Interexaminer Reliability of Three Methods of Combining Test Results to Determine Side of Sacral Restriction, Sacral Base Position, and Innominate Bone Position

Henry C. Tong, MD; Oscar G. Heyman, DO; Derek A. Lado, DO; and Mara M. Isser, DO

Context: Sacroiliac joint dysfunction is diagnosed based on the combined results of several palpatory examinations. Previous studies have compared the interexaminer reliability of only one of these methods of diagnosis.

Objective: To compare the interexaminer reliability of three methods of combining palpatory examinations to determine the side of sacroiliac joint dysfunction, sacral base position, and innominate bone position.

Design: Blinded single-cohort reliability study.

Methods: Patients with low back pain underwent two identical sets of palpatory examinations given by two physicians, separately, at a university spine center. The results of each set were compiled and interpreted by three methods: using the test result with the highest interexaminer reliability (method 1), requiring at least one test result to be abnormal for the variable to be abnormal (method 2), and requiring all test results to be abnormal for the variable to be abnormal (method 3). The \( \kappa \) was calculated for each method.

Results: There were 24 subjects (mean age, 68.3 years), of which 15 (62%) were women. The \( \kappa \) was consistently higher with method 1, at 0.47, 0.08, and 0.32 for the sacral position, innominate bone position, and side of sacroiliac joint dysfunction, respectively. Corresponding values for method 2 were 0.09, 0.4, and 0.16, and for method 3 were 0.16, 0.1, and -0.33.

Conclusion: Using the results of the most reliable examination consistently has the best interexaminer reliability.

A recent review article on the validity and reliability of tests for low-back dysfunction concluded that no single test has been adequately studied to be able to determine its validity and reliability. A similar judgment was expressed in two other review articles, which concluded that sacroiliac joint (SIJ) mobility tests were not proven to be reliable.

It has been suggested that interexaminer reliability may be improved by combining results from several tests into a composite multitest score (MTS). Haas noted, based on probability calculations, that the expected rate of agreement is lowest when a middle threshold value is used (eg, three of five tests are required to yield positive results before the MTS is considered positive). Thus, the \( \kappa \) (kappa) statistic is theoretically more likely to be greater when middle threshold values are used. However, this concept has only been evaluated in a few studies. Two of these studies evaluated one method of combining the results of four tests to determine the presence of SIJ dysfunction, and had conflicting findings. The method used by Cibulka et al required positive results from at least three of four tests before results of the MTS were considered conclusive. The authors showed that a cluster of four tests had substantial interexaminer reliability (\( \kappa = 0.88 \)). When the same four tests were reevaluated in a multicenter study by Freburger and Riddle, the interexaminer reliability was found to be fair (\( \kappa = 0.23 \)). A study evaluating the MTSs of four SIJ provocation tests noted substantial reliability (\( \kappa = 0.7 \)) when using a method that required five tests, with the results of at least three being positive prior to diagnosis. However, these studies did not adequately evaluate the benefit of the MTS because they only presented the reliability of the resulting composite scores and not the interexaminer reliability of the individual tests for comparison.

Two studies did present the reliability of individual tests and the MTS. Keating et al evaluated 46 subjects and showed only slightly stronger reliability with the MTS. Boline et al did not show improvement with the MTS. These two studies suggest that MTSs do not improve interexaminer reliability when compared with the individual test results. However, these studies only looked at one method of combining the individual test results.

The effect of MTSs on the interexaminer reliability of diagnoses for the sacral and innominate bone positions was not examined by any of the studies mentioned previously. When treating patients using a directed manual treatment program, it is not enough to simply determine the presence of an SIJ dysfunction. The results of at least two of the individual
**Sacroiliac Joint Dysfunction Tests**

<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>seated flexion test</strong></td>
<td>Used to determine the side of the dysfunction of the sacrum moving on the innominate bone. With the subject sitting and the evaluator seated behind the subject, the evaluator palpates the inferior slopes of the bilateral posterior superior iliac spines. As the subject bends forward, the evaluator’s thumbs follow the motion of the posterior superior iliac spines cephalad. If one side moves more cephalad than the other side by more than 1 cm, the side that moves more is considered abnormal. The dysfunction is recorded as symmetric, left, or right.</td>
</tr>
<tr>
<td><strong>standing stork test</strong></td>
<td>Used to determine the side of the dysfunction of the innominate bone moving on the sacrum. With the subject standing and the evaluator seated behind the subject, the evaluator’s left thumb palpates the inferior slope of the posterior superior iliac spines (PSISs), and the right thumb palpatates the midline of the sacrum at the same level. The subject then flexes the left hip and knee to a minimum of 90 degrees. The sacroiliac joint motion is considered normal if the thumb on the PSISs moves caudal and abnormal if the thumb on the PSISs does not move or if it rises. The test is then repeated on the right side. The dysfunction is recorded as bilateral, left, or right.</td>
</tr>
<tr>
<td><strong>standing flexion test</strong></td>
<td>With the subject standing and the evaluator seated behind the subject, the evaluator palpates the inferior slopes of the bilateral posterior superior iliac spines. As the subject bends forward to touch the floor, the evaluator’s thumbs follow the posterior superior iliac spines cephalad. If one side moves more cephalad than the other side by greater than 1 cm, that is considered abnormal. The dysfunction is recorded as symmetric, left, or right.</td>
</tr>
</tbody>
</table>

**Innominate Bone and Sacral Base Position Tests**

<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>sacral base position with trunk flexion and trunk extension</strong></td>
<td>With the subject sitting and the evaluator seated behind the subject, the evaluator palpates the sacral base with the subject’s trunk forward flexed and backward flexed. The results of this test are abnormal if one side of the sacral base is anterior or posterior relative to the other side with trunk flexion or trunk extension. The dysfunction is recorded as symmetric, left-base anterior (flexion) or posterior (extension), or right-base anterior (flexion) or posterior (extension).</td>
</tr>
<tr>
<td><strong>supine ASIS symmetry</strong></td>
<td>With the patient lying in the supine position after bridging and traction are applied to straighten the body and legs, the evaluator palpates the inferior slope of the anterior superior iliac spine. The findings of the standing stork test, the standing flexion test, or both establish any restricted range of motion. If the anterior superior iliac spine on the side of dysfunction is more cephalad or caudal than the contralateral side by greater than 1 cm, that side is abnormal. The dysfunction is recorded as symmetric, left-side cephalad, or right-side cephalad.</td>
</tr>
<tr>
<td><strong>supine medial malleolus symmetry</strong></td>
<td>With the patient lying in the supine position after bridging and traction are applied to straighten the body and legs, the evaluator palpates the inferior aspect of the bilateral medial malleoli. If the medial malleolus on the side of dysfunction is more cephalad or caudal than the contralateral side by greater than 1 cm, that side is considered abnormal. The dysfunction is recorded as symmetric, left-side cephalad, or right-side cephalad.</td>
</tr>
</tbody>
</table>

**Figure.** Palpatory examinations evaluating the presence of SIJ dysfunction, the sacral position, and the innominate position. Adapted from Greenman.11

Palpatory examinations need to be combined to obtain diagnoses for the sacral and innominate bone positions.11

The purpose of the current study is to compare the interexaminer reliability of three methods of combining the results of osteopathic palpatory examinations to determine the side of SIJ dysfunction, sacral base position, and innominate bone position.

**Methods**

Between December 2002 and April 2003, new patients seen at a university spine center each underwent two separate evaluations by two physicians performing identical palpatory examinations. The first evaluator (H.C.T.) was aware of the patients’ clinical histories, while a second evaluator (O.G.H., D.A.L., or M.M.I.) was blinded to these histories and current medical status as well as the results obtained by the first examiner. Subjects were included in the study if they had a chief complaint of low back pain. Subjects were excluded from the study if they could not tolerate the physical examination as a result of pain. Demographic variables, including age, sex, height, and weight, were recorded.
Palpatory examinations were conducted to evaluate three variables: the presence and side of SIJ dysfunction, the sacral position, and the innominate bone position (Figure). For all of the tests, examiners used their dominant eyes as recommended by Greenman.11 The institutional review board of the University of Michigan Medical School in Ann Arbor approved the study, and informed consent was obtained from all subjects.

### Diagnostic Methods

- **Method 1** used the examination that had the best interexaminer reliability to determine the diagnoses for the sacral and innominate bone positions.
- **Method 2** required that at least one palpatory examination reveal dysfunction for a variable to be considered abnormal.
- **Method 3** required that all of the palpatory examinations reveal dysfunction for a variable to be considered abnormal.

The resulting data from the three different models were then combined to determine the side of SIJ dysfunction, sacral position, and innominate bone position.

### Statistical Analysis

Data were analyzed using SPSS software (version 10.1; SPSS Inc, Chicago, Ill). Because all of the tests and the diagnoses for the sacral and innominate bone positions are categorical variables, the value of $\kappa$ was calculated to determine interexaminer reliability.12 The $\kappa$ statistic reports the amount of agreement seen after adjusting for the amount of agreement that is expected to occur by chance alone. Landis and Koch14 recommended using a $\kappa$ coefficient of 0.2 as the lower limit for fair, 0.4 for moderate, 0.6 for substantial, and 0.8 for almost perfect reliability.

### Results

Twenty-four subjects were chosen chronologically. No subject was excluded from the study as a result of exclusion criteria. The demographics of the subjects are described in Table 1.
mean age of participants was 68.3 years, and 15 patients (62%) were women. The mean height of participants was 1.7 m, the mean weight was 78.3 kg, and the mean body mass index was 27.8 kg/m².

The interexaminer reliability of each palpatory examination is summarized in Table 2. Of the examinations for SIJ dysfunction, the standing stork test had the best interexaminer reliability, with a $\kappa$ of 0.27 ($P = .07$), and the seated flexion test had the worst interexaminer reliability, with a $\kappa$ of -0.06 ($P = .68$). The standing flexion test had a $\kappa$ of 0.14 ($P = .37$). Sacral base position with trunk flexion ($\kappa = .37; P = .002$) had better interexaminer reliability than sacral base position with trunk extension ($\kappa = .05; P = .26$). For the innominate bone position tests, the medial malleolus symmetry test ($\kappa = .21; P = .3$) had better interexaminer reliability than the supine anterior superior iliac spine symmetry test ($\kappa = .15; P = .48$).

The interexaminer reliability of the resulting sacral position diagnosis and innominate bone position diagnosis for the three diagnostic methods are summarized in Table 3. The interexaminer reliability of sacral position when divided into all nine possible categories was incalculable due to the large number of categories and small number of subjects. Consequently, we calculated the interexaminer reliability of sacral position as characterized by the two components that determine sacral position: sacral base position (normal, anterior, or posterior) and side of dysfunction (left, right, or bilateral).

Table 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sacral position†</td>
<td>0.08</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td>Base position</td>
<td>0.32</td>
<td>0.40</td>
<td>0.10</td>
</tr>
<tr>
<td>Side of dysfunction</td>
<td>0.47</td>
<td>0.16</td>
<td>0.32</td>
</tr>
<tr>
<td>Innominant position</td>
<td></td>
<td></td>
<td>-0.33</td>
</tr>
</tbody>
</table>

*Method 1: For each variable, use the test result with the highest interexaminer reliability; Method 2: A variable is abnormal if at least one test result is abnormal; Method 3: A variable is abnormal only if all test results for that variable are abnormal.

† The interexaminer reliability of sacral position when divided into all nine possible categories was incalculable due to the large number of categories and small number of subjects. Consequently, we calculated the interexaminer reliability of sacral position as characterized by the two components that determine sacral position: sacral base position (normal, anterior, or posterior) and side of dysfunction (left, right, or bilateral).

Comment

Several different palpatory examinations are used to evaluate joint motion as well as joint position to detect any abnormality. It has been suggested that combining results from several tests to form a composite MTS will increase interexaminer reliability. In fact, Haas recommends that if an MTS is used, based on expected random chance agreements, an intermediate threshold score should be used to “ensure moderate agreement.” However, this theory has not been tested.
OREIGNAL CONTRIBUTION

with empiric data to ascertain if it maximizes the interexaminer reliability of MTS.

Maximizing interexaminer reliability is essential, as most of the studies evaluating palpatory examinations of the sacrum and pelvis have shown poor to fair interexaminer reliability.1 When they noted poor interexaminer reliability, Flynn et al15 correctly decided not to include palpatory examinations in their analyses. Dreyfuss et al16 noted poor interexaminer reliability with the palpatory examinations yet still used their results in the analyses. However, they did not explain whether they used the physician's findings or the chiropractor's findings, calling into question the validity of their conclusion that the palpatory examination had poor sensitivity and specificity in determining SIJ pain.16

The current study found that using the test with the best interexaminer reliability (method 1) consistently yielded the score with the highest interexaminer reliability. For example, the standing stork test had the best interexaminer reliability in testing for SIJ dysfunction. Using looser criteria (method 2) had slighter better interexaminer reliability to determine the side of SIJ dysfunction. However, this method also had significantly worse reliability when determining the innominate bone position. Using stricter criteria (method 3) consistently had the worst interexaminer reliability for both sides of SIJ dysfunction and innominate bone position.

Previously mentioned studies either did not give the interexaminer reliability of the individual tests5,7,17 or they evaluated the interexaminer reliability of only one method of combining multiple palpatory examination results.9,18 Our study does not support the recommendation by Haas4 that an interexaminer reliability of the individual tests5,7,17 or they evaluated the interexaminer reliability of the individual tests5,7,17 for each patient.

Our findings suggest that maximizing the interexaminer reliability is a prerequisite to conducting studies that truly evaluate the sensitivity and specificity of the palpatory examination, and thus validate this aspect of osteopathic medicine. In addition, maximizing interexaminer reliability is important for clinical care because prescribed manual treatments are based on the results of the palpatory examination. By using the most reliable method to diagnose the cause of low back pain, the physician can be more confident in his or her treatment decisions.

Our study has several strengths. Interexaminer reliability was evaluated for a variety of palpatory examinations and diagnostic methods. In addition, we examined the interexaminer reliability of examinations that detect sacral position and innominate bone position in addition to SIJ.

When interpreting the results of this study, several limitations should be considered. First, the results need to be replicated by other studies. A larger study with more tests to further evaluate this issue is planned by the authors. Also, even though the medial malleolus symmetry test has better reliability than the supine anterior superior iliac spine symmetry test, the former test may not be a valid measure of innominate bone position if the subject has a significant leg length discrepancy. Finally, interexaminer reliability is not the only factor to determine what integration method should be used to diagnose structural dysfunction. Sensitivity and specificity may take precedence over reliability in certain instances.

Our study shows that the maximum interexaminer reliability occurs when only the result of the most reliable test is used to determine the side of SIJ dysfunction, sacral base position, and innominate bone position. Therefore, this method should be used when making clinical management decisions to ensure that the most appropriate treatment is implemented for each patient.

References