systematic reviews and meta-analysis, it does not address directly or in a detailed manner the conduct of systematic reviews, for which other guidelines are available. Table 3 shows a checklist

Table 3. *Compliance with individual PRISMA checklist items.*

Section/ Topic	#	Checklist item	Total Completion 1.0	Partial Completion 0.5	Non-complete 0
TITLE					
Title	1	Identify the report as a systematic review, meta-analysis, or both.			
ABSTRACT					
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, particIpants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.			
INTRODUCTION					
Rationale	3	Describe the rationale for the review in the context of what is already known.			
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).			
METHODS					
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.			
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.			
Information sources	7	Describe all information sources (e.g., data-bases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.			
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.			
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).			
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.			
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.			
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.			
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).			
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I2) for each meta-analysis.			
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).			
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.			
RESULTS					
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.			
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.			

Table 3 con't. *Compliance with individual PRISMA checklist items.*

Section/ Topic	#	Checklist item	Total Completion 1.0	Partial Completion 0.5	Non-complete 0
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).			
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.			
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.			
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).			
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).			
DISCUSSION					
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).			
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).			
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.			
FUNDING					
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.			

Source: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). http://prisma-statement.org/ (55)

of items to include when reporting a systematic review with or without meta-analysis. PRISMA may be used for a systematic review of randomized and nonrandomized trials.

PRISMA scoring has been published extensively (78,79). Previous authors of interventional techniques (16,54) and opioid disorders (79) utilized scoring of each item rated as Yes for total compliance, Unclear for partial compliance, or No for noncompliance. Score values corresponded to 1, 0.5, or 0. Possible range of PRISMA scores for each systematic review was 0 to 27. To assess the compliance with individual PRISMA items, as in the previous systematic reviews (16,54,79), high compliance was utilized as 90% to 100%, medium compliance as 70% to 89%, low compliance as 30% to 69%, and 0% to 29% as very low compliance.

AMSTAR

AMSTAR was published in 2007 (80). The final items included 11 items developed from 11 components (Table 4). The tool with 11 items was shown to have good face and content validity for measuring the methodologic quality of systematic reviews (80). A critical appraisal of AMSTAR with description of challenges, limitations, and potential solutions from the perspective of an assessor showed that there was some methodological limitations of the AMSTAR checklist, as well as challenges involved in evaluation of the checklist items. Some items of the AMSTAR checklist seem to assess quality of reporting of a systematic review more than its methodologic quality. In addition, some items may be difficult to interpret, hindering accurate assessment. Potential solutions were presented to improve each AMSTAR item with the aim of allowing a more thorough assessment of the systematic reviews.

Overall, AMSTAR has been utilized with a lesser frequency than PRISMA in assessing reporting and methodologic quality.

SIGN

Methodologic quality assessment by utilizing SIGN (76) was seen for interventional techniques (16,54). Manchikanti et al (16) and Cho et al (29) utilized SIGN

Table 4. *Compliance with individual AMSTAR checklist items.*

AMSTAR Items	YES	NO	NA
1. Was a priori design provided (protocol established before the conduct of review)?			
2. Was there duplicate study selection and data extraction?			
3. Was a comprehensive literature search performed?			
4. Was the status of publication (ie, gray literature) used as an inclusion criterion?			
5. Was a list of studies (included and excluded) provided?			
6. Were the characteristics of the included studies provided?			
7. Was the scientific quality of the included studies assessed and documented?			
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?			
9. Were the methods used to combine the findings of studies appropriate?			
10. Was the likelihood of publication bias assessed?			
11. Was the conflict of interest included, both for the systematic review authors and included studies' authors?			

Source: A Measurement Tool to Assess systematic Reviews (AMSTAR). https://amstar.ca/ (56).

in assessing treatment outcomes for patients with post lumbar surgery syndrome, which also included percutaneous adhesiolysis, along with comparative analysis of spinal cord stimulation. Further, Manchikanti et al (16) utilized SIGN to assess effectiveness of percutaneous adhesiolysis in post lumbar surgery syndrome in performing a systematic analysis of findings of systematic reviews. The quality assessment was based on 3 options: ++ indicated all or most of the standards were met, + indicated some of the standards were met, and – indicated all standards were not met as shown in Table 5. Grading is shown in Table 6.

By comparing all the studies previously assessed and all the systematic reviews previously published, Manchikanti et al (16) identified a multitude of flaws in the analysis and application of the criteria in systematic reviews.

IOM Standards

The IOM developed standards for systematic reviews (72). It described the function/purpose of a systematic review as a tool to identify, select, assess, and synthesize the findings of similar but separate studies, and to help clarify what is known and not known about the potential benefits and harms of drugs, devices, and other health care services. In developing standards for systematic reviews, the IOM committee defined a "standard" as "a process, action, or procedure for performing systematic reviews that is deemed essential to producing scientifically valid, transparent, and reproducible results."

The IOM developed standards for initiating a systematic review (Table 7), standards for finding and assessing individual studies (Table 8), standards for synthesizing body of evidence (Table 9), and standards for reporting systematic reviews (Table 10).

Table 5. *Degree of evidence as described by SIGN.*

1++	- High quality meta-analysis and systematic review conducted by randomized clinical trials - Randomized controlled trials with a very low risk of bias
1+	- Well-designed meta-analysis and systematic review conducted by randomized or non-randomized clinical trials - Randomized or non-randomized clinical trials with a low risk of bias
1-	- Meta analysis and systematic review conducted by randomized or non-randomized clinical trials - Randomized or non-randomized clinical trials with a high risk of bias
2++	High-quality systematic review conducted by a patient control study, cohort study, or diagnosis analytic study - High-quality patient control study, cohort study, or diagnosis analytic study of very low risk of confounding, bias or contingency, or a high possibility of cause and effect relationship
2+	- High-quality patient control study, cohort study, or diagnosis analytic study of the low risk of a confounding, bias or contingency, or the normal possibility of a cause and effect relationship
2-	- Patient control study, cohort study, or diagnosis analytic study of the high risk of a confounding bias or contingency, or the low possibility of a cause and effect relationship
3	- Non-analytic studies, e.g., before-and-after study, case series, case report
4	- Expert opinion

Source: Cho JH, Lee JH, Song KS, et al. Treatment outcomes for patients with failed back surgery. Pain Physician 2017; 20:E29-E43 (29).

The IOM Committee concluded that systematic reviews should be used to inform health care decision-

Table 6. *Recommendation grade.*

A	- At least one metaanalysis, systematic review, or RCT rated as 1 + + and directly applicable to the target population or - A systematic review of RCTs or a body of evidence consisting principally of studies rated as 1 + directly applicable to the target population and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2 + + directly applicable to the target population and demonstrating overall consistency of results or - Extrapolated evidence from studies rated as 1 + + or 1 +
C	- A body of evidence including studies rated as 2 + directly applicable to the target population and demonstrating overall consistency of results or - Extrapolated evidence from studies rated as 2 + +
D	- Evidence level 3 or 4 or - Extrapolated evidence from studies rated as 2 +

Source: Cho JH, Lee JH, Song KS, et al. Treatment outcomes for patients with failed back surgery. Pain Physician 2017; 20:E29-E43 (29).

Table 7. *Standards for initiating a systematic review.*

STANDARD 1:	Establish a team with appropriate expertise and experience to conduct the systematic review
STANDARD 2:	Manage bias and conflict of interest (COI) of the team conducting the systematic review
STANDARD 3:	Ensure user and stakeholder input as the review is designed and conducted
STANDARD 4:	Manage bias and COI for individuals providing input into the systematic review
STANDARD 5:	Formulate the topic for the systematic review
STANDARD 6:	Develop a systematic review protocol
STANDARD 7:	Submit the protocol for peer review
STANDARD 8:	Make the final protocol publicly available, and add any amendments to the protocol in a timely fashion

Table 8. *Standards for finding and assessing individual studies.*

STANDARD 1:	Conduct a comprehensive systematic search for evidence
STANDARD 2:	Take action to address potentially biased reporting of research results
STANDARD 3:	Screen and select studies
STANDARD 4:	Document the search
STANDARD 5:	Manage data collection
STANDARD 6:	Critically appraise each study

Table 9. *Standards for synthesizing the body of evidence*

STANDARD 1:	Use a prespecified method to evaluate the body of evidence
STANDARD 2:	Conduct a qualitative synthesis
STANDARD 3:	Decide if, in addition to a qualitative analysis, the systematic review will include a quantitative analysis (meta-analysis)
STANDARD 4:	If conducting a meta-analysis 4.1 Use expert methodologists to develop, execute, and peer review the meta-analyses 4.2 Address the heterogeneity among study effects 4.3 Accompany all estimates with measures of statistical uncertainty 4.4 Assess the sensitivity of conclusions to changes in the protocol, assumptions, and study selection (sensitivity analysis)

NOTE: The order of the standards does not indicate the sequence in which they are carried out.

Table 10. *Standards for reporting systematic reviews*

STANDARD 1:	Prepare final report using a structured format
STANDARD 2:	Peer review the draft report
STANDARD 3:	Publish the final report in a manner that ensures free public access

makers about what is known and not known about the effectiveness of health interventions (75). Patients expect that their doctors and other health care providers know what type of treatment to recommend. In reality, however, the evidence that informs current health care decisions is often incomplete and may be biased, and there are no standards in place to ensure that systematic reviews of the evidence are objective, transparent, and scientifically valid (75). Higher-quality systematic reviews have the potential to improve the decisions made by clinicians, to better inform patient choice, and to provide a more trustworthy basis for decisions by payers and policy makers.

Assessment of Clinical Studies

Evidence synthesis is based on systematic reviews and meta-analysis. Systematic reviews and meta-analyses are performed from RCTs, nonrandomized studies, and diagnostic accuracy studies. However, if systematic reviews are not available, the authors of guidelines must conduct systematic reviews of the available studies or update the systematic reviews with any new available evidence.

RCTs

RCTs are considered the gold standard of EBM. However, RCTs have some limitations, especially when investigating the treatment options for pain management because of too strict inclusion and exclusion criteria and confounding factors (age, gender, past therapy, etc.) that do not allow generalization.

In RCTs, patients receive either an investigational (targeted) or placebo/control treatment, the order of treatments is randomized, and patients and clinicians are blinded to allocation. However, there is extensive debate regarding placebo controls and study design. Hill, writing on the issues related to the debate about using placebos, described that the essential medical question at issue is how the new treatment compares with the old one, not whether the new treatment is better than nothing (80,81). There are extensive discussions also in reference to the role of placebo in RCTs along with nocebo effects in determining the results of placebo-controlled trials. Consequently, an appropriate placebo design is crucial (82,83).

The null hypothesis in randomized trials assumes that there is no difference between a new and a placebo/control treatment. Properly conducted RCTs minimize bias by acknowledging known and unknown factors, which may affect the treatment effect.

Consolidated Standards of Reporting Trials (CONSORT) acknowledges that properly conducted RCTs provide the best evidence on the efficacy of a health care intervention. The biased interpretation of the results, in favor or against a particular treatment, and incorrect understanding of randomized trials leads to poor quality data interpretation or drawn conclusions. The CONSORT statement for reporting randomized trials (84) provides a checklist of items to include study question, study population, randomization, blinding, interventions, outcomes, statistical analysis, results, discussion, and funding. Thus reports of RCTs should contain the CONSORT flow diagram available at: www.consort-statement.org/consort-statement/flow-diagram.

Randomization is the process of dividing patients to experimental (treatment) or control groups, assuming that each patient has an equal chance to be assigned to one of these groups (85).

Randomization procedure refers to generating a random sequence of allocations, which may be a simple random assignment of patients to any of the equally probable trial groups. However, randomization may have some advantages of eliminating bias in assignment of treatments, and by facilitating blinding, and permits the use of probability theory to express the likelihood that any outcome difference between intervention groups merely reflects chance (86). Thus preventing selection and confounding biases is the most important advantage of randomization (86). Simple randomization is based on a single sequence of random assignments. For simple patient randomization, a random number table or computer-generated random numbers can be used. Restricted randomization describes any procedure that helps achieve balance between groups in size or characteristics by blocking, stratification, or covariate adaptation. Blocking ensures a close balance of the numbers in each group during any time of the trial. However, certain covariates among patients, such as comorbid medical conditions, could undermine proper comparison of the groups and negatively influence the results. Stratification ensures that patients receiving each intervention are well-matched per baseline characteristics, such as age and stage of the disease. Stratified randomization may be difficult to implement if numerous covariates must be controlled (86).

Allocation concealment is a technique in which the allocation sequence is hidden from those assigning patients to the intervention groups. This prevents researchers inadvertently, or otherwise, from influencing the assigning process. Allocation concealment includes 4 standard methods: central randomization; pharmacy controlled; sequentially numbered, opaque, sealed envelopes; and sequentially numbered containers.

Because most studies have very strict eligibility criteria on which they choose treatment groups, conclusions may only be applicable to those who fit such criteria. Placebo-controlled RCTs may fail to apply to the general patient population. Related to the disadvantages of placebo-controlled RCTs, some physicians focus on practical clinical trials, which ensure that patients are diverse, that they come from a heterogeneous group of practice settings and geographic locations, and that they reflect the underlying affected population, at the same time focusing on that the trial's endpoints reflect a broad range of meaningful clinical outcomes.

The choice of a proper control group is critical when designing a clinical trial because it affects conclusions drawn from the generated results, the degree of minimizing bias in the study, types of patients who could be recruited, credibility of the results, and interpretation of the study. The control group provides us with the conclusion of what would have happened if patients had not received the test treatment or comparison if they had received a different treatment (86).

In placebo-controlled trials, patients are randomly allocated to receive either a treatment or identical-appearing drug-free product. These trials are always double-blind, and groups like control or placebo serve to investigate potential influences on the course of the

disease, other than those derived from the pharmacologic action of the test drug. The implementation of randomization and blinding in a placebo-controlled trial minimizes patient and investigator bias but remains vulnerable to breaking blind through the perception of the pharmacologic effects of one treatment. There are usually no ethical issues comparing the new treatment to placebo when a treatment is tested for a condition for which no effect is yet known. Advantages of a randomized, double-blind, placebo-controlled trial include exploring the effectiveness of a treatment, freedom from assumptions, and the ability to distinguish adverse effects caused by a drug or procedure from those resulting from the underlying disease. Disadvantages include ethical concerns, lack of generalizability, and lack of effectiveness comparativeness (86).

Active (positive) control is used when a new treatment is compared with an active control treatment group. Such trials are usually, but not always, double-blind and demonstrate the efficacy of the tested treatment by showing that it is as good as a known effective treatment or superior to the active control. They may also be used to compare the efficacy and/or safety of the 2 newly developed treatments. A randomized, blinded, active control trial generally minimizes patient and investigator bias. Active controls pose fewer ethical and practical problems than placebo-controlled trials. Disadvantages relate to problems in quantitating safety outcomes, lack of direct evaluation of effect size, and need for large sample sizes tested.

In no-treatment control groups, patients are randomly assigned to a test treatment or to a no-treatment study group. In this type of controlled design, neither patients nor the investigators are blind to the treatment assignment and make it applicable only when it is too difficult or impossible to double-blind. This type of control is used both in interventional pain management and surgery.

In a dose-response control design, patients are randomized to one out of several fixed-dose groups, initially assigning them on either fixed-dose or titrating them to a certain dose gradually. These trials are usually double-blind and may include placebo and/or an active group. If the study is blinded, bias is minimized, and this design may be acceptable in interventional pain management or surgery trials. Advantages include efficiency and a possible ethical advantage and enable the determination of efficacy and safety in situations in which a placebo-controlled trial has similar credibility. Disadvantages include the necessity to recognize that a positive dose-response without significant pairwise differences may cause confusion in the determination of which dose, other than the highest, is genuinely effective.

Placebo-controlled trials, active-controlled trials, and dose-response controlled trials elicit and provide various aspects of understanding of the effectiveness of a modality, as shown in Table 11 (81). Consequently, a placebo-controlled trial provides information on the existence of effect and measurable absolute effect size; however, it does not provide any dose-response relationship or compare therapies. In contrast, an active control trial shows existence of effect and also comparative effectiveness but does not measure absolute effect size nor does it show dose-response relationship. To obtain the results on more than 2 aspects of the existence of the effect, effect size, dose relationship, and comparativeness of therapies, multiple combinations of designs may be utilized (Table 11). Multiple manuscripts have described these variations in the results; however, even experts at AHRQ often ignore these crucial aspects of design (7,12-14,26).

Table 11. *Usefulness of specific control types in various situations*

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

Y=Yes, N=No, P=Possible, depending on whether there is historical evidence of sensitivity to drug effects.
Source: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials E10. July 20, 2000 (80).

Table 11 demonstrates the crucial aspect in designing and conducting of the trials or studies, systematic reviews and meta-analysis, and finally, evidence synthesis and its clinical applications.

However, it has been described that many RCTs may not be justified or have limited value or may be misleading (87,88). This was shown in a cross-sectional analysis of the ethics and science of RCTs (87). It has shown that randomized trials are associated with risks to patients, and they are also associated with significant costs (87). Consequently, it has been postulated that researchers, regulators, funding agencies, and publishers should be able to determine when an RCT is ethically justified, at least in virtue of its scientific potential, meeting when it is necessary over other forms of research or over no research at all. It has been shown that as much as 85% of clinical research may be wasteful (88). Further, there is a strong ethical, financial, and scientific impetus to develop such criteria, for if a trial is not scientifically justified, it follows that it cannot be ethically justified on the grounds that patients are put at risk by redundant research (86). It was also described that CONSORT guidelines do not elaborate on how to go about evaluating rationale (84). In this study of justification of RCTs (87), authors included 208 RCT articles and 199 protocols. They showed that only 44% of combined texts of RCT articles and protocols showed a clearly stated hypothesis, some form of uncertainty, and cited a relevant systematic review or meta-analysis. Other aspects of RCTs include blinding.

The details for conducting a study are defined by the study protocol. Prior to the beginning of a trial, the institutional review board evaluates the investigation to assess study design quality, ethical questions, safeguards for patients, and a review of the informed consent statement. The CONSORT statement for reporting randomized trials alleviates the problem of inadequate reporting of RCTs (84), and further addresses equivalence and noninferiority trials. Equivalence trials serve to determine if one intervention is therapeutically similar to another, whereas a noninferiority trial seeks to determine if a treatment is no worse than the reference treatment (86). Critical appraisal of a clinical trial's quality is often incomplete and confounded by poor methodology. All trials must be registered with the US National Institutes of Health Clinical Trial Registry of the United States at www.clinicaltrials.gov.

In the study of interventional or surgical trials, various types of controls are utilized. These range from no-treatment group, placebo control, and sham interventions. In RCTs, treatments are typically tested by comparing the efficacy in an active treatment arm with the efficacy in a placebo arm. The failure to detect significant differences between active treatments and placebos is one of the main sources of uncertainty in the RCTs, especially as negative RCTs continue to explode across spinal therapeutics. In fact, the failure rates were one of the highest in musculoskeletal diseases (83). Placebo response is the most commonly provided explanation leading to a smaller difference between the effect of the intervention and placebo. The placebo-nocebo phenomenon is subject to often increasing debate with extensive research—often controversial (82). A multitude of contextual factors of placebo and nocebo responses and definitions continue to evolve.

Multiple issues related to control trials include placebo and nocebo response, masking, blinding, statistical analysis, and finally interpretation of the results and their application to clinical settings. Thus placebo-control trials of pharmacologic treatments are typically conducted double-blind. In these studies, the process of masking treatment assignment is considered ethically acceptable provided that shared decision-making was made, and the consent process established the nature of the study. However, in circumstances in which a surgical or interventional procedure itself constitutes the treatment; a randomized, placebo-control trial raises different issues (82). In these settings, only the patient is blinded, whereas the clinician can distinguish active from inactive treatment. Thus the gold standard of clinical research, namely double-blind, randomized placebo-controlled trial, is not applicable in a multitude of settings. Consequently, in interventional techniques, instead of a placebo, a "sham" procedure is utilized, which is considered controversial. However, a sham procedure itself is not the same as placebo and creates a multitude of other issues.

In drug trials, a placebo is administered with preservation of double-blind nature and also concealment of allocation. Consequently, placebo response has been defined as "the reduction in symptoms as a result of factors related to patient's perception of the therapeutic intervention" (82). However, with the development of multiple modes of placebo intervention, the placebo also has been defined as a "psychobiology phenomenon occurring in the patient's brain after the administration of an inert substance, or of a sham physical treatment such as sham surgery, along with verbal suggestions (or any other cue of clinical benefit)" (82). Apart from placebo, there is also nocebo activity, which

has been omitted from the literature and clinical assessments to a great extent as an uncomfortable truth (82). The term was coined to denote the counterpart of the use of the placebo, with a description of worsening of the symptoms or reduction of the beneficial effect by the administration of an inactive or active treatment.

Placebo has been administered in interventional trials without appropriate care by administering an assumed inert substance, which may not be an inert substance into an active structure or by administration of active substances into so-called inert structures, which may not be inert.

In contrast, in the no-treatment group, conservative management is provided. There is no blinding of the treatment. Consequently, this avoids placebo effect because there are substances administered. However, placebo effect due to education and other cues continues to persist. In sham surgery, an incision is made, and the procedure is carried out until the final intervention. The true placebo nature of such interventions is questionable because of the induction of multiple psychological phenomena (82). Consequently, placebo and sham interventions may be similar to active interventions. Thus placebo or sham interventions in studying efficacy of vertebral augmentation procedures are met with multiple flaws. Multiple failures of placebo interventions with interventional techniques with injection of sodium chloride solution into epidural space and facet joints has been described in the past, along with the illogical conversation of active controls into placebo controls by overenthusiastic academicians. Further, there is overwhelming evidence of the effectiveness of local anesthetics of epidural injections and facet joint nerve blocks, similar to local anesthetics with steroids (12,14). The same disadvantages in a multiplied format are as a result of sham intervention in comparing the efficacy of vertebral augmentation.

Even then, a placebo is not the inert substance alone, but rather its administration within a complex psychosocial context, with a whole ritual of the therapeutic act. Sham intervention exceeds the placebo intervention and provides not only complex psychosocial context, but also physical and therapeutic context.

There is significant confusion in the word placebo and placebo effect for clinicians, methodologists, and scientists. A clinician is interested in any improvement that may take place in the group of patients who either take the inert substance or receive a sham treatment, and this improvement may be attributed to a multitude of factors including spontaneous remission, regression of the mean, and patients' expectation of the benefit. Contrary to the clinician, the scientist is only interested in the improvement that derives from the patients' expectations, namely, an active process occurring in the patient's brain. However, a methodologist is mostly interested in assessing the difference in the statistical results without consideration to clinical or scientific aspects.

Observational Studies

Most of the research in clinical practice comes from observational studies (89). Proponents of observational studies describe them as being just as effective as RCTs, however, from a methodological perspective, these 2 studies are considered complementary rather than opposing. Reporting of observational research is often insufficient to make statements about the strengths and weaknesses of the investigations. Incorporation of high-quality evidence into clinical practice to improve the effectiveness and safety of patient care is not always limited to randomized, double-blind, placebo-controlled trials.

There are 3 main types of observational studies: cohort, case-control, and cross-sectional designs, and these represent different approaches in an attempt to investigate the occurrence of health-related events in a given population and time period (90). The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement (89) with explanation and elaboration for reporting observational studies addresses the 3 main types of observational studies and ensures a clear presentation of what was done in any observational study. The STROBE checklist is available at: www.strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE_checklist_v4_combined.pdf.

Cohort studies are the best methods for determining the incidence and natural history of the condition (90). In a prospective cohort study, individuals without the disease, but who may or may not have been exposed, are selected and followed over a period of time. In a retrospective cohort study, information on exposure and disease is already collected and is used to evaluate the relationship between exposure and disease over a period of time. Cohort studies describe incidence or natural history, analyze risk factors, and thereby calculate relative risk.

In case–control studies, 2 groups are selected, either diseased or nondiseased, exposure is measured in both of the groups, and the association of exposure to the disease is calculated. Even though they are described as the only retrospective by some (89), they also can

be prospective and are particularly useful for studying infrequent events. They are simple to organize, useful for hypothesis generation, and allow assessment of the influence of predictors on outcome via calculation of an odds ratio.

Cross-sectional studies determine prevalence. They are relatively quick, do not differentiate between cause and effect, and can study multiple outcomes.

Diagnostic Accuracy Studies

In the continued world of dynamic growth of diagnostic test development, exaggerated and often biased results from poorly designed and reported diagnostic studies can affect the utilization of health care interventions. In studies of diagnostic accuracy, the outcomes from one or more tests under evaluation are compared with outcomes from reference standards, both measured in individuals who are suspected of having the condition of interest. The test refers to any method for obtaining additional information on a patients' health state (91). Consequently, this information can be obtained from history and physical examination, laboratory tests, imaging tests, functional tests, interventional diagnostic procedures, and histopathology. Interventional diagnostic techniques are utilized in diagnosis of spinal pain requiring further diagnosis, which is not attainable from physical examination, nerve conduction studies, and imaging. These tests have been considered as highly controversial because of the nature of outcome parameters utilized, namely the pain relief.

The term accuracy refers to the amount of agreement between the information from the test under evaluation, referred to as the index test, and the reference standard. In addition, sensitivity and specificity, likelihood ratios, diagnostic odds ratios, and the area under a receiver operator characteristic curve may also be assessed. Multiple diagnostic reports have been published in reference to the prevalence and accuracy of diagnostic techniques of facet joint pain, sacroiliac joint pain, and discogenic pain.

Methodologic Quality of Risk of Bias Assessment

Key recommendations included transparency and reproducibility of judgments, separating risk of bias from other constructs such as applicability and precision, and evaluating the risk of bias per outcomes.

RCTs

Multiple instruments have been developed over the years to assess the methodological quality, along with bias, in RCTs (50,53). The Cochrane review editorial board has developed its own criteria, which has been utilized extensively and has been modified over the years. Table 12 and Appendix Table 1 shows Cochrane review criteria (53), and Table 13 and Appendix Table 2 shows criteria developed by interventional pain physicians with a specific item checklist for the assessment of RCTs of interventional pain management techniques (50). The criteria developed by Manchikanti et al (50) is

Table 12. *Sources of risk of bias.*

Bias Domain	Source of Bias	Possible Answers
Selection	(1) Was the method of randomization adequate?	Yes/No/Unsure
Selection	(2) Was the treatment allocation concealed?	Yes/No/Unsure
Performance	(3) Was the patient blinded to the intervention?	Yes/No/Unsure
Performance	(4) Was the care provider blinded to the intervention?	Yes/No/Unsure
Detection	(5) Was the outcome assessor blinded to the intervention?	Yes/No/Unsure
Attrition	(6) Was the drop-out rate described and acceptable?	Yes/No/Unsure
Attrition	(7) Were all randomized participants analyzed in the group to which they were allocated?	Yes/No/Unsure
Reporting	(8) Are reports of the study free of suggestion of selective outcome reporting?	Yes/No/Unsure
Selection	(9) Were the groups similar at baseline regarding the most improtant prognosis indications?	Yes/No/Unsure
Performance	(10) Were cointerventions avoided or similar?	Yes/No/Unsure
Performance	(11) Was the compliance acceptable in all groups?	Yes/No/Unsure
Detection	(12) was the timing of the outcome assessment similar in all groups?	Yes/No/Unsure
Other	(13) Are the other sources of potential bias unlikely?	Yes/No/Unsure

Source: Furlan AD, Malmivaara A, Chou R, et al; Editorial Board of the Cochrane Back, Neck Group. 2015 Updated Method Guideline for Systematic Reviews in the Cochrane Back and Neck Group. Spine (Phila Pa 1976) 2015; 40:1660-1673 (53).

Table 13. *Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM – QRB.*

		Scoring
1.	Trial Design and Guidance Reporting	0-3
2.	Type and Design of Trial	0-3
3.	Setting/Physician	0-2
4.	Imaging	0-3
5.	Sample Size	0-3
6.	Statistical Methodology	0-1
7.	Inclusiveness of Population	0-2
8.	Duration of Pain	0-2
9.	Previous Treatments	0-2
10.	Duration of Follow-up with Appropriate Interventions	0-3
11.	Outcomes Assessment Criteria for Significant Improvement	0-4
12.	Analysis of all Randomized Participants in the Groups	0-2
13	Description of Drop Out Rate	0-2
14.	Similarity of Groups at Baseline for Important Prognostic Indicators	0-2
15.	Role of Co-Interventions	0-1
16.	Method of Randomization	0-2
17.	Concealed Treatment Allocation	0-2
18.	Patient Blinding	0-1
19.	Care Provider Blinding	0-1
20.	Outcome Assessor Blinding	0-1
21.	Funding and Sponsorship	-3-+3
22.	Conflicts of Interest	-3-+3
	TOTAL MAXIMUM	48

Source: Manchikanti L, et al. Assessment of methodologic quality of randomized trials of interventional techniques: Development of an interventional pain management specific instrument. *Pain Physician* 2014; 17:E263-E290 (50).

specific to interventional techniques and assesses the design, patient inclusion criteria, conduct, outcomes, randomization, allocation concealment, blinding, and conflicts of interest. Consequently, the instrument developed by Manchikanti et al will be the most appropriate instrument for utilization in studies relating to regenerative medicine to assess their quality. These instruments have been utilized in multiple systematic reviews and meta-analyses.

Although Cochrane criteria are universally accepted and were included in multiple trials, this was not specific for interventional techniques. In contrast, Interventional Pain Management Techniques – Quality Appraisal of Reliability and Risk of Bias Assessment (IPM-QRB) was specifically developed for interventional techniques, specifically in patients suffering with chronic spinal pain. This checklist includes various types of criteria, including trial design and guidance report, along with setting, physician, imaging, chronicity of pain, previous treatments, and multiple other appropriate criteria. It has been shown to be more robust than Cochrane review criteria and were considered as providing better information than Cochrane review criteria when compared head-to-head with both Cochrane review criteria and IPM-QRB. Literature pertaining to the SIGN (29,76) is not extensive, even though it has been reported in some studies related to interventional techniques (16,29).

Nonrandomized Studies

Similar to the checklist for RCTs, Manchikanti et al (51) developed a comprehensive instrument that is helpful in assessing the methodological quality of nonrandomized trials and is specific to interventional techniques (Table 14 and Appendix Table 3).

IPM checklist with Interventional Pain Management Techniques – Quality Appraisal of Reliability and Risk of Bias Assessment for Nonrandomized Studies (IPM-QRBNR) has been evaluated in multiple assessments. With the rapid development of RCTs, observational studies are not as frequently used. Further, methodologic quality assessment for these is not utilized.

SIGN also has developed an instrument to assess the methodologic quality and risk of bias assessment in observational studies (16,29). In contrast to RCTs, observational studies have not been methodologically assessed as frequently. Further instruments for assessment are also limited.

QUADAS and QAREL

Quality Assessment of Diagnostic Accuracy Studies (QUADAS) and Quality Appraisal Tool for Studies of Diagnostic Reliability (QAREL) were developed to assess diagnostic accuracy studies (92,93). In developing QAREL, key principles for the quality of studies of diagnostic reliability were identified. These included epidemiologic principles, existing quality appraisal checklist, and the Standards for Reporting of Diagnostic Accuracy (STARD) and QUADAS resources. They developed a checklist of 11 items that explored 7 principles. Items covered the spectrum of subjects, the spectrum of examiners, examiner blinding, order effects of examination, suitability of the time interval among

repeated measurements, appropriate test application and interpretation, and appropriate statistical analysis as shown in Table 15. In addition, the QAREL checklist was also assessed for reliability (93,94). Authors found that QAREL was a reliable assessment tool for studies of diagnostic reliability when raters agreed on criteria for the interpretation of each item. Nine out of 11 items had good or moderate reliability, and 2 items achieved fair reliability. The heterogeneity in the test included in this study may have resulted in an underestimation of the reliability of these 2 items.

Literature Search

Comprehensive searches should be performed from at least 3 sources without language restrictions, with appropriately developed search terminology of search. Widely available sources include:

1. PubMed from 1966
   www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed
2. EMBASE from 1980
   www.embase.com
3. Cochrane Library
   www.thecochranelibrary.com/view/0/index.html
4. Google Scholar
   https://scholar.google.com/
5. US National Guideline Clearinghouse (NGC)
   www.guideline.gov
6. Clinical Trials
   www.clinicaltrials.gov
7. Interventional Trial Registries
8. Previous systematic reviews and cross references.

Measurement of Treatment Effect in Data Syntheses (Meta-Analyses)

Data were summarized using meta-analyses when at least 5 studies per type of disorder were available meeting the inclusion criteria.

Conclusions of both qualitative and quantitative outcome measures were evaluated. Qualitative (the direction of a treatment effect) and quantitative (the magnitude of a treatment effect) conclusions were evaluated. Random-effects meta-analyses to pool data were also used (95,96).

The minimum amount of change in pain score to be clinically meaningful has been described as a 2-point change on a scale of 0 to 10 (or 20 percentage points), based on findings in trials studying general chronic pain, chronic musculoskeletal pain, and chronic low back pain (97). However, recent studies evaluating interventional techniques have used 50% or more pain relief as the cutoff threshold for clinically meaningful improvement in pain relief or functional status. Consequently, for analysis in these systematic reviews, we utilized clinically meaningful pain relief of at least a 3-point change on an 11-point scale of 0 to 10, or 50% pain relief from the baseline, and/or a functional status improvement of 40% or more as clinically significant.

Outcomes may be assessed between the groups or in the same group from baseline to posttreatment; however, some methodologists tend to focus only on between the groups. This essentially provides lack of improvement or lack of difference between the groups in an active control trial, noninferiority, or equivalence trial. Thus it is essential that outcomes be monitored pre- and posttreatment rather than between the groups or utilizing both methodologies. Consequently, in all the systematic reviews and the evidence assessment for interventional pain management, the outcomes have been assessed based on the design between the groups and in the same group pre- and posttreatment.

Table 14. *IPM checklist for assessment of nonrandomized or observational studies of IPM techniques utilizing IPM-QRBNR.*

		Scoring
1.	Study Design Guidance and Reporting	0-4
2.	Study Design and Type	0-4
3.	Setting/Physician	0-2
4.	Imaging	0-3
5.	Sample Size	0-4
6.	Statistical Methodology	0-2
7.	Inclusiveness of Population	1-4
8.	Duration of Pain	0-2
9.	Previous Treatments	0-2
10.	Duration of Follow-up with Appropriate Interventions	1-4
11.	Outcomes Assessment Criteria for Significant Improvement	0-4
12.	Description of Drop Out Rate	0-2
13	Similarity of Groups at Baseline for Important Prognostic Indicators	0-2
14.	Role of Co-Interventions	1-2
15.	Method of Assignment of Participants	1-4
16.	Funding and Sponsorship	-3-+3
	TOTAL MAXIMUM	48

Source: Manchikanti L, et al. Development of an interventional pain management specific instrument for methodologic quality assessment of nonrandomized studies of interventional techniques. *Pain Physician* 2014; 17: E291-E317 (51).

Table 15. *Quality Appraisal of Diagnostic Reliability (QAREL) checklist.*

Item	Yes	No	Unclear	N/A
1. Was the test evaluated in a spectrum of subjects representative to patients who would normally receive the test in clinical practice?				
2. Was the test performed by examiners representative of those who would normally perform the test in practice?				
3. Were raters blinded to the reference standard for the target disorder being evaluated?				
4. Were raters blinded to the findings of other raters during the study?				
5. Were raters blinded to their own prior outcomes of the test under evaluation?				
6. Were raters blinded to clinical information that may have influenced the test outcome?				
7. Were raters blinded to additional cues, not intended to form part of the diagnostice test procedure?				
8. Was the order in which raters examined subjects varied?				
9. Were appropriate statistical measures of agrement used?				
10. Was the application and interpretation of the test appropriate?				
11. Was the time interval between measurements suitable in relation to the stability of the variable being measured?				
12. If there were dropouts from the study, was this less than 20% of the sample?				
TOTAL				

Source: Lucas N, et al. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010; 63:854-861 (94).

Outcome of the Studies

Randomized trials were judged to be positive if the intervention was clinically relevant and effective, either with a placebo control or active control. This indicates that the difference in effect for the primary outcome measure is statistically significant on the conventional 5% level. In a negative study, no significant difference between the treatment groups or no improvement from baseline is identified.

Observational studies were judged to be positive if the intervention was effective, with outcomes reported at 1 month, 3 months, 6 months, and 1 year.

The outcomes were judged as improvement in at least 40% of patients at distinct reference points with positive or negative results reported at 1 month, 3 months, 6 months, and 1 year.

Outcomes included the prevalence of pain and false-positive rate. Based on the earlier-described parameters, the reliability of the data derived from each study were assessed.

The advantages and disadvantages of various methodologies available are too extensive to be described in this manuscript. These have been described in various other manuscripts in the past (32,39,59).

Analysis of Evidence

Evidence analysis was performed based on US Preventive Task Force (USPSTF) criteria as illustrated in Table 16, which has been utilized by multiple authors (98).

The analysis was conducted using 3 levels of evidence ranging from good, fair, and limited or poor.

Grading or Rating of the Quality or Strength of Evidence

The grading of evidence must be based on systematic reviews, meta-analyses, and evidence development by other guidance from RCTs, observational studies, and other clinical reports. The grading of evidence is based on the best evidence synthesis developed as a modified approach to the grading of evidence by the American Society of Interventional Pain Physicians (ASIPP) as shown in Table 17.

This methodology specifies the level of scientific evidence and offers a transparent approach to grading quality of evidence and strength of recommendations. The CDC has adopted the GRADE method (99). AHRQ also has recommended a similar strength of recommendation (70).

Table 17 shows the qualitative modified approach to the grading of evidence providing a rating for strength of evidence, whereas Table 18 shows guidance for the strength of recommendations. Level I provides strong or significant evidence, with high confidence

Table 16. *Method for grading the overall strength of the evidence for an intervention.*

Grade	Definition
Good	Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes (at least 2 consistent, higher-quality RCTs or studies of diagnostic test accuracy).
Fair	Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, size, or consistency of included studies; generalizability to routine practice, or indirect nature of the evidence on health outcomes (at least one higher-quality trial or study of diagnostic test accuracy of sufficient sample size; 2 or more higher-quality trials or studies of diagnostic test accuracy with some inconsistency; at least 2 consistent, lower-quality trials or studies of diagnostic test accuracy, or multiple consistent observational studies with no significant methodological flaws.)
Limited or poor	Evidence is insufficient to assess effects on health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality trials, important flaws in trial design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes.

Adapted from methods developed by U.S. Preventive Services Task Force (98).

that the available evidence reflects the true magnitude and direction of the net effect, and further research is very unlikely to change either the magnitude or direction of this net effect. Level II provides moderate or intermediate evidence with moderate confidence that the available evidence reflects the true magnitude and direction of the net effect. Thus further research may be unlikely to alter the direction of the net effect but may alter the magnitude of the net effect.

Levels III to V provide weak evidence with low confidence that the available evidence reflects the true magnitude and direction of the net effect. Consequently, further research may change the magnitude and/or the direction of this net effect.

Assessment and Recommendations of Benefits and Harms

The guidelines clearly describe the potential benefits and harms for the interventions and explicitly link the information to specific recommendations. Guidelines supporting documents must summarize the relevant supporting evidence and explicitly link this information to recommendations.

Table 17. *Qualitative modified approach to grading of evidence.*

Level I	Strong	Evidence obtained from multiple relevant high quality randomized controlled trials for effectiveness
Level II	Moderate	Evidence obtained from at least one relevant high quality randomized controlled trial or multiple relevant moderate or low quality randomized controlled trials
Level III	Fair	Evidence obtained from at least one relevant high quality nonrandomized trial or observational study with multiple moderate or low quality observational studies
Level IV	Limited	Evidence obtained from multiple moderate or low quality relevant observational studies
Level V	Consensus based	Opinion or consensus of large group of clinicians and/or scientists for effectiveness as well as to assess preventive measures, adverse consequences, effectiveness of other measures.

Adapted from: Manchikanti L, Falco FJE, Benyamin RM, Kaye AD, Boswell MV, Hirsch JA. A modified approach to grading of evidence. *Pain Physician* 2014; 17:E319-E325 (52).

Table 18. *Guide for strength of recommendations.*

Strong	There is high confidence that the recommendation reflects best practice. This is based on: a) strong evidence for a true net effect (e.g., benefits exceed harms); b) consistent results, with no or minor exceptions; c) minor or no concerns about study quality; and/or d) the extent the panelists' agreement. Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a strong recommendation.
Moderate	There is moderate confidence that the recommendation reflects best practice. This is based on: a) good evidence for a true net effect (e.g. benefits exceed harms); b) consistent results, with minor and/or few exceptions; c) minor and/or few concerns about study quality; and/or d) the extent of panelists' agreement. Other compelling considerations (discussed in the guideline's literature review and analyses) may also warrant a moderate recommendation.
Weak	There is some confidence that the recommendation offers the best current guidance for practice. This is based on: a) limited evidence for a true net effect (e.g., benefits exceed harms); b) consistent results, but with important exceptions; c) concerns about study quality; and/or d) the extent of panelists' agreement. Other considerations (discussed in the guideline's literature review and analyses) may also warrant a weak recommendation.

Source: National Guideline Clearinghouse Extent Adherence to Trustworthy Standards (NEATS) instrument (70).

IOM standards demand that for each recommendation, a rating of the strength of the recommendation in light of benefits and harms, available evidence, and the confidence in the underlying evidence should be provided. In preparation of these guidelines, the rating schemes recommended by NEATS were utilized as shown in Table 18 (70).

Grading Recommendations

As recommended by the IOM, for each recommendation, information was provided with an explanation of the reasoning underlying the recommendation, including a clear description of potential benefits and harms; a summary of the relevant available evidence; description of the quality, quantity, and consistency of the aggregate available evidence; an explanation of the part played by values, opinion, theory, and clinical experience in deriving the recommendations; a rating of the level of confidence; a rating of the strength of recommendation; and a description and explanation of any differences of opinion regarding the recommendation.

In grading recommendations, the grading of recommendations from USPSTF was utilized (Table 16).

Specificity of Recommendations

Guideline recommendations must, to a great extent, be specific and unambiguous, providing guidance on what actions should or should not be taken in various situations of chronic opioid therapy for various population groups.

External Review

Guidelines preferably should be subjected to external peer review. In addition, the guidelines must also be posted on organization's website, and in the newsletters, et cetera, to obtain comments from stakeholders, scientific and clinical experts, organizations, patients, and representation of the public.

Updating of Guidelines

The guidelines must be updated in a window of 3 to 5 years based on significant changes in the evidence, public policy, or adverse events.

Conclusions

In this manuscript, we described comprehensively the evidence synthesis from systematic reviews to observational studies, including methodologic quality and bias assessment. The manuscript described various methods utilized in the assessment of the quality of systematic reviews, RCTs, diagnostic accuracy studies, and observational studies. We also describe various factors that continue to impede the development of appropriate clinical practice guidelines. These impediments include biases due to a variety of conflicts and confluence of interest, inappropriate and inadequate and insufficient methodologic quality, poor writing, and ambiguous presentation, projecting a view that these do not apply to individual patients or too restrictive with the elimination of clinical autonomy, and overzealous and inappropriate recommendations, either positive, negative, or noncommittal.

Acknowledgments

The authors wish to thank Vidyasagar Pampati, MSc, for statistical assistance; and Bert Fellows, MA, Director Emeritus of Psychological Services at Pain Management Centers of America, for manuscript review; and Tonie M. Hatton and Diane E. Neihoff, transcriptionists, for their assistance in preparation of this manuscript. We would like to thank the editorial board of Pain Physician for review and criticism in improving the manuscript.

Author Contributions

The review was designed by LM, MB, ADK, and JH.

All authors contributed to preparation to the manuscript, reviewed, and approved the content with final version.

Author Affiliations

Dr. Manchikanti is Co-Director, Pain Management Centers of America, Paducah, KY, Clinical Professor, Anesthesiology and Perioperative Medicine, University of Louisville, Louisville, KY, and Professor of Anesthesiology-Research, Department of Anesthesiology, School of Medicine, LSU Health Sciences Center, New Orleans and Shreveport, LA, USA

drlm@thepainmd.com

Dr. Atluri is Medical Director, Tri-State Spine Care Institute, Cincinnati, OH, USA

saiatluri@gmail.com

Dr. Boswell is Professor and Chairman, Pain Medicine, Department of Anesthesiology, University of Toledo College of Medicine and Life Sciences, Toledo, OH, USA

mark.boswell2@utoledo.edu; mvboswell01@gmail.com

Dr. Calodney is Director of Clinical Research, Precision Spine Care, Tyler, TX and Department of Anesthesiology, Louisiana State University, Shreveport, LA, USA
aaroncalodney@me.com

Dr. Diwan is President, Park Avenue Spine and Pain, and Associate Clinical Professor, Albert Einstein College of Medicine, New York, NY, USA
sudhir.diwan63@gmail.com

Dr. Gupta is a Consultant in Pain Medicine, Bradford Teaching Hospital NHS Foundation Trust, Bradford Royal Infirmary, Bradford, United Kingdom

Dr. Kaye is Professor, Pain Fellowship Director, Provost, Chief Academic Officer, Vice Chancellor of Academic Affairs, Departments of Anesthesiology and Pharmacology, Toxicology, and Neurosciences, LSU Health Sciences Center, Shreveport, Oshcher Shreveport Hospital and Pain Clinic Feist-Wieller Cancer Center, Shreveport, LA, USA
akaye@lsuhsc.edu, alankaye44@hotmail.com

Dr. Knezevic is Vice Chair for Research and Education; Department of Anesthesiology; Advocate Illinois Masonic Medical Center, Chicago, IL and Clinical Professor, Department of Anesthesiology, and Clinical Professor Department of Surgery, College of Medicine, University of Illinois, Chicago, IL, USA
nick.knezevic@gmail.com

Dr. Candido is Chairman, Department of Anesthesiology, Advocate Illinois Masonic Medical Center and Professor of Clinical Anesthesia and Clinical Surgery, University of Illinois College of Medicine, Chicago, IL, USA
kdcandido1@gmail.com; kdcandido@yahoo.com

Dr. Abd-Elsayed is Medical Director, UW Health Pain Services, and Associate Professor, Anesthesiology, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA
alaawny@hotmail.com; abdelsayed@wisc.edu

Dr. Pappolla is Medical Director, St. Michael's Pain and Spine Clinics, Houston, TX, and Professor of Neurology, Univeristy of Texas Medical Branch, Galveston, TX, USA
pappolla@aol.com

Dr. Racz is Professor and Chairman Emeritus Anesthesiology, Texas Tech University Health Sciences Center, Lubbock, TX, and Preston Surgery Center, Frisco, TX, USA
gbracz@yahoo.com

Dr. Sachdeva is Associate Professor of Anesthesiology and Pain Medicine, Vice-Chair, Pain Medicine Director, Pain Division and Fellowship Program, University of Cincinnati College of Medicine, Cincinnati, OH, USA
harsh.sachdeva@uc.edu

Dr. Sanapati is Co-Director, Pain Management Centers of America, Evansville, IN, and Assistant Professor of Anesthesiology and Research, School of Medicine, LSU Health Sciences Center, New Orleans, LA, USA
msanapati@gmail.com

Dr. Shah is Vice-Chair, Department of Anesthesiology, and Enterprise Director, Pain Services, University of California Irvine, Orange, CA, USA
shahshalini@gmail.com

Dr. Singh is Clinical Associate Professor, Department of Anesthesiology, Peri-operative, and Pain Medicine, Stanford University, Stanford, CA, USA
vanila@stanford.edu

Dr. Soin is Medical Director, Ohio Pain Clinic, and Clinical Assistant Professor of Surgery at Wright State University, Dayton, OH, USA
drsoin@gmail.com

Dr. Hirsch is Vice Chair and Service Line Chief of Neurointerventional Radiology, Chief of Neurointerventional Spine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
jahirsch@mgh.harvard.edu

Appendix tables are available at www.painphysicianjournal.com

References

1. Manchikanti L, Falco FJE, Singh V, et al. An update of comprehensive evidence-based guidelines for interventional techniques of chronic spinal pain. Part I: Introduction and general considerations. *Pain Physician* 2013; 16:S1-S48.
2. Manchikanti L, Kaye AD, Soin A, et al. Comprehensive evidence-based

guidelines for facet joint interventions in the management of chronic spinal pain: American Society of Interventional Pain Physicians (ASIPP) guidelines. *Pain Physician* 2020; 23:S1-S127.

3. Manchikanti L, Centeno CJ, Atluri S, et al. Bone marrow concentrate (BMC) therapy in musculoskeletal disorders: Evidence-based policy position statement of American Society of Interventional Pain Physicians (ASIPP). *Pain Physician* 2020; 23:E85-E131.
4. Navani A, Manchikanti L, Albers SL, et al. Responsible, safe, and effective use of biologics in the management of low back pain: American Society of Interventional Pain Physicians (ASIPP) guidelines. *Pain Physician* 2019; 22:S1-S74.
5. Kaye AD, Manchikanti L, Novitch MB, et al. Responsible, safe, and effective use of antithrombotics and anticoagulants in patients undergoing interventional techniques: American Society of Interventional Pain Physicians (ASIPP) guidelines. *Pain Physician* 2019; 22:S75-S128.
6. Manchikanti L, Helm S 2nd, Benyamin RM, Hirsch JA. A critical analysis of Obamacare: Affordable care or insurance for many and coverage for few? *Pain Physician* 2017; 20:111-138.
7. Manchikanti L, Benyamin RM, Falco FJE, Caraway DL, Datta S, Hirsch JA. Guidelines warfare over interventional techniques: Is there a lack of discourse or straw man? *Pain Physician* 2012; 15:E1-E26.
8. Obama B. United States health care reform: Progress to date and next steps. *JAMA* 2016; 316:525-532.
9. Blumenthal D, Collins SR, Fowler EJ. The Affordable Care Act at 10 Years - Its coverage and access provisions. *N Engl J Med* 2020; 382:963-969.
10. Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E (eds); Committee on Standards for Developing Trustworthy Clinical Practice Guidelines; Institute of Medicine. *Clinical Practice Guidelines We Can Trust.* The National Academies Press, Washington, DC, 2011.
11. Field MJ, Lohr KN (eds). Committee to Advise the Public Health Service on *Clinical Practice Guidelines, Institute of Medicine.* Clinical Practice Guidelines. Directions for a New Program. National Academy Press, Washington, 1990.
12. Manchikanti L, Knezevic NN, Boswell MV, Kaye AD, Hirsch JA. Epidural injections for lumbar radiculopathy and spinal stenosis: A comparative systematic review and meta-analysis. *Pain Physician* 2016; E365-E410.
13. Chou R, Hashimoto R, Friedly J, et al. Pain Management Injection Therapies for Low Back Pain. Technology Assessment Report ESIB0813. (Prepared by the Pacific Northwest Evidence-based Practice Center under Contract No. HHSA 290-2012-00014-I.) Rockville, MD: Agency for Healthcare Research and Quality; July 10, 2015. Accessed 12/14/2020. https://www.cms.gov/medicare/coverage/determinationprocess/downloads/id98ta.pdf
14. Boswell MV, Manchikanti L. Appropriate design and methodologic quality assessment, clinically relevant outcomes are essential to determine the role of epidural corticosteroid injections. Commentary RE: Chou R, Hashimoto R, Friedly J, Fu R, Bougatsos C, Dana T, Sullivan SD, Jarvik J. Epidural corticosteroid injections for radiculopathy and spinal stenosis: A systematic review and meta-analysis. Ann Intern Med 2015; 163:373-381. Evid *Based Med* 2016; 21:89.
15. Manchikanti L, Knezevic NN, Sanapati SP, Sanapati MR, Kaye AD, Hirsch JA. Is percutaneous adhesiolysis effective in managing chronic low back and lower extremity pain in post-surgery syndrome: A systematic review and meta-analysis. *Curr Pain Headache Rep* 2020; 24:30.
16. Manchikanti L, Soin A, Boswell MV, Kaye AD, Sanapati M, Hirsch JA. Effectiveness of percutaneous adhesiolysis in post lumbar surgery syndrome: A systematic analysis of findings of systematic reviews. *Pain Physician* 2019; 22:307-322.
17. Manchikanti L, Knezevic NN, Sanapati MR, et al. Effectiveness of percutaneous adhesiolysis in managing chronic central lumbar spinal stenosis: A systematic review and meta-analysis. *Pain Physician* 2019; 22:E523-E550.
18. Manchikanti L, Knezevic NN, Sanapati J, Kaye AD, Sanapati M, Hirsch JA. Is epidural injection of sodium chloride solution a true placebo or an active control agent? A systematic review and meta-analysis. *Pain Physician* 2021; 24:41-59.
19. Manchikanti L, Knezevic NN, Parr A, Kaye AD, Sanapati M, Hirsch JA. Does epidural bupivacaine with or without steroids provide long-term relief? A systematic review and meta-analysis. *Curr Pain Headache Rep* 2020; 24:26.
20. Knezevic N, Manchikanti L, Urits I, et al. Lack of superiority of epidural injections with lidocaine with steroids compared to without steroids in spinal pain: A systematic review and meta-analysis. *Pain Physician* 2020; 4S:S239-S270.
21. Lee JH, Shin KS, Park SJ, et al. Comparison of clinical efficacy between transforaminal and interlaminar epidural injections in lumbosacral disc herniation: A systematic review and meta-analysis. *Pain Physician* 2018; 21:433-448.
22. Lee JH, Kim DH, Kim DH, et al. Comparison of clinical efficacy of epidural injection with or without steroid in lumbosacral disc herniation: A systematic review and meta-analysis. *Pain Physician* 2018; 21:449-468.
23. Lee JH, Shin KH, Bahk SJ, et al. Comparison of clinical efficacy of transforaminal and caudal epidural steroid injection in lumbar and lumbosacral disc herniation: A systematic review and meta-analysis. *Spine J* 2018; 18:2343-2353.
24. Oliveira CB, Maher CG, Ferreira ML, et al. Epidural corticosteroid injections for lumbosacral radicular pain. *Cochrane Database Syst Rev* 2020; 4:CD013577.
25. Shanthanna H, Busse J, Wang L, et al. Addition of corticosteroids to local anaesthetics for chronic non-cancer pain injections: A systematic review and meta-analysis of randomised controlled trials. *Br J Anaesth* 2020; 125:779-801.
26. Chou R, Atlas SJ, Loeser JD, Rosenquist RW, Stanos SP. Guideline warfare over interventional therapies for low back pain: Can we raise the level of discourse? *J Pain* 2011; 12:833-839.
27. The Cochrane Collaboration. Evidence Health Care. Accessed 12/114/2020. https://www.cochrane.org/evidence
28. Pinto RZ, Maher CG, Ferreria ML, et al. Epidural corticosteroid injections in the management of sciatica: A systematic review and meta-analysis. *Ann Intern Med* 2012; 157:865-877.
29. Cho JH, Lee JH, Song KS, et al. Treatment outcomes for patients with failed back surgery. *Pain Physician* 2017; 20:E29-E43.
30. Cappola AR, FitzGerald GA. Confluence, not conflict of interest: Name change necessary. *JAMA* 2015;

314:1791-1792.

31. Fung J. The corruption of evidence based medicine- Killing for profit. Medium, April 10, 2018. Accessed 12/14/2020. https://medium.com/@drjasonfung/the-corruption-of-evidence-based-medicine-killing-for-profit-41f2812b8704
32. Wilson K. Evidence-based medicine. The good the bad and the ugly. A clinician's perspective. *J Eval Clin Pract* 2010; 16:398-400.
33. Khodambashi S, Nytrø Ø. Reviewing clinical guideline development tools: features and characteristics. *BMC Med Inform Decis Mak* 2017; 17:132.
34. Eisner DA. Reproducibility of science: Fraud, impact factors and carelessness. *J Mol Cell Cardiol* 2018; 114:364-368.
35. Vandvik PO, Brandt L, Alonso-Coello P, et al. Creating clinical practice guidelines we can trust, use, and share: A new era is imminent. *Chest* 2013; 144:381-389.
36. Rafiq M, Boccia S. Application of the GRADE approach in the development of guidelines and recommendations in genomic medicine. *Genomics Insights* 2018; 11:1178631017753360.
37. Schünemann HJ, Santesso N, Vist GE, et al. Using GRADE in situations of emergencies and urgencies: Certainty in evidence and recommendations matters during the COVID-19 pandemic, now more than ever and no matter what. *J Clin Epidemiol* 2020;127:202-207.
38. Gad El-Rab W, Zaïane OR, El-Hajj M. Formalizing clinical practice guideline for clinical decision support systems. *Health Informatics J* 2017; 23:146-156.
39. Walter HJ, Bukstein OG. Forward to the future: Clinical updates and clinical practice guidelines. *J Am Acad Child Adolesc Psychiatry* 2017; 56:811-812.
40. Sniderman AD, LaChapelle KJ, Rachon NA, Furberg CD. The necessity for clinical reasoning in the era of evidence-based medicine. *Mayo Clin Proc* 2013; 88:1108-1114.
41. Grannis FW Jr. Methodologists are not qualified to write clinical guidelines. *Lancet* 2009; 374:1325-1326.
42. Mahlknecht P, Glechner A, Gartlehner G. Guideline development: Going from evidence to recommendations. Challenges and opportunities--a methodologist's view. *Z Evid Fortbild Qual Gesundhwes* 2015; 109:319-329.
43. Isaacs D, Fitzgerald D. Seven alternatives to evidence based medicine. *BMJ* 1999; 319:1618.
44. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: Rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015; 350:h870.
45. Argyriou C, Georgiadis GS, Georgakarakos EI, et al. Applying evidence-based medicine in actual clinical practice: Can we bridge the gap? A review of the literature. *Hellenic J Cardiol* 2015; 56:373-378.
46. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: The significance and presentation of recommendations. *J Clin Epidemiol* 2013; 66:719-725.
47. Schünemann HJ, Cuello C, Akl EA, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *J Clin Epidemiol* 2019; 111:105-114.
48. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: Development and pilot validation. *Syst Rev* 2014; 3:82.
49. Smith MT, McCrae CS, Cheung J, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: An american academy of sleep medicine clinical practice guideline. *J Clin Sleep Med* 2018; 14:1231-1237.
50. Manchikanti L, Hirsch JA, Cohen SP, et al. Assessment of methodologic quality of randomized trials of interventional techniques: Development of an interventional pain management specific instrument. *Pain Physician* 2014; 17:E263-E290.
51. Manchikanti L, Hirsch JA, Heavner JE, et al. Development of an interventional pain management specific instrument for methodologic quality assessment of nonrandomized studies of interventional techniques. *Pain Physician* 2014; 17:E291-E317.
52. Manchikanti L, Falco FJE, Benyamin RM, Kaye AD, Boswell MV, Hirsch JA. A modified approach to grading of evidence. *Pain Physician* 2014; 17:E319-E325.
53. Furlan AD, Malmivaara A, Chou R, et al; Editorial Board of the Cochrane Back, Neck Group. 2015 updated method guideline for systematic reviews in the Cochrane Back and Neck Group. *Spine (Phila Pa 1976)* 2015; 40:1660-1673.
54. Riado Minguez D, Kowalski M, Vallve Odena M, et al. Methodological and reporting quality of systematic reviews published in the highest ranking journals in the field of pain. *Anesth Analg* 2017; 125:1348-1354.
55. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Accessed 12/14/2020. http://prisma-statement.org/
56. A Measurement Tool to Assess Systematic Reviews (AMSTAR). Accessed 12/14/2020. https://amstar.ca/
57. Saarni SI, Gylling HA. Evidence based medicine guidelines: A solution to rationing or politics disguised as science? *J Med Ethics* 2004; 30:171-175.
58. 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.
59. Timmermans S, Mauck A. The promises and pitfalls of evidence-based medicine. *Health Aff* 2005; 24:18-28.
60. Manchikanti L, Falco FJE, Boswell MV, Hirsch JA. Facts, fallacies, and politics of comparative effectiveness research: Part 2. Implications for interventional pain management. *Pain Physician* 2010; 13:E55-E79.
61. Manchikanti L, Sanapati MR, Pampati V, Boswell MV, Kaye AD, Hirsch JA. Update on reversal and decline of growth of utilization of interventional techniques in managing chronic pain in the Medicare population from 2000 to 2018. *Pain Physician* 2019; 22:521-536.
62. Manchikanti L, Sanapati MR, Soin A, et al. An updated analysis of utilization of epidural procedures in managing chronic pain in the Medicare population from 2000 to 2018. *Pain Physician* 2020; 12:111-126.
63. Manchikanti L, Sanapati MR, Pampati V, et al. Update of utilization patterns of facet joint interventions in managing spinal pain from 2000 to 2018 in the US fee-for-service Medicare population. *Pain Physician* 2020; 23:E133-E149.
64. Manchikanti L, Manchikanti MV, Vanaparthy R, Kosanovic R, Pampati V. Utilization patterns of sacroiliac joint injections from 2000 to 2018 in fee-for-service Medicare population. *Pain Physician* 2020; 23:439-450.
65. Manchikanti L, Kosanovic R, Pampati V, Kaye AD. Declining utilization patterns of percutaneous adhesiolysis

procedures in the fee-for-service (FFS) Medicare population. *Pain Physician* 2021; 24:17-29.

66. Dieleman JL, Cao J, Chapin A, et al. US health care spending by payer and health condition, 1996-2016. *JAMA* 2020; 323:863-884.
67. Keehan SP, Cuckler GA, Poisal JA, et al. National health expenditure projections, 2019-28: Expected rebound in prices drives rising spending growth. *Health Aff (Millwood)* 2020; 39:704-714.
68. U.S. Department of Health and Human Services Agency for Healthcare Research and Quality (AHRQ). Accessed 12/14/2020. www.ahrq.gov
69. National Institute for Health and Clinical Excellence (NICE). Accessed 12/14/2020. www.nice.org.uk
70. National Guideline Clearinghouse Extent Adherence to Trustworthy Standards (NEATS) instrument.
71. Zähringer J, Schwingshackl L, Movsisyan A, et al. Use of the GRADE approach in health policymaking and evaluation: A scoping review of nutrition and physical activity policies. *Implement Sci* 2020; 15:37.
72. BMJ. What is GRADE? 2017. Accessed 12/14/2020. https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/
73. Gupta S, Manohar M. Survey of Awareness of the Guidelines Published by the Faculty of Pain Medicine and the British Pain Society. Top 5 Abstracts from the 11th Annual North England Pain Medicine Group Meeting, 2019. *Br J Pain* 2019; 13:1-6. 74. Gupta S, Brown M. Guidance Documents: A guide to development, dissemination and review. Transmitter. Published by the Faculty of Pain Medicine of the Royal College of Anaesthetists, London, Spring 2020, pp 12-13.
75. Eden J, Levit L, Berg A, Morton S (eds); Committee on Standards for Systematic Reviews of Comparative Effectiveness Research; Institute of Medicine. *Finding What Works in Health Care.* Standards for Systematic Reviews. The National Academies Press, Washington, DC, 2011.
76. Baird AG, Lawrence JR. Guidelines: is bigger better? A review of SIGN guidelines. *BMJ Open* 2014; 4:e004278.
77. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009; 62:e1-34.
78. Kattan S, Lee SM, Kohli MR, Setzer FC, Karabucak B. Methodological quality assessment of meta-analyses in endodontics. *J Endod* 2018; 44:22-31.
79. Ross A, Rankin J, Beaman J, et al. Methodological quality of systematic reviews referenced in clinical practice guidelines for the treatment of opioid use disorder. *PLoS One* 2017; 12:e0181927.
80. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; 7:10.
81. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials E10. July 20, 2000.
82. Manchikanti L, Boswell MV, Kaye AD, Helm II S, Hirsch JA. Therapeutic role of placebo: Evolution of a new paradigm in understanding research and clinical practice. *Pain Physician* 2017; 20:363-386.
83. Carlino E, Vase L. Can knowledge of placebo and nocebo mechanisms help improve randomized clinical trials? *Int Rev Neurobiol* 2018; 138:329-357.
84. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 2012; 11:MR000030.
85. Kao LS, Tyson JE, Blakely ML, et al. Clinical research methodology I: Introduction to randomized trials. *J Am Coll Surg* 2008; 206:361-369.
86. Manchikanti L, Hirsch JA, Smith HS. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 2: Randomized controlled trials. *Pain Physician* 2008; 11:717-773.
87. De Meulemeester J, Fedyk M, Jurkovic L, et al. Many randomized clinical trials may not be justified: A cross-sectional analysis of the ethics and science of randomized clinical trials. *J Clin Epidemiol* 2018; 97:20-25.
88. Lund H, Brunnhuber K, Juhl C, et al. Towards evidence based research. *BMJ* 2016; 355:i5440.
89. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Int J Surg* 2014; 12:1495-1499.
90. Mann CJ. Observational research methods. Research design II: Cohort, cross sectional, and case-control studies. *Emerg Med J* 2003;20:54-60.
91. Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Fam Pract* 2004; 21:4-10.
92. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: A tools for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:25.
93. Lucas N, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol* 2013; 13:111.
94. Lucas N, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010; 63:854-861.
95. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327:557-560.
96. Harbord R, Higgins J. METAREG: Stata module to perform meta-analysis regression. Boston College Department of Economics, Boston, MA. Accessed 12/14/2020. http://econpapers.repec.org/software/bocbocode/s446201.htm
97. Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003; 12:12-20.
98. Harris RP, Helfand M, Woolf SH, et al; Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force. *Am J Prevent Med* 2001; 20:21-35.
99. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain – United States, 2016. *JAMA* 2016; 315:1624-1645.

Appendix Table 1. *Sources of risk of bias and Cochrane Review rating system.*

Bias Domain	Source of Bias		Possible Answers
Selection	(1) Was the method of randomization adequate?	A random (unpredictable) assignment sequence. Examples of adequate methods are coin toss (for studies with 2 groups), rolling a dice (for studies with 2 or more groups), drawing of balls of different colors, drawing of ballots with the study group labels from a dark bag, computer-generated random sequence, preordered sealed envelopes, sequentially-ordered vials, telephone call to a central office, and preordered list of treatment assignments.	Yes/No/ Unsure
		Examples of inadequate methods are: alternation, birth date, social insurance/security number, date in which they are invited to participate in the study, and hospital registration number.	
Selection	(2) Was the treatment allocation concealed?	Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.	Yes/No/ Unsure
Performance	(3) Was the patient blinded to the intervention?	Index and control groups are indistinguishable for the patients or if the success of blinding was tested among the patients and it was successful.	Yes/No/ Unsure
Performance	(4) Was the care provider blinded to the intervention?	Index and control groups are indistinguishable for the care providers or if the success of blinding was tested among the care providers and it was successful.	Yes/No/ Unsure
Detection	(5) Was the outcome assessor blinded to the intervention?	Adequacy of blinding should be assessed for each primary outcome separately. This item should be scored "yes" if the success of blinding was tested among the outcome assessors and it was successful or:	Yes/No/ Unsure
		for patient-reported outcomes in which the patient is the outcome assessor (e.g., pain, disability): the blinding procedure is adequate for outcome assessors if participant blinding is scored "yes"	
		for outcome criteria assessed during scheduled visit and that supposes a contact between participants and outcome assessors (e.g., clinical examination): the blinding procedure is adequate if patients are blinded, and the treatment or adverse effects of the treatment cannot be noticed during clinical examination	
		for outcome criteria that do not suppose a contact with participants (e.g., radiography, magnetic resonance imaging): the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed when assessing the main outcome	
		for outcome criteria that are clinical or therapeutic events that will be determined by the interaction between patients and care providers (e.g., cointerventions, hospitalization length, treatment failure), in which the care provider is the outcome assessor: the blinding procedure is adequate for outcome assessors if item "4" (caregivers) is scored "yes"	
		for outcome criteria that are assessed from data of the medical forms: the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed on the extracted data	
Attrition	(6) Was the drop-out rate described and acceptable?	The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias a "yes" is scored (N.B. these percentages are arbitrary, not supported by literature).	Yes/No/ Unsure
Attrition	(7) Were all randomized participants analyzed in the group to which they were allocated?	All randomized patients are reported/analyzed in the group they were allocated to by randomization for the most important moments of effect measurement (minus missing values) irrespective of noncompliance and cointerventions.	Yes/No/ Unsure

Appendix Table 1 con't. *Sources of risk of bias and Cochrane Review rating system.*

Bias Domain	Source of Bias		Possible Answers
Reporting	(8) Are reports of the study free of suggestion of selective outcome reporting?	All the results from all prespecified outcomes have been adequately reported in the published report of the trial. This information is either obtained by comparing the protocol and the report, or in the absence of the protocol, assessing that the published report includes enough information to make this judgment.	Yes/No/ Unsure
Selection	(9) Were the groups similar at baseline regarding the most important prognostic indicators?	Groups have to be similar at baseline regarding demographic factors, duration and severity of complaints, percentage of patients with neurological symptoms, and value of main outcome measure(s).	Yes/No/ Unsure
Performance	(10) Were cointerventions avoided or similar?	If there were no cointerventions or they were similar between the index and control groups.	Yes/No/ Unsure
Performance	(11) Was the compliance acceptable in all groups?	The reviewer determines if the compliance with the interventions is acceptable, based on the reported intensity, duration, number and frequency of sessions for both the index intervention and control intervention(s). For example, physiotherapy treatment is usually administered for several sessions; therefore it is necessary to assess how many sessions each patient attended. For single-session interventions (e.g., surgery), this item is irrelevant.	Yes/No/ Unsure
Detection	(12) Was the timing of the outcome assessment similar in all groups?	Timing of outcome assessment should be identical for all intervention groups and for all primary outcome measures.	Yes/No/ Unsure
Other	(13) Are other sources of potential bias unlikely?	Other types of biases. For example: When the outcome measures were not valid. There should be evidence from a previous or present scientific study that the primary outcome can be considered valid in the context of the present. Industry-sponsored trials. The conflict of interest (COI) statement should explicitly state that the researchers have had full possession of the trial process from planning to reporting without funders with potential COI having any possibility to interfere in the process. If, for example, the statistical analyses have been done by a funder with a potential COI, usually "unsure" is scored.	Yes/No/ Unsure

Source: Furlan AD, et al; Editorial Board of the Cochrane Back, Neck Group. 2015 updated method guideline for systematic reviews in the Cochrane back and neck group. *Spine (Phila PA)* 1976) 2015; 40:1660-1673 (53).

Appendix Table 2. *Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM – QRB.*

Appendix Table 2. *Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM – QRB.*

		Scoring
I.	**TRIAL DESIGN AND GUIDANCE REPORTING**	
1.	CONSORT or SPIRIT	
	Trial designed and reported without any guidance	0
	Trial designed and reported utilizing minimum criteria other than CONSORT or SPIRIT criteria or trial was conducted prior to 2005	1
	Trial implies it was based on CONSORT or SPIRIT without clear description with moderately significant criteria for randomized trials or the trial was conducted before 2005	2
	Explicit use of CONSORT or SPIRIT with identification of criteria or trial conducted with high level reporting and criteria or conducted before 2005	3
II.	**DESIGN FACTORS**	
2.	Type and Design of Trial	
	Poorly designed control group (quasi selection, convenient sampling)	0
	Proper active-control or sham procedure with injection of active agent	2
	Proper placebo control (no active solutions into active structures)	3
3.	Setting/Physician	
	General setting with no specialty affiliation and general physician	0
	Specialty of anesthesia/PMR/neurology/radiology/ortho, etc.	1
	Interventional pain management with interventional pain management physician	2
4.	Imaging	
	Blind procedures	0
	Ultrasound	1
	CT	2
	Fluoro	3
5.	Sample Size	
	Less than 50 participants in the study without appropriate sample size determination	0
	Sample size calculation with less than 25 patients in each group	1
	Appropriate sample size calculation with at least 25 patients in each group	2
	Appropriate sample size calculation with 50 patients in each group	3
6.	Statistical Methodology	
	None or inappropriate	0
	Appropriate	1
III.	**PATIENT FACTORS**	
7.	Inclusiveness of Population	
7a.	For epidural procedures:	
	Poorly identified mixed population	0
	Clearly identified mixed population	1
	Disorders specific trials (i.e. well defined spinal stenosis and disc herniation, disorder specific, disc herniation or spinal stenosis or post surgery syndrome)	2
7b.	For facet or sacroiliac joint interventions:	
	No diagnostic blocks	0
	Selection with single diagnostic blocks	1
	Selection with placebo or dual diagnostic blocks	2
8.	Duration of Pain	
	Less than 3 months	0

Appendix Table 2 con't. *Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM – QRB.*

		Scoring
	3 to 6 months	1
	> 6 months	2
9.	Previous Treatments	
	Conservative management including drug therapy, exercise therapy, physical therapy, etc.	
	Were not utilized	0
	Were utilized sporadically in some patients	1
	Were utilized in all patients	2
10.	Duration of Follow-up with Appropriate Interventions	
	Less than 3 months or 12 weeks for epidural or facet joint procedures, etc. and 6 months for intradiscal procedures and implantables	0
	3 to 6 months for epidural or facet joint procedures, etc., or 1 year for intradiscal procedures or implantables	1
	6 months to 17 months for epidurals or facet joint procedures, etc., and 2 years or longer for discal procedures and implantables	2
	18 months or longer for epidurals and facet joint procedures, etc., or 5 years or longer for discal procedures and implantables	3
IV.	**OUTCOMES**	
11.	Outcomes Assessment Criteria for Significant Improvement	
	No descriptions of outcomes OR < 20% change in pain rating or functional status	0
	Pain rating with a decrease of 2 or more points or more than 20% reduction OR functional status improvement of more than 20%	1
	Pain rating with decrease of ≥ 2 points AND ≥ 20% change or functional status improvement of ≥ 20%	2
	Pain rating with a decrease of 3 or more points or more than 50% reduction OR functional status improvement with a 50% or 40% reduction in disability score	2
	Significant improvement with pain and function ≥ 50% or 3 points and 40% reduction in disability scores	4
12.	Analysis of all Randomized Participants in the Groups	
	Not performed	0
	Performed without intent-to-treat analysis without inclusion of all randomized participants	1
	All participants included with or without intent-to-treat analysis	2
13.	Description of Drop Out Rate	
	No description of dropouts, despite reporting of incomplete data or ≥ 20% withdrawal	0
	Less than 20% withdrawal in one year in any group	1
	Less than 30% withdrawal at 2 years in any group	2
14.	Similarity of Groups at Baseline for Important Prognostic Indicators	
	Groups dissimilar with significant influence on outcomes with or without appropriate randomization and allocation	0
	Groups dissimilar without influence on outcomes despite appropriate randomization and allocation	1
	Groups similar with appropriate randomization and allocation	2
15.	Role of Co-Interventions	
	Co-interventions were provided but were not similar in the majority of participants	0
	No co-interventions or similar co-interventions were provided in the majority of the participants	1
V.	**RANDOMIZATION**	

Appendix Table 2 con't. *Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM – QRB.*

		Scoring
16.	Method of Randomization	
	Quasi randomized or poorly randomized or not described	0
	Adequate randomization (coin toss, drawing of balls of different colors, drawing of ballots)	1
	High quality randomization (Computer generated random sequence, pre-ordered sealed envelopes, sequentially ordered vials, telephone call, pre-ordered list of treatment assignments, etc.)	2
VI.	**ALLOCATION CONCEALMENT**	
17.	Concealed Treatment Allocation	
	Poor concealment of allocation (open enrollment) or inadequate description of concealment	0
	Concealment of allocation with borderline or good description of the process with probability of failure of concealment	1
	High quality concealment with strict controls (independent assignment without influence on the assignment sequence)	2
VII.	**BLINDING**	
18.	Patient Blinding	
	Patients not blinded	0
	Patients blinded adequately	1
19.	Care Provider Blinding	
	Care provider not blinded	0
	Care provider blinded adequately	1
20.	Outcome Assessor Blinding	
	Outcome assessor not blinded or was able to identify the groups	0
	Performed by a blinded independent assessor with inability to identify the assignment-based provider intervention (i.e., subcutaneous injection, intramuscular distant injection, difference in preparation or equipment use, numbness and weakness, etc.)	1
VIII.	**CONFLICTS OF INTEREST**	
21.	Funding and Sponsorship	
	Trial included industry employees	-3
	Industry employees involved; high levels of funding with remunerations by industry or an organization funded with conflicts	-3
	Industry or organizational funding with reimbursement of expenses with some involvement	0
	Industry or organization funding of expenses without involvement	1
	Funding by internal resources only with supporting entity unrelated to industry	2
	Governmental funding without conflict such as NIH, NHS, AHRQ	3
22.	Conflicts of Interest	
	None disclosed with potential implied conflict	0
	Marginally disclosed with potential conflict	1
	Well disclosed with minor conflicts	2
	Well disclosed with no conflicts	3
	Hidden conflicts with poor disclosure	-1
	Misleading disclosure with conflicts	-2
	Major impact related to conflicts	-3
	TOTAL	48

Source: Manchikanti L, et al. Assessment of methodologic quality of randomized trials of interventional techniques: Development of an interventional pain management specific instrument. Pain Physician 2014; 17:E263-E290 (50).

Appendix Table 3. *IPM checklist for assessment of nonrandomized or observational studies of IPM techniques utilizing IPM-QRBNR.*

		Scoring
I.	**STUDY DESIGN AND GUIDANCE REPORTING**	
1.	STROBE or TREND Guidance	
	Case Report/Case Series	0
	Study designed without any guidance	1
	Study designed with minimal criteria and reporting with or without guidance	2
	Study designed with moderately significant criteria or implies it was based on STROBE or TREND without clear description or the study was conducted before 2011 or similar criteria utilized with study conducted before 2011	3
	Designed with high level criteria or explicitly uses STROBE or TREND with identification of criteria or conducted prior to 2011	4
II.	**DESIGN FACTORS**	
2.	Study Design and Type	
	Case report or series (uncontrolled – longitudinal)	0
	Retrospective cohort or cross-sectional study	1
	Prospective cohort case-control study	2
	Prospective case control study	3
	Prospective, controlled, nonrandomized	4
3.	Setting/Physician	
	General setting with no specialty affiliation and general physician	0
	Specialty of anesthesia/PMR/neurology, etc.	1
	Interventional pain management with interventional pain management physician	2
4.	Imaging	
	Blind procedures	0
	Ultrasound	1
	CT	2
	Fluoro	3
5.	Sample Size	
	Less than 100 participants without appropriate sample size determination	0
	At least 100 participants in the study without appropriate sample size determination	1
	Sample size calculation with less than 50 patients in each group	2
	Appropriate sample size calculation with at least 50 patients in each group	3
	Appropriate sample size calculation with 100 patients in each group	4
6.	Statistical Methodology	
	None	0
	Some statistics	1
	Appropriate	2
III.	**PATIENT FACTORS**	
7.	Inclusiveness of Population	
7a.	For epidural procedures:	
	Poorly identified mixed population	1
	Poorly identified mixed population with large sample (≥ 200)	2
	Clearly identified mixed population	3
	Disorders specific trials (i.e. well defined spinal stenosis and disc herniation, disorder specific, disc herniation or spinal stenosis or post surgery syndrome)	4
7b.	For facet or sacroiliac joint interventions:	

Appendix Table 3 con't. *IPM checklist for assessment of nonrandomized or observational studies of IPM techniques utilizing IPM-QRBNR.*

		Scoring
	No specific selection criteria	1
	No diagnostic blocks based on clinical symptomatology	2
	Selection with single diagnostic blocks	3
	Selection with placebo or dual diagnostic blocks	4
8.	Duration of Pain	
	Less than 3 months	0
	3 to 6 months	1
	> 6 months	2
9.	Previous Treatments	
	Conservative management including drug therapy, exercise therapy, physical therapy, etc.	
	Were not utilized	0
	Were utilized sporadically in some patients	1
	Were utilized in all patients	2
10.	Duration of Follow-up with Appropriate Interventions	
	Less than 3 months or less for epidural or facet joint procedures, etc., and 6 months for intradiscal procedures and implantables	1
	3-6 months for epidural or facet joint procedures, etc., or one year for intradiscal procedures or implantables	2
	6-12 months for epidurals or facet joint procedures, etc., and 2 years or longer for discal procedures and implantables	3
	18 months or longer for epidurals and facet joint procedures, etc., or 5 years or longer for discal procedures and implantables	4
IV.	**OUTCOMES**	
11.	Outcomes Assessment Criteria for Significant Improvement	
	No descriptions of outcomes OR < 20% change in pain rating or functional status	0
	Pain rating with a decrease of 2 or more points or more than 20% reduction OR functional status improvement of more than 20%	1
	Pain rating with decrease of ≥2 points AND ≥ 20% change or functional status improvement of ≥ 20%	2
	Pain rating with a decrease of 3 or more points or more than 50% reduction OR functional status improvement with a 50% or 40% reduction in disability score	2
	Significant improvement with pain and function ≥ 50% or 3 points and 40% reduction in disability scores	4
12.	Description of Drop Out Rate	
	No description despite reporting of incomplete data or more than 30% withdrawal	0
	Less than 30% withdrawal in one year in any group	1
	Less than 40% withdrawal at 2 years in any group	2
13.	Similarity of Groups at Baseline for Important Prognostic Indicators	
	No groups or groups dissimilar with significant influence on outcomes	0
	Groups dissimilar without significant influence on outcomes	1
	Groups similar	2
14.	Role of Co-Interventions	
	Dissimilar co-interventions or similar co-interventions in some of the participants	1
	No co-interventions or similar co-interventions in majority of the participants	2

Appendix Table 3 con't. *IPM checklist for assessment of nonrandomized or observational studies of IPM techniques utilizing IPM-QRBNR.*

		Scoring
V.	**ASSIGNMENT**	
15.	Method of Assignment of Participants	
	Case report/case series or selective assignment based on outcomes or retrospective evaluation based on clinical criteria	1
	Prospective study with inclusion without specific criteria	2
	Retrospective method with inclusion of all participants or random selection of retrospective data	3
	Prospective, well-defined assignment of methodology and inclusion criteria (quasi randomization, matching, stratification, etc.)	4
VI.	**CONFLICTS OF INTEREST**	
16.	Funding and Sponsorship	
	Trial included industry employees with or without proper disclosure	-3
	Industry employees involved; high levels of funding with remunerations by industry or an organization funded with conflicts	-3
	Industry or organizational funding with reimbursement of expenses with some involvement or no information available	0
	Industry or organization funding of expenses without involvement	1
	Funding by internal resources only	2
	Governmental funding without conflict such as NIH, NHS, AHRQ	3
	TOTAL MAXIMUM	48

Source: Manchikanti L, et al. Development of an interventional pain management specific instrument for methodologic quality assessment of nonrandomized studies of interventional techniques. Pain Physician 2014; 17:E291-E317 (51).