Background: A major component of a systematic review is an assessment of the methodological quality and bias of randomized trials. The most commonly utilized methodological quality assessment and bias assessment for randomized trials is by the Cochrane Review Group. While this is not a “gold standard,” it is an indication of the current state-of-the-art review methodology. There is, however, no specific instrument to assess the methodological quality of manuscripts published for interventional techniques.

Objectives: Our objective was to develop an instrument specifically for interventional pain management, to assess the methodological quality of randomized trials of interventional techniques.

Methods: Item generation for the instrument was based on a definition of quality, to the extent to which the design and conduct of the trial were congruent with the objectives of the trial. Applicability was defined as the extent to which the trial produced procedures could be applied with contemporary interventional pain management techniques. Multiple items based on Cochrane review criteria were utilized along with specific requirements for interventional techniques.

Results: A total of 22 items were developed which formed IPM-QRB or Interventional Pain Management Techniques – Quality Appraisal of Reliability and Risk of Bias Assessment tool. This included 9 of the 12 items from the Cochrane review criteria with deletion of some items that were repetitive or duplicate, and the addition of 13 new items.

The results were compared for inter-rater reliability of Cochrane review criteria and IPM-QRB, and inter-instrument reliability.

The assessment was performed in multiple stages with an initial learning curve. The final assessment was for 4 randomized controlled trials (RCTs) utilizing both Cochrane review criteria and IPM-QRB criteria. The inter-rater agreement for Cochrane review criteria with overall intra-class correlation coefficient was 0.407 compared to an intra-class correlation coefficient of 0.833 for IPM-QRB criteria. The inter-rater agreement worse for Cochrane review criteria despite twice the number of items, and the addition of 13 new items.

Limitations: Limited validity or accuracy assessment of the instrument and the large number of items to be scored.

Conclusion: We have developed a new comprehensive instrument to assess the methodological quality of randomized trials of interventional techniques. This instrument is superior to Cochrane review methodology criteria in that it provides more extensive and specific information for interventional techniques that will be useful in assessing the methodological quality and bias of interventional techniques.

Key words: Methodological quality assessment, evidence-based medicine, comparative effectiveness research, Cochrane Reviews, interventional techniques, risk of bias assessment.

Pain Physician 2014; 17:E263-E290
Health care research, practice, and policy focus on improving the organization, delivery, and outcomes of care (1,2). These objectives are achieved by the appropriate development of guidelines based on currently available knowledge generated through research in combination with professional experience and consideration differences between individual patients (1-13). Evidence synthesis and the development of guidelines through systematic reviews is a dynamic process, and has resulted in the Institute of Medicine (IOM) re-engineering its recommendations for the development of clinical guidelines and systematic reviews in 2011 (12,13). IOM redefined clinical practice guidelines as, “statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options.” IOM (13) also described the function and purpose of a systematic review as, “to identify, select, assess, and synthesize the findings of similar but separate studies and to help clarify what is known and not known about the potential benefits and harms of drugs, devices, and other health care services.” Systematic reviews can be helpful for clinicians who want to integrate research findings into their daily practices, for patients to make well-informed choices about their own care, and for professional medical societies and other organizations that develop clinical practice guidelines. However, a number of challenges have arisen in implementing the IOM systematic review standards. Chang et al (14) showed that these standards based on a mix of theoretical principles, empirical evidence, and commonly considered best practices, set a high bar for authors of systematic reviews. They also showed that based on over 15 years of experience conducting systematic reviews, the Agency for Healthcare Research and Quality (AHRQ) – Evidence-Based Practice Center (EPC) program has examined the EPCs adherence to and agreement with the IOM standards. Even a program as large such as AHRQ, with a large infrastructure as well as resource and support from the government, found challenges in implementing all of the IOM standards. Young and Greenberg (15) also showed that the IOM failed to follow its own standards for 4 out of 8, partially followed 2 of 8, and fully complied with only 2 standards.

Multiple manuscripts have been published about conducting systematic reviews and assessment of the methodological quality and risk of bias of the included studies (16-32). The Quality of Reporting of Meta-analyses (QUOROM) statement was developed in 1999 to improve standards for the reporting of systematic reviews (33). Almost all journals have adapted the QUOROM recommendations for the reporting of various criteria for systematic reviews. Subsequently, as an update to QUOROM, Preferring Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) was developed and subsequently adopted by leading journals (34). Even with QUOROM and the transformation to PRISMA, many reviews in the field of spinal pain, specifically interventional pain management, are of low methodological quality and lack essential components (1,4,22-42). Quality assessment of trials has been used to determine a minimum quality threshold for the selection of primary studies for systematic reviews, and assist in determining the strength of inferences (43). It has repeatedly been shown that the different design features of a trial can have a substantial impact on estimates of treatment effects (44-48). The method of randomization, inadequate allocation concealment, follow-up proportions, and industry sponsorship have all been shown to influence the results of trials and may lead to biased or inaccurate results and conclusions in systematic reviews and meta-analysis (43-55). The impact of primary trial bias on evidence synthesis has been recognized for years and the approach to quality assessment has been inconsistent and controversial (56). Many tools including those of Jadad, Chalmers, Consolidated Standards of Reporting Trials (CONSORT), Delphi List, and Cochrane review criteria have been used to determine the quality of randomized controlled trials (RCTs) in different health areas (16-18,20,21,43,56). Thus, there is no agreement regarding which tools are optimal to accurately determine trial quality. Most tools have not been developed using scientifically rigorous methods, lack of reliability, and/or have not been fully validated including those of Cochrane Review (43,56). In fact, the use of different tools for evaluating the quality of primary research can lead to different end results (43,57-59). Thus, a clinical trial may be rated on a quality scale disparate by different measurement tools. Discrepancies in an evaluation of the quality of research may skew interpretation, reporting, and as a result, could potentially impact recommendations for clinical care. Furthermore, an understanding of the rater and the bias embedded either intended or unintended, may also exert a significant influence on the final evidence and recommendations. Finally, the tools include different items, some of which relate more to the detail of...
reporting rather than methodological quality.

As a result of these shortcomings with existing tools and methods for quality assessment, there has been a shift in the traditional scoring approach to the assessment of trial quality. To address this, Cochrane Back Review Group (CBRG) Editorial Board published method guidelines for systematic reviews in the field of spinal disorders in 1997 (41). The 1997 guidelines were updated in 2003 to address the main steps in conducting a systematic review including literature search, inclusion criteria, assessing methodological quality, data extraction, and data analysis (42). In 2009, Cochrane method guidelines for systematic reviews were further updated (21). These updated guidelines included various recommendations divided into 7 categories: objectives, literature search, inclusion criteria, risk of bias assessment, data extraction, data analysis, and updating the review. They classified each recommendation into minimum criteria (mandatory) and further guidance (optional). Even then, most of the empirical evidence regarding the relationship between trial components and treatment effect estimates comes from RCTs in the area of medicine and is based only on dichotomous outcomes (45,47,48).

There have not been studies conducted in other health areas such as interventional pain management. When compared to drug trials conducted in medicine, physical therapy or surgery, RCTs conducted in interventional pain management have unique features. Interventional pain management is classified as an evolving specialty and is comprised of diverse facets that may affect trial results, such as treatment setting and physician performing the procedure, understanding of the placebo reactions, the trial design whether it is active control or placebo control and misinterpretation of active controls as placebo controls, and blinding of the physician and/or the patients may not always be possible. Consequently, it is necessary that empirical evidence be expanded in the area of interventional pain management in order to determine the factors that affect treatments and estimates in these trials, but in order to provide accurate results and recommendations as well. Due to an explosion of the number of systematic reviews and guidelines, it is not only mandatory, but urgent to develop methodology for assessment of various types of trials so that clear and concrete evidence of interventional techniques without misinterpretation and misappropriation can be provided.

Marin et al (22), along with other coauthors and the editorial board of the CBRG, summarized 15 years of the CBRG activities at the marking of the twentieth anniversary of the Cochrane Collaboration. They covered a broad range of interventions from conservative therapy such as exercise and massage, to invasive procedures such as disc replacement and interventional techniques. This review (22) showed that there have been 62 Cochrane reviews devoted to these topics currently published in the Cochrane library. CBRG also published 56 additional articles in independent medical journals. A Cochrane review of 2013 showed that there have been a total of 5,804 full reviews and 2,386 protocols for reviews in production (60). The majority of them, or 81%, examined nonpharmacological treatments for back and neck pain. Instead of recommending levels of evidence, CBRG adapted the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach to determine the overall quality of evidence for important patient-centered outcomes (61).

Furlan et al (21) updated method guidelines for systematic reviews and were very emphatic in their preparation of these guidelines that they should not be construed as a “gold standard,” but rather as an indication of current state-of-the-art review methods. These guidelines should be used to plan, conduct, or evaluate systematic reviews in the field of spinal pain within and outside the framework of the Cochrane reviews. The Cochrane review criteria remains the most widely used and respected criteria, but they are not devoid of criticism (1,22-35,38-42,62,63). An assessment of Cochrane review criteria showed low agreement between 2 reviewers in a recent manuscript (64).

In addition to methods guidelines for systematic reviews, it has been recommended that randomized trials also be conducted in accordance with rigorous criteria designed by consensus, such as those established by CONSORT - Consolidated Standards of Reporting Trials guidelines (65-67). In the early 1990s, journal editors, investigators, and methodologists independently published recommendations on the reporting of trials (68,69). The CONSORT statement was the consolidation of these recommendations (70,71). Guidelines and quality assessment scales have also been published for diagnostic and observational studies (11,48,72-77). The CONSORT guidelines were widely adapted by leading journals because the scientific community of medicine depends on the transparent reporting of clinical trials to avoid bias and provide evidence for effectiveness (9,22-26,29,32,48,67,78-80). The CONSORT statement comprises a checklist of items that should be included in reports of RCTs, along with a diagram for document-
ing the flow of participants in a trial. As described by CONSORT, biased results from poorly designed trials can mislead decision-making in health care at all levels, including the formulation of national public health policies.

In assessing the reporting and methodological quality of systematic reviews in the orthopedic literature, Gagnier and Kellam (32) reviewed 76 systematic reviews and meta-analysis and arrived at the conclusion that reporting and methodological quality in the top 5 orthopedic journals was poor. In view of the fact that the clinical relevance and generalization of published orthopedic systematic reviews appears to be questionable and their contribution to clinical decision-making suboptimal, the authors cautioned that clinicians should be careful when interpreting the findings (79). The use of PRISMA and Assessment of Multiple Systematic Reviews (AMSTAR) guidelines in designing, implementing, and writing systematic reviews was recommended as a way to improve the quality of systematic reviews and meta-analyses in orthopedic journals (34,81).

These attempts to improve the quality of the reporting of RCTs are what led to the development and widespread acceptance of CONSORT guidelines (68-71). Most of CONSORT is also relevant to a wider class of study designs, such as non-inferiority, equivalence, factorial, cluster, and crossover trials.

Similar to CONSORT, Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) was developed in order to improve the quality of clinical trial protocols by defining an evidence-based set of items to address in a protocol (9,23). The authors of the SPIRIT 2013 statement (11) asserted that existing guidelines for protocol content vary greatly in their scope, seldom describe how they were developed, and rarely provide full disclosure or empirical evidence to support their recommendations (82). The problems that underlie protocol deficiencies can lead to avoidable protocol amendments, poor trial conduct, and inadequate reporting in publications (83-85). In response to these shortcomings, SPIRIT was initiated in 2007 to improve the completeness of trial protocols by producing evidence-based recommendations for a minimum set of items to be included. The SPIRIT 2013 statement includes a 33-item checklist and a diagram.

Deficiencies persist despite the recent proliferation of guidelines, specifically with regard to the lack of individualized criteria regarding the selection of patients and the conduct of procedures. Consequently, the assessment of interventional techniques may require additional items to properly evaluate the quality of study performance, outcomes assessment, and risk of bias. In a recent manuscript, Bicket et al (86) have described a separate instrument for quality assessment for epidural injections. This instrument has not been widely utilized due to its recent publication and limitations. However, their analysis showed some significant variations compared to Cochrane review criteria.

Multiple factors influence the quality of trials and assessments; yet, appropriately rating studies is critical to our assessment of efficacy for interventional procedures. This instrument is intended to provide investigators with the capability of rating interventional studies in pain medicine.

Our objective is to develop a unique tool for assessment of the methodological quality of randomized trials for interventional techniques with a modification of the Cochrane assessment tool and the addition of multiple other items for the American Society of Interventional Pain Physicians (ASIPP).

**Methods**

In this manuscript, we report the development of Interventional Pain Management techniques Quality Appraisal of Reliability and Risk of Bias Assessment -- IPM-QRB for use for systematic reviews of interventional pain management techniques. Methodologic quality has been defined as “the extent to which all aspects of the study’s design and conduct can be shown to protect against systematic bias, nonsystematic bias, and inferential error” (20). For purposes of this instrument, we hold quality to the extent to which a study’s design, conduct, and analysis have minimized selection, measurement, and confounding biases with assessment of specific requirements for interventional techniques. Methodological procedures have been described in the design of quality appraisal tools (17,74,75). Generally, quality appraisal tools have used a numeric scoring system to rate individual studies with an overall quality score (16,19). However, there is ongoing discussion in relation to the importance of each item and weighing of the scores (48,58,87-90). For example, Cochrane methodological review criteria or risk of bias assessment has changed over the years, with and without numeric scoring systems for each item as shown in Table 1 (21,22,33,38,41,42,83-86,91-99). However, the analysis based on Cochrane review criteria has been appropriately used with multiple modifications and application of self-impressions, only to result in inappropriate conclusions and recommendations (1,2,4,35,39,99-106).
Methodologic Quality Assessment of Randomized Trials of Interventional Techniques

For the development of the present interventional scoring system, we sought to use a numeric scoring for each individual item, which when combined, results in an overall quality score. Item generation was based on the congruence between the design and conduct of the trial and the study’s objectives, as well as the perceived extent to which the trial procedures could be applied to interventional pain management techniques.

The investigators (LM and JH) produced a list of 30 individual items designed to investigate each of the principles. After discussions with 2 other investigators (RB and SC), they were reduced to 25. Each item was explained with references for justification. This list was circulated to the reference group with instructions to indicate if each item should remain on the list and to decide on appropriate numeric scoring for each item. The reviewers received basic written instructions for the instrument. Each item on the checklist can be rated as “yes,” “no,” or “unclear” and certain items can be rated as not applicable. Responses from the group were collated and discussed via the internet and at a guidelines meeting in person with the reviewers participating in the assessment and additional authors involved in guideline preparation. Two of the reviewers wanted more detail on each item in the questionnaire itself; however, the majority agreed the explanations were adequate.

While the majority of the issues were resolved without conflict, the language in reference to outcomes assessment was a significant issue to one of the authors (SC). He was also supported by another author (HH). They both believed that providing a cutoff with outcomes was inappropriate and it has to be in a different format. SC also stated that he consulted with a neurologist providing advice to the Food and Drug Administration (FDA) who also did not agree with the present outcomes assessment determination. However, the remaining authors, including the guideline group, understood that the outcomes we are assessing are appropriate and fit the clinical criteria with ability to determine significant outcomes. It was also understood that functional outcome and pain ratings may be measured with various instruments. There was also discussion in reference to the conflicts of interest and disclosure of the information. Thus, all the conflicts were resolved with over 90% agreement. Subsequently, the number of items was changed to 24 and, finally, to 22, which was subjected to analysis.

Analysis of Data

Each response option was recorded as a category, including unclear and not applicable. Data were analyzed for intraclass correlation coefficient (ICC) derived from a 2-way random model with absolute agreement. An ICC is measured on a scale of 0 to 1, 1 presents perfect reliability, whereas 0 indicates no reliability (100). The kappa statistic, sensitivity, specificity, and predictive positive and negative values were calculated to mea-
sure agreement. Kappa (101) is a chance corrected measure of inter-rater reliability and ranges from minus 1 to plus 1, with plus 1 being perfect agreement, minus 1 being perfect disagreement, and 0 being agreement no better than chance. In this study, kappa was interpreted as unreliable ($\kappa < 0.00$), poor ($\kappa = 0.01 – 0.20$), fair ($\kappa = 0.21 – 0.40$), moderate ($\kappa = 0.41 – 0.60$), good ($\kappa = 0.61 – 0.80$), and very good ($\kappa = 0.81 – 1.00$). A 95% confidence interval for kappa was computed using the test-based standard error. For this study, reliability was considered acceptable if it was moderate or higher. Comparison was conducted between Cochrane review criteria and IPM-QRB. Correlation of inter-rater reliability was assessed for both. All computations were performed using SPSS statistical software version 22 (IBM, New York, NY, USA).

**Results**

The final list of items as shown in Table 2 included 9 of 12 items from Cochrane review criteria (21) and 13

<table>
<thead>
<tr>
<th>I. CONSORT OR SPIRIT</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Trial Design Guidance and Reporting</td>
<td></td>
</tr>
<tr>
<td>Trial designed and reported without any guidance</td>
<td>0</td>
</tr>
<tr>
<td>Trial designed and reported utilizing minimum criteria other than CONSORT or SPIRIT criteria or trial was conducted prior to 2005</td>
<td>1</td>
</tr>
<tr>
<td>Trial implies it was based on CONSORT or SPIRIT without clear description with moderately significant criteria for randomized trials or the trial was conducted before 2005</td>
<td>2</td>
</tr>
<tr>
<td>Explicit use of CONSORT or SPIRIT with identification of criteria or trial conducted with high level reporting and criteria or conducted before 2005</td>
<td>3</td>
</tr>
</tbody>
</table>

II. DESIGN FACTORS

<table>
<thead>
<tr>
<th>II. DESIGN FACTORS</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Type and Design of Trial</td>
<td></td>
</tr>
<tr>
<td>Poorly designed control group (quasi selection, convenient sampling)</td>
<td>0</td>
</tr>
<tr>
<td>Proper active-control or sham procedure with injection of active agent</td>
<td>2</td>
</tr>
<tr>
<td>Proper placebo control (no active solutions into active structures)</td>
<td>3</td>
</tr>
<tr>
<td>3. Setting/Physician</td>
<td></td>
</tr>
<tr>
<td>General setting with no specialty affiliation and general physician</td>
<td>0</td>
</tr>
<tr>
<td>Specialty of anesthesia/PMR/neurology/radiology/ortho, etc.</td>
<td>1</td>
</tr>
<tr>
<td>Interventional pain management with interventional pain management physician</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. PATIENT FACTORS</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Inclusiveness of Population</td>
<td></td>
</tr>
<tr>
<td>7a. For epidural procedures:</td>
<td></td>
</tr>
<tr>
<td>Poorly identified mixed population</td>
<td>0</td>
</tr>
</tbody>
</table>
### Table 2 (cont.). Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM-QRB.

<table>
<thead>
<tr>
<th>Item</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clearly identified mixed population</td>
<td>1</td>
</tr>
<tr>
<td>Disorders specific trials (i.e. well defined spinal stenosis and disc herniation, disorder specific, disc herniation or spinal stenosis or post surgery syndrome)</td>
<td>2</td>
</tr>
<tr>
<td>7b. For facet or sacroiliac joint interventions:</td>
<td></td>
</tr>
<tr>
<td>No diagnostic blocks</td>
<td>0</td>
</tr>
<tr>
<td>Selection with single diagnostic blocks</td>
<td>1</td>
</tr>
<tr>
<td>Selection with placebo or dual diagnostic blocks</td>
<td>2</td>
</tr>
<tr>
<td>8. Duration of Pain</td>
<td></td>
</tr>
<tr>
<td>Less than 3 months</td>
<td>0</td>
</tr>
<tr>
<td>3 to 6 months</td>
<td>1</td>
</tr>
<tr>
<td>&gt; 6 months</td>
<td>2</td>
</tr>
<tr>
<td>9. Previous Treatments</td>
<td></td>
</tr>
<tr>
<td>Conservative management including drug therapy, exercise therapy, physical therapy, etc.</td>
<td></td>
</tr>
<tr>
<td>Were not utilized</td>
<td>0</td>
</tr>
<tr>
<td>Were utilized sporadically in some patients</td>
<td>1</td>
</tr>
<tr>
<td>Were utilized in all patients</td>
<td>2</td>
</tr>
<tr>
<td>10. Duration of Follow-up with Appropriate Interventions</td>
<td></td>
</tr>
<tr>
<td>Less than 3 months or 12 weeks for epidural or facet joint procedures, etc. and 6 months for intradiscal procedures and implantables</td>
<td>0</td>
</tr>
<tr>
<td>3 to 6 months for epidural or facet joint procedures, etc., or 1 year for intradiscal procedures or implantables</td>
<td>1</td>
</tr>
<tr>
<td>6 months to 17 months for epidurals or facet joint procedures, etc., and 2 years or longer for discal procedures and implantables</td>
<td>2</td>
</tr>
<tr>
<td>18 months or longer for epidurals and facet joint procedures, etc., or 5 years or longer for discal procedures and implantables</td>
<td>3</td>
</tr>
<tr>
<td>IV. OUTCOMES</td>
<td></td>
</tr>
<tr>
<td>11. Outcomes Assessment Criteria for Significant Improvement</td>
<td></td>
</tr>
<tr>
<td>No descriptions of outcomes OR &lt; 20% change in pain rating or functional status</td>
<td>0</td>
</tr>
<tr>
<td>Pain rating with a decrease of 2 or more points or more than 20% reduction OR functional status improvement of more than 20%</td>
<td>1</td>
</tr>
<tr>
<td>Pain rating with decrease of ≥ 2 points AND ≥ 20% change or functional status improvement of ≥ 20%</td>
<td>2</td>
</tr>
<tr>
<td>Pain rating with a decrease of 3 or more points or more than 50% reduction OR functional status improvement with a 50% or 40% reduction in disability score</td>
<td>2</td>
</tr>
<tr>
<td>Significant improvement with pain and function ≥ 50% or 3 points and 40% reduction in disability scores</td>
<td>4</td>
</tr>
<tr>
<td>12. Analysis of all Randomized Participants in the Groups</td>
<td></td>
</tr>
<tr>
<td>Not performed</td>
<td>0</td>
</tr>
<tr>
<td>Performed without intent-to-treat analysis without inclusion of all randomized participants</td>
<td>1</td>
</tr>
<tr>
<td>All participants included with or without intent-to-treat analysis</td>
<td>2</td>
</tr>
<tr>
<td>13. Description of Drop Out Rate</td>
<td></td>
</tr>
<tr>
<td>No description of dropouts, despite reporting of incomplete data or ≥ 20% withdrawal</td>
<td>0</td>
</tr>
<tr>
<td>Less than 20% withdrawal in one year in any group</td>
<td>1</td>
</tr>
<tr>
<td>Less than 30% withdrawal at 2 years in any group</td>
<td>2</td>
</tr>
<tr>
<td>14. Similarity of Groups at Baseline for Important Prognostic Indicators</td>
<td></td>
</tr>
<tr>
<td>Groups dissimilar with significant influence on outcomes with or without appropriate randomization and allocation</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 2 (cont.). Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM-QRB.

<table>
<thead>
<tr>
<th>Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM-QRB.</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups dissimilar without influence on outcomes despite appropriate randomization and allocation</td>
<td>1</td>
</tr>
<tr>
<td>Groups similar with appropriate randomization and allocation</td>
<td>2</td>
</tr>
<tr>
<td><strong>15. Role of Co-Interventions</strong></td>
<td></td>
</tr>
<tr>
<td>Co-interventions were provided but were not similar in the majority of participants</td>
<td>0</td>
</tr>
<tr>
<td>No co-interventions or similar co-interventions were provided in the majority of the participants</td>
<td>1</td>
</tr>
</tbody>
</table>

**VI. RANDOMIZATION**

<table>
<thead>
<tr>
<th>Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM-QRB.</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>16. Method of Randomization</strong></td>
<td></td>
</tr>
<tr>
<td>Quasi randomized or poorly randomized or not described</td>
<td>0</td>
</tr>
<tr>
<td>Adequate randomization (coin toss, drawing of balls of different colors, drawing of ballots)</td>
<td>1</td>
</tr>
<tr>
<td>High quality randomization (computer generated random sequence, pre-ordered sealed envelopes, sequentially ordered vials, telephone call, pre-ordered list of treatment assignments, etc.)</td>
<td>2</td>
</tr>
</tbody>
</table>

**VI. ALLOCATION CONCEALMENT**

<table>
<thead>
<tr>
<th>Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM-QRB.</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>17. Concealed Treatment Allocation</strong></td>
<td></td>
</tr>
<tr>
<td>Poor concealment of allocation (open enrollment) or inadequate description of concealment</td>
<td>0</td>
</tr>
<tr>
<td>Concealment of allocation with borderline or good description of the process with probability of failure of concealment</td>
<td>1</td>
</tr>
<tr>
<td>High quality concealment with strict controls (independent assignment without influence on the assignment sequence)</td>
<td>2</td>
</tr>
</tbody>
</table>

**VII. BLINDING**

<table>
<thead>
<tr>
<th>Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM-QRB.</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>18. Patient Blinding</strong></td>
<td></td>
</tr>
<tr>
<td>Patients not blinded</td>
<td>0</td>
</tr>
<tr>
<td>Patients blinded adequately</td>
<td>1</td>
</tr>
<tr>
<td><strong>19. Care Provider Blinding</strong></td>
<td></td>
</tr>
<tr>
<td>Care provider not blinded</td>
<td>0</td>
</tr>
<tr>
<td>Care provider blinded adequately</td>
<td>1</td>
</tr>
<tr>
<td><strong>20. Outcome Assessor Blinding</strong></td>
<td></td>
</tr>
<tr>
<td>Outcome assessor not blinded or was able to identify the groups</td>
<td>0</td>
</tr>
<tr>
<td>Performed by a blinded independent assessor with inability to identify the assignment-based provider intervention (i.e., subcutaneous injection, intramuscular distant injection, difference in preparation or equipment use, numbness and weakness, etc.)</td>
<td>1</td>
</tr>
</tbody>
</table>

**VIII. CONFLICTS OF INTEREST**

<table>
<thead>
<tr>
<th>Item checklist for assessment of randomized controlled trials of IPM techniques utilizing IPM-QRB.</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>21. Funding and Sponsorship</strong></td>
<td></td>
</tr>
<tr>
<td>Trial included industry employees</td>
<td>-3</td>
</tr>
<tr>
<td>Industry employees involved; high levels of funding with remunerations by industry or an organization funded with conflicts</td>
<td>-3</td>
</tr>
<tr>
<td>Industry or organizational funding with reimbursement of expenses with some involvement</td>
<td>0</td>
</tr>
<tr>
<td>Industry or organization funding of expenses without involvement</td>
<td>1</td>
</tr>
<tr>
<td>Funding by internal resources only with supporting entity unrelated to industry</td>
<td>2</td>
</tr>
<tr>
<td>Governmental funding without conflict such as NIH, NHS, AHRQ</td>
<td>3</td>
</tr>
<tr>
<td><strong>22. Conflicts of Interest</strong></td>
<td></td>
</tr>
<tr>
<td>None disclosed with potential implied conflict</td>
<td>0</td>
</tr>
<tr>
<td>Marginally disclosed with potential conflict</td>
<td>1</td>
</tr>
<tr>
<td>Well disclosed with minor conflicts</td>
<td>2</td>
</tr>
<tr>
<td>Well disclosed with no conflicts</td>
<td>3</td>
</tr>
<tr>
<td>Hidden conflicts with poor disclosure</td>
<td>-1</td>
</tr>
<tr>
<td>Misleading disclosure with conflicts</td>
<td>-2</td>
</tr>
<tr>
<td>Major impact related to conflicts</td>
<td>-3</td>
</tr>
</tbody>
</table>

**TOTAL MAXIMUM** | 48       |
new items. Cochrane review criteria had some duplicity with a certain number of items. Consequently, only 9 of the 12 items were utilized. These 22 items were separated into 8 main categories, with multiple subcategories.

To assess the ability of the various reviewers to quantify the methodologic assessment, initially 4 manuscripts were assigned to 4 authors blindly, for assessment with Cochrane review criteria and IPM-QRB criteria (102-105). There was poor correlation among the inter-rater reliability of both instruments; however, there was good inter-instrument reliability.

Following this, with additional discussions, 4 additional manuscripts (106-109) were sent to all the reviewers for scoring. The scores of these manuscripts were considered as final. The IPM-QRB instrument also was fine tuned with some alterations; however, all the items remained intact. The results of inter-rater agreement and inter-instrument agreement are detailed below.

**Overall Rating of the Quality**

The data from 16 reviews were collated showing inter-rater correlation criteria in Table 3 for Cochrane review criteria and inter-rater correlation criteria for IPM-QRB in Table 4.

Manuscripts were rated based on both instruments as high quality if they achieved the scores of 8 or more of 12 on Cochrane review criteria and 32 of 48 or more on IPM-QRB criteria assessment. The scores were utilized only if they agreed with assessment by 3 authors, with consensus.

**Overall Inter-Rater Reliability**

Table 3 shows inter-reliability of Cochrane review criteria with 85% correlation, another item with 95% correlation, and the remaining with 100% correlation for the first blinded manuscript (109). The second manuscript by Wilson-MacDonald et al (108) was published in 2005, even though the study was completed in 1999. The manuscript showed 100% correlation above 65%; however, above 80% correlation, there were only 25% of the items meeting this level of criteria. Overall this manuscript met moderate correlation criteria.

### Table 3. Inter-rater agreement of Cochrane review criteria.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the method of randomization adequate?</td>
<td>100%</td>
<td>100%</td>
<td>75%</td>
<td>81%</td>
<td>89%</td>
</tr>
<tr>
<td>2. Was the treatment allocation concealed?</td>
<td>100%</td>
<td>94%</td>
<td>81%</td>
<td>81%</td>
<td>89%</td>
</tr>
<tr>
<td>3. Was the patient blinded to the intervention?</td>
<td>100%</td>
<td>75%</td>
<td>81%</td>
<td>81%</td>
<td>84%</td>
</tr>
<tr>
<td>4. Was the care provider blinded to the intervention?</td>
<td>100%</td>
<td>50%</td>
<td>75%</td>
<td>100%</td>
<td>81%</td>
</tr>
<tr>
<td>5. Was the outcome assessor blinded to the intervention?</td>
<td>75%</td>
<td>69%</td>
<td>56%</td>
<td>50%</td>
<td>63%</td>
</tr>
<tr>
<td>6. Was the drop-out rate described and acceptable?</td>
<td>100%</td>
<td>69%</td>
<td>94%</td>
<td>88%</td>
<td>88%</td>
</tr>
<tr>
<td>7. Were all randomized participants analysed in the group to which they were allocated?</td>
<td>100%</td>
<td>63%</td>
<td>88%</td>
<td>94%</td>
<td>86%</td>
</tr>
<tr>
<td>8. Are reports of the study free of suggestion of selective outcome reporting?</td>
<td>100%</td>
<td>75%</td>
<td>88%</td>
<td>75%</td>
<td>84%</td>
</tr>
<tr>
<td>9. Were the groups similar at baseline regarding the most important prognostic indicators?</td>
<td>31%</td>
<td>75%</td>
<td>100%</td>
<td>100%</td>
<td>76%</td>
</tr>
<tr>
<td>10. Were co-interventions avoided or similar?</td>
<td>81%</td>
<td>75%</td>
<td>44%</td>
<td>81%</td>
<td>70%</td>
</tr>
<tr>
<td>11. Was the compliance acceptable in all groups?</td>
<td>100%</td>
<td>69%</td>
<td>94%</td>
<td>88%</td>
<td>88%</td>
</tr>
<tr>
<td>12. Was the timing of the outcome assessment similar in all groups?</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement for score of &lt; 8</td>
<td>0%</td>
<td>19%</td>
<td>19%</td>
<td>37%</td>
<td>19%</td>
</tr>
<tr>
<td>Agreement for score of ≥ 8</td>
<td>100%</td>
<td>81%</td>
<td>81%</td>
<td>63%</td>
<td>81%</td>
</tr>
<tr>
<td>Agreement for &gt; 60% of items</td>
<td>92% (11)</td>
<td>92% (11)</td>
<td>83% (10)</td>
<td>92% (11)</td>
<td>100%</td>
</tr>
<tr>
<td>Agreement for at least 80% of items</td>
<td>83% (10)</td>
<td>25% (3)</td>
<td>67% (8)</td>
<td>83% (10)</td>
<td>92%</td>
</tr>
<tr>
<td>Intra-class correlation coefficient (single reviewer Absolute Agreement)</td>
<td>0.407</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(95% CI 0.144 – 0.911)
The third manuscript, by Carette et al (106) of facet joint injections, was published in 1991 of a trial performed from 1987 to 1989. This manuscript correlated with 83% of the reviewers’ assessment for 60% of the items. At above 80% items correlation, only 67% criteria were met.

The fourth manuscript by Ackerman and Ahmad (107) met 92% reliability criteria with above 60 item inter-rater reliability and 83% with inclusion of 80% criteria.

Overall the intra-class correlation coefficient was 0.407 (95% CI; 0.144 – 0.911).

Table 4 shows inter-rater reliability of IPM-QRB criteria. The manuscript by Manchikanti et al (109) showed 91% correlation with over 80% criteria rating them in the same direction. In contrast, the manuscript by Wilson-MacDonald et al (108), above 80% correlation was detected for only 36% of the items, whereas above 60% correlation was detected for two-thirds of the items, or for only 36% of the items, with no significant changes noted between 60% or 80% criteria acceptance.

Similarly, for Carette et al (106), 80% of the reviewers had the same opinion for 32% of the items, whereas...
64% had a similar rating with 14 items correlating by more than 60% of the reviewers. Finally, for Ackerman and Ahmad (107), 59% of the items met the criteria with 80% of the reviewers correlating. There was 73% agreement when rating agreement of 60% or greater was utilized as a standard.

The overall intra-class correlation coefficient was 0.833 (95% CI; 0.592 – 0.986), significantly higher than the Cochrane review coefficient of 0.407 (95% CI; 0.144 – 0.911). Inter-relationship criteria between the 2 instruments showed Manchikanti et al (109) was rated as high-quality with both instruments with scores above 8 of 12 on Cochrane review criteria and at least 32 of 48 on IPM-QRB criteria. However, Wilson-MacDonald et al (108), with Cochrane review criteria scoring was scored high-quality or 8 of 12, whereas on IPM-QRB criteria, 95% of the reviewers scored it below 32. Similarly, Carette et al (106) was scored as high-quality by 85% of the reviewers for Cochrane review criteria, whereas 65% of the reviewers scored it as high quality utilizing IPM-QRB criteria.

Finally, Ackerman and Ahmad (107) was scored as high-quality by 55% of the reviewers, meeting 8 of the 12 Cochrane review criteria, compared to all of the reviewers providing a score below 32 utilizing IPM-QRB criteria. Thus, there was only one manuscript (108) which scored equally on both instruments. This shows the need for the present instrument.

**Individual Criteria Assessment**

As shown in Table 4, inter-rater agreement of IPM-QRB criteria was variable for individual items. Overall, of the 22 items resulting in a total maximum potential score of 48, 4 items had an average inter-rater agreement of 48%, 48%, 44%, and 45%. The remaining 18 items showed agreement above 60% ranging from 61% to 92%.

The items with low agreement were related to item 1 describing adherence to CONSORT or SPIRIT guidelines, item number 11 describing outcomes assessment criteria for significant improvement, item 20 describing outcome assessor blinding, and finally, item 22 describing conflicts of interest. Overall agreement for all 4 manuscripts as rated by 16 reviewers was over 82% with an intra-class correlation coefficient of 0.833 with a 95% CI of 0.592 - 0.986.

**DISCUSSION**

In this study, we designed an instrument to assess the methodologic quality of RCTs of interventional techniques. This instrument is specific for spinal interventional techniques including minimally invasive interventions describing disc interventions, augmentation procedures, and implantables. The assessment was assessed only for interventional techniques commonly performed which included those of epidurals and facet joint interventions. In this evaluation, we assessed the reliability of individual items on the IPM-QRB checklist in the area of trial assessment and compared the inter-rater reliability with Cochrane review criteria.

This assessment showed an intra-class correlation coefficient of 0.833 (95% CI; 0.592 – 0.986) with more than 82% for individual items with > 60% agreement among the reviewers providing overall very good agreement. Further, comparison between Cochrane review criteria and IPM-QRB criteria showed superior agreement among the reviewers for IPM-QRB criteria. Overall, the intra-class correlation coefficient for Cochrane review criteria for all 4 trials was 0.407 (95% CI; 0.144 – 0.911) with fair reliability. Thus, IPM-QRB criteria showed very good intra-class correlation coefficient (0.833) compared to fair intra-class correlation coefficient of the Cochrane review instrument (0.407). Cochrane review criteria have been extensively used over the years, thus IPM-QRB criteria’s superior rating of reliability shows the value of this instrument despite extensive expansion. In addition, utilizing Cochrane review criteria, 3 trials were considered as high quality meeting 8 of 12 criteria. In contrast, only one trial met the criteria of the high quality achieving a score of 32 out of 48 by IPM-QRB criteria. These differences are significant in that utilizing extensive assessment specific for interventional techniques indicates general application of criteria from Cochrane review may not be optimal for interventional techniques.

As described earlier, Cochrane review has been utilized extensively. However, the majority of the experts utilizing Cochrane review criteria were quality assessment content experts rather than clinical experts. Further, there was no involvement of interventional pain physicians.

The quality assessment of content experts may be biased by prior opinions, and it may be desirable to have both a clinical content expert and a non-expert with methodologic background assess the quality of the studies. In practice, methodologic quality assessments are performed by methodologists without clinical knowledge and with very little time invested along with an inherent bias based on financial incentives or interventional pain management experts who also may have certain inherent biases in addition to financial
incentives.

This instrument is unique with the use of Cochrane review criteria for bias and the addition of multiple items specific for interventional techniques. These added items are practical and improve the quality of reporting, study design, outcomes assessment, and most importantly, conflicts of interest assessment. Multiple factors should be taken into consideration in interpretation of this instrument and its application. This is the first time such an instrument has been developed. Reviewers will face some difficulties with immediate understanding of the instrument; however, once the learning curve has passed, reviewers will not only become comfortable but will start appreciating various insights of this instrument.

Assessment of individual items was provided with clarity for interventional techniques. All 22 items included in IPM-QRB with the background description and rater agreement for scoring are described below.

I. ADHERENCE TO CONSORT OR SPIRIT GUIDELINES

Item 1: Trial Design Guidance and Reporting

Adherence to CONSORT or SPIRIT guidelines is crucial (9,11,32,66). Critical appraisal of the quality of clinical trials is possible only if the design, conduct, and analysis of RCTs are thoroughly and accurately described in the report. Moher et al (66), in their elaboration on CONSORT guidelines, showed that reporting of RCTs are often incomplete (110-112) and compounded by problems of poor methodology (113-116). Numerous other reviews have documented deficiencies in reports of clinical trials (66,91).

A Cochrane database systematic review in 2012 assessing the completeness of reporting of over 6,000 RCTs (9) found that the characteristics of the populations were variable, resulting in heterogeneity between included evaluations. Validity assessments of the included studies also revealed unclear judgments. The results revealed that 81% more RCTs published in CONSORT-endorsing journals adequately described allocation concealment compared to those published in non-endorsing journals. Allocation concealment was reported adequately in 45% of CONSORT-endorsing journals versus 22% of RCTs in non-endorsing journals. Other outcomes with significantly different results based on whether or not the journal endorsed CONSORT guidelines included scientific rationale and background in the introduction, sample size, method used for sequence generation, and an aggregate score over reported CONSORT items -- “total sum score.”

Due to a multitude of problems with CONSORT that underlie the protocol deficiencies which may in turn lead to avoidable protocol amendments, poor trial conduct, and inadequate reporting in trial publications (11), SPIRIT was launched in 2007. This international project aims to improve the completeness of trial protocols while producing evidence-based recommendations for a minimum set of items to be addressed in protocols. The authors of SPIRIT also provide similar data as CONSORT with multiple deficiencies.

The scoring system of IPM-QRB adds flexibility in reference to the study on the trials conducted prior to 2005 since CONSORT guidance was not available until 2001. Further, appropriate scoring was also provided

<table>
<thead>
<tr>
<th>Trial Objective</th>
<th>Type of Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure absolute effect size</td>
<td>Placebo Control</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Show existence of effect</td>
<td>Placebo Control</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Show dose-response relationship</td>
<td>Placebo Control</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Compare therapies</td>
<td>Placebo Control</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Y = Yes, N = No, P = Possible, depending on whether there is historical evidence of sensitivity to drug effects. Adapted and modified from: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials E10, July 20, 2000 (118).
for the trials describing moderately significant criteria for randomized trials or trials conducted with high level reporting and criteria, even without CONSORT or SPIRIT reporting, as long as the reporting criteria were of high quality and included all the elements. In this assessment, adherence to CONSORT or SPIRIT guidelines is considered to be of paramount importance, which is reflected in the scoring system.

II. DESIGN FACTORS

Multiple design factors include study design, setting where the procedure is performed, type of physician performing the procedure, imaging used to confirm accuracy, sample size, and statistical methodology.

Item 2: Type of Design of Trial

The World Health Organization (WHO) defines a clinical trial as, “any research study that prospectively subjects human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes” (117).

Two critical components of a randomized trial are randomization and a control group. However, there are multiple types of control designs as shown in Table 5 (118). These include placebo-control, active-control, dose-response, placebo plus active, placebo plus dose-response, active plus dose-response, and active plus active plus dose-response, which may be the most robust design. The most commonly utilized designs in clinical research are placebo-control and active-control. In interventional pain management, due to various difficulties with regard to blinding, ethics, and regulatory issues, designing a true placebo-controlled study is difficult, so that most types of studies utilize an active-control design. A placebo-control measures absolute effect size and can establish “efficacy.” In contrast, an active-control trial design may show the existence of an effect, and also compares therapies. The difference between a placebo-control and an active-control is the latter fails to measure absolute effect size, otherwise known as efficacy, which is extremely important in early assessments.

Many researchers misinterpret active-controls as placebo-controls. For example, clinical and experimental studies have repeatedly shown that nerve blocks and joint injections done with local anesthetics and steroids both provide long-term improvement (1,35,86,103,109,119-142). Considering local anesthetic as a true placebo in a clinical trial may lead to the false conclusion that an intervention (such as a steroid injection) is ineffective. The reliability and accuracy of the placebo design is also important. It has been shown that placebo studies are susceptible to response and other types of bias. The therapeutic effect of placebos has been underestimated by methodologists and clinicians who do not support interventions. The effects of placebo, nocebo, and pure and fake placebo have been extensively investigated (102,103,106,143-157). Inactive solutions when injected into active structures often exert therapeutic effects, as illustrated in multiple studies (1,2,149-156). It is therefore essential to understand what constitutes a true placebo effect. To date, there have been few studies that have documented an appropriate placebo design (105,106,158,159). However, there have been multiple trials with inappropriate placebo design (102,104,154,160). Further misinterpretation continues to prevail of local anesthetics injected into an active structure as placebo in the systematic reviews (2,4,36-39,99,134,135,155-157).

Item 3: Setting/Physician

The training of physicians performing these procedures is crucial in assessing the quality of a randomized trial (2,161-164). Studies designed by non-specialists are more likely to include inappropriate candidates for studies (e.g. patients with mechanical pain being enrolled in an epidural steroid injection study), and procedures may not be performed in an optimal manner (e.g. maximizing lesion size for radiofrequency denervation or appropriate adhesiolysis). The optimum setting would be a fellowship-trained interventional pain physician performing an image-guided procedure, while procedures performed by general physicians may be less likely to yield benefit. The scoring system reflects the likelihood of a study demonstrating efficacy stratified by the people performing the procedures.

Item 4: Imaging

Imaging is mandatory for certain techniques such as facet joint interventions and transforaminal epidural injections. Physicians have performed these without fluoroscopy in many cases including epidural injections, sacroiliac joint injections, and facet joint interventions (2,102,104,105,108,165). The disadvantages of epidural injections with inferior results performed without fluoroscopy have been extensively described (2,135,165).

Proponents of ultrasound claim equal accuracy compared to fluoroscopy but this has not been proven (102,166-172). Further, ultrasound is recommended for peripheral nerve blocks, plexus blocks, and intraarticular
joint injections, but spinal structures are generally too deep to be properly visualized with ultrasound. Therefore, procedures performed under ultrasound will receive a score of 1.

Computed tomography (CT) also has been utilized. CT yields excellent visualization of the surrounding anatomy for procedures such as celiac plexus block, but its benefits are less apparent for other techniques. In a systematic review, Bui and Bogduk (173) showed a lack of effectiveness for transforaminal epidural injections performed under CT guidance. Atluri et al (174), in assessing complications resulting in fatalities of transforaminal epidural injections, showed an unusually high number of complications associated with CT-guided procedures. Further, radiation exposure is unnecessarily high with CT-guided procedures, which take longer to perform and consume greater resources (175-177). Thus, procedures performed under CT guidance are provided with a score of 2.

Fluoroscopy is the most appropriate imaging modality for performing interventional techniques and is conferred a score of 3.

**Item 5: Sample Size**

Sample sizes are generally calculated based on assumptions of benefit, garnered from previous studies or pilot study results (178-181). However, they are often inappropriately calculated (9,11,32,65-71). Some trials include very small sample sizes making it difficult to assess negative results. For interventional techniques, very small sample sizes have been utilized (182,183) in some high quality trials.

**Item 6: Statistical Methodology**

The importance of statistical methodology has been emphasized in guidance from CONSORT, SPIRIT, QUOROM, PRISMA, and Cochrane methodologic review criteria (9,11,21,22,23,25,28,29,33,34,37,38,40,59,60,65-67,71,76,79,93,94,96).

**III. PATIENT FACTORS**

**Item 7: Inclusiveness of Population**

Studies may be conducted with different populations. Clinical relevance is important in assessing methodologic quality of assessment and risk of bias. For facet joint and sacroiliac joint interventions in particular, therapeutic selection of criteria is crucial to eliminate false positive results and optimize outcomes. Despite the ongoing controversy surrounding the optimal number of diagnostic blocks (i.e. 1 or 2) or the threshold for designating a block as positive (e.g. > 50% or > 80% pain relief), when conducting clinical trials it is of paramount importance to eliminate placebo-responders and patients with other pain etiologies (i.e. false-positive results) (2,106,119-121,161,182-189).

**Item 8: Duration of Pain**

Most acute pain episodes resolve within 3 months, after which pain is designated as chronic. Thus, whenever patients with less than 3 months or even those with 3 to 6 months are included in a study, the results may be affected by the natural progression of the disease, which results in high response rates in the “control” group. The literature shows better efficacy for nearly all interventions in patients with acute pain. Paradoxically, disease burden in general, and long duration of pain in particular, is also associated with poorer treatment results. Considering the practice of interventional pain management being mostly chronic pain of months or years duration (2,103,109,119-133,190), significant priority has been given to the patients with at least chronic established pain of 6 months and longer.

**Item 9: Previous Treatments**

The use of conservative interventions prior to incorporating interventional techniques in a management algorithm of chronic pain patients is not only crucial, but also mandatory (103,109,119-133,191-194). In addition, conservative management, including structured exercise programs, behavioral rehabilitation programs, physical therapy, occupational therapy, chiropractic management, and drug therapy, has been shown to provide pain relief and improve physical and functional status at least in some patients (195-205). Further, these interventions also provide proper body mechanics, instructions for structured exercise programs, behavior modifications, and avoid the placebo effects. Consequently, assessment of conservative interventions and their appropriateness is considered as crucial in this assessment in the modern practice of interventional pain management.

**Item 10: Duration of Follow-up with Appropriate Interventions**

Duration of follow-up with interventional techniques is crucial. Studies utilizing only less than 3 months of follow-up do not provide any significant information about long-term outcomes, though they increase the multiplicity of the trials. Thus, short-term tri-
als, specifically with placebo controls, are important for efficacy assessment but not so for long-term outcomes and practical application in clinical settings. The majority of the placebo trials are only relevant for 3 months, specifically if the interventions are limited and not provided as required based on appropriate relief criteria for each intervention. The majority of the drug trials are of short-term duration (195,196,206), or short-term with less than 3 months of outcomes assessment. Consequently, the majority of the placebo controlled trials are only relevant for 3 month follow-up.

In this assessment, duration of follow-up with appropriate interventions (i.e., repeating interventions as needed within appropriate criteria without exceeding safety criteria) has been considered as crucial in managing chronic pain and providing accountable and value-based interventional pain management (103,109,119-133,207).

IV. OUTCOMES

Item 11: Outcomes Assessment Criteria for Significant Improvement

Outcomes assessment has been judged to be crucial in interventional pain management. The primary outcome in general is pain and the secondary outcome is function. In the past, functional outcomes were highly variable with minute changes considered as significant (208-213). However, recent publications in interventional pain management have illustrated robust measures in assessing significant improvement (2,103,109,119-133,161,163,164).

In fact, the CBRG guidelines also provided a set of component outcomes for low back pain trials. Deyo et al (214) and Bombardier (215) made recommendations for standardized measures to facilitate comparison of results among studies and ensure more complete reporting of relevant outcomes. They recommended psychometrically sound instruments for investigators who have sufficient resources to collect and analyze such data utilizing the use of Oswestry Disability Index (ODI) or the Roland-Morris Disability Questionnaire to measure functional status. CBRG authors and editors have also examined what it means to show a clinically meaningful change on patient important outcomes for back pain. Ostelo et al (216) proposed cutoffs for minimally important change for both pain and functional status. They concluded that a 30% improvement from baseline may be considered clinically meaningful improvement when comparing before and after measures for individual patients.

The minimally clinical important difference (MCID) was defined as the smallest difference in score in the domain of interest which the patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in a patient’s management. Copay et al (211) reported that the achieved commonly calculated MCIDs on one test does not appear to correlate well with achieving the same MCID on another. Gatchel and Mayer (212) challenged one MCID method based on expert consensus supposing > 30% improvement on subjective patient reports is the MCID threshold. They demonstrated that, at least in their Workers’ Compensation population, achieving this MCID correlates poorly with the objective extended outcomes of health care utilization work status. Parker et al (213), in determination of MCID in pain, disability, and quality of life after revision fusion for symptomatic pseudoarthrosis, showed variations by as much as 400% based on calculation technique. MCID was suggested to be as low as 2 points for ODI and 3 points for SF-12. These wide variations and low value of MCID question the face validity of such calculation techniques, especially when applied to heterogeneous disease and patient groups with a multitude of psychosocial confounders such as failed back syndromes. Gatchel et al (209) in studying the validation of a consensus-based MCID threshold using an objective functional external anchor found that a 30% or greater improvement on the self-report measure was significantly associated with improvement in physical function on progressive isoinertial lifting evaluation obtained after treatment. In conducting multiple trials, Manchikanti et al (103,109,119-135,163,164) and others (161,182) described more robust outcomes both in the neck as well as low back pain with at least 50% improvement in pain, as well as at least 40% to 50% improvement in disability as a combined outcome measure. In the development of IPM-QRB, outcomes assessment was the contentious issue for 2 reviewers, SC and HH. However, with extensive assessment of the other 18 authors of the instrument development and an additional 20 members of the guideline development group and with consultation with the statistician and 2 outside authors involved in conducting RCTs, the criteria utilized in the development of the outcomes assessment criteria of significant improvement was considered as appropriate.

Item 12: Analysis of All Randomized Participants in the Groups
Item 12 of the analysis of all randomized participants in the groups was adapted from Cochrane review criteria. Furlan et al (21), in their 2009 updated guidelines for systematic reviews, described the following criteria in assessing the analysis of randomized participants in the groups.

“All randomized patients are reported/analyzed in the group they were allocated to by randomization for the most important moments of the effect measurements (minus missing values) irrespective of non-compliance and co-interventions.”

As alluded to in Cochrane review criteria, analysis of all randomized participants in the groups is crucial. The authors should describe how data was accounted for, and it has to be deemed reasonable (i.e., does not overestimate the effect size).

**Item 13: Description of Dropout Rate**

Item 13, providing the description of dropout rates, was adapted from Cochrane review criteria (21).

In their 2009 updated guidelines for systematic reviews, Furlan et al (21) described dropout rate assessment as follows:

“The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias a ‘yes’ is scored. (However, these percentages are arbitrary, not supported by literature).”

Dropout rate, along with descriptions of how it was accounted for, is crucial in all trials; however, more so for interventional techniques. This aspect should be taken into consideration with overall assessment rather than individual scores, as all other scores. Essentially, in placebo-controlled trials, there will be significant dropouts early on, whereas in active-controlled trials, the threshold of 20% and 30% may be achieved in a significant proportion of the trials.

**Item 14: Similarity of Groups at Baseline for Important Prognostic Indicators**

Item 14, or similarity of groups at baseline, was adapted from Cochrane review criteria, with some modification. Furlan et al (21), in their 2009 updated guidelines for systematic reviews, described similarity of groups as follows:

“In order to receive a “yes”, groups have to be similar at baseline regarding demographic factors, duration and severity of complaints, percentage of patients with neurological symptoms, and value of main outcome measure(s).”

For this assessment, groups dissimilar on baseline variables with or without proper randomization and allocation concealment with the potential to significantly influence outcome were considered as inappropriate (e.g., disease severity, disease duration).

However, groups dissimilar on baseline variables despite proper randomization and allocation concealment are expected to not likely affect outcomes. Finally, groups similar on baseline measures were considered as the most appropriate measure.

**Item 15: Role of Co-Interventions**

Item 15, in relation to the role of cointerventions, was adapted from Cochrane review criteria. Furlan et al (21), in their 2009 updated guidelines for systematic reviews, described the role of co-interventions as follows:

“This item should be scored ‘yes’ if there were no co-interventions or they were similar between the index and control groups.”

The role of cointerventions for interventional techniques is crucial. Almost all patients do receive some type of cointervention; however, it is important to keep the cointerventions similar in both groups but, within the groups, individuals may differ to the extent of their activities and work status, etc. These were not considered as cointerventions. Continued structured interventions and required drug therapy, if provided to all participants in the group, are considered as appropriate.

**V. RANDOMIZATION**

**Item 16: Method of Randomization**

Item 16, describing the method of randomization, was adapted from Cochrane review criteria. Furlan et al (21), in their updated guidelines for systematic reviews, described method of randomization as follows:

“A random (unpredictable) assignment sequence. Examples of adequate methods are coin toss (for studies with 2 groups), rolling a dice (for studies with 2 or more groups), drawing of balls of different colors, drawing of ballots with the study group labels from a dark bag, computer-generated random sequence, pre-ordered sealed envelopes, sequentially-ordered vials, telephone call to a central office, and pre-ordered list of treatment assignments. Examples of inadequate methods are: alternation, birth date, social insurance/
security number, date in which they are invited to participate in the study, and hospital registration number."

Randomization is crucial; however, reviewers also should understand that despite adequate randomization and allocation concealment, baseline characteristics may differ (217). Thus, differences in baseline characteristics may not reflect the process of randomization in allocation concealment.

VI. ALLOCATION CONCEALMENT

**Item 17: Concealed Treatment Allocation**

Concealed treatment allocation in Item 17 was adapted from Cochrane review criteria. In their 2009 updated guidelines for systematic reviews, Furlan et al (21) described allocation concealment as follows:

“Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.”

Studies in which allocation is ineffectively concealed using a short, alternating sequence (e.g. every even number patient is allocated to one group), whereby the investigator or other personnel, including the patient, can appreciate the group assignment, were considered inappropriate.

VII. BLINDING

The blinding criteria in section 7 describing patient blinding, care provider blinding, and outcome assessor blinding reflected in Items 18, 19, and 20 were adapted from Cochrane review criteria. Furlan et al (21), in their 2009 updated guidelines, provided appropriate guidance to assess these items.

**Item 18: Patient Blinding**

Furlan et al (21), in their 2009 updated guidelines for systematic reviews, described patient blinding as follows:

“This item should be scored ‘yes’ if the index and control groups are indistinguishable for the patients or if the success of blinding was tested among the patients and it was successful.”

Some treatments are more difficult to blind than others (e.g. transforaminal epidural injections are more difficult to blind than interlaminar epidural injections because patients often experience radicular pain during injection). The difficulties inherent in blinding radiofrequency denervation studies include masking lesion-related pain and neuritis, which can be overcome to some extent with local anesthetic and corticosteroid, and the sound of the radiofrequency generator during treatment, which can be lowered but not turned off. For placebo-controlled trials, multiple difficulties are encountered with subcutaneous injections compared to a caudal epidural injection or an interlaminar epidural injection, injections provided in a distinct region, even if it is intramuscular, the effects of the medication with local anesthetic with numbness and weakness, increased soreness with pure injection of sodium chloride solution or steroid, or radiofrequency neurotomy may lead patients to guess the group assignment and also may incorporate nocebo effects in final outcomes. Overall, adequate blinding is crucial in all interventional trials.

**Item 19: Care Provider Blinding**

Furlan et al (21), in their 2009 updated guidelines for systematic reviews, described care provider blinding as follows:

“This item should be scored ‘yes’ if the index and control groups are indistinguishable for the care providers or if the success of blinding was tested among the care providers and it was successful.”

Care provider blinding sometimes may be difficult related to identification of the solution injected, sham lesioning injection in a different location, etc. However, it may be easier in active-controlled trials, specifically provided as part of routine treatment for many patients and the patients involved in the trials are mixed with the other patients.

**Item 20: Outcome Assessor Blinding**

Furlan et al (21), in their 2009 updated guidelines for systematic reviews, described allocation concealment as follows:

“Adequacy of blinding should be assessed for the primary outcomes. This item should be scored ‘yes’ if the success of blinding was tested among the outcome assessors and it was successful or:

–for patient-reported outcomes in which the patient is the outcome assessor (e.g., pain, disability): the blinding procedure is adequate for outcome assessors if participant blinding is scored ‘yes’

–for outcome criteria assessed during scheduled visit and that supposes a contact between participants and outcome assessors (e.g., clinical examination): the blinding procedure is adequate if patients are
blinded, and the treatment or adverse effects of the treatment cannot be noticed during clinical examination

–for outcome criteria that do not suppose a contact with participants (e.g., radiography, magnetic resonance imaging): the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed when assessing the main outcome

–for outcome criteria that are clinical or therapeutic events that will be determined by the interaction between patients and care providers (e.g., co-interventions, hospitalization length, treatment failure), in which the care provider is the outcome assessor: the blinding procedure is adequate for outcome assessors if item “4” (caregivers), it is scored ‘yes’

–for outcome criteria that are assessed from data of the medical forms: the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed on the extracted data.”

Outcome assessor blinding is difficult. It all depends on appropriate patient blinding, and when the outcome assessor is independent, may derive information from patient impressions. Even though outcome assessors, as well as the patients, are blinded, their curiosity and potential indications to identify the treatment, may affect the judgment of the outcome assessor along with the patient.

**VIII. CONFLICTS OF INTEREST**

**Item 21: Funding and Sponsorship**

The relationship between industry and the research community is under increasing scrutiny. Studies have shown that industry-sponsored studies are 3.6 times more likely than non-industry sponsored studies to yield a positive result, a finding that is magnified by publication bias and the fact that a disproportion number of negative industry-sponsored studies are never submitted for publication. Industry has funded a substantial proportion of research published in all medical journals (206,218-235), with the effect being magnified in high-impact journals (206). Multiple studies have raised questions in reference to the evidence generated from industry-funded studies, and the level of evidence from industry-funded studies is widely acknowledged to be lower than that for findings obtained from studies funded by governments, foundations, or universities. It has been asserted that improving the quality of industry-funded research might increase the quality of evidence for making clinical decisions. In an effort to reduce bias and enhance transparency, a Sunshine Act proposal that went into effect as part of the Affordable Care Act (ACA) requires manufacturers of drugs, medical devices, and biologicals that participate in the U.S. federal health care programs to track payments and items of value given to physicians and teaching hospitals (232,233).

Lundh et al (229) showed that industry sponsored studies more often reported favorable findings and less adverse events than non-industry sponsored studies. They also showed that the results of industry-sponsored studies tend to be more heterogeneous than other studies. They concluded that “industry bias” cannot be fully accounted for by standard “risk of bias” assessments. Amiri et al (220) evaluated how sources of funding and conflicts of interest influence the outcomes and quality of spinal research. They analyzed 1,356 papers, among which 864 were suitable for assessment, showing industry-funded studies showed favorable outcomes 88% of the time, compared to 73% and 74% of publicly funded and foundation-funded studies, respectively. Bhandari et al (226) also found the association between finding and funding source with industry-funded trials more likely to be associated with positive findings for both medical and surgical interventions.

In contrast, Khan et al (230) found no significant differences between the likelihood of a positive outcome and funding source. They also noted that while industry-funded RCTs were of higher methodological quality and had significantly more study centers and subjects, non-profit funded RCTs had longer follow-up periods and were more likely to study different treatment strategies.

Although the debate continues on both sides with pros and cons of industry findings and inherent difficulties in assessing industry bias, this instrument has provided significant impetus to the assessment of the studies based on industry funding.

**Item 22: Conflicts of Interest**

Conflicts of interest besides industry sponsorship have been extensively discussed and guidance has been provided for authors, peer reviewers, and editors (235-250). Conflicts of interest may be related not only to direct funding, but also indirect sponsorship through a professional society or even governmental organization, or by other potential sources of remuneration such as stocks, advisory boards, or speaker bureaus.
The effects of significant conflicts of interest have been discussed in many forums and publications, including guideline preparation and promotion.

In this assessment, significant importance has been provided to assessing the conflict of interest information; however, it is realized that it is difficult to assess such conflicts and the resulting bias.

Limitations of this assessment include a difficult learning curve and final assessment of only 4 trials, though by a larger number of raters for the reliability. The major advantage of this instrument is that we were able to compare it with Cochrane review criteria.

Thus, further evaluation is warranted to assess the reliability of IPM-QRB, as it will be utilized in systematic reviews and development of contemporary guidelines for interventional techniques.

**Conclusion**

In this assessment, we presented a new instrument, namely Interventional Pain Management techniques - Quality Appraisal of Reliability (IPM-QRB) and assessed its reliability as an assessment tool for the methodologic quality and risk of bias of interventional RCTs. Very good reliability and inter-rater correlation was appreciated.

**Acknowledgments**

The authors wish to thank Dr. N Bala Krishna, MSc, PhD, for statistical analysis, Alvaro F. Gómez, MA, Laurie Swick, BS, for manuscript review, and Tonie M. Hatton and Diane E. Neihoff, transcriptionists, for their assistance in preparation of this manuscript. We would like to thank the editorial board of Pain Physician for review and criticism in improving the manuscript.

**Author Affiliations**

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

Dr. Hirsch is Vice Chief of Interventional Care, Chief of Minimally Invasive Spine Surgery, Service Line Chief of Interventional Radiology, Director of Endovascular Neurosurgery and Neuroendovascular Program, Massachusetts General Hospital; and Associate Professor, Harvard Medical School, Boston, MA.

Dr. Cohen is Professor, Department of Anesthesiology and Critical Care Medicine, Pain Management Division, Johns Hopkins School of Medicine, Baltimore, MD and Professor, Department of Anesthesiology, Uniformed Services University of the Health Sciences, Bethesda, MA.

Dr. Heavner is Professor, Departments of Anesthesiology and Physiology, Director, Anesthesia and Pain Research, Texas Tech University Health Sciences Center, Lubbock, TX.

Dr. Falco is Medical Director of Mid Atlantic Spine & Pain Physicians, Newark, DE; Director, Pain Medicine Fellowship Program, Temple University Hospital, Philadelphia, PA; and Adjunct Associate Professor, Department of PM&R, Temple University Medical School, Philadelphia, PA.

Dr. Diwan is Executive Director of Manhattan Spine and Pain Medicine, New York, NY.

Dr. Boswell is Chairman, Department of Anesthesiology and Perioperative Medicine, University of Louisville, Louisville, KY.

Dr. Candido is Chairman and Professor, Department of Anesthesiology, Advocate Illinois Masonic Medical Center, Chicago, IL.

Dr. Onyewu is Attending Physician, Mid Atlantic Spine & Pain Physicians, Newark, DE, and Elkton, MD; Faculty, Pain Medicine Fellowship Program, Temple University Hospital, Philadelphia, PA; and Adjunct Assistant Professor, Temple University Medical School, Philadelphia, PA.

Dr. Zhu is Adjunct Assistant Professor, Temple University Medical School, Philadelphia, PA; Faculty, Pain Medicine Fellowship Program, Temple University Hospital, Philadelphia, PA; and Attending Physician, Mid Atlantic Spine & Pain Physicians, Newark, DE and Elkton, MD.

Dr. Sehgal is Medical Director, Interventional Pain Program, and Associate Professor and Director Interventional Pain Management, Department of Orthopedics and Rehabilitation Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI.

Dr. Alan D. Kaye is Chairman and Professor, Department of Anesthesia, LSU Health Science Center, New Orleans, LA.

Dr. Benyamin is Medical Director, Millennium Pain Center, Bloomington, IL and Clinical Assistant Professor of Surgery, College of Medicine, University of Illinois, Urbana-Champaign, IL.

Dr. Helm is Medical Director, The Helm Center for Pain Management, Laguna Hills, CA.

Dr. Singh is Medical Director, Spine Pain Diagnostics Associates, Niagara, WI.

Dr. Datta is Medical Director, Laser Spine & Pain Institute, New York, NY, and Professorial Lecturer, Mount Sinai School of Medicine, Department of Anesthesiology, New York, NY.

Dr. Abdi is Professor and Chair, Department of Pain
Conflict of Interest:

Dr. Cohen served as a paid co-investigator on a Department of Defense study. He is also a consultant for Halozyme and Kimberly Clark. Dr. Falco is a consultant for St. Jude Medical Inc. and Joimax Inc. Dr. Kaye is a speaker for Depomed, Inc. Dr. Benyamin is a consultant and lecturer for Boston Scientific and Kimberly Clark. Dr. Helm is a clinical investigator with Epimed and receives research support from Cephalon/Teva, AstraZeneca, and Purdue Pharma, LP. He has attended an advisory group meeting for Activas. Dr. Datta receives research support from Sucampo Pharmaceuticals and an honorarium from Smith and Nephew. Dr. Vallejo receives research support from Cephalon/Teva, BioDelivery Sciences International, Inc., Mundipharma Research GmbH & Co., AstraZeneca, Purdue Pharma, LP, and Theravance, and is a consultant for Nevro and Kimberly-Clark. Dr. Racz is a Consultant for and has family ownership of Epimed International, is a Consultant to Cosman RF Company, and has Medtronic patent issues.

References


Methodologic Quality Assessment of Randomized Trials of Interventional Techniques


14. Young BK, Greenberg PB. Are the In-...


143. Hróbjartsson A, Kaptchuk TJ, Miller FG. Placebo effect studies are susceptible to response bias and to other types of bias: A meta-analysis of the whole truth, may do patients harm: Placebo and nocebo in interventional medicine: A friend or a foe – or simply foes? Pain Physician 2011; 14:E157-E174.


148. Howick J, Bishop FL, Hirsch JA. Placebo effect studies are susceptible to response bias and to other types of bias: A meta-analysis of the whole truth, may do patients harm: Placebo and nocebo in interventional medicine: A friend or a foe – or simply foes? Pain Physician 2011; 14:E157-E174.


176. Redberg RF. Cancer risks and radiation exposure from computed tomographic scans: How can we be sure that the benefits outweigh the risks? Arch Intern Med 2009; 169:2049-2050.


190. National Government Services, Inc. LCD for Pain Management (L3829). Effective Date 03/01/2009.

191. Cigna Government Services. LCD for Pain Management (L3284). Revision Effective Date: 01/03/2012.

192. Palmetto GBA. Local Coverage Determination (LCD): Paravertebral Facet Joint Block (L37176). Effective Date: 03/10/2011.


198. Sharma R, Haas M, Stano M, Specman


245. Lo B, Ott C. What is the enemy in CME, conflicts of interest or bias? JAMA 2013; 310:1019-1020.


