The diagnostic accuracy of the Kemp’s test: a systematic review

Kent Stuber, BSc, DC, MSc
Caterina Lerede, BSc, DC
Kevyn Kristmanson, BSc, DC
Sandy Sajko, BPE, DC, MSc, RCCSS
Paul Bruno, DC, PhD

Background: The objective of this review was to evaluate the existing literature regarding the accuracy of the Kemp’s test in the diagnosis of facet joint pain compared to a reference standard.

Methods: Several databases were searched. All diagnostic accuracy studies comparing the Kemp’s test with an acceptable reference standard were included. Included studies were scored for quality and internal validity.

Results: Five articles met the inclusion criteria of this review. Two studies had a low risk of bias, and three had a low concern regarding applicability. Pooling of data from studies using similar methods revealed that the test’s negative predictive value was the only diagnostic accuracy measure above 50% (56.8%, 59.9%).

Contexte: L’objectif de cette étude était d’évaluer la documentation scientifique publiée traitant de l’exactitude du test de Kemp dans le diagnostic de la douleur des facettes articulaires par rapport à une référence normative.

Méthodologie : Des recherches ont été faites dans plusieurs bases de données. Toutes les études sur l’exactitude des diagnostics comparant le test de Kemp à une référence normative acceptable ont été incluses. Les études retenues ont été notées sur une échelle de qualité et de validité interne.

Résultats : Cinq articles ont satisfait les critères d’inclusion dans cette étude. Deux études présentaient un faible risque de biais, alors que trois autres avaient un manque d’intérêt quant à l’applicabilité. Les données recueillies d’études utilisant des méthodologies semblables ont révélé que la valeur négative prédictive du test présentait l’unique mesure de l’exactitude de diagnostic supérieure à 50 % (56,8 % ; 59,9 %).
Introduction
Zygapophyseal (facet) joint pain has been defined as pain originating from any structure related to the facet joints, including the fibrous capsule, synovial membrane, hyaline cartilage, and bone.\textsuperscript{1,2} Facet joint pain may be localized to its associated spinal region or referred to distant sources, and multiple studies have demonstrated characteristic pain referral patterns for the cervical, thoracic, and lumbar facet joints.\textsuperscript{1,3-6} Reports of facet joint pain prevalence rates vary widely in the literature, due in part to the variety of methods used to confirm the diagnosis.\textsuperscript{1} It is generally accepted that the most reliable and valid method of diagnosing facet joint pain is through the use of anesthetic injections to the facet joints (intra-articular) or their nerve supply (medial branch blocks).\textsuperscript{1,4-6} Due to the moderately high false-positive rates of single block studies, double block studies using comparative or control injections are recommended to achieve a definitive diagnosis.\textsuperscript{1,3-6} Studies using double diagnostic blocks have demonstrated facet joint pain prevalence rates of 36-67\% in chronic neck pain patients, 34-48\% in chronic thoracic pain patients, and 15-45\% in chronic low back pain patients.\textsuperscript{3,6}

Due to the cost and risk of complications associated with diagnostic blocks, it would be beneficial to establish clinical screening procedures that can reliably and validly diagnose facet joint pain.\textsuperscript{7} Some studies\textsuperscript{10-15} have shown a possible association between certain clinical features and a positive response to facet joint anesthesia. However, the collective literature in this area generally suggests that there are no historic or physical examination findings that can reliably predict this response and therefore accurately diagnose facet joint pain.\textsuperscript{1,6,7}

Conclusions: Currently, the literature supporting the use of the Kemp’s test is limited and indicates that it has poor diagnostic accuracy. It is debatable whether clinicians should continue to use this test to diagnose facet joint pain.

Key words: Kemp test, joint, facet, pain, diagnosis, chiropractic

Conclusions : À l’heure actuelle, il n’y a pas suffisamment de documents scientifiques appuyant l’utilisation du test de Kemp, ce qui laisse prévoir une faible précision diagnostique du test. Il y a lieu de se demander si les cliniciens devraient poursuivre l’utilisation de ce test par le diagnostic des douleurs de facettes articulaires.

Mots clés : test de Kemp, articulation, facette, douleur, diagnostic, chiropratique

K Stuber, C Lerede, K Kristmanson, S Sajko, P Bruno
The diagnostic accuracy of the Kemp’s test: a systematic review

Figure 1a:
Cervical Kemp’s test – start position.

Figure 1b:
Cervical Kemp’s test – finishing position.

Figure 2a:
Lumbar Kemp’s test – start position.

Figure 2b:
Lumbar Kemp’s test – finishing position.
is present” and “A negative Kemp’s test is a strong indicator that facet syndrome is not present”. Interestingly, 41.5% of respondents considered a positive Kemp’s test to be a strong indicator of the presence of a lateral disc lesion, compared to 26.8% who considered it to be a strong indicator of facet syndrome.

As with any clinical test, the accuracy of Kemp’s test in diagnosing its target condition (in this case facet joint pain) needs to be considered by clinicians in order for the test to be applied most appropriately in practice. This seems particularly pertinent considering the test’s apparently high usage rate and perceived usefulness in diagnosing facet joint pain amongst chiropractors. Therefore, the purpose of the current study was to systematically review the evidence related to the diagnostic accuracy of the Kemp’s test in the diagnosis of facet joint pain compared to a reference standard (i.e. diagnostic block).

Methods

Study design
The methods of this systematic review were decided a priori and adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.20 The PRISMA statement includes a 27-item checklist designed to improve reporting of systematic reviews and meta-analyses.

Search strategy
Several online databases (EMBASE, PubMed, MEDLINE, CINAHL, PEDro, Index to Chiropractic Literature) were searched in all languages from their date of commencement to October 2013. Reference searching of any retrieved articles was also employed. The search strategy employed in EMBASE, PubMed, MEDLINE, and CINAHL can be seen in Box 1. A further keyword search was conducted in PEDro and the Index to Chiropractic Literature databases and included the terms facet syndrome AND diagnostic accuracy, as well as Kemp's OR Quadrant OR extension rotation AND diagnostic accuracy. Individual searches for Kemp’s OR Quadrant OR extension rotation were also undertaken.

Study selection
Two authors (CL and KS) independently reviewed the electronic database search results (title and abstract) independently. Any titles and abstracts that appeared to meet inclusion criteria were selected for full text review. The inclusion and exclusion criteria for this review can

### Box 1. Search strategy

1. (facet OR zygapophyseal) AND diagnosis AND pain
2. (Kemp’s OR Quadrant OR extension rotation) AND (spine OR back OR neck)

### Box 2. Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Any article published in English.</td>
</tr>
<tr>
<td>2. Articles published in a peer reviewed journal or abstracts from scientific conference.</td>
</tr>
<tr>
<td>3. Diagnostic accuracy study comparing the Kemp’s test (or Quadrant test or extension-rotation test) with an acceptable reference standard, preferably facet joint injections.</td>
</tr>
<tr>
<td>4. Any setting.</td>
</tr>
<tr>
<td>5. Adult participants with cervical or lumbar facet joint pain of any duration (acute, subacute, chronic, recurrent) and any intensity (no minimum or maximum score on a pain scale).</td>
</tr>
<tr>
<td>6. The outcomes in the comparison studies include those that either require complete relief of facet joint pain symptoms after injection or a minimum subjective numerical decrease (such as a percentage) in pain upon injection.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Articles not published in English.</td>
</tr>
<tr>
<td>2. Articles not published in a peer-reviewed journal.</td>
</tr>
<tr>
<td>3. Studies that did not employ a comparison or reference standard test.</td>
</tr>
<tr>
<td>4. Studies that reported on patients with a condition other than facet joint pain, including but not limited to degenerative joint disease, degenerative disc disease, malignancies, infections, pregnancy, or neurological conditions.</td>
</tr>
</tbody>
</table>
be seen in Box 2. Any disagreements were resolved by discussion between the two authors. The same two authors independently conducted the full text review of the retrieved articles comparing them with the inclusion and exclusion criteria. Any disagreements were resolved by discussion, with a third author (PB) consulted if resolution was not achieved, to produce the final articles for inclusion. A data extraction form was prepared with one author (CL) independently extracting data from the selected studies. A second author (KS) reviewed the completed form for accuracy, with any disagreements resolved by a third author (PB).

**Quality assessment**
All full text journal articles that met the inclusion criteria of the review were independently scored for quality and internal validity by two authors (SS and KS) using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS 2). Any disagreements were resolved by discussion.

**Diagnostic accuracy measures**
For the quantitative assessment, statistical measures of diagnostic accuracy (sensitivity, specificity, positive and negative likelihood ratios, and positive and negative pre-
dictive values) from each included study were calculated from two by two (2 × 2) tables completed by one of the authors (KS) and confirmed by another author (PB), and further confirmed with direct reporting from each included study where applicable. Data was pooled from studies deemed to be sufficiently similar (in terms of methods and minimum pain relief values), and the same statistical measures were calculated from cumulative 2 × 2 tables.

Results

Study selection

Figure 3 depicts the flow of articles through the review process. Five articles, evaluating a total of 616 patients, met the inclusion criteria for this review. Of these articles, two were identified through the electronic database search,\textsuperscript{12,22} while the remaining three articles\textsuperscript{13,23,24} were identified by reference searching. All five included studies specifically assessed lumbar facet joint pain; none evaluated cervical facet joint pain. None of the included studies specifically named the test as “Kemp’s” or “Quadrant”; rather, they all referred to the test as the extension-rotation test.

Study descriptions

Table 1 presents the characteristics of the studies included in the review. Three studies\textsuperscript{12,22,24} employed double block injections as the reference standard, while the remaining two studies employed single block injections as the reference standard.\textsuperscript{13,23} Interestingly, in two of the studies\textsuperscript{13,22} the authors considered the absence of pain exacerbation on extension-rotation to be a positive result.

Diagnostic accuracy measures

Table 2 presents the statistical measures of diagnostic accuracy in the included studies as calculated from 2 × 2 tables. None of the included studies had both sensitivity and specificity measures of at least 50% for the extension-rotation test. Sensitivity was found to be 100% in two studies (Laslett, 2006\textsuperscript{12} using a 95% pain relief standard; Schwarzer, 1994\textsuperscript{24}) and 85.7% (Laslett, 2006\textsuperscript{12} using a 75% pain relief standard). The highest specificity was 67.3% by Manchikanti et al.\textsuperscript{22} The highest positive likelihood ratio was 1.29 (Laslett, 2006\textsuperscript{12} using a 95% pain relief standard), while the lowest negative likelihood ratio was 0.00 (Laslett, 2006\textsuperscript{12} 95% pain relief standard; Schwarzer, 1994\textsuperscript{24}). The highest positive predictive value was 43.5%
The diagnostic accuracy of the Kemp’s test: a systematic review

Table 2.
Diagnostic Accuracy Measures of Included Studies

<table>
<thead>
<tr>
<th>Author, year of publication</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>LR+</th>
<th>LR-</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laslett, 200612</td>
<td>100%</td>
<td>22.3%</td>
<td>1.29</td>
<td>0.00</td>
<td>13.0%</td>
<td>100%</td>
</tr>
<tr>
<td>Laslett, 200612</td>
<td>85.7%</td>
<td>21.8%</td>
<td>1.10</td>
<td>0.66</td>
<td>26.1%</td>
<td>82.6%</td>
</tr>
<tr>
<td>Manchikanti, 200022</td>
<td>32.1%</td>
<td>67.3%</td>
<td>0.98</td>
<td>1.01</td>
<td>43.5%</td>
<td>55.8%</td>
</tr>
<tr>
<td>Revel, 199223</td>
<td>31.8%</td>
<td>22.2%</td>
<td>0.41</td>
<td>3.07</td>
<td>33.3%</td>
<td>21.1%</td>
</tr>
<tr>
<td>Revel, 199813</td>
<td>23%</td>
<td>51.7%</td>
<td>0.48</td>
<td>1.49</td>
<td>17.7%</td>
<td>60%</td>
</tr>
<tr>
<td>Schwarzer, 199424</td>
<td>100%</td>
<td>11.6%</td>
<td>1.13</td>
<td>0.00</td>
<td>17.6%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Legend: LR+ = Positive likelihood ratio; LR- = Negative likelihood ratio; PPV = Positive Predictive Value; NPV = Negative Predictive Value.

Table 3.
Pooled Study Diagnostic Accuracy Measures

<table>
<thead>
<tr>
<th>Study parameters</th>
<th>Single block injections with minimum 75% improvement in symptoms (Revel, 199823; Revel, 199223)</th>
<th>Double block injections with minimum 75% improvement in symptoms (Laslett, 200612; Manchikanti, 200022)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>34.5%</td>
<td>45.5%</td>
</tr>
<tr>
<td>Specificity</td>
<td>47.2%</td>
<td>46.9%</td>
</tr>
<tr>
<td>LR+</td>
<td>0.65</td>
<td>0.86</td>
</tr>
<tr>
<td>LR−</td>
<td>1.39</td>
<td>1.16</td>
</tr>
<tr>
<td>PPV</td>
<td>26.3%</td>
<td>33.2%</td>
</tr>
<tr>
<td>NPV</td>
<td>56.8%</td>
<td>59.9%</td>
</tr>
</tbody>
</table>

Legend: LR+ = Positive likelihood ratio; LR− = Negative likelihood ratio; PPV = Positive Predictive Value; NPV = Negative Predictive Value.

Table 4.
QUADAS 2 Methodological Quality Data for the Included Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Risk of Bias</th>
<th>Applicability Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient Selection</td>
<td>Index Test</td>
</tr>
<tr>
<td>Laslett, 200612</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Manchikanti, 200022</td>
<td>Y</td>
<td>?</td>
</tr>
<tr>
<td>Revel, 199223</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Revel, 199813</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Schwarzer, 199424</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Legend: Y = low risk; N = high risk; ? = Unsure.
(Manchikanti, 2000), while the highest negative predictive value was 100%, found in two studies (Laslett, 2006; Schwarzer, 1994).

Results from two studies (Revel, 1998; Revel, 1992) employing single block injections and requiring at least 75% improvement in pain symptoms were pooled and statistical measures of diagnostic accuracy were calculated in 2 x 2 tables and are reported in Table 3. Similarly, two studies (Laslett, 2006; Manchikanti, 2000) that employed double block injections and requiring at least 75% improvement in pain symptoms were pooled and statistical measures of diagnostic accuracy were calculated in 2 x 2 tables and are reported in Table 3. These pooled results revealed that only negative predictive value produced results above 50%.

Methodological quality
Table 4 presents the QUADAS 2 assessment results for each included study. Two studies were found to have low risk of bias (Laslett, 2006; Revel, 1998), while three studies were deemed to be at risk of bias (Manchikanti, 2000; Revel, 1992; Schwarzer, 1994). Three studies were deemed to have low concern regarding applicability (Laslett, 2006; Revel, 1998; Manchikanti, 2000), whereas two studies had concerns regarding applicability (Revel, 1992; Schwarzer, 1994).

Discussion
Interpretation of results and clinical relevance of the findings
The evidence supporting the diagnostic accuracy of the Kemp’s test in the diagnosis of facet joint pain is limited. Only two studies met the inclusion criteria for this review and were judged to have a low risk of bias and low concern regarding applicability (Table 4). Of these, only the study by Laslett et al employed double diagnostic block injections and required at least 75% improvement in pain symptoms after the confirmatory block. With these considerations in mind, we synthesized the data from those studies whose methods were suitably similar (Table 3). The calculated specificity and positive predictive values were generally quite low, indicating that the value of a positive Kemp’s test result in diagnosing facet joint pain is highly dubious. This suggestion is supported by two studies in which patients underwent radio-frequency denervation treatment following a positive response to a single diagnostic block injection; the results demonstrated that pain exacerbation by extension-rotation was significantly correlated with treatment failure. Although the calculated sensitivity values were also <50%, the negative predictive values were ~60%, suggesting that a negative Kemp’s test result may have moderate clinical value in eliminating the facet joint as a source of pain.

Although the Kemp’s test in isolation appears to be of limited usefulness in the diagnosis of facet joint pain, a positive or negative test result may have value as part of a prediction rule or serial approach to diagnosis that incorporates other clinical variables to establish a diagnosis. Laslett et al described five “optimal” clinical prediction rules created from a multitude of clinical variables, four of which included “positive extension/rotation test” as one of the rule variables. Mirroring the results of our review, these four rules generally demonstrated poor to moderate specificity and positive predictive values, and very high sensitivity and negative predictive values. These results also support the potential usefulness of a negative Kemp’s test result.

An important point to consider is that all of these studies only consider the ability of the Kemp’s test to discriminate between individuals who respond to diagnostic joint blocks from those who do not. Such procedures can be used therapeutically to reduce or eliminate an individual’s back pain; however, the ability of the Kemp’s test to discriminate between individuals who would respond to other forms of treatment (e.g. spinal manipulation) is unknown. Clinical prediction rules to indicate back pain patients who are more or less likely to respond to spinal manipulation have been proposed. Although these prediction rules do not include a positive or negative Kemp’s test result as a predictor variable, the authors of these studies do not specifically describe the Kemp’s tests as being one of the potential variables that was considered for inclusion in any of the prediction rules.

Limitations
Our literature search only yielded five studies that met the a priori inclusion criteria. Since these studies varied in terms of several important aspects of the methods used (e.g. single vs. double joint blocks, differences in the minimum pain reduction required to be considered a “posi-
tive” response to joint block), conclusions drawn from pooling of the data must be tempered. In addition, the general quality of the included articles was moderate with three studies deemed to be a risk of bias and two studies had concerns regarding applicability.

Our database search only revealed two of the studies included in our review. The remaining three were only identified by reference searching of the included studies and other reviews elicited by our search. Combined with the fact that non-English articles were excluded from our review, it is possible that additional studies investigating the diagnostic accuracy of the Kemp’s test have not been included herein.

Finally, the current accepted “gold standard” for diagnosing facet joint pain is through the use of anesthetic injections to the facet joints (intra-articular) or their nerve supply (medial branch blocks). As such, this was used as the reference standard by which the accuracy of the Kemp’s test was evaluated in our review. However, the limitations of the current “gold standard” as being definitive evidence of facet joint pain would potentially affect the diagnostic accuracy values reported herein.

Suggestions for future research
It is vital that any studies investigating the diagnostic accuracy of the “Kemp’s test” or its potential inclusion in a clinical prediction rule related to the categorization of treatment responders/non-responders use a consistent terminology to describe the procedure. Since “extension-rotation test” seems to be the most common term that has been used in the literature to date, we propose that future researchers (and clinicians) adopt this term when referring to this procedure. Future studies investigating the diagnostic accuracy of this (or any) clinical test at diagnosing facet joint pain should use double joint injections and require a consistent level of improvement (we suggest 75%) in pain symptoms as the reference standard.

Conclusions
The literature supporting the use of the Kemp’s test to diagnose facet joint pain is limited and generally indicates that the test has poor diagnostic accuracy. There is preliminary evidence that a negative test result, either in isolation or as part of a clinical prediction rule, may have some clinical value in eliminating the facet joint as a source of pain. Until more evidence is established regarding this test, however, it is debatable whether clinicians should continue to use it to diagnose facet joint pain.

Acknowledgments
The authors thank and acknowledge Mary Chipanshi of the University of Regina for her assistance in formulating the search strategy and conducting the literature search.

References


