Research Report

Evaluation of the Presence of Sacroiliac Joint Region Dysfunction Using a Combination of Tests: A Multicenter Intertester Reliability Study

Background and Purpose. The authors examined the intertester reliability of assessments made based on a composite of 4 tests of pelvic symmetry or sacroiliac joint (SIJ) movement that are advocated in the literature for identifying people with SIJ region dysfunction. "Sacroiliac joint region dysfunction" is a term used to describe pain in or around the region of the joint that is presumed to be due to malalignment or abnormal movement of the SIJs. Subjects. Sixty-five patients with low back pain and unilateral buttock pain were seen in 1 of 11 outpatient clinics. Methods. Thirty-four therapists, randomly paired for each subject, served as examiners. Kappa coefficients and observed proportions of positive (P_{pos}) and negative (P_{neg}) agreement were calculated to estimate reliability. Results. For the composite test results, percentages of agreement ranged from 60% to 69%, kappa coefficients varied from .11 to .23, and P_{pos} was lower than 50%. Discussion and Conclusion. Reliability of measurements obtained with the 4 tests appears to be too low for clinical use. Given the measurement error found in this study, the authors suspect it is likely that either the proper treatment technique will not be chosen based on the test results or the intervention will be applied to the wrong side. The 4 tests probably should not be used to examine patients suspected of having SIJ region dysfunction, although the role of therapist training in use of the procedures is unclear. [Riddle DL, Freburger JK, North American Orthopaedic Rehabilitation Research Network. Evaluation of the presence of sacroiliac joint region dysfunction using a combination of tests: a multicenter intertester reliability study. *Phys Ther.* 2002;82:772–781.]

Key Words: *Kappa, Measurement, Reliability, Sacroiliac joint.*

Daniel L Riddle, Janet K Freburger, North American Orthopaedic Rehabilitation Research Network*

Many diagnostic tests have been developed to identify what is thought to be a dysfunction of the region of the sacroiliac joint (SIJ). "Sacroiliac joint region dysfunction" is a term used to describe pain in or around the region of the joint[1] that is presumed to be due to malalignment or abnormal movement of the SIJs.[2] Magee,[3] for example, described 31 tests that have appeared in the literature for use on patients suspected of having SIJ region dysfunction. Studies designed to determine the psychometric properties of diagnostic tests for the SIJ region began to appear in the literature in 1985[4] and have begun to appear more frequently.[5–9] We reviewed this literature and found there is evidence for the reliability and weak evidence for the diagnostic validity of data obtained with some measures designed to provoke pain from patients suspected of having SIJ region dysfunction.[10] Cibulka and colleagues[11] provided the only data we found to support the reliability of data obtained with measures designed to determine the alignment or movement of the SIJs on patients suspected of having SIJ region dysfunction.

Cibulka and colleagues[11] defined SIJ region dysfunction as being present if at least 3 of the following 4 tests were positive: the standing flexion test, the prone knee flexion test, the supine long sitting test, and palpation of posterior superior iliac spine (PSIS) heights in a sitting position. Two therapists with an unspecified amount of training in the test procedures examined 26 patients with low back pain or buttock pain. Intertester agreement for determining the presence of SIJ region dysfunction was high (κ=.88).

Cibulka and colleagues implied in articles published in 1988[11] and 1999[12] that tests were classified simply as positive or negative, regardless of whether the tests indicated dysfunction on the right or left side and regardless of the type of asymmetry present (ie, whether the tests indicated the possibility of an anteriorly or posteriorly rotated innominate). For example, the supine long sitting test could be graded as positive for

* Participating clinics from the North American Orthopaedic Rehabilitation Research Network were Conroy Orthopaedic & Sports Physical Therapy, Flossmoor, Ill, Life Care Medical Center, Glassboro, NJ, Appalachian Physical Therapy Inc, Dahlonega, Ga, Physiotherapy on Bay, Toronto, Ontario, Canada, Pro Active Physiotherapy, Hamilton, Ontario, Canada, West End Physiotherapy Clinic, Hamilton, Ontario, Canada, Canadian Sport Rehabilitation Institute, Calgary, Alberta, Canada, Rehab Plus Associates, Midlothian, Va, Walser Physiotherapy, Thunder Bay, Ontario, Canada, Sooke Evergreen Physiotherapy, Sooke, British Columbia, Canada, and St Joseph's Hospital, Hamilton, Ontario, Canada.

DL Riddle, PT, PhD, is Associate Professor, Department of Physical Therapy, Medical College of Virginia Campus, Virginia Commonwealth University, 1200 E Broad St, Richmond, VA 23298-0224 (USA) (driddle@hsc.vcu.edu). Address all correspondence to Dr Riddle.

JK Freburger, PT, PhD, is NRSA Postdoctoral Research Fellow, Cecil G Sheps Center for Health Services Research, and Assistant Professor, Division of Physical Therapy, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Dr Riddle and Dr Freburger provided concept/research design, writing, and data collection and analysis. Dr Riddle provided project management, and Dr Freburger provided fund procurement. The North American Orthopaedic Rehabilitation Research Network provided subjects, facilities/equipment, and institutional liaisons. Carissa A Bennett and Andrew C Gallo provided clerical support.

This study was approved by the Institutional Review Board of Virginia Commonwealth University.

This work was supported, in part, by a National Research Service Award Postdoctoral Traineeship from the Agency for Healthcare Research and Quality and sponsored by the Cecil G Sheps Center for Health Services Research Grant T32-HS00032.

This article was submitted May 2, 2001, and was accepted April 3, 2002.

any 1 of the following 4 conditions: right innominate posteriorly rotated, left innominate posteriorly rotated, right innominate anteriorly rotated, and left innominate anteriorly rotated. Therapists, therefore, may have agreed that 3 or more tests were positive without agreeing on the side involved or the type of asymmetry present. Cibulka and colleagues did not describe whether these types of disagreements were addressed and implied that tests were graded simply as positive or negative. We suspect this is the case because the manipulative intervention advocated by Cibulka and colleagues was designed for use regardless of the type of asymmetry that was present.[13]

Several authors[11,14–19] have suggested that examination and management of people with SIJ region dysfunction sometimes require identification of the involved side, type of asymmetry present, and correction of the asymmetry. For example, mobilization techniques designed to treat what is thought to be a posteriorly rotated innominate on the right side are different from those techniques designed to treat a suspected left posteriorly rotated innominate.[20] It would appear to be important to know the degree of agreement, not only for judgments of the presence or absence of SIJ region dysfunction, but also for the type of asymmetry thought to be present.

The study of Cibulka and colleagues[11] is especially important because their study provides the only evidence that suggests that assessments of innominate alignment or motion, when used in combination, have clinical utility. In our experience, tests requiring the assessment of innominate bone symmetry or movement are commonly done in practice. We believe that a study that is more generalizable than that of Cibulka et al would provide clinicians with additional information that could be used to determine appropriate examination strategies for SIJ region dysfunction. The purposes of our study were: (1) to replicate the study of Cibulka and colleagues[11] on a larger group of patients and with a larger group of therapists and (2) to examine the degree of agreement between therapists by taking into account the side of the presumed dysfunction and the type of asymmetry present.

Method

Examiners

The examiners were 34 therapists working in 11 clinics located in either the United States or Canada. Only therapists who regularly treated patients with low back pain were included in the study. Table 1 presents descriptive information on the participating therapists.

Each of the participating therapists was given a written description of the 4 examination procedures and photographs of the procedures. The photographs illustrated the beginning and ending positions of the patient and the position of the therapist for each test. Participating therapists were instructed to practice the examination procedures on each other and then on patients. All therapists in each clinic had to indicate that they felt comfortable they were conducting the tests properly before data collection began in that clinic. No other information or advice was given to the therapists.

Table 1.
Therapist Characteristics (n=34)

Therapist Characteristic[a]	$\bar{X}$	SD	Range
No. of years as a therapist	11.4	7.3	1–30
No. of years treating patients with LBP	10.1	6.6	1–28
Percentage of caseload represented by patients with LBP	33.4	16.3	5–85
Percentage of caseload represented by patients with SIJ region dysfunction	11.6	10.0	0–50
No. of continuing education courses taken that dealt with SIJ region dysfunction	3.1	1.8	0–8

[a] LBP=low back pain, SIJ=sacroiliac joint.

Subjects

A total of 65 patients participated in the study. To be included in the study, patients had to: (1) be between 18 and 65 years of age, (2) be referred for treatment of a low back problem, (3) have unilateral or bilateral low back pain, (4) be a new patient or a patient who was currently receiving treatment for a low back problem, (5) have discomfort reported in the area of the buttock at the time of admission to the study, and (6) be able to reach at least the level of the patellae with their fingertips when flexing the lumbar spine while standing with the knees extended. This motion was necessary to complete 2 of the tests that were studied. The region of the buttock was defined as having the following boundaries: the iliac crest superiorly, the gluteal fold inferiorly, the sacral spinous processes medially, and the greater trochanter laterally. Pain also could be reported anywhere in the involved lower extremity. We admitted only patients with unilateral buttock pain so that therapists could describe their test results relative to the symptomatic side. Patients were excluded if they: (1) had lumbar surgery within the year prior to the study and (2) reported lower-extremity parasthesias or muscle weakness. Characteristics of the patients are presented in Table 2.

Procedure

After completing an institutional review board-approved consent form, each patient recorded his or her age, height, weight, and sex on a form. In addition, patients indicated the duration of their back problem and whether their work status (on the job or at home) was

Table 2.
Patient Characteristics (n=65)

Patient Characteristic	
Age (y)	
$\bar{X}$	47.4
SD	14.0
Range	18–81
Sex	
Female	42 (65%)
Male	23 (25%)
Height (cm)	
$\bar{X}$	169.6
SD	10.7
Range	152.4–193.0
Weight (kg) (n=64)	
$\bar{X}$	72.3
SD	13.2
Range	49.0–108.9
Body mass index[a]	
$\bar{X}$	25.0
SD	3.4
Range	19.2–36.3
Pain rating on 10-cm visual analog scale	
$\bar{X}$	3.4
SD	2.1
Range	0–7.7
Duration of pain (wk)	
$\bar{X}$	45.2
SD	80.4
Range	1–330
Pain distribution (n=58)	
Back	52 (89%)
Buttock	54 (93%)
Thigh	16 (28%)
Leg	11 (19%)
Foot	3 (5%)
Pain interfering with job or housework (n=60)	
Yes	44 (73%)
No	16 (27%)

[a] Body mass index calculated as: [weight (kg)/height(m)^2].

affected by their back problem. The physical therapist who identified the eligible patient (evaluating physical therapist) also completed a form that indicated the distribution of the patient's pain and the pain intensity by use of a visual analog scale. The evaluating physical therapist then identified the retest physical therapist from a random list of the participating therapists for that clinic. The evaluating physical therapist conducted the examinations first, out of sight of the retest physical therapist. The evaluating physical therapist conducted the 4 procedures in the following order and recorded the results on a form: standing flexion test, prone knee flexion test, supine long sitting test, and sitting PSIS test. This order was chosen because Cibulka et al[11] appeared to conduct the tests in this order in the original study that described reliability for the measures. The procedures for each of the tests were done as described by Cibulka et al.[11] We asked the therapists to identify whether the test was positive on the left side or the right side, and we asked them to identify the type of asymmetry if one was found. The possible findings for each test are summarized in Table 3.

A few minutes after the evaluating physical therapist completed the measurements, the retest therapist conducted the examinations. The retest therapist first had the patient rate his or her pain intensity using a visual analog scale and then conducted the same tests in the same order as the evaluating physical therapist.

Data Analysis

Results of the 4 tests advocated by Cibulka and colleagues[11,12] were combined, and if 3 of the 4 tests were positive, the patient was considered to have SIJ region dysfunction. The Figure illustrates the 3 approaches we used for examining composite scores from the 4 tests. First, the test results can be dichotomized and rated as positive or negative, independent of whether they indicate that the same impairment is present on the same side. This is the method that Cibulka et al[11] appeared to use. For our first analysis, we collapsed all positive ratings (independent of the side and type of asymmetry determined to be present) and determined the extent of agreement when paired therapists rated 3 or more tests as positive or negative.

For our second composite analysis, we examined whether therapists agreed on 1 of the following 3 judgments: 2 or more tests were negative, 3 or more tests indicated dysfunction on the right side, or 3 or more tests indicated dysfunction on the left side. In this analysis, therapists did not necessarily have to agree on the type of asymmetry present (ie, anteriorly or posteriorly rotated innominate), just the side that was involved. For example, if a therapist concluded that the supine long sitting test indicated an anteriorly rotated innominate on the left side, the prone knee flexion test indicated a posteriorly rotated innominate on the left side, and the standing flexion test was positive on the left side (presumably indicating a hypomobile left SIJ), the composite score was positive left.

In our third analysis, we determined the extent of agreement for a 5-category scale (anteriorly rotated on the right side, anteriorly rotated on the left side, negative, posteriorly rotated on the right side, and posteriorly rotated on the left side). We chose this scale because 3 of the 4 tests we examined are used to determine whether an innominate was rotated relative to the other innominate (Tab. 3). Therefore, if 3 tests are positive, at least 2 of the 3 tests will be indicative of a rotated innominate. For the third analysis, therapists had to agree on the side involved (right, left, or none) and the type of asymmetry that was present (anteriorly or posteriorly rotated

Table 3.
Description of the Interpretation of the Possible Findings for Each Diagnostic Test[a]

Diagnostic Test	Possible Findings	Interpretation
Standing flexion test	Negative	PSISs appear to move equally
	Positive on the right side	Right PSIS moves cranially more than left PSIS (right SIJ hypomobile)
	Positive on the left side	Left PSIS moves cranially more than right PSIS (left SIJ hypomobile)
Prone knee flexion test	Negative	No relative change in leg lengths between the 2 test positions
	Posteriorly rotated innominate on the right side	Symptoms are on the right side, the right leg appears shorter than the left leg in the prone knee extended position, and the right leg appears to be about equal to or longer than the left leg in the prone knee flexed position
	Posteriorly rotated innominate on the left side	Symptoms are on the left side, the left leg appears shorter than the right leg in the prone knee extended position, and the left leg appears to be about equal to or longer than the right leg in the prone knee flexed position
	Anteriorly rotated innominate on the right side	Symptoms are on the right side, the right leg appears to be longer than left leg in the prone knee extended position, and the right leg appears to be about equal to or shorter than the left leg in the prone knee flexed position
	Anteriorly rotated innominate on the left side	Symptoms are on the left side, the left leg appears longer than the right leg in the prone knee extended position, and the left leg appears to be about equal to or shorter than the right leg in the prone knee flexed position
Supine long sitting test	Negative	No relative change in leg lengths between the 2 test positions
	Posteriorly rotated innominate on the right side	Symptoms are on the right side, the right leg appears shorter than the left leg in supine position, and the right leg appears to be about equal to or longer than the left leg in long sitting position
	Posteriorly rotated innominate on the left side	Symptoms are on the left side, the left leg appears shorter than the right leg in supine position, and the left leg appears to be about equal to or longer than the right leg in long sitting position
	Anteriorly rotated innominate on the right side	Symptoms are on the right side, the right leg appears longer than the left leg in supine position, and the right leg appears to be about equal to or shorter than the left leg in long sitting position
	Anteriorly rotated innominate on the left side	Symptoms are on the left side, the left leg appears longer than the right leg in supine position, and the left leg appears to be about equal to or shorter than the right leg in long sitting position
Sitting PSIS test	Negative	PSISs are symmetrical
	Positive on the right side	Right PSIS lower than left PSIS (left anteriorly rotated innominate if pain on left side; right posteriorly rotated innominate if pain on right side)
	Positive on the left side	Left PSIS lower than right PSIS (right anteriorly rotated innominate if pain on right side; left posteriorly rotated innominate if pain on left side)

[a] PSIS=posterior superior iliac spine, SIJ=sacroiliac joint.

innominate) for at least 3 tests. For example, if a therapist concluded that the supine long sitting test indicated the presence of a posteriorly rotated innominate on the left side, the sitting PSIS test was positive on the left side (indicating the presence of a posteriorly rotated innominate on the left side), and the standing flexion test was positive on the left side (indicating a hypomobile left SIJ), the composite score was posteriorly rotated innominate on the left side.

Percentages of agreement and Cohen kappa statistic (κ) coefficients were calculated for the individual tests and for the 3 composite test results. Because we suspected that the distribution of our data would be skewed, we also calculated the maximum kappa (κ_{max}) and kappa/kappa maximum (κ/κ_{max}) values.[21] The latter value indicates the proportion of agreement achieved by the therapists, taking into account the maximum kappa value possible.[21] The maximum kappa value can be useful when the kappa value is low despite a high observed proportion of agreement.[22] In our study, we suspected a proportionally large number of negative findings because 3 of the 4 tests needed to be positive to indicate a positive composite result. A large proportion of negative results would increase the likelihood of agreement by chance and subsequently reduce the kappa value.

We calculated the observed proportion of positive agreement (P_{pos}) and the observed proportion of negative agreement (P_{neg}).[23] These indices indicate whether disagreements are more likely for positive or negative

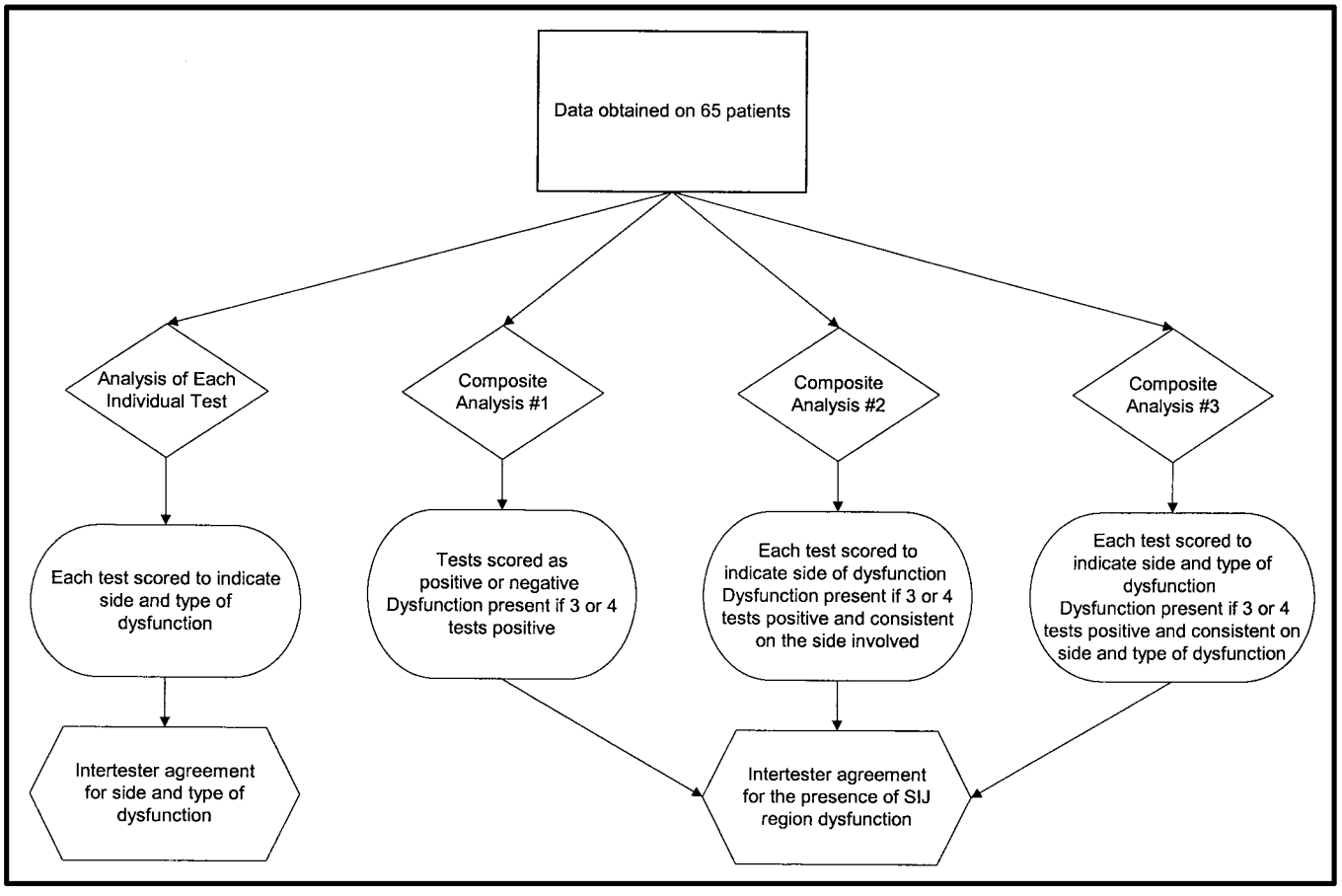

Figure.
Depiction of the approaches used to assess reliability for the composite scores and the individual scores of the 4 tests. SIJ=sacroiliac joint.

judgments, thus helping to resolve the paradoxical results of a high proportion of agreement but a low kappa.

Cicchetti and Feinstein[23] provided several examples to illustrate how P_{pos} and P_{neg} can help clarify the meaning of a low kappa coefficient and a high percentage of agreement. In one example, the percentage of agreement for a set of dichotomous data was 85% and the kappa coefficient was .70. The corresponding P_{pos} was .84, and the P_{neg} was also high at .86. A second example had an identical percentage of agreement of 85%, but the kappa coefficient was .32. The corresponding P_{pos} was .91, but the P_{neg} was much lower at .40. One reason for the relatively low kappa coefficient in the second example was that the raters frequently disagreed on judgments of negative test results.

Typically, a generalized kappa statistic is used to describe the degree of agreement corrected for chance when many potential pairs of raters participate in the study, a scenario consistent with our study.[24] We chose to calculate the Cohen kappa coefficients because we found no methods in the literature for calculating a maximum kappa coefficient from a generalized kappa coefficient. Cicchetti (personal communication, 2001) also suggested that the Cohen kappa statistic should be used when calculating P_{pos} and P_{neg}.

To determine whether the use of the Cohen kappa statistic in place of the generalized kappa statistic was appropriate, we calculated both a Cohen kappa coefficient and a generalized kappa coefficient for each of the 3 composite analyses. If these coefficients were essentially equal for each of the analyses, we believed it was acceptable to use the Cohen kappa statistic in place of the generalized kappa statistic. The 2 forms of kappa coefficients were identical for the first composite analysis (κ=.18) and the second composite analysis (κ=.11) and differed by .04 for the third composite analysis (Cohen κ=.23, generalized κ=.27). We therefore considered it appropriate to use the Cohen kappa statistic in place of the generalized kappa statistic for all analyses.

Results

Kappa coefficients for individual tests varied from .19 (SE=.09) to .37 (SE=.10), and percentages of agreement varied from 44.6% to 63.1%. The therapists

Table 4.
Intertester Reliability of the Individual Tests[a]

Test	% Agreement	κ (SE)	κ_{max}	κ/κ_{max}	P_{pos}	P_{neg}
Standing flexion test	55.4	.32 (.09)	.79	40.5	56.7	51.5
Prone knee flexion test	60.0	.26 (.10)	.91	28.6	40.9	69.8
Supine long sitting test	44.6	.19 (.09)	.90	21.1	35.7	48.6
Sitting posterior superior iliac spine test	63.1	.37 (.10)	.93	39.8	55.6	68.4

[a] κ=kappa coefficient, κ_{max}=maximum kappa coefficient, κ/κ_{max}=kappa coefficient divided by kappa maximum coefficient, P_{pos}=observed proportion of positive agreement, P_{neg}=observed proportion of negative agreement.

Table 5.
Intertester Reliability for Composite Results of the Four Tests[a]

Finding	% Agreement	κ (SE)	κ_{max}	κ/κ_{max}	P_{pos}	P_{neg}
Composite analysis 1: 3 of 4 tests (+ or −)	61.5	.18 (.12)	.89	20.2	49.0	69.1
Composite analysis 2: 3 of 4 tests (+ right, + left, negative)	60.0	.11 (.11)	.90	12.2	30.0	68.9
Composite analysis 3: 3 of 4 tests (anterior right, anterior left, negative, posterior right, posterior left)	69.2	.23 (.12)	.85	27.1	33.3	80.0

[a] κ=kappa coefficient, κ_{max}=maximum kappa coefficient, κ/κ_{max}=kappa coefficient divided by kappa maximum coefficient, P_{pos}=observed proportion of positive agreement, P_{neg}=observed proportion of negative agreement.

achieved between 21.1% and 40.5% of the maximum kappa value for each of the 4 tests (Tab. 4).

For the composite test results, kappa coefficients varied from .11 (SE=.11) to .23 (SE=.12), and percentages of agreement varied from 60% to 69.2%. Therapists achieved between 12.2% and 27.1% of the maximum kappa value, P_{pos} varied from 30% to 49%, and P_{neg} varied from 68.9% to 80% (Tab. 5).

Discussion

Based on the percentages of agreement, the kappa values, and the kappa/kappa maximum values, we found what we consider to be poor reliability for the individual tests. Potter and Rothstein[4] reported slightly lower percentages of agreement for the same 4 tests (23.5%-43.8%). We are unsure why our percentages of agreement were slightly higher than those reported by Potter and Rothstein. We believe, however, that error on the order of 40% or more that is not corrected for chance agreement is unacceptable for individual patient decision making. We also consider the kappa values for each of the 4 tests to be unacceptable for clinical use, especially in light of the kappa/kappa maximum values. Using the P_{pos} values, therapist agreement was less than 60% when one therapist found a positive test result. We therefore agree with the recommendations of Potter and Rothstein[4] and Cibulka et al,[11] who discouraged the use of these tests in isolation.

Reliability exists along a continuum from no agreement (eg, κ=0) to perfect agreement (eg, κ=1). Landis and Koch[25] suggested that kappa values from .21 to .40 indicate "fair" agreement, an admittedly arbitrary label that does not take into account how a measurement is used and the consequences of a wrong decision. Although our data indicate that agreement for the individual tests exceeded that expected due to chance, we contend that reliability is too low for making treatment decisions on individual patients.

Many of the various interventions proposed for patients with SIJ region dysfunction typically require the therapist to identify the type of dysfunction present or the side of involvement.[11,16–20] We believe that therapists who use the 4 tests we examined to identify the type of dysfunction or the side of involvement are likely to deliver interventions to individuals who do not have a dysfunction or to deliver interventions incorrectly (either the proper technique will not be chosen or the intervention will be applied to the wrong side). In the latter case, the individual's problem, theoretically, could be exacerbated following the intervention. For example, if the cause of the individual's buttock pain is an anteriorly rotated innominate on the left, but the therapist determines that the individual's innominate is posteriorly rotated on the left, interventions to correct the posteriorly rotated innominate, theoretically, could exacerbate the problem.

More research is needed to guide clinicians on the choice of examination procedures and interventions for patients with pain that may be arising from the SIJ region. Until that research is done, alternative test

procedures such as pain provocation tests would likely provide therapists with more reliable and, theoretically, more useful information than tests of SIJ alignment or movement.

We also found what we consider to be poor reliability for the composite results from the 4 tests classified as positive or negative. Our kappa coefficient for these dichotomized judgments was .18, and the kappa/kappa maximum value was 20.2%.

In contrast, Cibulka et al[11] reported a kappa coefficient of .88. One factor that can lower the kappa coefficient is a low prevalence of the condition of interest. In our study, a relatively small number of patients had a composite score of positive, indicating the presence of SIJ region dysfunction based on Cibulka and colleagues' criteria. A total of 38% of all dichotomous composite judgments in our study were rated as positive. However, as can be seen by the kappa maximum, this relatively small percentage of positive test results was not the primary reason for the low kappa value. One likely explanation for the low kappa value was the very low P_{pos} of 49%. That is, when one therapist rated a composite score of positive, the other therapist rated the same patient as positive 49% of the time, a number essentially equal to chance. Reliability for the composite scores also appears to us to be too low for clinical use.

It is not clear why our results differed so dramatically from those of Cibulka et al.[11] One potential explanation is that only 2 therapists participated in the study of Cibulka et al, and these therapists worked together and practiced the procedures prior to the study. The therapists also developed the approach. Cibulka et al did not describe the nature and quality of the therapists' training, so it is unclear how this training may have influenced reliability. The therapists in our study did not undergo extensive training. They were instructed to practice the procedures on each other and on patients until they felt ready to use the procedures on patients. The spectrum of patients was different between the study of Cibulka et al and our study. The majority of patients in the study of Cibulka et al reportedly had pain localized to the lumbar area. No patients reportedly had pain below the knee. Patients were admitted to our study only if they reported unilateral buttock pain, a symptom commonly associated with patients thought to have SIJ region dysfunction.[5,6] In addition, approximately 20% of our patients reported pain below the knee, a complaint that is apparently not unusual in patients with SIJ region dysfunction.[5] It is unclear what affect differences in patients' pain distribution may have had on the results of the 2 studies.

We believe our data are more generalizable than those of Cibulka et al.[11] We had 34 therapists participating in our study, whereas Cibulka et al had 2 examiners. We examined 65 patients, whereas Cibulka et al studied 26 patients. Finally, we contend that most therapists who use these techniques likely apply the methods in ways that are similar to those used by the therapists in our study.

The general background and experience of the therapists who participated in our study was extensive (Tab. 2). They had a mean of 10.1 years (SD=6.6, range=1–28) of experience treating patients with low back pain, and they estimated that on average 11.6% (SD=10.0%, range=0% to 50%) of their caseload consisted of patients suspected of having dysfunction of the SIJ region. In addition, therapists reported attending a mean of 3.1 (SD=1.8, range=0–8) continuing education courses that were solely on the evaluation and treatment of the SIJ or that included a section on the SIJ. We believe it is likely that most therapists in our study had seen or had used the tests examined in the study because 3 of the 4 tests are commonly described in many textbooks and, in our experience, are commonly used in practice. However, we did not collect these data. It is our contention that these tests are well defined and that therapists with clinical experiences similar to those of the therapists in our study should be able to conduct these procedures reasonably well.

We examined our data to determine whether we could account for the large amount of error. We examined the pain intensity data to determine whether the patients' reported pain intensity varied between repeated tests. Pain that varies could result in the patient performing repeated tests differently and in therapists finding different results. We calculated an intraclass correlation coefficient (ICC [2,1])[26] to describe the reliability of visual analog scale pain ratings taken by each therapist just prior to taking measurements on a patient. The ICC (2,1) was .97 (95% confidence interval=.95-.98). These data indicate that pain intensity did not vary appreciably between repeated tests and was not a source of error.

We also determined whether reliability differed for patients who were overweight. When patients are overweight, bony landmarks around the pelvis may be more difficult to palpate and could lead to additional error. A total of 31 of our patients had a body mass index (BMI) higher than 25, the criterion for grade 1 obesity.[27] The kappa value for these patients was .21 (SE=.18) for composite judgments of positive or negative test results (composite test 1). The kappa value for patients who were not obese (BMI<25) was .14 (SE=.17). These data strongly suggest that being overweight was not a source of error in the study.

One limitation of our study was the inclusion of data from 4 patients who apparently did not report buttock pain prior to testing. In addition, data indicating pain distribution were missing for 7 patients. Pain distribution was important because therapists were instructed to interpret the supine long sitting test and the prone knee flexion test results relative to the painful side. We conducted an a posteriori analysis of patients with documented unilateral buttock pain (n=54) to determine whether the 11 subjects who did not have confirmed buttock pain influenced the results. The kappa values for the patients with confirmed buttock pain were .17 (SE=.13) for the first composite test, .11 (SE=.12) for the second composite test, and .27 (SE=.12) for third composite test. Reliability was not appreciably affected by inclusion of data from the 11 subjects who may not have had unilateral buttock pain (Tab. 5).

In reliability studies, researchers attempt, among other things, to reduce the error associated with measurements.[23] We were unable to attribute the substantial error in our study to either the therapists or the patients. We believe the most likely source of error related to the nature of the phenomena these measures are designed to assess. The magnitude of rotatory movement in the SIJ is, on average, on the order of only a few degrees.[28–31] We contend that this small amount of movement combined with the inherent variability in size and shape of the innominate bone landmarks[32,33] makes it highly unlikely that most therapists can make reliable judgments based on palpation of bony landmarks on the pelvis.

Although we question whether therapists can make reliable judgments given the variability in bony anatomy and the small amount of SIJ motion, the findings of Cibulka et al[11] suggest that training may have contributed to the high reliability they reported. Unintentional therapist bias is also a possible explanation for their findings. In our study, we used multiple combinations of therapists. We contend that the use of many therapists may have decreased the potential effects of therapist bias on the results. Multiple combinations of paired therapists, however, also limit, to some degree, conclusions about intertester reliability. We conducted a multicenter study, and we randomly paired therapists at each clinic. For practical reasons, we did not examine all possible intertester combinations (ie, all therapists who participated in the study did not evaluate all patients). The results of our study may have differed had we conducted the study in this manner. We also controlled the order in which the 4 tests were conducted. Reliability may have differed with a different order of testing.

Conclusion

The intertester reliability of assessments of the presence of SIJ region dysfunction using a composite of 4 diagnostic tests was poor and was not dependent on the method of classifying the nature of the test results. Reliability for the individual tests was slightly higher than for the composite scores, but we still consider it to be inadequate for clinical use. Given our results and the limited generalizability of the work of Cibulka et al,[11] we recommend an alternative approach for identifying patients suspected of having SIJ region dysfunction. Tests designed to provoke a patient's pain appear to have more support for use in identifying patients who may have SIJ region dysfunction than do tests presumed to measure SIJ alignment or movement.[8]

References

1 Fortin JD, Aprill CN, Ponthieux B, Pier J. Sacroiliac joint: pain referral maps upon applying a new injection/arthrography technique, part II: clinical evaluation. *Spine.* 1994;19:1483–1489.

2 Dreyfuss P, Dreyer S, Griffin J, et al. Positive sacroiliac screening tests in asymptomatic adults. *Spine.* 1994;19:1138–1143.

3 Magee DJ. *Orthopaedic Physical Assessment.* 3rd ed. Philadelphia, Pa: WB Saunders Co; 1997:434–458.

4 Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *Phys Ther.* 1985;65:1671–1675.

5 Dreyfuss P, Michaelsen M, Pauza K, et al. The value of medical history and physical examination in diagnosing sacroiliac joint pain. *Spine.* 1996;21:2594–2602.

6 Maigne JY, Aivaliklis M, Pfefer F. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. *Spine.* 1996;21:1889–1892.

7 Slipman CW, Sterenfeld EB, Chou LH, et al. The predictive value of provocative sacroiliac joint stress maneuvers in the diagnosis of sacroiliac joint syndrome. *Arch Phys Med Rehabil.* 1998;79:288–292.

8 Broadhurst NA, Bond MJ. Pain provocation tests for the assessment of sacroiliac joint dysfunction. *J Spinal Disorders.* 1998;11:341–345.

9 Laslett M, Williams M. The reliability of selected pain provocation tests for sacroiliac joint pathology. *Spine.* 1994;19:1243–1249.

10 Freburger JK, Riddle DL. Using published evidence to guide the examination of the sacroiliac joint region. *Phys Ther.* 2001;81:1135–1143.

11 Cibulka MT, Delitto A, Koldehoff RM. Changes in innominate tilt after manipulation of the sacroiliac joint in patients with low back pain: an experimental study. *Phys Ther.* 1988;68:1359–1363.

12 Cibulka MT, Koldehoff RM. Clinical usefulness of a cluster of sacroiliac joint tests in patients with and without low back pain. *J Orthop Sport Phys Ther.* 1999;29:83–92.

13 Stoddard A. *Manual of Osteopathic Technique.* 3rd ed. London, England: Hutchinson; 1980:83–85.

14 Bernard TN, Cassidy JD. The sacroiliac joint syndrome, pathophysiology, diagnosis, and management. In: Frymoyer JW, ed. *The Adult Spine: Principles and Practice.* 2nd ed. Philadelphia, Pa: Lippincott-Raven; 1997:2343–2366.

15 Cyriax JH. *Textbook of Orthopaedic Medicine.* 11th ed. London, England: Baillière Tindall; 1984.

16 Cibulka MT. The treatment of the sacroiliac joint component to low back pain: a case report. *Phys Ther*. 1992;72:917–922.

17 Dontigny RL. Anterior dysfunction of the sacroiliac joint as a major factor in the etiology of idiopathic low back pain syndrome. *Phys Ther*. 1990;70:250–265.

18 Maitland GD. *Vertebral Manipulation*. 5th ed. Boston, Mass: Butterworths; 1986:314–319.

19 Mitchell FL Jr, Moran PS, Pruzzo NA. *An Evaluation and Treatment Manual of Osteopathic Muscle Energy Techniques*. Valley Park, Mo: Mitchell, Moran and Pruzzo Associates; 1979.

20 Hertling D. The sacroiliac joint and the lumbar-pelvic-hip complex. In: Hertling D, Kessler RM, eds. *Management of Common Musculoskeletal Disorders: Physical Therapy Principles and Methods*. 3rd ed. Philadelphia, Pa: JB Lippincott Co; 1990:698–736.

21 Cohen J. A coefficient of agreement for nominal scales. *Educ Psych Meas*. 1960;20:37–46.

22 Feinstein AR, Cicchetti DV. High agreement but low kappa, I: the problems of two paradoxes. *J Clin Epidemiol*. 1990;43:543–549.

23 Cicchetti DV, Feinstein AR. High agreement but low Kappa, II: resolving the paradoxes. *J Clin Epidemiol*. 1990;43:551–558.

24 Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378–382.

25 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.

26 Shrout PE, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.

27 *American College of Sports Medicine Guidelines for Exercise Testing and Prescription*. 5th ed. Baltimore, Md: Williams & Wilkins; 1995.

28 Sturesson B, Uden A, Vleeming A. A radiostereometric analysis of movements of the sacroiliac joints during the standing hip flexion test. *Spine*. 2000;25:364–368.

29 Sturesson B, Uden A, Vleeming A. A radiostereometric analysis of the movements of the sacroiliac joints in the reciprocal straddle position. *Spine*. 2000;25:214–217.

30 Sturesson B, Selvik G, Uden A. Movements of the sacroiliac joint: a roentgen stereophotogrammetric analysis. *Spine*. 1989;14:162–165.

31 Egund N, Olsson TH, Schmid H, Selvik G. Movements in the sacroiliac joints demonstrated with roentgen stereophotogrammetry. *Acta Radiol Diag*. 1978;19:833–846.

32 Bernard TN, Cassidy JD. Sacroiliac joint syndrome: pathophysiology, diagnosis, and treatment. In: Frymoyer JM, ed. *The Adult Spine*. New York, NY: Raven Press; 1993:2107–2130.

33 Vix VA, Ryu CY. The adult symphysis pubis: normal and abnormal. *Am J Roentgenol Radium Ther Nucl Med*. 1971;112:517–525.