Pain Physician 2009; 12:517-540 • ISSN 1533-3159

Evidence-Based Medicine

Evidence-Based Medicine, Systematic Reviews, and Guidelines in Interventional Pain Management: Part 5. Diagnostic Accuracy Studies

Laxmaiah Manchikanti, MD[1], Richard Derby, MD[2], Lee Wolfer, MD[2], Vijay Singh, MD[3], Sukdeb Datta, MD[4], and Joshua A. Hirsch, MD[5]

From: [1]Pain Management Center of Paducah, Paducah, KY; [2]Spinal Diagnostics & Treatment Center, Daly City, CA; [3]Pain Diagnostics Associates, Niagara, WI; [4]Vanderbilt University Medical Center, Nashville, TN; and [5]Massachusetts General Hospital and Harvard Medical School, Boston, MA.

Dr. Manchikanti is Medical Director of the Pain Management Center of Paducah, Paducah, KY. Dr. Derby is Medical Director of Spinal Diagnostics & Treatment Center, Daly City, CA, and Associate Professor, Department of Physical Medicine and Rehabilitation, Stanford University, Stanford, CA. Dr. Wolfer is with the Spinal Diagnostics & Treatment Center, Daly City, CA. Dr. Singh is Medical Director of Pain Diagnostics Associates, Niagara, WI. Dr. Datta is Director, Vanderbilt University Interventional Pain Program, Associate Professor, Dept. of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN. Dr. Hirsch is Chief of Minimally Invasive Spine Surgery, Depts. of Radiology and Neurosurgery, Massachusetts General Hospital and Assistant Professor of Radiology, Harvard Medical School, Boston, MA.

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

Disclaimer: There was no external funding in the preparation of this manuscript.
Conflict of interest: Dr. Hirsch is a consultant and minor shareholder for Cardinal Healthcare and Medtronic..

Manuscript received: 01/22/2009
Accepted for publication: 02/25/2009

Free full manuscript:
www.painphysicianjournal.com

Diagnosis is a critical component of health care. The world of diagnostic tests is highly dynamic. New tests are developed at a fast pace and technology of existing tests is continuously being improved. However, clinicians, policy makers, and patients routinely face a range of questions regarding diagnostic tests. Well designed diagnostic test accuracy studies can help in making these decisions, provided that they transparently and fully report their participants, tests, methods, and results (as facilitated). For example, by the standards for the reporting of diagnostic accuracy studies (STARD) statement.

Exaggerated and biased results from poorly designed and reported diagnostic test studies can trigger their premature dissemination and lead physicians into making incorrect treatment decisions. Thus, a diagnostic test is useful only to the extent that it distinguishes between conditions or disorders that might otherwise be confused. While almost any test can differentiate healthy persons from severely affected ones, appropriate diagnostic tests should differentiate mild and moderate forms of disease.

Shortcomings in a study design and interpretation can affect estimates of diagnostic accuracy. Thus, quality diagnostic studies are essential in medicine in general and interventional pain management in particular. The STARD initiative was developed to improve the accuracy and completeness in the reporting of studies of diagnostic accuracy and provide guidance to assist in reducing the potential for bias in the study and to evaluate a study's generalizability.

In the practice of interventional pain management, in addition to diagnostic tests which include laboratory tests, imaging tests, and physical examination, diagnostic interventional techniques are crucial. Interventional techniques as a diagnostic tool in painful conditions is important due to multiple challenging clinical situations, which include the purely subjective nature of pain and underdetermined and uncertain pathophysiology in most painful spinal conditions. Precision diagnostic blocks are used to clarify these challenging clinical situations in order to determine the pathophysiology of clinical pain, the site of nociception, and the pathway of afferent neural signals.

Part 5 of evidence-based medicine (EBM) in interventional pain management describes the various aspects of diagnostic accuracy studies.

Key words: Evidence-based medicine, diagnostic studies, systematic reviews, randomized trials, interventional pain management, standards for the reporting of diagnostic accuracy studies (STARD)

Pain Physician 2009; 12:517-540

Making a diagnosis is a complex cognitive task that involves both logical reasoning and pattern recognition. Diagnostics is the process of identifying a medical condition or disease by its signs and symptoms from the results of a clinical examination and other evaluative procedures. Apart from signs and symptoms, clinicians use a variety of tests including imaging and biochemistry to identify physiological derangements, establish prognosis, monitor illness, and diagnose (1,2). Consequently, the term diagnostic criteria designates the combination of findings which allows the clinician to ascertain the diagnosis of the respective disease (3).

Clinicians often use diagnostic criteria as a package or strategy (1). The clinician generally formulates a hypothesis of likely diagnosis and in many cases will obtain further testing to confirm or clarify the diagnosis, before suggesting definitive treatment. Consequently, a clinician often thinks of evaluating or recommending not a single test, but a diagnostic strategy.

In modern medicine, the diagnosis of illness, along with diagnostic accuracy of individual or combined diagnostic tests, serves as the basis for decisions on the treatment strategies, referrals, disability assessments, reimbursement, and more (3).

The world of diagnostic tests is highly dynamic. New tests are developed at a fast rate and technology of existing tests is continuously being improved. Exaggerated and biased results from poorly designed and reported diagnostic test studies can trigger their premature dissemination and lead physicians into making incorrect treatment decisions (4).

A diagnostic test is useful only to the extent that it distinguishes between conditions or disorders that might otherwise be confused. Almost any test can differentiate healthy persons from severely affected ones; this ability however tells us nothing about the clinical utility of a test. The true pragmatic value of a test is therefore established only in a study that closely resembles clinical practice. Usually, when clinicians think about diagnostic tests, they focus on accuracy (sensitivity and specificity), that is, how well the test classifies patients correctly as having or not having a disease. The underlying assumption is, however, that obtaining a better idea of whether a target condition is present or absent will result in improved outcome. Diagnostic accuracy can be expressed in many ways, including sensitivity and specificity, likelihood ratios, diagnostic odds ratio, and the area under a receiver-operator characteristics (ROC) curve (5-8).

There are several potential threats to the internal and external validity of a study of diagnostic accuracy. A survey of studies of diagnostic accuracy published in 4 medical journals between 1978 and 1993 revealed that the methodologic quality was mediocre at best (9). In addition, the absence of critical information about the design and conduct of diagnostic studies has been confirmed in meta-analyses (10,11). It was also shown that diagnostic studies with specific design features are associated with biased, optimistic estimates of diagnostic accuracy compared to studies without such deficiencies (12). In an evaluation of the assessment of neck pain and its associated disorders, it was shown that there was little information on the validity or utility of self-reported history in evaluating neck pain (13).

Shortcomings in study design can affect estimates of diagnostic accuracy (14-37). It has been reported that many issues in the design and conduct of diagnostic accuracy studies can lead to bias or variation (38). In a recent manuscript evaluating systematic reviews of diagnostic test accuracy, it was shown that the challenges that remain are the poor reporting of original diagnostic test accuracy studies and difficulties with the interpretation of the results of diagnostic test accuracy research (39). A literature survey of sample sizes of studies on diagnostic accuracy concluded that few studies on diagnostic accuracy report considerations of sample size. Further, the number of participants in most studies on diagnostic accuracy is probably too small to analyze variability of the measure of accuracy across patient subgroups (40).

Multiple issues of diagnostic accuracy of interventional techniques have been described (19-37). Similar to the Consolidated Standards of Reporting Trials (CONSORT) (41-43), the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) (44,45), and the Standards for the Reporting of Diagnostic Accuracy studies (STARD) initiative were established to improve accuracy and completeness in the reporting of studies of diagnostic accuracy (4). Further, the Grading Recommendations Assessment, Development and Education (GRADE) for diagnosis and strategies also has been published (1).

The objective of this manuscript is to review the importance of design, implementation, and reporting of diagnostic accuracy trials in interventional pain management.

1.0 An Introduction to Diagnostic Studies

Studies to determine the diagnostic accuracy of a test are a vital part in the evaluation process of a patient (14-39,46-48). The term *test* refers to any method for obtaining additional information on a patient's health status. It includes information from the history and physical examination, laboratory tests, imaging tests, functional assessments, psychological evaluation, and finally, histopathology. The results of diagnosis provide the basis for further clinical actions, further diagnostic testing, or the initiation, modification, or termination of treatment (4).

The condition of interest or target condition can refer to a particular disease or to any other identifiable condition that may prompt clinical actions. In this framework, the *reference standard* is considered to be the best available method for establishing the presence or absence of the condition of interest. The reference standard can be a single method, or a combination of methods to establish the presence of the target condition. It can include laboratory tests, imaging tests, and pathology, but also the dedicated clinical follow-up of subjects (4).

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. Consequently, diagnostic accuracy may be expressed in multiple ways, not limited to sensitivity and specificity, but also as likelihood ratios, odds ratio, and finally, the area under an ROC curve (6-8).

Studies of diagnostic tests have commonly addressed one of the 2 main objectives (47). The first, and traditionally the most common, aim of diagnostic test evaluation is to establish the diagnostic accuracy of the test. This is usually done in observational studies reporting test parameters such as sensitivity, specificity, likelihood ratios, and the predictive value of the test. Measures of sensitivity and specificity can be used together with estimates of pre-test probability of disease to produce estimates of the post-test probability of disease. The second objective of a diagnostic accuracy test is to evaluate the impact of one or more diagnostic strategies on therapy decisions and/or patient outcomes. This trend is in response to the recognition that increasing diagnostic accuracy is of little use if there is no resulting change or improvement in patient care. Such studies tend to be either randomized controlled trials (RCTs) or non-experimental comparative studies, both of which are valid under given circumstances (47). These study designs may also be used in combination, for example, in the comparison of 2 tests, all patients could receive both tests in randomized sequence, with the second test performed without knowledge of the results of the first. However, various other methods have been developed in order to evaluate diagnostic tests in relation to their clinical or therapeutic impact (49-52). Further, the decision on analytic methods has also been proposed as a means of establishing the impact of a diagnostic technology on patient outcomes and costs (53,54).

2.0 The Differences Between Studies Of Diagnosis And Therapy

Diagnostic tests are different from therapeutic trials or evaluations. In studies of diagnostic accuracy, results from one or more tests are compared with the results obtained with the reference standard on the same subjects. Several factors threaten the internal and external validity of a study of diagnostic accuracy. These factors are all different from the quality of randomized or observational studies. Some of these factors have to do with the design of such studies, others with the selection of patients, the execution of the tests, or the analysis of the data. Consequently, systems to grade the strength of scientific evidence (55) assessed a multitude of articles and developed separate systems for randomized trials, observational studies, and diagnostic studies. Table 1 illustrates the differences between important quality features of RCTs and diagnostic studies (56-59). Thus, while randomization is the major criteria for a therapeutic trial, appropriate criterion or a gold standard is the major and most important criteria for a diagnostic study.

Others (60-62) also have observed that observational studies are generally the most appropriate for answering questions related to diagnostic accuracy, incidence, prevalence, and etiology.

3.0 Evolution Of Diagnostic Studies

Sackett and Haynes (63) proposed a system to classify various developmental stages of a diagnostic test from Phase I to IV studies. Table 2 illustrates the characteristics of a diagnostic study. Thus, clinicians should understand where the tests are along this evolutionary scientific continuum to make the best judgements about whether they should adapt these tests in practice (3).

Table 1. *Important quality features of selected study designs.*

Study design	Questions for ascertaining quality (validity)
Therapy (e.g., randomized controlled trials)	1. Were patients randomized? 2. Was concealment of allocation adequate? 3. Were patients analyzed in the groups to which they were randomized? 4. Were patients aware of group allocation? 5. Were clinicians aware of group allocation? 6. Were outcome assessors aware of group allocation? 7. Was follow-up complete?
Diagnosis (e.g., cross-sectional diagnostic study)	1. Was there a comparison with an independent, appropriate gold standard? 2. Did the included patients cover a wide patient spectrum likely to be encountered in a usual clinical practice setting? 3. Was the index test result interpreted without the knowledge of gold standard, and vice versa? 4. Did the study prospectively recruit consecutive patients suspected to have the disease of interest?
Harm (e.g., cohort or case-control study)	1. Did the investigators demonstrate similarity in all known determinants of outcome (e.g., confounders)? Did they adjust for differences in the analysis? 2. Were exposed patients equally likely to be identified in the two groups? 3. Were the outcomes measured in the same way in the groups being compared? 4. Was follow-up sufficiently complete?
Prognosis (e.g., cohort study)	1. Was the sample of patients representative? 2. Were the patients sufficiently homogenous with respect to prognostic risk? 3. Was follow-up sufficiently complete? 4. Were objective and unbiased outcome criteria used?

Adapted from Guyatt GH, Rennie D. *Users' Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice.* The Evidence-Based Medicine Working Group. AMA Press, Chicago, 2002 (57).

Table 2. *Research characteristics of a diagnostic study.*

Disease Present/Absent	Disease +/-	Gold Standard Yes/No	Blinding Yes/No	Utility Yes/No
Phase I	+/-	No	No	No
Phase II	+	No	No	No
Phase III	+/-	Yes	Yes	No
Phase IV	+	No	Yes	Yes

Adapted from Sackett and Haynes. The architecture of diagnostic research. *BMJ* 2002; 324:539-541 (63).

3.1 Phase 1 Studies of Diagnostic Tests

The singular purpose of Phase I studies is to answer the question: Do test results in affected patients differ from those in normal individuals? Such studies are typically conducted among patients known to have the disease and a group of individuals definitely known not to have the disease. If a test is found to be very rarely positive in healthy, normal controls with no suggestion of the disease, this is a good sign and the first step for future investigations in more clinically relevant settings in Phases II to IV, as shown in Table 2. The Phase I study is the most basic assessment, and while an encouraging Phase I study, does not confirm the diagnostic validity a negative Phase I study will likely not have further diagnostic value (3).

3.2 Phase II Studies of Diagnostic Tests

A Phase II study is also designed to answer a singular question: Are patients with certain test results more likely to have the target disorder? Consequently, Phase II studies compare the range of test results of groups of patients who already have the established diagnosis. The fundamental question here is whether certain values of test results are able to predict the

presence of the disease more than are other values. This testing strategy only includes patients for whom the clinician already has diagnostic certainty and the clinician is performing the test to categorize the range of results seen in this condition. Thus, Phase II diagnostic studies do not confirm the validity and require evaluation in Phase III and IV designs before they can be recommended for widespread clinical adaptation (3).

3.3 Phase III Studies of Diagnostic Tests

Phase III studies provide multiple answers with a singular question: Do the test results distinguish patients with and without the target disorder among those in whom it is clinically sensible to suspect the disorder? Phase III studies require showing the presence or absence of the disease, comparison with a gold standard, and blinded assessment of the test. Phase III tests are essential as a diagnostic test and may perform well in completely normal subjects with 100% negative results, but may be positive in an unacceptable proportion of subjects without the disease who have similar symptoms. The first key feature of the Phase III diagnostic study is that the test must be conducted in a clinical study population in which the disease status is uncertain. The second key feature refers to blinding, that is the results of the test must be independently interpreted from a recognized gold standard (3).

3.4 Phase IV Studies of Diagnostic Tests

Phase IV studies are also designed to provide multiple answers with a singular question: Do patients undergoing a specific diagnostic test fair better in their health outcomes than similar patients who have not been exposed to the test? Consequently, a Phase IV study tests the clinical utility of the test, thus, a test may be valid but has no impact on outcomes if there is no effective treatment available or may even adversely affect the patient who has the test done, particularly if risky tests are performed and the treatments are highly ineffective. This study design is a continuation of Phase III patients who have used the experimental test in their evaluation and those who have not. A diagnostic test with high utility will demonstrate better health outcomes when the test is used compared to when it is not used. Consequently, this requires a randomized design with all the requirements of a standardized protocol with blinded interpretation (3).

4.0 Why Quality Diagnostic Studies?

A rigorous evaluation process of diagnostic tests before introduction into clinical practice could not only reduce the number of unwanted clinical consequences related to misleading estimates of test accuracy, but also limit health care costs by preventing unnecessary testing. Consequently, testing of the diagnostic accuracy and the production of quality diagnostic tests are vital parts in the modern health care environment. There are numerous threats to the internal and external validity of a study of diagnostic accuracy. Numerous evaluations (7-10,17,18,38-40,46,47) have shown overall poor methodologic quality with lack of information on key elements of design, conduct, and analysis. In a study evaluating the association between compliance with methodological standards of diagnostic research and reported test accuracy, a low sensitivity was demonstrated with a lack of adherence to standards such as STARD and instruments such as the Quality Assessment tool for Diagnostic Accuracy Studies (QUADAS) (46,64). Since the development of STARD statements for the reporting of studies of diagnostic accuracy (4,65), a study of the reproducibility of the STARD checklist (66) concluded that overall reproducibility of the quality of reporting on diagnostic accuracy studies using the STARD statement was good. However, they also found substantial disagreements for specific items, which were caused by difficulties in assessing the reporting of these items due to a lack of clarity within the articles. However, the same authors also reported less than optimal results prior to 2000, with only 41% of the articles reporting 50% of the STARD items (67).

An explanatory document of the STARD statement for reporting studies of diagnostic studies (65) facilitates the use, understanding, and dissemination of checklists. The document contains a clarification of the meaning, rationale, and optimal use of each item on the checklist, as well as a short summary of the available evidence on bias and applicability.

The quality of any study can be considered in terms of internal validity, external validity, and the quality of data analysis and reporting. Internal validity can be defined as the degree to which estimates of diagnostic accuracy produced in a study have not been biased as a result of study design, conduct, analysis, or presentation. This includes various aspects of the accuracy study including sample selection, problems with the reference standard, and non-independent assessment.

External validity concerns the degree to which the results of a study can be applied to patients in practice, and is affected by factors such as the spectrum of disease or non-disease, setting, other patient characteristics, how the diagnostic test was conducted, the threshold or cutoff point used, and the reproducibility of the test.

Several methodological reviews have evaluated the quality of reporting of diagnostic studies and have developed multiple standards.

Poor methodologic quality has been empirically proven to affect the results of controlled trials and meta-analyses of intervention or treatment studies (68,69), which also has been shown for meta-analyses of diagnostic studies (12). Consequently, it has been a well-known fact that the study quality should be assessed in any attempt to use results of published studies of diagnostic evaluation (7,70-73). For example, statistical methods have been developed to account for some degree, verification bias (74) and methods to evaluate tests for which there is no or only an imperfect reference standard available (75,76).

5.0 Accuracy of Diagnostic Studies

Three primary features of a diagnostic test are reliability or reproducibility, validity or accuracy, and predictive value in different populations.

5.1 Reliability

For a test to be valid it must first be shown to be reliable, that is, a test should consistently give the same result when it is repeated on the same person under the same conditions in a set time frame (3). Differences in the results upon repetition of a test, even under the same conditions, can arise for several reasons, the commonest being normal biologic variations in the test subject, individual observer inconsistencies (intraobserver variability), difference across observers (interobserver variability), as well as level of experience in applying the test, and differences in the underlying technology of the test equipment.

5.2 Validity

The validity or accuracy of a diagnostic test is typically demonstrated by comparing it to a gold or criterion standard. A criterion standard is a well-accepted and commonly applied method of identifying the disease or clinical entity of interest. Sensitivity of a test is the proportion of people with the disease who will have a positive result, whereas specificity is the proportion of people without the disease who will have a negative test result (77). In simple terms, the validity of a diagnostic test refers to its ability to correctly identify people with a condition (positive for condition or at risk for that condition) or absence of the condition (negative for the condition or not a risk for the condition).

5.2.1 Concept Validity

Concept validity is that the procedure appears in theory to have a reasonable anatomical or physiological basis. Diagnostic blocks have concept validity on the grounds that it sounds reasonable and that if a structure is a source of pain, anesthetizing it will relieve that pain (78). Thus, the thrust of concept validity is the theoretical basis of the test.

5.2.2 Content Validity

Content validity essentially defines the test accurately and ensures that the procedure is performed consistently in the same manner (78). Thus, content validity does not render the procedure itself valid, but it ensures that the name of the procedure is used consistently to mean the same thing.

5.2.3 Face Validity

For a diagnostic block to have face validity it must be shown that the diagnostic test actually does what it is supposed to do in an anatomical or physiological sense (78). If a particular structure is said to be the target, it must be shown that the structure is anesthetized. Face validity can be tested and established either by a study which has replicated the results or testing for face validity in each and every case.

5.2.4 Construct Validity

Construct validity is considered as the most critical of all the subtypes of validity. It establishes if the test actually achieves what it is supposed to achieve by measuring the extent to which a test correctly distinguishes the presence, but also the absence, of the condition that the test is supposed to detect. Construct validity measures if the test actually works or not, and how well it works (78).

5.3 Criterion or Gold Standard

Evaluation of the diagnostic accuracy tests is an important, dynamic, and emerging part of medicine. Testing a test involves comparing, in the same sample of patients, the results of a test with unknown validity

with the results of some other test whose validity is beyond question or a criterion standard formerly also known as the gold standard.

The criterion standard, also known as the reference standard, is the test used to measure the presence or absence of the target condition. To assess the diagnostic accuracy of the test, its results are compared with the results of the reference criterion standard. The criterion standard is therefore an important determinant of the diagnostic accuracy of a test. It may be obtained in many ways, including laboratory tests, imaging tests, function tests, and pathology, but also clinical follow-up by participants, contrary to the widely held belief that there has to be a biopsy of the tissue.

The decision of which reference standard to use depends on the definition of the target condition and the purpose of the study. If no single reference standard is available, the most likely state of the patients can be derived from careful clinical follow-up or a consensus between observers (79) or modeled from results of 2 or more index tests (an index test is the test under evaluation) (80-82). It is not well appreciated that the criterion or reference standard is a proxy for the target condition and therefore often not perfect (47). Reference standard error bias occurs when errors of imperfect reference standard(s) bias the measurement of diagnostic accuracy of the test under evaluation (80,83). Further estimates of test performance are based on the assumption that the test is being compared to a reference standard that is 100% sensitive and specific. If this is not the case, as most often it is not, it may be that the index test classifies results correctly that have been incorrectly classified by the reference standard. Consequently, this would provide an underestimation of the performance of the index test (test and evaluation).

6.0 Quality Assessment of Diagnostic Studies

Assessment of methodologic quality is crucial in all types of studies (84-88). There are several instruments for methodologic quality assessment of diagnostic studies. West et al (55) in the Agency for Healthcare Research and Quality (AHRQ) evidence report of technology assessment titled "Systems to Rate the Strength of Scientific Evidence," provided pertinent evidence for rating the quality of individual articles including studies of diagnostic tests. They identified 15 non-Evidence-Based Practice Centers (EPCs) systems for assessing the quality of diagnostic studies, with 6 of them being checklists. Five domains are key for making judgements about the quality of diagnostic test reports: study population, adequate description of the test, appropriate reference standard, blinded comparison of test and reference, and avoidance of verification bias. The AHRQ panel identified 15 non-EPC systems with inclusion of 6 checklists and 3 EPC systems. In summary, they developed 5 key domains for making judgements about the quality of diagnostic test reports: study population, adequate description of the test, appropriate reference standard, blinded comparison of test and reference, and avoidance of verification bias. As illustrated in Table 3 this scoring

Table 3. *Modified AHRQ methodologic assessment criteria for diagnostic interventions.*

Criterion	Weighted Score (points)
1. Study Population	15
Subjects similar to populations in which the test would be used and with a simlar spectrum of disease	
2. Adequate Description of Test	10
Details of test and its administration sufficient to allow for replication of study	
3. Appropriate Reference Standard	30
Appropriate reference standard (gold standard) used for comparison	
Reference standard reproducible	
4. Blinded Comparison of Test	30
Evaluation of test without knowledge of disease status, if possible	15
Independent, blind interpretation of test and reference	15
5. Avoidance of Verification Bias	15
Decision to perform reference standard not dependent on results of test under study	
Total Score	100

Adapted and modified from West S et al. S*ystems to Rate the Strength of Scientific Evidence, Evidence Report*, Technology Assessment No. 47. AHRQ Publication No. 02-E016 (55).

Table 4. *The QUADAS tool*.

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	()	()	()
2. Were selection criteria clearly described?	()	()	()
3. Is the reference standard likely to correctly classify the target condition?	()	()	()
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	()	()	()
5. Did the whole sample or a random selection of a sample, receive verification usig a reference standard of diagnosis?	()	()	()
6. Did patients receive the same reference standard regardless of the index test result?	()	()	()
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	()	()	()
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	()	()	()
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	()	()	()
10. Were the index test results interpreted without knowledge of the results of the reference standard?	()	()	()
11. Were the reference standard results interpreted without knowledge of the results of the index test?	()	()	()
12. Were the same clincial data available when the test results were interpreted as would be available when the tst is used in practice?	()	()	()
13. Were uninterpretable/intermediate test results reported?	()	()	()
14. Were withdrawals from the study explained?	()	()	()

Adapted from Whiting P et al. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:25 (59).

has been applied in multiple systematic reviews with weighted scoring (21-37,89).

A tool for the QUADAS (59) was developed by combining empirical evidence and expert opinion in a formal consensus method. The QUADAS tool is presented together with guidelines for scoring each of the items included in the tool (Table 4). The QUADAS tool consists of 14 items with yes, no, or unclear answers. No weighted scoring system has been developed thus far.

6.1 Strength of Evidence

A hierarchy of strength of evidence for treatment decisions provided by Guyatt and Drummond (90) does not include diagnostic accuracy studies. However, diagnostic accuracy studies also fall into the purview of evidence-based medicine (EBM) and best evidence synthesis is essential. Guyatt and Drummond (90) in their introduction to the philosophy of EBM begin with the assertion that EBM is a shift in medical paradigms and about solving clinical problems (91,92). Further, EBM acknowledges that intuition, unsystematic clinical experience, and pathophysiologic rationale are insufficient grounds for clinical decision-making, and stresses the examination of evidence from clinical research. Above all, EBM suggests that a formal set of rules must complement medical training and common sense for clinicians to interpret the results of clinical research effectively. Consequently, knowing the results of evidence-based practice is necessary but not sufficient for delivering the highest quality of patient care. In addition, there is no philosophy of hierarchy of evidence in guiding diagnostic accuracy tests. Even then, the philosophy of hierarchy of evidence in guiding therapy, though not absolute, has created an emphasis on the importance of randomized trials, which has been applied to diagnostic accuracy studies rather inaccurately. This has led to the critics of the EBM characterizing EBM as a stick by which policy makers and academicians beat clinicians (93-97). Further, it has been stated that, "the hierarchy of evidence (of which RCT has the highest priority) has done nothing more than glorify the results of imperfect experimental designs on unrepresentative populations in controlled research environments above all other sources of evidence which may be equally valid or far more applicable in given clinical circumstances" (98). There have been multiple publications in recent years on EBM, including diagnostic accuracy studies, some of which have accurately portrayed EBM while others have not (4,20-37,86-88,99-110).

6.2 Bias and Variation in Studies of Diagnostic Accuracy

In a classic diagnostic accuracy study, a consecutive series of patients who are suspected of having the target condition, undergo the index test. All patients are then verified by the same reference standard. The index test and reference standard are then read by persons blinded to the results of each, and various measures of agreement are calculated, which include sensitivity, specificity, likelihood ratios, and diagnostic odds ratios. However, the classic design has many variations, including differences in the way patients are selected for the study, in test protocol, in the verification of patients, and in the way the index test and reference standard are read. Some of the differences may bias the results of a study, whereas others may limit the applicability of the results (38).

Bias is said to be present in a study if distortion is introduced as a consequence of defects in the design or conduct of a study. Therefore, a biased diagnostic accuracy study will produce estimates of test performance that differ from the true performance of the test.

Variations arise from the differences among studies, for example, in terms of population, setting, test protocol, or definition of the target disorder (65). However, the variability does not lead to biased estimates of the test performance; rather, it limits the applicability of results. Consequently, it is an important consideration when evaluating studies of diagnostic accuracy.

While bias and variation are different, the distinctions are not. For example, diagnostic accuracy is higher both with sensitivity and specificity when patients with a target condition and healthy volunteers are compared. This is considered as spectrum bias. However, strictly speaking, one could argue that it is a form of variability; sensitivity and specificity have been measured correctly within the study, and thus there is no bias. However, the results cannot be applied to the clinical setting as the results lack generalizability (111). On the other hand, others have argued that when the goal of a study is to measure the accuracy of a test in the clinical setting, an error in the method of patient selection is made that will lead to biased estimates of test performance (38).

In a study of evidence of bias and variation in diagnostic accuracy studies (17), the authors found that the largest overestimation of accuracy was found in studies that included severe cases and healthy controls. The design features associated with significant overestimations of diagnostic accuracy were non-consecutive inclusion of patients and retrospective data collection. In addition, random inclusion of eligible patients and differential verification also resulted in higher estimates of diagnostic accuracy. The selection of patients on the basis of whether they had been referred for the index test, rather than on clinical symptoms, was significantly associated with lower estimates of accuracy.

7.0 Diagnostic Tests in Interventional Pain Management

In the practice of interventional pain management there are 3 types of diagnostic tests, laboratory tests, imaging tests, and interventional diagnostic tests. There are multiple ways of looking at the usefulness of diagnostic tests. The hierarchical evaluation uses 6 possible endpoints to determine the utility of a test;. the more criteria in the scheme that are fulfilled, the more useful the test. Tests that fulfill fewer criteria have only limited usefulness (112). These criteria include 1) technical aspects which include reliability, accuracy, and feasibility; 2) diagnostic accuracy with validity; 3) diagnostic thinking whether the test is going to make a change in the diagnosis of therapy; 4) therapeutic effectiveness with either change in the management as a result of the outcome of the test or the diagnostic test may result in initiation or cessation of therapy; 5) outcomes with the ability to improve patient outcomes or at least provide diagnosis; and 6) societal outcomes which essentially translates and raises the question if the test is effective for the society as a whole.

Using interventional techniques as a diagnostic tool in painful conditions is important due to multiple challenging clinical situations, which include the purely subjective nature of pain and undetermined and uncertain pathophysiology in most painful spinal conditions. Precision diagnostic blocks are used to clarify these challenging clinical situations in order to determine the pathophysiology of clinical pain, the site of nociception, and pathway of afferent neural signs.

Precise anatomical diagnosis in low back pain has been described not only as elusive, and the diagnostic evaluation is often frustrating for both physicians and patients (3,113-127). History, physical examination, and imaging provide limited information It is stated that in spinal pain the diagnosis can be provided with certainty in only approximately 15% or so of cases

without disc herniation or radiculitis. Precision diagnostic blocks have changed this substantially. In fact, Nachemson (122) reported that in only 15% of cases could a pathoanatomical explanation be found for patients with chronic low back pain of more than 3 months and. he stated, "probably very little can be done at our present state of ignorance to treat these patients and improve their natural histories." Thus, when a source of pain is not obvious, diagnosis often depends on who makes the diagnosis and sets the reference standards by which the diagnosis is proven.

In a study of the evaluation of relative contributions of various structures in patients with chronic low back pain who failed to respond to conservative modalities of treatments, including physical therapy, chiropractic, and drug therapy, with lack of radiological evidence to indicate disc protrusion or radiculopathy, the diagnosis was established in approximately 70% to 80% of the patients by utilizing controlled, comparative double-diagnostic blocks (128). Utilizing controlled diagnostic blocks, facet joint pain has been demonstrated in 36% to 67% in the cervical spine, 34% to 48% in the thoracic spine, and 16% to 40% in the lumbar spine (21-27,128-133). Similarly, prevalence of discogenic pain has been demonstrated in 26% to 39% of patients (24,28-32,128,134-136), whereas sacroiliac joint pain has been established in 10% to 26% of patients (33-35,128,137).

Hancock et al (20) performed a systematic review of tests to identify the disc, sacroiliac joint, and facet joint as the source of low back pain. They found 353 potentially eligible articles of which only 41 articles met inclusion criteria. While they stated that overall the quality of studies was moderate with average 8.8 positive results from a possible 14, the item which scored worst was the spectrum of patients where only 7 of 41 (17%) studies scored positive. Other items which were generally poor included time between index and reference test (27% positive), availability of clinical data (29% positive), and reporting of uninterpretable results (22% positive).

Rubinstein and van Tulder (19) evaluated scientific evidence for diagnostic procedures for neck and low back pain. They commented that it was quite remarkable that while many named orthopedic tests of the neck and low back are often illustrated in orthopedic textbooks, there is little evidence to support their diagnostic accuracy, and therefore their use in clinical practice. Consistent with clinical experience, many studies have demonstrated that the physical examination serves primarily to confirm suspicions that arose during the history.

They also showed that individual red flags do not necessarily mean the presence of serious pathology. Red flags have not been evaluated comprehensively in any systematic review. However, the incidence of spinal tumors is very low. They also concluded that plain spinal radiography is not a valuable tool for non-specific neck or low back pain. They also concluded that there is strong evidence for the diagnostic accuracy of facet joint blocks in evaluating spinal pain, and moderate evidence for transforaminal epidural injections as well as sacroiliac joint injections for diagnostic purposes.

Szadek et al (138) in a systematic review of the diagnostic validity of criteria for sacroiliac joint pain focusing on the diagnostic validity of the International Association for the Study of Pain (IASP) criteria for diagnosing sacroiliac joint pain and concluded that there was no gold standard for sacroiliac joint pain diagnosis. Consequently, the diagnostic validity of tests related to the IASP criteria for sacroiliac joint pain should be regarded with care. They included 18 studies. Five studies examined the pattern of sacroiliac joint pain, whereas another 5 examined stress tests, specific for sacroiliac joint pain. None of the studies evaluated the diagnostic validity of the sacroiliac joint infiltration or the diagnostic validity of the IASP criteria set as a whole. In all studies, they concluded that the sacroiliac joint selective infiltration was used as a gold standard. However, the technique, medications, and required pain relief after infiltration varied considerably between the studies. Taking the double infiltration technique as a reference test, the pooled data of the thigh thrust test, compression test, and 3 or more positive stressing tests, showed discriminative power for diagnosing sacroiliac joint pain.

In contrast, Rupert et al (35), in a systematic appraisal of the literature, showed the indicated the level of evidence as II-2 for the diagnosis of sacroiliac joint pain utilizing controlled local anesthetic blocks (128,137,139-141). In this systematic review, they utilized 50% relief as the criterion standard rather then the 80% criterion standard of other systematic reviews of facet joint pain diagnosis (21-23). This study also showed the indicated level of evidence for accuracy of provocative maneuvers in the diagnosis of sacroiliac joint pain was limited (Level II-3).

In a systematic review of lumbar provocation discography in asymptomatic subjects with a meta-analysis of false-positive rates, Wolfer et al (32) identified

11 studies. They combined all extractable data, and arrived at a false-positive rate of 9.3%. This is in contrast to multiple studies published by Carragee questioning the validity of diagnostic discography (142-144). Similarly, Manchikanti et al (36) also evaluated, in a systematic review, the value of lumbar discography. Of the 69 studies identified, only 9 studies met inclusion criteria for evidence synthesis (128,144-151). The remaining studies did not utilize IASP criteria. Wolfer et al (32) and Manchikanti et al (36) concluded that lumbar discography had low false-positive rates with indicated evidence level of II-2 with regards to diagnostic accuracy. In contrast, two systematic reviews (24,31) of cervical and thoracic discography as a diagnostic for chronic spinal pain yielded lower evidence than for lumbar discography. Both these reviews (24,31) concluded that there was a paucity of literature, poor methodologic quality, and few studies were performed utilizing IASP criteria. In fact, of the 33 manuscripts considered for inclusion, only 3 studies met inclusion criteria for cervical discography (135,152,153). For thoracic discography, after review of multiple manuscripts, only 2 manuscripts met the inclusion criteria.

However, the systematic review of diagnostic and utility of cervical, thoracic, and lumbar facet joint nerve blocks provided a different picture (21-23). The evidence for diagnostic utility of cervical and lumbar facet joint nerve blocks is Level I or II-1, whereas for thoracic facet joint interventions, it is II-1. The evidence is high and positive despite utilization of stricter criteria by the authors of systematic reviews with 80% relief with controlled diagnostic blocks (22-23). Fortunately, multiple studies were available for the diagnosis of lumbar (128-131,154-156) and cervical (124,129,131,135,154,157-160) facet joint pain, which were considered to be of high quality. In contrast, for thoracic facet joint blocks, there were only 3 studies performed by only one group of investigators (129,131,132). Even then, all of them met inclusion criteria and provided reasonable evidence.

In a systematic review (161) of the accuracy of diagnostic tests of lumbar spinal stenosis, 24 articles were included with 15 related to imaging tests, 7 related to clinical tests, and 2 related to other diagnostic tests. The authors of the systematic review concluded that the overall quality was poor; with only 5 studies scoring positive on more than 50% of the quality items — only 20% of the included studies. They concluded that because of heterogeneity and overall poor quality, no firm conclusions about the diagnostic performance of the differences can be drawn. A previous study (162) concluded that published studies of the value of computed tomography (CT) and magnetic resonance imaging (MRI) for the diagnosis of lumbar stenosis lacked methodologic rigor and did not permit strong conclusions about the relative diagnostic accuracies of these procedures.

A systematic review of the diagnostic accuracy of the straight leg raising test in herniated disc (163) evaluated multiple studies. They identified 17 diagnostic publications. They showed verification bias in one study. They showed pooled sensitivity for the straight leg raising test of 0.91 with a pooled specificity of 0.26. They also showed that discriminative power was lower in recent studies, in studies with only inclusion of primary hernias, and with blind assessment of both the index test, straight leg raising test, and the reference standard surgery. Further, the cross straight leg raising test had a sensitivity of 0.9 with pooled specificity of 0.88. They concluded that the diagnostic accuracy of the straight leg raising test is limited by its low specificity.

7.1 Diagnostic Interventional Techniques

The theoretical basis of controlled diagnostic blocks is based on the fact that if a patient genuinely has pain from a particular target structure, complete relief of that pain should be obtained consistently whenever that structure is anesthetized. Further, there should not be relief if some other structure is anesthetized or if any inactive agent is used to block the target structure. If a patient responds to a first block, but fails to respond appropriately to subsequent controlled blocks, their initial response is deemed to have been false-positive (78).

For diagnostic interventional techniques to be accurate, apart from concept validity, content validity, and face validity, construct validity has to be established. Diagnostic blocks have concept validity on the grounds that it sounds reasonable and that if a structure is a source of pain, anesthetizing it will relieve the pain (78). Content validity defines the test accurately and ensures that the procedure is performed consistently in the same manner.

Face validity and construct validity are significant factors in maintaining the validity of diagnostic accuracy, interventional studies, and diagnostic accuracy tests. The face validity may be established by fluoroscopy and injection of contrast or by a physiological

approach utilizing a detectable and testable function other than pain, such as sympathetic block. To demonstrate the face validity, it is essential to show that potentially confounding targets are not affected in addition to the target structures. The essential element of face validity is to show that the intended target is selectively or discretely anesthetized. Consequently, flooding everything in the vicinity of a target structure or a nerve with local anesthetic does not secure face validity. Another disadvantage with interventional techniques is that despite aiming the injection to a particular structure, it cannot be certainly stated that either the structure will be anesthetized or that only that structure will be anesthetized. In most cases, the flow of injectate depends on the technique used. Consequently, fluoroscopic guidance may assist and it is one of the means available at present by which the face validity of diagnostic blocks can be demonstratedUtilizing an inappropriate technique and injecting low volumes also can corrupt face validity assumptions. Further, injection of inert substances into closed spaces such as joint cavities or epidural space or directly over the nerves is not established.

The next crucial step for diagnostic interventional techniques is maintaining the construct validity to avoid false-positives. Construct validity establishes that the test actually achieves what it is supposed to achieve by measuring the extent to which a test correctly distinguishes the presence, and also the absence, of the condition that the test is supposed to detect — namely false-positive results. Construct validity measures if the test actually works or not, and how well it works.

7.2 Types of Controls

Potentially there are 3 types of controls available for interventional diagnostic techniques. Anatomical controls involve deliberately anesthetizing some adjacent structure that is not the suspected source of pain. Construct validity is achieved if the patient obtains relief whenever the suspected source is anesthetized, but not when the adjacent structure is anesthetized. Thus, any other types of patterns are considered false-positive responses.

The second form of control involves using a placebo agent in which the protocol requires a sequence of 3 blocks. The first block must involve an active agent, in order to establish, prima facie, that the target structure does appear to be the source of pain. The other 2 agents are administered in randomized double-blind basis. Under these conditions, a true-positive response would be the one in which the patient obtained relief on each occasion that an active agent was used, but no relief when the inactive agent was used. However, the issue of an inert substance and its activity in closed surfaces and also injection over nerves is not resolved. Thus, to achieve the same effect and to demonstrate a placebo response, the inert substance, i.e., sodium chloride solution, should be injected without any potential effect on any of the structures. But, at the same time, the injection should be performed in a manner that the patient is not aware that the injection is not performed into the active structure. Consequently, injection into interspinous ligament for an epidural injection and injection away from a medial branch but not close to the medial branch will provide such an effect.

A third approach, most commonly utilized in the United States, and also more pragmatic, is to use comparative local anesthetic blocks. The blocks are performed on separate occasions using local anesthetic agents with different durations of action (164-171). In this approach, the consistency of response and the duration of response are tested. Failure to respond to the second block constitutes inconsistency, and indicates that first response was a false-positive. A response concordant with the expected duration of action of the agent used strongly suggests a genuine, physiologic response, even though lack of concordance does not invalidate the response. Comparative blocks are confounded by the peculiar properties of local anesthetics with prolonged effects (21-23,172-197). If the patient reports a concordant response, the chances of the response being false-positive are only 14%. If the responses are complete, but prolonged in duration, the chances of false-positive response are 35%, but in 65% of patients the response is likely to be genuine. While comparative blocks reduce the false-positives, they do not prove that the response is true-positive. However, if the number of repetitions is increased but the responses remain consistent, the probability that the responses are false becomes increasingly smaller (78).

7.3 Pitfalls of Interventional Diagnostic Techniques

Multiple pitfalls have been described with diagnostic interventional techniques including extensive criticism of the philosophical approach of controlled local anesthetic blocks and the determination of positive and negative responses (109,110).

7.3.1 Facet or Zygapophysial Joint Blocks

Diagnostic blocks of a facet or zygapophysial joint can be performed by anesthetizing the joint by injections of local anesthetic intraarticularly or on the medial branches of the dorsal rami that innervate the target joint to test whether the joint is the source of pain.

The rationale for using facet joint blocks for diagnosis is based on the fact that facet joints are capable of causing pain and they have a nerve supply (21-23,25-27). They have been shown to be a source of pain in patients using diagnostic techniques of known reliability and validity. The value, validity, and clinical effectiveness of diagnostic facet joint nerve blocks has been also illustrated by application of therapeutic modalities based on the diagnosis with controlled comparative local anesthetic blocks.

The face validity of lumbar and cervical medial branch or facet joint nerve blocks has been established by injecting small volumes of local anesthetic and contrast material onto the target points for these structures and by determining the spread of contrast medium in posteroanterior and lateral radiographs (21-23,25-27,78,198-200). Construct validity of facet joint blocks is important to eliminate a placebo effect as the source of confounding results and to secure true-positive results (156,160,201-204). The hypothesis that testing a patient first with lidocaine and subsequently with bupivacaine provides a means of identifying that the placebo response has been tested and proven (160,164-171).

Utilizing the modified criteria established by the IASP (205), false-positive rates varying from 27% to 63% were demonstrated (21-23). The minimal effect of sedation (89,206-208) and lack of influence of psychological factors on the validity of controlled lumbar diagnostic local anesthetic blocks of facet joints have also been demonstrated (209,210). Other variables were also evaluated (211-213).

The systematic reviews by Datta et al (22), Falco et al (23), and Atluri et al (21) utilizing strict criteria of controlled diagnostic blocks and methodologic assessment quality criteria showed the diagnostic accuracy of controlled local anesthetic blocks as Level I or II-1. Datta et al (22) considered 35 manuscripts and included only 9 studies and utilized 7 studies in the evidence synthesis. Atluri et al (21) considered 3 studies with all 3 meeting inclusion criteria. Falco et al (23) considered 14 studies for inclusion and only 9 studies met inclusion criteria. However, all the studies meeting inclusion criteria also met methodologic quality assessment.

7.3.2 Provocation Discography

Discography is a procedure that is used to characterize the pathoanatomy/architecture of the intervertebral disc and to determine if the intervertebral disc is a source of chronic spinal pain. Formal studies have shown that the discs are innervated and can be a source of pain that has pathomorphologic correlates (24,28,30-32,36). Even though the specific neurobiological events involved in how discography causes pain have not been elucidated, sound anatomic, histopathological, radiological, and biomechanical evidence suggests that lumbar discography may help to identify symptomatic and pathological intervertebral discs. Discography was compared with myelography, CT, MRI, and results of surgical and conservative management. CT discography was reported to be more accurate than myelography and plain CT and as good as MRI (24,28,30-32).

The systematic reviews of lumbar (36), cervical (24), and thoracic (31) discography have indicated level II-2 evidence for the accuracy of discography. Wolfer et al (32) in a systematic review of false-positives showed similar evidence and also illustrated that with appropriately performed provocation discography, the false-positive rate is low. In these systematic reviews, multiple studies were considered for inclusion but only a minority of studies met inclusion criteria. For lumbar discography, 11 studies were considered for inclusion criteria and 5 (142,148,214-216) met methodologic quality assessment criteria for evidence synthesis; for cervical discography, 3 studies were considered for inclusion criteria with all 3 (135,152,153) met methodologic quality assessment criteria for evidence synthesis; and for thoracic discography, 4 studies were considered for inclusion criteria and 2 (217,218) met methodologic quality assessment criteria for evidence synthesis.

7.3.3 Sacroiliac Joint Blocks

Due to the inability to make the diagnosis of sacroiliac joint pain with non-invasive tests, sacroiliac joint blocks appear to be the only evaluation to provide appropriate diagnosis. Further, controlled studies have established sacroiliac joints as a potential source of low back and lower extremity pain.

The indicated evidence was Level II-2 for the diagnosis of sacroiliac joint pain utilizing controlled diagnostic blocks as per Rupert et al (35).

8.0 How to Report Diagnostic Studis

The STARD statement was developed in 2002 and published in 2003 (4), along with an explanation and elaboration (66). The authors also published a document entitled "Development and Validation of Methods for Assessing and Quality of Diagnostic Accuracy Studies" (47). As shown in Table 5, they have published a 25-item checklist. The methodology includes participants, test methods, and statistical methods. Participant section includes the study population, participant recruitment, participant sampling, and data collection. Test methods encompass the reference standard, technical specifications of materials and methods, definition and rationale for units, cutoffs, and/or categories of the results of the index tests and the reference standard, the details of the personnel

Table 5. *STARD checklist for the reporting of studies diagnostic accuracy.*

Section and Topic	Item #		On page #
TITLE /ABSTRACT/ KEY WORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading "senstivity and specificity").	
INTRODUCTION	2	State the research questions or study aim, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups	
METHODS		Describe	
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Participant sampling: Was the study population a consectuive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	
	6	Data collection: Was data collection planned before the index test and reference standard were performed (Prospective study) or after (retrospective study)?	
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	
	9	Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.	
	10	The number, training and expertise of the persons executing and readig the index tests and the reference standard.	
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the othe rtest and describe any other clinical information available to the readers.	
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals)	
	13	Methods for calculating test reproducibility, if done	
RESULTS		Report	
Participants	14	When study was done, including beginning and ending dates of recruitment.	
	15	Clincial and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	
	16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why particpants failed to receive eithe rtest (a flow diagram is strongly recommended).	
Test results	17	Time interval fromthe index tests to the reference standard, and any treatment administred between.	
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Any adverse events from performing the index tests or the reference standard.	
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	
	22	How indeterminate results, missing responses and outliers of the index tests were handled.	
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	24	Estimates of test reproducibility, if done.	
DISCUSSION	25	Discuss the clinical applicability of the study findings.	

Adapted from Bossuyt PM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Ann Intern Med* 2003; 138:40-44.

involved in the execution of the reference standard, and whether or not the readers of the index test and reference standards were blind to the results of the other test and described any other clinical information available to the other test. Finally, statistical methods include methods for calculating or comparing measures of diagnostic accuracy, statistical methods used to quantify uncertainty, and methods for calculating test reproducibility. Figure 1 illustrates the participant flow diagram of diagnostic accuracy studies.

The guiding principle in the development of the STARD checklist was to select items that would help readers judge the potential for bias in the study and appraise the applicability of the findings. Other general considerations shaping the current content and format of the checklist were that (1) the STARD group

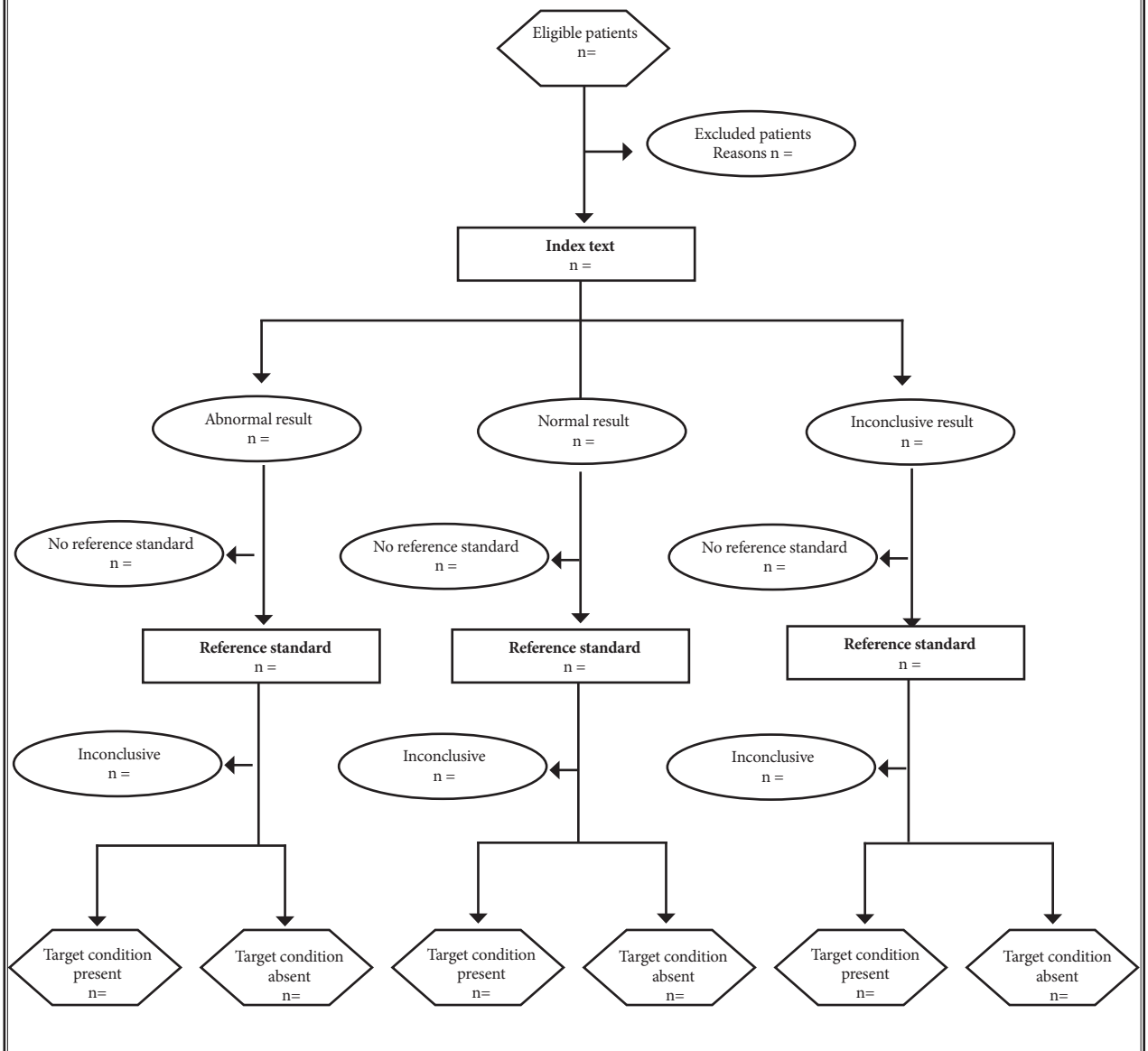

Fig. 1. ***Prototypical flow diagram of a diagnostic accuracy study.***

Adapted from Bossuyt PM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Ann Intern Med* 2003; 138:40-44.).

believes that one general checklist for studies of diagnostic accuracy, rather than different checklists for each field, is likely to be more widely disseminated and perhaps accepted by users. Further, a separate background document explains the meaning and rationale of each item and briefly summarizes the type and amount of evidence (109). STARD also considers that the flow diagram is an essential element in the CONSORT standards for reporting of randomized trials, thus, a comparable flow diagram has become an essential element of STARD. Flow diagrams in the reports of diagnostic accuracy studies indicate the process of sampling and selecting participants (external validity), the flow of participants in relation to the timing and outcomes of tests, the number of subjects who failed to receive either the index test and/or the reference standard (potential for verification bias) (220-222), and the number of patients at each stage of the study, thus providing the correct denominator for proportions, namely internal consistency (221-223).

9.0 Reporting of Diagnostic Accuracy Studies

The STARD (4,65) provides a checklist of multiple items as shown in Table 5.

9.1 Title, Abstract, and Key Words

The STARD recommends that the use of the term "diagnostic accuracy" in the title or abstract of the report that compares the results of one or more index tests with the results of a reference standard.

The first section also describes a structured abstract and provides key words. Based on the requirements of the journal, the article will be published.

9.2 Introduction

The introduction should describe the scientific background, previous work on the subject, the remaining uncertainty, and, hence, the rationale for study. The research question should be clearly specified which will assist the readers in judging the appropriateness of the study design and data analysis.

9.3 Methods

This section includes the description about the participants, test methods, and statistical methods.

9.3.1 Participants

The report should include a clear and concise description of the targeted population. The eligibility criteria described the targeted patient population, including additional exclusion criteria used for reasons of safety or feasibility. It is essential that readers understand whether or not the study excluded patients with a specific condition known to adversely affect the way the test works, which would inflate diagnostic accuracy, also known as limited challenge bias (65,224). Further, the methods section should describe the setting and location, including where the patients were recruited and where the test and reference standard were performed. It is also important to describe how eligible subjects were identified. While participant recruitment in diagnostic studies can start at different points, frequently, the study enrolls consecutive patients clinically suspected of the target condition because of presenting symptoms or referral by another health care professional. Consequently, these patients then undergo the index tests, as well as the reference standard.

The targeted study population includes all patients that satisfy the criteria for inclusion and are not disqualified by one or more of the exclusion criteria. The included patients may be either a consecutive series of patients presenting at the study center or a subselection. The subselection may or may not be truly random.

The report should also clearly describe the data collection.

9.3.2 Test Methods

The test methods include the reference standard; technical specifications; definition of and rationale for the units; cutoffs and/or categories of the results of index tests and the reference standard; the number, training, and expertise of the persons executing and reading the index test and the reference standard; and whether or not the readers of the index tests and reference standard were blind to the results of the other tests and describe any other clinical information available to the readers. Thus, the authors should clearly define the reference standard and how the choice of the reference standard relates to the study question. Further, authors should describe the methods involved in the execution of the index test and reference standard in sufficient detail to allow other researchers to replicate the study. The descriptions also must be provided in reference to the definition of, and rationale for, the units, cutoffs, and/or categories of the results of the index tests and the reference standards.

Variability in the manipulation, processing, or reading of the index test or reference standard will af-

fect measures of diagnostic accuracy (225,226). Many studies have shown reader variability, especially in the field of imaging (227,228).

Finally, the methodology should describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of other tests and describe any other clinical information available to the readers. Blinding or masking of readers of the tests is important to avoid exacerbation of results.

9.3.3 Statistical Methods

The authors should describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty, such as 95% confidence intervals (CIs). In general, estimates of diagnostic accuracy are subject to chance variation, with larger studies usually resulting in more precise estimates.

The authors also should describe the methods for calculating test reproducibility if performed. In the real world, neither the index test nor the reference standard is perfect. Consequently, their reproducibility varies and limited reproducibility adversely affects diagnostic accuracy (229).

9.4 Results

The results section includes reporting of participants, test results, and estimates.

9.4.1 Participants

This section includes when the study was done, including beginning and ending dates of recruitment; clinical and demographic characteristics of the study population; and the number of participants satisfying the criteria for inclusion that did or did not undergo the index test and/or the reference standard and descriptions of why participants failed to receive either test as illustrated in Fig. 1.

9.4.2 Test Results

In this section, the report should include the time interval from the index tests to the reference standard, and any treatment administration between; distribution of severity of disease; a cross-sectional tabulation of results of the index tests by the results of the reference standard for conclusion results, the distribution of the test results by the results of the reference standard; and any adverse events from performing the index test or the reference standard.

9.4.3 Estimates

This section clearly describes the estimates of diagnostic accuracy and measures of statistical uncertainty with 95% CIs, the procedure of handling of indeterminate results, missing responses, etc.; estimates of variability of diagnostic accuracy between subgroups of participants, readers, or centers, if done; and estimates of test reproducibility, if done.

9.5 Discussion

The discussion should describe interpretation of the results, generalizability of the results, and overall evidence.

A structured format must be provided for the results. Based on the recommendations of the Annals of Internal Medicine (230), the authors should structure the discussion section by presenting:

- A brief synopsis of the key findings
- Consideration of possible mechanisms and explanation
- Comparison with relevant findings from other published studies
- Limitations of the present study and methods used to minimize and compensation for those limitations
- A brief section that summarizes the clinical and research implications of the work, as appropriate.

It is of particular importance to discuss the weaknesses and limitations of the study. Along with the limitations, discussion of any impression of the results is essential to be included in the weaknesses. Imprecision may arise in connection with several aspects of a study, including measurement of a primary diagnosis (41). It will also be worthwhile to describe the difference between statistically significant and clinical importance.

10.0 Discussion

Diagnosis is a critical component of health care, and clinicians, policy makers, and patients routinely face a range of questions regarding diagnostic tests. Well designed diagnostic test accuracy studies can help in making these decisions, provided that they transparently and fully report their participants, tests, methods, and results as facilitated by the STARD statement.

Interventional pain management is an evolving specialty with multiple stumbling blocks in conduct-

ing diagnostic accuracy studies. Limitations of interventional diagnostic studies include criterion standard without tissue biopsy; however, clinical follow-up can be used for validation of interventional diagnostic studies. Thus, appropriate design and interpretation of interventional diagnostic studies is crucial.

Acknowledgments

The authors wish to thank Vidyasagar Pampati, MSc, statistician; Sekar Edem for assistance in search of literature; and Tonie M. Hatton and Diane E. Neihoff, transcriptionists, for their assistance in the preparation of this manuscript.

References

1. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams JW Jr, Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt GH; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; 336:1106-1110.
2. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; 323:157-162.
3. Nordin M, Carragee EJ, Hogg-Johnson S, Weiner SS, Hurwitz EL, Peloso PM, Guzman J, Van Der Velde G, Carroll LJ, Holm LW, Cote P, Cassidy JD, Haldeman S. Assessment of neck pain and its associated disorders: Results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and its Associated Disorders. *Spine* 2008; 33:S101-S122.
4. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Ann Intern Med* 2003; 138:40-44.
5. Guyatt G, Drummond R. Part 1. The Basics: Using the Medical Literature. 1C2. Diagnostic tests. In: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. American Medical Association, Chicago, 2002, pp 121-140.
6. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. *Ann Intern Med* 1981; 94:557-592.
7. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. The selection of diagnostic tests. In: *Clinical Epidemiology, A Basic Science for Clinical Medicine*, 2nd ed. Little, Brown and Company, Boston, 1991, pp. 47-57.
8. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978; 8:283-298.
9. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995; 274:645-651.
10. Nelemans PJ, Leiner T, de Vet HC, van Engelshoven JM. Peripheral arterial disease: Meta-analysis of the diagnostic performance of MR angiography. *Radiology* 2000; 217:105-114.
11. Devries SO, Hunink MGM, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis of the diagnostic performance of duplex ultrasonography in peripheral arterial disease. *Acad Radiol* 1996; 3:361-369.
12. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061-1066.
13. Carragee EJ. Validity of self-reported history in patients with acute back and neck pain after motor vehicle accident. *Spine J* 2008; 8:311-319.
14. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *Can Med Assoc J* 1986; 134:587-594.
15. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11:88-94.
16. Kent DL, Larson EB. Disease, level of impact, and quality of research methods. Three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Invest Radiol* 1992; 27:245-254.
17. Rutjes AW, Reitsma JB, Di Nisio M, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; 174:469-476.
18. Shunmugam M, Azuara-Blanco A. The quality of reporting of diagnostic accuracy studies in glaucoma using the Heidelberg retina tomograph. *Invest Ophthalmol Vis Sci* 2006; 47:2317-2323.
19. Rubinstein SM, van Tulder M. A best-evidence review of diagnostic procedures for neck and low-back pain. *Best Pract Res Clin Rheumatol* 2008; 22:471-482.
20. Hancock MJ, Maher CG, Latimer J, Spindler MF, McAuley JH, Laslett M, Bogduk N. Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. *Eur Spine J* 2007; 16:1539-1550.
21. Atluri S, Datta S, Falco FJE, Lee M. Systematic review of diagnostic utility and therapeutic effectiveness of thoracic facet joint interventions. *Pain Physician* 2008; 11:611-629.
22. Datta S, Lee M, Falco FJE, Bryce DA, Hayek SM. Systematic assessment of diagnostic accuracy and therapeutic utility of lumbar facet joint interventions. *Pain Physician* 2009; 12:437-460.
23. Falco FJE, Erhart S, Wargo BW, Bryce DA, Atluri S, Datta S, Hayek SM. Systematic review of diagnostic utility and therapeutic effectiveness of cervical facet joint interventions. *Pain Physician* 2009; 12:323-344.
24. Manchikanti L, Dunbar EE, Wargo BW, Shah RV, Derby R, Cohen SP. Systematic review of cervical discography as a diagnostic test for chronic spinal pain. *Pain Physician* 2009; 12:305-321.
25. Boswell MV, Singh V, Staats PS, Hirsch JA. Accuracy of precision diagnostic blocks in the diagnosis of chronic spinal pain of facet or zygapophysial joint origin: A systematic review. *Pain Physician* 2003; 6:449-456.
26. Sehgal N, Shah RV, McKenzie-Brown A, Everett CR. Diagnostic utility of facet (zygapophysial) joint injections in chronic spinal pain: A systematic review of evidence. *Pain Physician* 2005; 8:211-224.
27. Sehgal N, Dunbar EE, Shah RV, Colson JD. Systematic review of diagnostic utility of facet (zygapophysial) joint injections in chronic spinal pain: An update. *Pain Physician* 2007; 10:213-228.
28. Shah RV, Everett CR, McKenzie-Brown AM, Sehgal N. Discography as a diagnostic test for spinal pain: A systemat-

ic and narrative review. *Pain Physician* 2005; 8:187-209.

29. Cohen SP, Larkin TM, Barna SA, Palmer WE, Hecht AC, Stojanovic MP. Lumbar discography: A comprehensive review of outcome studies, diagnostic accuracy, and principles. *Reg Anesth Pain Med* 2005; 30:163-183.
30. Buenaventura RM, Shah RV, Patel V, Benyamin R, Singh V. Systematic review of discography as a diagnostic test for spinal pain: An update. *Pain Physician* 2007; 10: 147-164.
31. Singh V, Manchikanti L, Shah RV, Dunbar EE, Glaser SE. Systematic review of thoracic discography as a diagnostic test for chronic spinal pain. *Pain Physician* 2008; 11:631-642.
32. Wolfer L, Derby R, Lee JE, Lee SH. Systematic review of lumbar provocation discography in asymptomatic subjects with a meta-analysis of false-positive rates. *Pain Physician* 2008; 11:513-538.
33. McKenzie-Brown AM, Shah RV, Sehgal N, Everett CR. A systematic review of sacroiliac joint interventions. *Pain Physician* 2005; 8:115-125.
34. Hansen HC, McKenzie-Brown AM, Cohen SP, Swicegood JR, Colson JD, Manchikanti L. Sacroiliac joint interventions: A systematic review. *Pain Physician* 2007; 10:165-184.
35. Rupert MP, Lee M, Manchikanti L, Datta S, Cohen SP. Evaluation of sacroiliac joint interventions: A systematic appraisal of the literature. *Pain Physician* 2009; 12:399-418.
36. Manchikanti L, Glaser SE, Wolfer L, Derby R, Cohen SP. Systematic review of lumbar discography as a diagnostic test for chronic low back pain. *Pain Physician* 2009; 12:541-559.
37. Datta S, Everett CR, Trescot AM, Schultz DM, Adlaka R, Abdi S, Atluri SL, Smith HS, Shah RV. An updated systematic review of diagnostic utility of selective nerve root blocks. *Pain Physician* 2007; 10:113-128.
38. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: A systematic review. *Ann Intern Med* 2004; 140:189-202.
39. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic accuracy. *Ann Intern Med* 2008; 149:889-897.
40. Bachmann LM, Puhan MA, ter Riet G, Bossuyt PM. Sample sizes of studies on diagnostic accuracy: Literature survey. *BMJ* 2006; 332:1127-1129.
41. Moher D, Schulz KF, Altman D; CONSORT Group. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; 285:1987-1991.
42. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134:663-694.
43. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA* 2006; 295:1152-1160.
44. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M,; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Ann Intern Med* 2007; 147:W163-W194.
45. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Ann Intern Med* 2007; 147:573-577.
46. Stengel D, Bauwens K, Rademacher G, Mutze S, Ekkernkamp A. Association between compliance with methodological standards of diagnostic research and reported test accuracy: Meta-analysis of focused assessment of US for trauma. *Radiology* 2005; 236:102-111.
47. Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004; 8:1-234.
48. Freedman LS. Evaluating and comparing imaging techniques: A review and classification of study designs. *Br J Radiol* 1987; 60:1071-1081.
49. Somoza E, Mossman D. Comparing and optimizing diagnostic tests: An information theoretical approach. *Med Decision Making* 1992; 12:179-188.
50. Plasencia CM, Alderman BW, Baron AE, Rolfs RT, Boyko EJ. A method to describe physician decision thresholds and its application in examining the diagnosis of coronary artery disease based on exercise treadmill testing. *Med Decision Making* 1992; 12:204-212.
51. Moons KG. *Diagnostic Research: Theory and Application*.: Erasmus University, Rotterdam, 1996.
52. Moons KG, Stijnen T, Michel BC, Buller HR, Van Es GA, Grobbee DE, Habbema JD. Application of treatment thresholds to diagnostic-test evaluation: An alternative to the comparison of areas under receiver operating characteristic curves. *Med Decision Making* 1997; 17:447-454.
53. Thornbury JR. Eugene W. Caldwell lecture. Clinical efficacy of diagnostic imaging: Love it or leave it. *AJR Am J Roentgenol* 1994; 162:1-8.
54. Simon DG, Lubin MF. Cost-effectiveness of computerized tomography and magnetic resonance imaging in dementia. *Med Decision Making* 1985; 5:335-354.
55. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L. *Systems to Rate the Strength of Scientific Evidence*, Evidence Report, Technology Assessment No. 47. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Health care Research and Quality, 2002. www.thecre.com/pdf/ahrq-system-strength.pdf
56. Pai M, McCulloch M, Gorman JD, Pai N, Enanoria W, Kennedy G, Tharyan P, Colford JM Jr. Systematic reviews and meta-analyses: An illustrated, step-by-step guide. *Natl Med J India* 2004; 17:86-95.
57. Guyatt GH, Rennie D. *Users' Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice*. The Evidence-Based Medicine Working Group. AMA Press, Chicago, 2002.
58. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323:42-46.
59. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:25.
60. Eden J, Wheatley B, McNeil B, Sox H. Systematic reviews: The central link

between evidence and clinical decision making. In: *Knowing What Works in Health Care: A Roadmap for the Nation*. National Academies Press, Washington, DC, 2008, pp 81-120.

61. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med* 2005; 142:1090-1099.
62. Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, Lau J. Challenges in systematic reviews of diagnostic technologies. *Ann Inter Med* 2005; 142:1048-1055.
63. Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ* 2002; 324:539-541.
64. Westwood ME, Whiting PF, Kleijnen J. How does study quality affect the results of a diagnostic meta-analysis? *BMC Med Res Methodol* 2005; 5:20.
65. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HC, Lijmer JG; Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Ann Inter Med* 2003; 138:W1-W12.
66. Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Bossuyt PM, Reitsma JB, Bouter LM, de Vet HC. Reproducibility of the STARD checklist: An instrument to assess the quality of reporting of diagnostic accuracy studies. *BMC Med Res Methodol* 2006; 15:12.
67. Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Bossuyt PM, Reitsma JB, Bouter LM, de Vet HC. Quality of reporting of diagnostic accuracy studies. *Radiology* 2005; 235:347-353.
68. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does the quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352:609-613.
69. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412.
70. Vamvakas EC. Meta-analyses of studies of the diagnostic accuracy of laboratory tests. A review of the concepts and methods. *Arch Pathol Lab Med* 1998; 122:675-686.
71. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995; 48:119-130.
72. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994; 271:389-391.
73. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA* 1994; 271:703-707.
74. Zhou XH. Correcting for verification bias in studies of a diagnostic test's accuracy. *Stat Methods Med Res* 1998; 7:337-353.
75. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. *Stat Methods Med Res* 1998;7:354-370.
76. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007; 11:ix-51.
77. Altman DG, Bland JM. Diagnostic tests. I. Sensitivity and specificity. *BMJ* 1994; 308:1552.
78. Bogduk N. Principles of diagnostic blocks. In: Slipman C, Derby R, Simeone FA, Mayer TG (eds). *Interventional Spine: An Algorithmic Approach*. Saunders Elsevier, Philadelphia, 2008, pp 187-192.
79. Weller SC, Mann NC. Assessing rater performance without a "gold standard" using consensus theory [published erratum appears in *Med Decis Making* 1997;17:240]. *Med Decis Making* 1997; 17:71-79.
80. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: A review. *J Clin Epidemiol* 1988; 41:923-937.
81. Begg CB, Metz CE. Consensus diagnoses and "gold standards" [published erratum appears in *Med Decis Making* 1990;10:149]. *Med Decis Making* 1990; 10:29-30.
82. Henkleman RM, Kay I, Bronskill MJ. Receiver operating characteristic (ROC) analysis without truth. *Med Decis Making* 1990; 10:24-29.
83. Buck AA, Gart JJ. Comparison of a screening test and a reference test in epidemiologic studies. I. Indices of agreement and their relation to prevalence. *Am J Epidemiol* 1966; 83:586-592.
84. Manchikanti L, Boswell MV, Giordano J. Evidence-based interventional pain management: Principles, problems, potential, and applications. *Pain Physician* 2007; 10:329-356.
85. Manchikanti L. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 1: Introduction and general considerations. *Pain Physician* 2008; 11:161-186.
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. Manchikanti L, Benyamin RM, Helm S, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 3: Systematic reviews and meta-analysis of randomized trials. *Pain Physician* 2009; 12:35-72.
88. Manchikanti L, Singh V, Smith HS, Hirsch JA. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 4: Observational studies. *Pain Physician* 2009; 12:73-108.
89. Smith HS, Chopra P, Patel VB, Frey ME, Rastogi R. Systematic review on the role of sedation in diagnostic spinal interventional techniques. *Pain Physician* 2009; 12:195-206.
90. Guyatt G, Drummond R. Part 1. The Basics: Using the Medical Literature. 1A. Introduction: The philosophy of evidence-based medicine. In: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. American Medical Association, Chicago, 2002, pp 3-12.
91. Napodano RJ. *Values in Medical Practice*. Human Sciences Press, New York, 1986.
92. Haynes RB, Sackett RB, Gray JM, Cook DC, Guyatt GH. Transferring evidence from research into practice: 1. The role of clinical care research evidence in clinical decisions. *ACP J Club* 1996; 125: A14-A16.
93. Hotopf M. The pragmatic randomized controlled trial. *Adv Psychiatr Treat* 2002; 8:326-333.
94. Williams DD, Garner J. The case against 'the evidence': A different perspective on evidence-based medicine. *Br J Psychiatry* 2002; 180:8-12.
95. Hotopf M, Lewis G, Normand C. Putting trials on trial: The costs and conse-

quences of small trials in depression: A systematic review of methodology. *J Epidemiol Community Health* 1997; 51:354-358.

96. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998; 317:1181-1184.
97. Hotopf M, Churchill R, Lewis G. Pragmatic randomized controlled trials in psychiatry. *Br J Psychiatry* 1999; 175:217-223.
98. Miles A, Charlton B, Bentley P, Polychronis A, Grey J, Price N. New perspectives in the evidence-based healthcare debate. *J Eval Clin Pract* 2000; 6:77-84.
99. Carragee EJ, Hurwitz EL, Cheng I, Carroll LJ, Nordin M, Guzman J, Peloso P, Holm LW, Côté P, Hogg-Johnson S, van der Velde G, Cassidy JD, Haldeman S; Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. Treatment of neck pain: Injections and surgical interventions: Results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* 2008; 33:S153-S169.
100. American College of Occupational and Environmental Medicine (ACOEM). Low Back Disorders. In: *Occupational Medicine Practice Guidelines: Evaluation and Management of Common Health Problems and Functional Recovery of Workers*, Second Edition. American College of Occupational and Environmental Medicine, Elk Grove Village, 2007.
101. American College of Occupational and Environmental Medicine. Chronic Pain Chapter (revised 2008). In: *Occupational Medicine Practice Guidelines: Evaluation and Management of Common Health Problems and Functional Recovery of Workers,* Second Edition. American College of Occupational and Environmental Medicine, Elk Grove Village, Epublished August 14, 2008.
102. Manchikanti L, Singh V, Derby R, Helm S, Trescot AM, Staats PS, Prager JP, Hirsch JA. Review of occupational medicine practice guidelines for interventional pain management and potential implications. *Pain Physician* 2008; 11:271-289.
103. Manchikanti L, Singh V, Helm S, Trescot AM, Hirsch JA. A critical appraisal of 2007 American College of Occupational and Environmental Medicine (ACOEM) practice guidelines for interventional pain management: An independent review utilizing AGREE, AMA, IOM, and other criteria. *Pain Physician* 2008; 11:291-310.
104. Chou R. Using evidence in pain practice: Part I: Assessing quality of systematic reviews and clinical practice guidelines. *Pain Med* 2008; 9:518-530.
105. Chou R. Using evidence in pain practice: Part II: Interpreting and applying systematic reviews and clinical practice guidelines. *Pain Med* 2008; 9:531-541.
106. Niemisto L, Kalso E, Malmivaara A, Seitsalo S, Hurri H. Cochrane Collaboration Back Review Group. Radiofrequency denervation for neck and back pain: A systematic review within the framework of the Cochrane collaboration back review group. *Spine* 2003, 28:1877-1888.
107. Levin JH. Prospective, double-blind, randomized placebo-controlled trials in interventional spine: What the highest quality literature tells us. *Spine J* 2008; Sep 11 [Epub ahead of print].
108. Manchikanti L, Singh V, Derby R, Schultz DM, Benyamin RM, Prager JP, Hirsch JA. Reassessment of evidence synthesis of occupational medicine practice guidelines for interventional pain management. *Pain Physician* 2008; 11:393-482.
109. Carragee EJ, Haldeman S, Hurwitz E. The pyrite standard: The Midas touch in the diagnosis of axial pain syndromes. *Spine J* 2007; 7:27-31.
110. Bogduk N. In defense of King et al: The validity of manual examination in assessing patients with neck pain. *Spine J* 2007; 7:749-752; author reply (Carragee EJ) 752-753.
111. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. WB Saunders Co., Philadelphia, 1985.
112. Pearl WS. A hierarchical outcomes approach to test assessment. *Ann Emerg Med* 1999; 33:77-84.
113. Boswell MV, Trescot AM, Datta S, Schultz DM, Hansen HC, Abdi S, Sehgal N, Shah RV, Singh V, Benyamin RM, Patel VB, Buenaventura RM, Colson JD, Cordner HJ, Epter RS, Jasper JF, Dunbar EE, Atluri SL, Bowman RC, Deer TR, Swicegood JR, Staats PS, Smith HS, Burton AW, Kloth DS, Giordano J, Manchikanti L. Interventional techniques: Evidence-based practice guidelines in the management of chronic spinal pain. *Pain Physician* 2007; 10:7-111.
114. Bogduk N, McGuirk B. Causes and sources of chronic low back pain. In: *Medical Management of Acute and Chronic Low Back Pain. An Evidence-Based Approach: Pain Research and Clinical Management,* Vol. 13. Elsevier Science BV, Amsterdam, 2002, pp 115-126.
115. Bogduk N, McGuirk B. An algorithm for precision diagnosis. In: Bogduk N, McGuirk B, eds. *Medical Management of Acute and Chronic Low Back Pain. An Evidence-Based Approach: Pain Research and Clinical Management*.: Elsevier Science BV, Amsterdam, 2002; 13:177-186.
116. Deyo RA, Weinstein JN. Low back pain. *N Engl J Med* 2001; 344:363-370.
117. Deyo RA. Fads in the treatment of low back pain. *N Engl J Med* 1991; 325:1039-1040.
118. Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? *JAMA* 1992; 268:760-765.
119. van Tulder MWV, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 1997; 22:2128-2156.
120. Cohen SP, Argoff CE, Carragee EJ. Management of low back pain. *BMJ* 2008; 338:100-106.
121. Bogduk N. Low back pain. In: *Clinical Anatomy of Lumbar Spine and Sacrum*, 4th edition. Churchill Livingstone, New York, 2005, pp 183-216.
122. Nachemson AL. The natural course of low back pain. In: White A, Gordon SL (eds). *Symposium on Idiopathic Low Back Pain*. Mosby, St. Louis, 1982, pp 46-51.
123. Mooney V. Where is the pain coming from? *Spine* 1987; 12:754-759.
124. Spitzer WO, LeBlanc FE, Dupuis M. Scientific approach to the assessment and management of activity-related spinal disorders: A monograph for clinicians. Report of Quebec Task Force on Spinal Disorders. *Spine* 1987; 12:S1-S59.
125. Bogduk N, McGuirk B. Sources and causes of neck pain. In *Management of Acute and Chronic Neck Pain. An Evidence-Based Approach*. Elsevier, 2006, pp 9-20.
126. Bogduk N, McGuirk B. Acute neck pain: Algorithm for acute neck pain. In: *Management of Acute and Chronic Neck Pain. An Evidence-Based Approach*. Elsevier, 2006, pp 69-77.
127. Bogduk N, Lord S. Cervical zygapophysial joint pain. *Neurosurg Q* 1998; 8:107-117.

128. Manchikanti L, Singh V, Pampati V, Damron K, Barnhill R, Beyer C, Cash K. Evaluation of the relative contributions of various structures in chronic low back pain. *Pain Physician* 2001; 4:308-316.
129. Manchukonda R, Manchikanti KN, Cash KA, Pampati V, Manchikanti L. Facet joint pain in chronic spinal pain: An evaluation of prevalence and false-positive rate of diagnostic blocks. *J Spinal Disord Tech* 2007; 20:539-545.
130. Manchikanti L, Manchukonda R, Pampati V, Damron KS, McManus CD. Prevalence of facet joint pain in chronic low back pain in postsurgical patients by controlled comparative local anesthetic blocks. *Arch Phys Med Rehabil* 2007; 88:449-455.
131. Manchikanti L, Boswell MV, Singh V, Pampati V, Damron KS, Beyer CD. Prevalence of facet joint pain in chronic spinal pain of cervical, thoracic, and lumbar regions. *BMC Musculoskelet Disord* 2004; 5:15.
132. Manchikanti L, Singh V, Pampati V, Beyer CD, Damron KS. Evaluation of the prevalence of facet joint pain in chronic thoracic pain. *Pain Physician* 2002; 5:354-359.
133. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, Bogduk N. Clinical features of patients with pain stemming from the lumbar zygapophysial joints. Is the lumbar facet syndrome a clinical entity? *Spine* 1994; 19:1132-1137.
134. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, Bogduk N. The relative contributions of the disc and zygapophyseal joint in chronic low back pain. *Spine* 1994; 19:801-806.
135. Yin W, Bogduk N. The nature of neck pain in a private pain clinic in the United States. *Pain Med* 2008; 9:196-203.
136. Aprill C, Dwyer A, Bogduk N. The prevalence of cervical zygapophyseal joint pain patterns II: A clinical evaluation. *Spine* 1990; 15:458-461.
137. Irwin RW, Watson T, Minick RP, Ambrosius WT. Age, body mass index, and gender differences in sacroiliac joint pathology. *Am J Phys Med Rehabil* 2007; 86:37-44.
138. Szadek KM, van der Wurff P, van Tulder MW, Zuurmond WW, Perez RS. Diagnostic validity of criteria for sacroiliac joint pain: A systematic review. *J Pain* 2009; 10:354-368.
139. Maigne JY, Aivakiklis A, Pfefer F. Results of sacroiliac joint double block and value of sacroiliac pain provocation test in 54 patients with low back pain. *Spine* 1996; 21:1889-1892.
140. Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: A validity study of a McKenzie evaluation and sacroiliac provocation tests. *Aust J Physiother* 2003; 49:89-97.
141. van der Wurff P, Buijs EJ, Groen GJ. A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. *Arch Phys Med Rehabil* 2006; 87:10-14.
142. Carragee EJ, Tanner CM, Khurana S, Hayward C, Welsh J, Date E, Truong T, Rossi M, Hagle C. The rates of false-positive lumbar discography in select patients without low back symptoms. *Spine* 2000; 25:1373-1380.
143. Carragee EJ, Chen Y, Tanner CM, Hayward C, Rossi M, Hagle C. Can discography cause long-term back symptoms in previously asymptomatic subjects? *Spine* 2000; 25:1803-1808.
144. Carragee EJ, Alamin TF, Miller J, Grafe M. Provocative discography in volunteer subjects with mild persistent low back pain. *Spine J* 2002; 2:25-34.
145. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, Bogduk N. The prevalence and clinical features of internal disc disruption in patients with chronic low back pain. *Spine* 1995; 20:1878-1883.
146. Carragee EJ, Lincoln T, Parmar VS, Alamin T. A gold standard evaluation of the "discogenic pain" diagnosis as determined by provocative discography. *Spine* 2006; 31:2115-2123.
147. Derby R, Kim BJ, Chen Y, Seo KS, Lee SH. The relation between annular disruption on computed tomography scan and pressure-controlled diskography. *Arch Phys Med Rehabil* 2005; 86:1534-1538.
148. Carragee EJ, Tanner CM, Yang B, Brito JL, Truong T. False-positive findings on lumbar discography. Reliability of subjective concordance assessment during provocative disc injection. *Spine* 1999; 24:2542-2547.
149. Cohen SP, Larkin T, Fant GV, Oberfoell R, Stojanovic M. Does needle insertion site affect diskography results? A retrospective analysis. *Spine* 2002; 27:2279-2283.
150. Derby R, Howard MW, Grant JM, Lettice JJ, Van Peteghem PK, Ryan DP. The ability of pressure-controlled discography to predict surgical and nonsurgical outcomes. *Spine* 1999; 24:364-371.
151. Manchikanti L, Singh V, Pampati VS, Fellows B, Beyer C, Damron K, Cash KA. Provocative discography in low back pain patients with or without somatization disorder: A randomized prospective evaluation. *Pain Physician* 2001; 4:227-239.
152. Bogduk N, Aprill C. On the nature of neck pain, discography and cervical zygapophysial joint blocks. *Pain* 1993; 54:213-217.
153. Palit M, Schofferman J, Goldthwaite N, Reynolds J, Kerner M, Keaney D, Lawrence-Miyasaki L. Anterior discectomy and fusion for the management of neck pain. *Spine* 1999; 24:2224-2228.
154. Manchikanti L, Singh V, Pampati V, Damron KS, Beyer CD, Barnhill RC. Is there correlation of facet joint pain in lumbar and cervical spine? An evaluation of prevalence in combined chronic low back and neck pain. *Pain Physician* 2002; 5:365-371.
155. Schwarzer AC, Wang S, Bogduk N, McNaught PJ, Laurent R. Prevalence and clinical features of lumbar zygapophysial joint pain: A study in an Australian population with chronic low back pain. *Ann Rheum Dis* 1995; 54:100-106.
156. Manchikanti L, Hirsch JA, Pampati V. Chronic low back pain of facet (zygapophysial) joint origin: Is there a difference based on involvement of single or multiple spinal regions? *Pain Physician* 2003; 6:399-405.
157. Lord SM, Barnsley L, Wallis BJ, Bogduk N. Chronic cervical zygapophysial joint pain with whiplash: A placebo-controlled prevalence study. *Spine* 1996; 21:1737-1745.
158. Manchikanti L, Manchikanti K, Pampati V, Brandon D, Giordano J. The prevalence of facet joint-related chronic neck pain in postsurgical and non-postsurgical patients: A comparative evaluation. *Pain Pract* 2008; 8:5-10.
159. Speldewinde G, Bashford G, Davidson I. Diagnostic cervical zygapophyseal joint blocks for chronic cervical pain. *Med J Aust* 2001; 174:174-176.
160. Barnsley L, Lord S, Wallis B, Bogduk N. False-positive rates of cervical zygapophysial joint blocks. *Clin J Pain* 1993; 9:124-130.
161. de Graaf I, Prak A, Bierma-Zeinstra S, Thomas S, Peul W, Koes B. Diagnosis of lumbar spinal stenosis: A systematic review of the accuracy of diagnostic tests. *Spine* 2006; 31:1168-1176.
162. Kent DL, Haynor DR, Larson EB, Deyo RA. Diagnosis of lumbar spinal steno-

sis in adults: A meta-analysis of the accuracy of CT, MR, and myelography. *AJR Am J Roentgenol* 1992; 158:1135-1144.

163. Devillé WL, van der Windt DA, Dzaferagi A, Bezemer PD, Bouter LM. The test of Lasègue: Systematic review of the accuracy in diagnosing herniated discs. *Spine* 2000; 25:1140-1147.
164. Buckley FP. Regional anesthesia with local anesthetics. In: Loeser JD (ed). *Bonica's Management of Pain*, 3rd Edition. Lippincott Williams & Wilkins, Philadelphia, 2001, pp 1893-1952.
165. Bonica JJ. Local anesthesia and regional blocks. In: Wall PD, Melzack R (eds). *Textbook of Pain*, 2nd Edition. Churchill Livingstone, Edinburgh, 1989, pp 724-743.
166. Bonica JJ, Buckley FP. Regional analgesia with local anesthetics. In: Bonica JJ (ed). *The Management of Pain, Vol. II*. Lea & Febiger, Philadelphia, 1990, pp 1883-1966.
167. Boas RA, Nerve blocks in the diagnosis of low back pain. *Neurosurg Clin N Am* 1991; 2:807-816.
168. Bonica JJ, Butler SH. Local anaesthesia and regional blocks. In: Wall PD, Melzack R (eds). *Textbook of Pain*, 3rd Edition. Churchill Livingstone, Edinburgh, 1994, pp 997-1023.
169. Barnsley L, Lord S, Bogduk N. Comparative local anaesthetic blocks in the diagnosis of cervical zygapophysial joint pain. *Pain* 1993; 55:99-106.
170. Bogduk N. Diagnostic nerve blocks in chronic pain. *Best Pract Res Clin Anaesthesiol* 2002; 16:565-578.
171. Bogduk N. International Spinal Injection Society guidelines for the performance of spinal injection procedures. Part 1. Zygapophysial joint blocks. *Clin J Pain* 1997; 13:285-302.
172. Nath S, Nath CA, Pettersson K. Percutaneous lumbar zygapophysial (facet) joint neurotomy using radiofrequency current in the management of chronic low back pain. A randomized double-blind trial. *Spine* 2008; 33:1291-1297.
173. North RB, Han M, Zahurak M, Kidd DH. Radiofrequency lumbar facet denervation: Analysis of prognostic factors. *Pain* 1994; 57:77-83.
174. Manchikanti L, Pampati RR, Fellows B, Bakhit CE. The diagnostic validity and therapeutic value of medial branch blocks with or without adjuvants. *Curr Rev Pain* 2000; 4:337-344.
175. Manchikanti L, Damron KS, Cash KA, Manchukonda R, Pampati V. Therapeutic medial branch blocks in managing chronic neck pain: A preliminary report of a randomized, double-blind, controlled trial: Clinical Trial NCT00033272. *Pain Physician* 2006; 9:333-346.
176. Manchikanti L, Manchikanti KN, Manchukonda R, Pampati V, Cash KA. Evaluation of therapeutic thoracic medial branch block effectiveness in chronic thoracic pain: A prospective outcome study with minimum 1-year follow up. *Pain Physician* 2006; 9:97-105.
177. Manchikanti L, Manchikanti K, Manchukonda R, Cash KA, Damron KS, Pampati V, McManus CD. Evaluation of lumbar facet joint nerve blocks in the management of chronic low back pain: A preliminary report of a randomized, double-blind controlled trial. Clinical Trial NCT000355914. *Pain Physician* 2007; 10:425-440.
178. Manchikanti L, Singh V, Falco FJ, Cash KA, Pampati V. Lumbar facet joint nerve blocks in managing chronic facet joint pain: One-year follow-up of a randomized, double-blind controlled trial: Clinical Trial NCT00355914. *Pain Physician* 2008; 11:121-132.
179. Manchikanti L, Singh V, Falco FJE, Cash KA, Pampati V. Effectiveness of thoracic medial branch blocks in managing chronic pain: A preliminary report of a randomized, double-blind controlled trial: Clinical Trial NCT00355706. *Pain Physician* 2008; 11:491-504.
180. Manchikanti L, Singh V, Falco FJ, Cash KA, Fellows B. Cervical medial branch blocks for chronic cervical facet joint pain: A randomized double-blind, controlled trial with one-year follow-up. *Spine* 2008; 33:1813-1820.
181. Manchikanti L, Pampati V, Bakhit CE, Rivera JJ, Beyer CD, Damron KS, Barnhill RC. Effectiveness of lumbar facet joint nerve blocks in chronic low back pain: A randomized clinical trial. *Pain Physician* 2001; 4:101-117.
182. Manchikanti L, Manchikanti KN, Damron KS, Pampati V. Effectiveness of cervical medial branch blocks in chronic neck pain: A prospective outcome study. *Pain Physician* 2004; 7:195-201.
183. Abdi S, Datta S, Lucas LF. Role of epidural steroids in the management of chronic spinal pain: A systematic review of effectiveness and complications. *Pain Physician* 2005; 8:127-143.
184. Boswell MV, Hansen HC, Trescot AM, Hirsch JA. Epidural steroids in the management of chronic spinal pain and radiculopathy. *Pain Physician* 2003; 6:319-334.
185. Conn A, Buenaventura R, Datta S, Abdi S, Diwan S. Systematic review of caudal epidural injections in the management of chronic low back pain. *Pain Physician* 2009; 12:109-135.
186. Parr AT, Diwan S, Abdi S. Lumbar interlaminar epidural injections in managing chronic low back and lower extremity pain: A systematic review. *Pain Physician* 2009; 12:163-188.
187. Benyamin RM, Singh V, Parr AT, Conn A, Diwan S, Abdi S. Systematic review of the effectiveness of cervical epidurals in the management of chronic neck pain. *Pain Physician* 2009; 12:137-157.
188. Buenaventura RM, Datta S, Abdi S, Smith HS. Systematic review of therapeutic lumbar transforaminal epidural steroid injections. *Pain Physician* 2009; 12:233-251.
189. Abdi S, Datta S, Trescot AM, Schultz DM, Adlaka R, Atluri SL, Smith HS, Manchikanti L. Epidural steroids in the management of chronic spinal pain: A systematic review. *Pain Physician* 2007; 10:185-212.
190. Manchikanti L, Singh V. Are the results of a multicenter analysis of radiofrequency denervation success as a function of single diagnostic block reliable? *Spine J* 2008; Sep 12 [Epub ahead of print].
191. Manchikanti L. Interventional pain management: Past, present, and future. The Prithvi Raj lecture: Presented at the 4th World Congress-World Institute of Pain, Budapest, 2007. *Pain Pract* 2007; 7:357-371.
192. Manchikanti L, Cash KA, McManus CD, Pampati V, Smith HS. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 1. Discogenic pain without disc herniation or radiculitis. *Pain Physician* 2008; 11:785-800.
193. Manchikanti L, Singh V, Cash KA, Pampati V, Damron KS, Boswell MV. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 2. Disc herniation and radiculitis. *Pain Physician* 2008; 11:801-815.
194. Manchikanti L, Singh V, Cash KA, Pampati V, Datta S. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 3. Post surgery syndrome. *Pain Physician* 2008; 11:817-831.

195. Manchikanti L, Cash KA, McManus CD, Pampati V, Abdi S. Preliminary results of randomized, equivalence trial of fluoroscopic caudal epidural injections in managing chronic low back pain: Part 4. Spinal stenosis. *Pain Physician* 2008; 11:833-848.
196. Manchikanti L, Cash KA, Pampati V, McManus CD, Damron KS. Evaluation of fluoroscopically guided caudal epidural injections. *Pain Physician* 2004; 7:81-92.
197. Singh V, Manchikanti L. Role of caudal epidural injections in the management of chronic low back pain. *Pain Physician* 2002; 5:133-148.
198. Dreyfuss P, Schwarzer AC, Lau P, Bogduk N. Specificity of lumbar medial branch and L5 dorsal ramus blocks. *Spine* 1997; 22:895-902.
199. Kaplan M, Dreyfuss P, Halbrook B, Bogduk N. The ability of lumbar medial branch blocks to anesthetize the zygapophysial joint. *Spine* 1998; 23:1847-1852.
200. Barnsley L, Bogduk N. Medial branch blocks are specific for the diagnosis of cervical zygapophyseal joint pain. *Reg Anesth* 1993; 18:343-350.
201. Lord SM, Barnsley L, Bogduk N. The utility of comparative local anesthetic blocks versus placebo-controlled blocks for the diagnosis of cervical zygapophysial joint pain. *Clin J Pain* 1995; 11:208-213.
202. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, Bogduk N. The false-positive rate of uncontrolled diagnostic blocks of the lumbar zygapophysial joints. *Pain* 1994; 58:195-200.
203. Manchikanti L, Pampati V, Fellows B, Pakanati RR. Prevalence of lumbar facet joint pain in chronic low back pain. *Pain Physician* 1999; 2:59-64.
204. Manchikanti L, Pampati V, Fellows B, Baha A. The inability of the clinical picture to characterize pain from facet joints. *Pain Physician* 2000; 3:158-166.
205. Merskey H, Bogduk N. *Classification of Chronic Pain. Descriptions of Chronic Pain Syndromes and Definition of Pain Terms*, 2nd ed. Task Force on Taxonomy of the International Association for the Study of Pain. IASP Press, Seattle, 1994.
205. Manchikanti L, Damron KS, Rivera J, McManus C, Jackson S, Barnhill R, Martin J. Evaluation of effect of sedation as a confounding factor in the diagnostic validity of lumbar facet joint pain: A prospective, randomized, double-blind, placebo-controlled evaluation. *Pain Physician* 2004; 7:411-417.
207. Manchikanti L, Pampati V, Damron KS, McManus CD, Jackson SD, Barnhill RC, Martin JC. The effect of sedation on diagnostic validity of facet joint nerve blocks: An evaluation to assess similarities in population with involvement in cervical and lumbar regions. *Pain Physician* 2006; 9:47-52.
208. Manchikanti L, Pampati V, Damron KS, McManus CD, Jackson SD, Barnhill RC, Martin JC. A randomized, prospective, double-blind, placebo-controlled evaluation of the effect of sedation on diagnostic validity of cervical facet joint pain. *Pain Physician* 2004; 7:301-309.
209. Manchikanti L, Pampati V, Fellows B, Rivera J, Damron K, Beyer C, Cash K. Influence of psychological factors on the ability of diagnose chronic low back pain of facet joint origin. *Pain Physician* 2001; 4:349-357.
210. Manchikanti L, Cash KA, Pampati V, Fellows B. Influence of psychological variables on the diagnosis of facet joint involvement in chronic spinal pain. *Pain Physician* 2008; 11:145-160.
211. Manchikanti L, Singh V, Fellows B, Pampati V. Evaluation of influence of gender, occupational injury, and smoking on chronic low back pain of facet joint origin: A subgroup analysis. *Pain Physician* 2002; 5:30-35.
212. Manchikanti L, Manchikanti K, Cash KA, Singh V, Giordano J. Age-related prevalence of facet joint involvement in chronic neck and low back pain. *Pain Physician* 2008; 11:67-75.
213. Manchikanti L, Manchikanti KN, Pampati V, Cash KA. Prevalence of side effects of prolonged low or moderate dose opioid therapy with concomitant benzodiazepine and/or antidepressant therapy in chronic non-cancer pain. *Pain Physician* 2009; 12:259-267.
214. Walsh TR, Weinstein JN, Spratt KF, Lehmann TR, Aprill C, Sayre H. Lumbar discography in normal subjects. A controlled prospective study. *J Bone Joint Surg Am* 1990; 72:1081-1088.
215. Derby R, Lee SH, Kim BJ, Chen Y, Aprill C, Bogduk N. Pressure-controlled lumbar discography in volunteers without low back symptoms. *Pain Med* 2005; 6:213-221.
216. Shin D, Kim H, Jung J, Sin D, Lee J. Diagnostic relevance of pressure-controlled discography. *J Korean Med Sci* 2006; 21:911-916.
217. Schellhas KP, Pollei SR, Dorwart RH. Thoracic discography. A safe and reliable technique. *Spine* 1994; 19:2103-2109.
218. Wood KB, Schellhas KP, Garvey TA, Aeppli D. Thoracic discography in healthy individuals. A controlled prospective study of magnetic resonance imaging and discography in asymptomatic and symptomatic individuals. *Spine* 1999; 24:1548-1555.
219. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *JAMA* 1997; 277:927-934.
220. Egger M, Jüni P, Bartlett C; CONSORT Group (Consolidated Standards of Reporting of Trials). Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001; 285:1996-1999.
221. Knottnerus JA. The effects of disease verification and referral on the relationship between symptoms and diseases. *Med Decis Making* 1987; 7:139-148.
222. Panzer RJ, Suchman AL, Griner PF. Workup bias in prediction research. *Med Decis Making* 1987; 7:115-119.
223. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987; 6:411-423.
224. Philbrick JT, Horwitz RI, Feinstein AR. Methodologic problems of exercise testing for coronary artery disease: Groups, analysis and bias. *Am J Cardiol* 1980; 46:807-812.
225. Brealey S, Scally AJ, Thomas NB. Review article: Methodological standards in radiographer plain film reading performance studies. *Br J Radiol* 2002; 75:107-113.
226. Elmore JG, Feinstein AR. A bibliography of publications on observer variability (final installment). *J Clin Epidemiol* 1992; 45:567-580.
227. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *N Engl J Med* 1994; 331:1493-1499.
228. Brealey S, Scally AJ. Bias in plain film reading performance studies. *Br J Radiol* 2001; 74:307-316.
229. Quinn MF. Relation of observer agreement to accuracy according to a two-receiver signal detection model of diagnosis. *Med Decis Making* 1989; 9:196-206.
230. *Annals of Internal Medicine*. Information for authors. Available at www.annals.org.