Sacroiliac joint dysfunction: Evidence-based diagnosis

Peter Huijbregts, PT, MSc, MHSc, DPT, OCS, MTC, FAAOMPT, FCAMT
Assistant Online Professor, University of St. Augustine for Health Sciences, St. Augustine, FL, USA
Consultant, Shelbourne Physiotherapy Clinic, Victoria, BC, Canada

Introduction

Low back pain (LBP) is a health problem with a major societal impact. Histology and injection[1-11] studies have established the nociceptive potential and clinical reality of LBP originating in the sacroiliac joint (SIJ) and its peri-articular tissues. Table 1 lists the pathological processes, which can involve the SIJ[12-20]. This article mainly deals with the diagnostic entity of sacroiliac joint dysfunction (SIJD). Paris[21] defined a joint dysfunction as a state of altered mechanics, characterized by an increase or decrease from the expected normal or by the presence of an aberrant motion. This positions SIJD as a patho-mechanical rather than pathological diagnosis[14,22].

The accepted gold standard or reference test for the diagnosis of SIJ-related pain is the fluoroscopically guided intra-articular anaesthetic injection or joint block[2-11,14.] Data on the prevalence of SIJ-related pain, therefore, is limited to highly selected populations of patients with chronic LBP referred for injection studies[4-6,11]. Schwarzer et al[4] found a 30% prevalence with single blocks. Maigne et al[5] reported a prevalence of 18.5% after double joint blocks. Dreyfuss et al[6] noted a 53% positive response to a single SIJ block and Laslett et al[11] confirmed SIJ-related pain in 33% of their subjects with single and double blocks.

A joint block is a highly specialized procedure, hardly available in everyday clinical practice; it is also not indicated for every patient with LBP. Generally, the only means available to the clinician to reach a diagnosis of SIJD are patient history and physical examination. SIJ physical examination comprises an active range of motion (AROM) examination consisting of cardinal and non-cardinal plane motions and special tests considered specific to the SIJ. These special tests fall in three categories[22-24]:

1. Positional palpation tests
2. Motion palpation tests
3. Provocation tests

For history items and physical tests to be clinically useful, the data they yield needs to be reliable, valid, and responsive to clinically relevant change[25]. The goal of this article is to discuss reliability and validity of history items and physical tests thought relevant for making a diagnosis of SIJD. To this end, we will first discuss definitions pertinent to the concepts of reliability and validity relating them to the diagnosis of SIJD. We will then review, in chronological order, research on reliability and validity of history items, AROM tests, individual special tests, multiple test regimens, and a comprehensive examination used for the diagnosis of

Table 1. Pathologies affecting the sacroiliac joint[12-20].

Traumatic conditions
- Fracture-dislocation
- Stress fracture
- Insufficiency fractures

Infectious conditions
- Bacterial infections (Staphylococcus aureus, Streptococcus, Pseudomonas, Cryptococcus neoformans)
- Tuberculosis
- Brucellosis

Inflammatory conditions
- Ankylosing spondylitis
- Psoriatic arthritis
- Reiter's syndrome
- Inflammatory bowel disease
- Undifferentiated spondylarthropathy
- (Juvenile) rheumatoid arthritis
- Systemic lupus erythematosus
- Behcet's disease
- Familial Mediterranean fever
- SAPHO syndrome
- Sjoegren's syndrome
- Sarcoidosis

Degenerative joint disease

Metabolic conditions
- (Pseudo) gout
- Paget's disease
- Osteomalacia
- Acromegaly
- Hyperparathyroidism
- Osteoporosis

Tumor and tumor-like conditions
- Lung, breast, kidney, and prostate metastases
- Pigmented villonodular synovitis
- Primary sacral tumors

Iatrogenic conditions
- Complications after bone graft harvesting

Sacroiliac joint syndrome

Miscellaneous conditions
- Osteitis condensans ilii
- Peri-partum pelvic instability

SIJD. We will conclude the article with a discussion of research validity of the studies reviewed and a conclusion with clinical implications.

Reliability and Validity

The two major types of reliability are test-retest and intra-rater/inter-rater reliability[25]. Test-retest reliability describes the consistency of measures repeated over time when there is no change in what is being measured[25]. Intra-rater reliability refers to the stability of measurements taken by one rater across two or more trials; inter-rater reliability is concerned with the level of agreement between findings of two or more raters measuring the same group of subjects[26]. Poor test-retest reliability can be a source of deficient intra- and inter-rater reliability. Changes in tissue response and mobility as a result of multiple tests may be a source of low intra- and inter-rater reliability for tests of the SIJ. Although traditionally reliability research has been emphasized as a precursor to validity research, Fritz and Wainner[27] made a case that its usefulness is best appreciated in conjunction with data from research examining diagnostic accuracy. Statistical measures used to establish reliability are percent agreement, variations of the κ-statistic, intra-class correlation coefficients, measures of correlation, and measures of clinical significance. Huijbregts[28] provided an in-depth discussion of the statistical validity of reliability studies reviewing these statistical measures.

The validity of a measurement is the degree to which a meaningful interpretation can be inferred from this measurement[25]. Validity has many aspects. Relevant to the studies discussed in this article are face validity, construct validity, and criterion-related validity. Face validity is the extent to which a test seems to measure what it proposes to measure[24-26]. With SIJD defined above as a painful joint dysfunction, AROM, motion palpation, and provocation tests for diagnosing SIJD have obvious face validity. However, face validity of positional palpation tests for detecting SIJD is less unequivocal. Cummings and Crowell[29] pointed out the influence leg length discrepancy (LLD) may have on falsely interpreting positional palpation findings as an indication of SIJD. Bony asymmetry may also provide for false-positive findings: a recent study reviewing 323 CT-scans unrelated to LBP found an asymmetry of over 5 mm for the acetabulum to iliac crest distance in 5.3% of subjects[30]. Mann et al[31] added muscle imbalances and congenital spinal abnormalities as reasons for abnormal positional palpation findings not related to SIJD.

Construct validity relates to the ability of a test to measure an abstract construct and to the degree with which this test measures all theoretical components of a construct[26]. SIJD defined as a painful joint dysfunction is a clear example of a construct. Levangie[32] verbalized two hypotheses regarding the pain associated with SIJD. One hypothesis holds that asymmetry of the pelvis (and associated asymmetry in the low back) cause a nociceptive mechanical stress on the structures attached to the innominates or within the SIJ. A second hypothesis holds that SIJ hypomobility, with or without positional abnormalities, places painful mechanical stresses on surrounding and intervening tissues, when one or both SIJ

fail in their function of dissipating force from the ground below or the trunk above. A third component of the SIJD construct holds that all SIJD is caused by failure of the form and/or force closure mechanism of this joint; hence, SIJ laxity is considered the underlying causative mechanism in all patients with SIJD[33]. Therefore, construct validity studies might try to correlate the different components hypothesized to be part of the construct of SIJD, i.e., LBP, positional abnormalities, hypomobility, and laxity.

Criterion-related validity indicates the extent to which a test can be used as a substitute measure for an established gold standard criterion test[26]. Concurrent criterion-related validity involves two tests performed at approximately the same time; this research evaluates whether the test studied could be used as a clinical alternative to the gold standard test[26]. Predictive validity studies attempt to establish to which degree a test can be used to predict a future criterion score[26].

Statistical measures used in validity studies include measures of diagnostic accuracy such as sensitivity, specificity, predictive values, likelihood ratios, but measures as odds ratios, and relative risk are also appropriate statistical measures for validity studies. Measures of correlation and statistical significance and even descriptive statistics are less appropriate, but have been used in the studies discussed below. The interested reader is referred to further resources on this topic[27,34-36].

History

The innervation pattern of the SIJ is extensive and variable, potentially resulting in very varied pain referral patterns[37]. Traditionally, pain due to SIJD has been described as typically unilateral, dull in character, and located over the buttocks. The pain might radiate down the posterior thigh, into the groin, or down the anterior thigh. Occasionally, pain might refer down the posterior or lateral calf into the foot and toes[38]. Etiology has been reported to involve a fall or lifting injury with torsional stresses, trauma transmitted through the hamstrings, sudden heavy lifting, prolonged lifting and bending, rising from a stooped position, or being involved in a rear-end motor vehicle accident with the ipsilateral foot on the brake[37,38]. Repeated torsional stresses as in figure skating, golfing, and bowling might have an etiological role[37]. SIJD-related pain has been reported aggravated with sitting or lying on the affected side, riding in a car, weight bearing on the affected side when standing or walking, Valsalva maneuver, and trunk flexion; the pain might be eased with weight bearing on the contralateral leg with the affected leg flexed[37]. Recent injection studies[2-4,6,9,10] have validated information gained from patient history with regards to pain location and aggravating and easing factors.

Location of Pain

Fortin et al[2,3] studied inter-rater reliability and concurrent criterion-related validity of location of pain in SIJD patients. First, the authors established an area of sensory changes, approximately 3 cm wide and 10 cm long, just inferior to the PSIS in ten asymptomatic subjects with fluoroscopically guided provocation arthrography to the right SIJ[2]. Two

subjects also noted sensory changes laterally to the greater trochanter and two had the area of hyperaesthesia extending further into the superior lateral thigh. The area of initial pain upon arthrography correlated well with this area of sensory changes. In a follow-up study[3] two medical doctors used a pain drawing; criteria for diagnosis of SIJD were the patient indicating a predominantly unilateral pain in the area just inferior to the PSIS described above. The subjects were 54 consecutive patients with LBP. Inter-rater reliability between two physicians on a diagnosis of SIJD yielded a κ-value of 0.96. All 16 patients thus identified had a provocation-positive fluoroscopically guided infiltration of the SIJ defined as pain generated within the distribution previously described by the patient. The study design did not allow for calculation of sensitivity, specificity, and predictive values[3]. Pointing to the area of pain referral established in these studies has since been introduced in the literature as the Fortin finger test[39].

Schwarzer et al[4] studied concurrent criterion-related validity of pain patterns. The raters were physicians. The subjects were 43 patients with chronic LBP below the L5-S1 level. The gold standard test was a fluoroscopically guided SIJ block with at least a 75% reduction of pain over the SIJ and buttock; L4-S1 facet joint infiltrations for all and discography for some patients served as control procedures. Statistics used were the χ^2- and Fisher exact tests to establish statistical significance of findings with diagnostic group assignment. The only statistically significant characteristic pain pattern in patients with SIJD was the presence of groin pain (P < 0.001). The prevalence of buttock, thigh, calf, and foot pain were not statistically different between SIJD and non-SIJD patients.

Dreyfuss et al[6] studied inter-rater reliability and concurrent criterion-related validity of selected pain referral patterns. Raters were a physician and a chiropractor. The subjects were 85 patients with LBP principally below L5. The gold standard test for the validity portion of the study was a 90-100% reduction in pain after a fluoroscopically guided SIJ block. Statistical measures used for the reliability study were percentage agreement and κ-values; the validity study used sensitivity, specificity, likelihood ratio, and a χ^2-test to establish significance between SIJD and non-SIJD patients. Intra-rater agreement for a pain drawing indicating SIJ, groin, or buttock pain was 92%, 87% and 91%, respectively with correspondent κ-values of 0.67, 0.70, and 0.71. Agreement on the patient pointing to within 2 inches of the PSIS as the area of maximal pain yielded 81% agreement with κ = 0.60. Sensitivity for a pain drawing indicating SIJ, groin, or buttock pain was 0.85, 0.19, and 0.80, respectively. Specificity was 0.08, 0.63, and 0.14, respectively. Likelihood ratios were 0.9, 0.5, and 0.9, respectively. Pointing to the PSIS (Fortin finger test[39]) yielded sensitivity, specificity, and likelihood ratios of 0.76, 0.47, and 1.4, respectively. Only this last test was significant at P = 0.04. The only difference in pain drawings between patient groups was the presence of pain above L5 in non-SIJD-patients; this was only present in two of the SIJD patients. The authors suggested further analysis of this possibly worthwhile diagnostic criterion.

Slipman et al[9] retrospectively studied concurrent criterion-related validity of pain referral patterns. The raters were physicians. Subjects were asked about pain referral zones and categorized into 18 potential pain referral zones. The subjects were 50 consecutive patients with LBP or buttock pain; patients with spondylarthropathies, lumbosacral radiculopathy, spondylolisthesis, and lumbar instability were excluded. The gold standard test was an 80% or greater reduction in pain after a fluoroscopically guided SIJ block. Statistical measures used were descriptive (percentages); t-tests and χ^2-tests were used to investigate relationships between pain patterns and patient age, sex, and symptom duration. Table 2 provides the frequency of pain referral patterns noted. The only statistically significant relationship reported was between pain distal to the knee and relative younger age. The authors suggested that this implied that older patients with pain distal to the knee should be suspected of spinal stenosis and neurogenic claudication rather than SIJD.

Table 2: Sacroiliac pain referral patterns in study by Slipman et al[9].

Anatomic region	Percentage	
Upper lumbar	06	
Lower lumbar	72	
Buttock	94	
Groin	14	
Abdomen	02	
Thigh	48	
• Posterior		• 30
• Lateral		• 20
• Anterior		• 10
• Medial		• 99
Lower leg	28	
• Posterior		• 18
• Lateral		• 12
• Anterior		• 10
• Medial		• 00
Ankle	14	
Foot	12	
• Lateral		• 08
• Plantar		• 04
• Dorsal		• 04
• Medial		• 00

Fukui et al[10] studied concurrent criterion-related validity of pain referral patterns. The raters were physicians. Patients were asked to indicate pain in five distinct anatomical regions. Subjects were 28 patients with LBP in whom the zygapophyseal joints and lumbosacral roots had been excluded as sources of pain by diagnostic blocks; 32 SIJs were injected. The gold standard test was a greater than 80% reduction in pain after a fluoroscopically guided SIJ block. Statistical measures used were descriptive only (percentages). All patients noted local pain over the SIJ; 68.7% reported pain in the medial buttock region; 37.5% indicated the trochanter and lateral thigh region; 31.2% the posterior thigh; and 9.3% the groin area.

Aggravating and Easing Factors

In the study mentioned above, Schwarzer et al[4] also studied concurrent criterion-related validity of certain history items. The rating scale was dichotomous: subjects were questioned whether pain was worse or relieved by sitting, standing, or walking. Statistics used were the χ^2- and Fisher exact tests to establish statistical significance of findings with diagnostic group assignment. None of the history items reached significance indicating none was able to discriminate between patient groups.

In the study discussed earlier, Dreyfuss et al[6] also studied concurrent criterion-related validity of selected history items. Patients were asked regarding pain with certain activities. The rating scale was dichotomous: same/worse or better. Statistical measures used were sensitivity, specificity, likelihood ratio, and a χ^2-test to establish significance between SIJD and non-SIJD patients. Table 3 reports sensitivity, specificity, and likelihood ratios. The authors concluded that no aggravating or relieving factor was of value for the diagnosis of SIJ-related pain.

Table 3. Concurrent validity of aggravating and easing factors in study by Dreyfuss et al[6].

	Sensitivity	Specificity	LR
Feeling better with:			
Standing	0.07	0.98	3.9
Walking	0.13	0.77	1.3
Sitting	0.07	0.80	1.2
Lying down	0.53	0.49	1.1
Same, with painful side up	0.75	0.23	1.0
Feeling worse with:			
Coughing/sneezing	0.45	0.47	0.9
Bowel movements	0.38	0.63	1.3
Heels/boots	0.26	0.56	0.8
Job activities	0.20	0.74	1.5

Active Range of Motion tests

With pain originating in the lumbar spine as the main differential diagnosis for SIJD, an AROM examination consisting of cardinal and non-cardinal plane trunk movements usually is part of the evaluation of a patient suspected of SIJD. Two injection studies[4,5] have provided information on the validity of these tests.

In the study discussed above, Schwarzer et al[4] also studied concurrent criterion-related validity of cardinal and non-cardinal plane AROM tests. The tests studied were trunk flexion, extension, bilateral rotation, and bilateral rotation combined with contralateral extension. The rating scale for these tests was dichotomous: complaints were either aggravated or not. Statistics used were the χ^2- and Fisher exact tests to establish statistical significance of findings with diagnostic group assignment. None of these AROM tests reached statistical significance and, therefore, the authors concluded none of the tests could be used to discriminate between patients with or without SIJD.

Maigne et al[5] studied concurrent criterion-related validity of cardinal plane AROM tests. The raters were physicians. The tests studied were trunk flexion, extension, and bilateral side bending. The rating scale was dichotomous: pain increased or not. The subjects were 54 patients with chronic pain and tenderness over the posterior aspect of the SIJ; relevant lumbar pathologies were ruled out. The gold standard test was a 75% or greater reduction of pain with a fluoroscopically guided double SIJ block. Statistical analysis was done using χ^2-tests. None of the cardinal plane AROM tests reached statistical significance (P = 0.15-0.48). The authors concluded that these test were not useful predictors of SIJ pain.

Special tests

Winkel[40] reviewed the literature and found 54 different special tests meant for the diagnosis of SIJD. As discussed, special tests for the SIJ fall into three different categories. Positional palpation tests attempt to diagnose SIJD by the detection of asymmetry in pelvic bony landmarks. Commonly used landmarks include the anterior (ASIS) and posterior superior iliac spines (PSIS), the iliac crests, the greater trochanters, the sacral sulcus (SS), and the inferior lateral angle of the sacrum (ILA). Motion palpation tests attempt to diagnose SIJD by the detection of abnormal relative motion of pelvic landmarks during active or passive motion tests or abnormal resistance to induced motion; some motion palpation tests use landmarks far removed from the pelvis, e.g., the supine-to-sit and the prone knee bend tests. Provocation tests aim to provoke the patient's specific pain complaint by stretching or compressing SIJ (peri-) articular structures. We will discuss reliability and validity studies of the individual special tests.

Positional Palpation Tests

Mann et al[31] studied intra- and inter-rater reliability of palpation and subsequent observation of iliac crest height in standing. The three-point rating scale consisted of equal, left lower, or right lower. The raters were three physical therapy students and eight physical therapists. The subjects were ten asymptomatic individuals; subjects with LBP on standing, SIJ hypermobility, and ilium deformities were excluded. The results were summarized descriptively. The authors concluded that iliac crest palpation and observation in standing was not a highly reliable test.

Potter and Rothstein[41] studied inter-rater reliability of pelvic landmark palpation in standing and sitting. The rating scale consisted of three points: left high, right high, or even. The raters were eight physical therapists. The subjects were 17 patients with mainly unilateral buttock pain; patients with neurological involvement or an acute lateral shift were excluded. The statistical measures used were percentage agreement and χ^2 goodness-of-fit analyses for 70 and 90% agreement levels. Palpation in standing of the iliac crest, PSIS, and ASIS levels yielded 35.29%; 35.29%; and 37.50% agreement, respectively. The same tests in sitting produced 41%.[18; 35,29]; and 43.75% agreement, respectively. None of the tests were significant for the goodness-of-fit tests.

Janos[42] studied concurrent criterion-related validity of PSIS palpation on prone subjects. The raters were

18 physical therapists. Subjects were asymptomatic. The gold standard test was an AP radiograph, to which the markings made were compared. Data were summarized descriptively. Twelve therapists correctly located both landmarks; six correctly located one and missed the other PSIS by, on average, 2 cm.

Richter and Lawall[43] studied intra and inter-rater reliability of ASIS and PSIS palpation in sitting and standing. The rating scale was dichotomous: pelvic torsion was considered present or absent. The raters were five medical doctors; ratings of four of them were collapsed into a hypothetical second rater. The inter-rater study used 35 patients with LBP; the intra-rater study used 26 patients. Kappa values were calculated. Inter-rater agreement for the presence of pelvic torsion in sitting yielded a κ-value of 0.48; in standing, the κ-value was 0.05. Intra-rater values were reported as 0.1 to 0.4 higher than inter-rater values.

Tullberg et al[44] studied concurrent criterion-related validity of palpation of the iliac crest, PSIS, and ASIS height with the patient standing, prone or supine and of the ILA in a prone position. The rating scale was dichotomous, judging presence or absence of asymmetry. The raters were two orthopaedic specialist physicians and a manual medicine physician. The subjects were ten patients with unilateral SIJD; agreement on this diagnosis between the three raters was a prerequisite for enrollment as a subject. The gold standard test was an assessment of three-dimensional SIJ position using Roentgen-stereophotogrammetric analysis (RSA) before and after a manipulation to the SIJ. Data were summarized descriptively. All three raters judged all positional tests indicative of asymmetry prior to manipulation and, with a few exceptions, normalized after manipulation. RSA showed no change in positional relationship pre and post-manipulation. The authors concluded that positional palpation tests did not provide a valid description of SIJ position.

Levangie[32] studied construct validity of positional palpation exploring the relationship between innominate torsional asymmetry and four motion palpation tests of SIJD,

i.e., the standing hip flexion, standing flexion, sitting flexion, and supine-to-sit tests. The rater was a physical therapist. Height of ASIS and PSIS was palpated and then measured rather than visually estimated; a 6 mm-difference was the cut-off point for a finding of innominate torsion. Subjects were 141 patients with LBP and 133 patients without LBP; subjects with leg length discrepancies, and pregnant, post-traumatic, and disk patients were excluded. Statistical measures used were sensitivity, specificity, positive and negative predictive values, and odds ratios with 95% confidence intervals (CI). Table 4 reports the results. The odds ratio for association between innominate torsion and two or more positive tests was 1.40 (95% CI: 0.72-2.71). The author concluded that neither the individual motion palpation tests, nor a composite of these tests was associated with innominate torsional asymmetry.

Levangie[45] studied construct validity of positional palpation exploring the association between innominate torsional asymmetry and LBP. Rater, technique of measuring pelvic landmarks, and subjects were similar to the study above[32]. Statistical measures used were odds ratios with 95% CI. The reference population consisted of subjects with 4 mm or less asymmetry. The odds ratio for the association of pelvic asymmetry with LBP was 0.80 (95% CI: 0.40-1.57) for subjects with 5-9 mm asymmetry; 0.65 (0.34-1.24) for those with 10-15 mm asymmetry; and 0.66 (0.34-1.29) for the subjects with >15 mm asymmetry). The author concluded that a substantive relationship between pelvic asymmetry and LBP was not supported by the study results. Only standing PSIS asymmetry had a weak positive association with LBP in the subgroup of men under age 35.

O'Haire and Gibbons[46] studied intra and inter-rater reliability of palpation of the PSIS, ILA, and SS in the prone position. The study used three-point rating scale: left higher, right higher, or equal. The raters were ten 5th year osteopathic students, who completed a one-hour training session prior to the study. The ten subjects were asymptomatic. The statistical measure used was a generalized κ-statistic (κ_g). Intra-rater agreement yielded a κ_g of 0.07-0.58 for PSIS palpation; 0.05-0.69 for ILA palpation;

Table 4. Association of motion palpation tests with innominate torsion in study by Levangie[32].

Test	OR (95%CI)	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Standing hip flexion	1.07 (0.42-2.74)	8%	93%	67%	35%
Standing flexion	0.81 (0.43-1.54)	17%	79%	61%	34%
Sitting flexion	1.01 (0.41-2.47)	9%	93%	78%	28%
Supine-to-sit	1.37 (0.80-2.33)	44%	64%	69%	38%

and 0.02-0.60 for SS palpation. Inter-rater agreement yielded κ_g-values of 0.04; 0.08; and 0.07, respectively.

Albert et al[47] studied inter-rater reliability of positional palpation tests of PSIS and ASIS. The rating scale was dichotomous: pelvic torsion present or absent. The raters were two physical therapists. The subjects were 34 women in the 33rd week of gestation. The statistical measures used were percent agreement and κ-values. Positional palpation yielded 91% agreement with a κ-value of 0.55. The authors also studied construct validity exploring the association between positional palpation findings and four different diagnostic groups of pelvic pain and a group without LBP. The diagnostic groups included patients with pain in all pelvic joints, the pubic symphysis, one SIJ or both. The raters were six physical therapists. The subjects were 2,269 women in the 33rd week of gestation. The statistical measures used were sensitivity and specificity. The sensitivity of positional palpation tests for detecting subjects with pain in all three pelvic joints, the symphysis, one SIJ, or both was reported as 0.26; 0.19; 0.32; and 0.46, respectively. Specificity was 0.77.

Riddle et al[48] studied inter-rater reliability of seated PSIS palpation. Ratings were on a three-point scale: negative, right positive, left positive. The raters were 11 physical therapists. The subjects were 65 patients with unilateral or bilateral LBP and unilateral buttock pain. Statistical measures used were the percent agreement, κ, standard error (SE), and κ/κ_{max}. The authors found 63.1% agreement, a κ-value of 0.37 (SE = 0.10), and a κ/κ_{max} value of 39.8. The authors concluded that the reliability of this test was poor and that this test should not form a basis for clinical decision-making.

Krawiec et al[49] (unintentionally) provided information on the construct validity of positional palpation exploring the correlation of LLD and innominate rotation position in subjects without LBP. Innominate rotation was determined with palpation followed by inclinometer measurement. The rater was an athletic trainer. The subjects were 44 asymptomatic collegiate athletes. Statistical measures used were the Pearson product moment correlation coefficient and descriptive statistics. Forty-two subjects (95%) had some degree of innominate rotation position. This study calls into question the relation between innominate positional abnormalities and LBP. The authors also excluded to some extent a causative role for LLD: they found only a weak association for LLD and innominate rotational asymmetry (r = 0.33-0.44), i.e., the leg length variation accounted for less than 19% of the variation in innominate rotation asymmetry.

Motion Palpation Tests

Wiles[50] studied inter-rater agreement of the standing hip flexion test (Figure 1): three variations of paired unilateral and bilateral manual contacts were studied. The rating scale used was a five-point scale for severity of restriction. The raters were six pairs of chiropractors. The subjects were 64 college students. The statistical measures used were the Pearson product-moment correlation coefficient, percentage agreement, and a t-test to reject or accept hypotheses related to sensitivity and specificity of the tests studied. Overall percentage agreement per hand contact ranged from 47%-64% (mean 55.2%) with r = 0.06-0.43. The overall correlation for all paired data yielded an r = 0.18. Collapsing the data to a dichotomous scale yielded an average percent agreement of 77.5% (range 54-93%). The inferior and the right bilateral manual contacts had the higher levels of agreement and correlation. Both hypotheses tested yielded non-significant P-values, but the author concluded that the P-value of 0.10 for the specificity hypothesis seemed to indicate that the tests are specific. He also concluded that a qualitative (dichotomous) rather than quantitative rating scale be used for these tests.

In the study mentioned above, Potter and Rothstein[41] also examined inter-rater reliability of motion palpation tests. The tests were the standing flexion (Figure 2), standing hip flexion, sitting flexion (Figure 3), supine-to-sit (Figure 4a and 4b), and prone knee flexion tests (Figure 5a and 5b). The three-point rating scale allowed for a choice of left or right positive or normal. Raters, subjects, and statistical measures were as discussed earlier. The inter-rater agreement for the standing flexion test was 43.75%; for the standing hip flexion test 46.67%; for the sitting flexion test 50.00%; for the supine-to-sit test 40.00%; and for the prone knee flexion test 23.53%. None of the test achieved statistical significance with the χ^2 goodness-of-fit tests for 70% or 90% agreement. The authors concluded motion palpation tests lacked sufficient reliability for clinical decision making.

Carmichael[51] studied the intra and inter-rater reliability for the standing hip flexion test; four variations of paired unilateral manual contacts were studied. The rating scale was dichotomous: fixation versus no fixation. The raters were ten chiropractic students; nine training sessions were done prior to the study for standardization. The subjects were 54 college students; moderate or greater leg, buttock, or LBP was a reason for exclusion. The statistical measures used were percentage agreement and κ-values. Mean aggregate intra-rater agreement on fixation was 89.2% (κ=0.180); mean individual intra-rater agreement was 89.9% (range 75.0-97.5%) with a mean κ-value of 0.314 (range -0.03-0.66). Inter-rater κ-values ranged from -0.0650 to 0.1930; mean percent agreement was 85.3%. The manual contacts in the upper portion of the SIJ yielded higher values for reliability. The author concluded that the standing hip flexion test was fairly reliable when used by a single examiner in repeated examinations of the same patient.

Herzog et al[52] studied intra- and inter-rater reliability of two variations of the standing hip flexion test, one with palpation on both PSIS and one with palpation on the PSIS and S2. A number of rating scales were used: fixation or no fixation; fixation left, right, or both. The raters were ten chiropractors: they received an instructional session for standardization purposes. The subjects were ten patients with SIJD and one asymptomatic control. The statistical measures used were percent agreement and a χ^2-test. Inter-rater agreement was 68%, 79%, and 72% for a positive finding, a negative finding, and identification of a positive finding on the correct side, respectively; all scores were significant at P < 0.01. The agreement on the question of

fixation or no fixation was 78%, 54%, 64%, and 65% for the first, second, third, and all three rating sessions combined, respectively. Only the second session did not reach statistical significance at P < 0.01. Agreement on the side of fixation yielded values of 60%, 60%, 64%, and 61% for the first, second, third, and all session combined; only the agreement for all sessions combined was significant at P < 0.01. Rater expertise and degree of perceived fixation did not affect percent agreement scores. Intra-rater agreement for the low expertise group was significant for both a positive finding (72%) and identification of the correct side (78%); the same scores for the high expertise group were non-significant at 64 and 67%, respectively. The authors noted that the tests studied were useful for re-evaluation by the same clinician, but also suggested that a multi-test regimen be used for inter-rater evaluations.

Mior et al[53] studied the intra and inter-rater reliability of an unspecified regimen of SIJ motion palpation tests. The rating scale was dichotomous: fixation versus no fixation. The raters were 74 chiropractic students divided in four groups receiving different forms of instruction in motion palpation procedures and a group of chiropractors. The subjects were 15 patients for the first session and ten patients and five subjects with radiographic evidence of SIJ fusion for the second session. The statistical measure used was κ. Mean κ-values for the students for the fist session ranged from 0.000 - 0.090 and for the second session from 0.013 - 0.300. Inter-rater agreement for the chiropractors yielded κ ranging from 0.000 - 0.167; intra-rater agreement for the chiropractors varied from κ = 0.15-1.00. The authors noted inconsistency of motion palpation tests regardless of experience or teaching methods.

In the study discussed above, Richter and Lawall[43] also studied the intra and inter-rater reliability of the standing hip flexion, sitting flexion, and sacral springing test. The rating scale was dichotomous: decreased or normal mobility. Raters and subjects were as reported above. Statistical measures used were κ-values for total agreement and agreement on hypomobility, both with 95% CI's. Table 5 consists of the reliability findings for the tests studied. The authors concluded that reliability of the SIJ tests studied was moderate to good, but still suggested a reliability study at the level of multi-test derived diagnosis.

Dreyfuss et al[54] studied construct validity of motion palpation tests exploring the relation between the absence of LBP and positive tests. The test included the standing and sitting flexion and standing hip flexion tests. The rating scale consisted of four points: right and or left positive or negative. The rater was a physical therapist. The subjects were 101 asymptomatic subjects. The statistical measures used were descriptive and the χ^2-test to investigate significance of differences between subgroups. Overall, 20% of subjects had a positive finding in at least one of the three tests with 13%, 8%, and 16% false positive results for the standing flexion, sitting flexion, and standing hip flexion tests, respectively. Women scored significantly more false positives on the standing hip flexion test; there were significantly more right-sided false positives for the seated flexion test in men and women and right-sided false positive standing flexion tests in women. The authors concluded

Figure 1. Standing hip flexion.

Figure 2. Standing flexion.
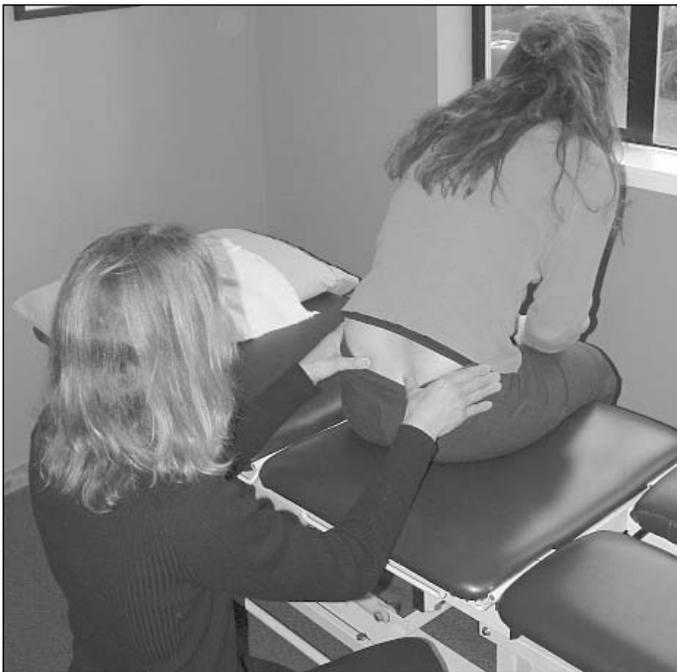
Figure 3. Sitting flexion.

Figure 4a. Supine to sit.

Figure 4b. Supine to sit.

Figure 5a. Prone knee flexion.

Figure 5b. Prone knee flexion.

that specificity of the tests studied was less than previously assumed and suggested that the examination not be limited to these tests when SIJ-related pain is suspected.

Bowman and Gribble[55] studied the inter-rater reliability of the standing flexion test. The study used a three-point rating scale. The raters were three physicians with osteopathic training. The subjects were seven asymptomatic volunteers and nine patients with LBP; acute LBP and nerve root involvement was a reason for exclusion. The statistical measures used were percentage agreement and κ. Inter-rater agreement was 52% and κ was 0.2333. The authors concluded that more reliable tests remained needed to resolve whether SIJD is clinically relevant.

Table 5. Intra and inter-rater agreement motion palpation tests (κ-values and 95% CI) in study by Richter and Lawall[43].

Tests		Intra-rater	Inter-rater
Sitting flexion test	κ (total)	0.83 (0.46-1.00)	0.54 (0.32-0.76)
	κ (decreased left)	0.92 (0.53-1.00)	0.51 (0.23-0.81)
	κ (decreased right)	0.92 (0.53-1.00)	0.64 (0.33-0.95)
Standing hip flexion test right	κ (total)	0.86 (0.56-1.00)	0.69 (0.40-0.97)
	κ (decreased right)	0.84 (0.46-1.00)	0.62 (0.29-0.95)
Standing hip flexion test left	κ (total)	0.93 (0.66-1.00)	0.65 (0.42-0.88)
	κ (decreased left)	0.90 (0.51-1.00)	0.48 (0.21-0.74)
Sacral springing right	κ (total)	0.81 (0.50-1.00)	0.47 (0.23-0.71)
	κ (decreased right)	0.75 (0.38-1.00)	0.46 (0.14-0.78)
Sacral springing left	κ (total)	0.74 (0.45-1.00)	0.47 (0.23-0.71)
	κ (decreased left)	0.83 (0.46-1.00)	0.46 (0.14-0.78)

In the study discussed earlier, Dreyfuss et al[6] also studied inter-rater reliability and concurrent criterion-related validity of SIJ motion palpation tests. The tests were the standing hip flexion and sacral base springing tests. The raters, subjects, and gold standard tests were as discussed earlier. The statistical measures used for the reliability study were percentage agreement and κ-values; the validity portion of the study used sensitivity, specificity, and likelihood ratio. Inter-rater reliability yielded 54% agreement for the standing hip flexion test (κ=0.22) and 60% for the sacral springing test (κ=0.15). Sensitivity, specificity, and likelihood ratio for the standing hip flexion test were 0.43; 0.68; and 1.3, respectively. The respective values for the sacral springing test were 0.75; 0.35; and 1.2. The authors concluded that the likelihood ratios for these tests were too close to 1.0 to significantly increase pre-test probability of SIJD.

Vincent-Smith and Gibbons[56] studied the intra and inter-rater reliability of the standing flexion test with bilateral manual contacts. The rating scale used three points: negative, right positive or left positive. The raters were nine osteopaths; a training session was held prior to the study for standardization. The subjects were nine asymptomatic volunteers. The statistical measures used were percentage agreement, κ, and an unspecified test to determine statistical significance. Inter-rater agreement yielded mean percentage agreement of 42% with a mean κ of 0.052. Intra-rater agreement ranged from 44% - 88% with a mean of 68%; κ ranged from 0.16-0.72 with a mean κ of 0.46. Only the inter-rater mean agreement was significant at P < 0.01. The authors concluded that the reliability of the standing flexion test remained questionable.

In the study mentioned earlier, Levangie[32] also researched the construct validity of four motion palpation tests determining the association between LBP and the individual tests. The tests were the standing hip flexion, standing flexion, sitting flexion, and supine-to-sit test. The rating scale was dichotomous: positive or negative. The statistical measure used was the odds ratio with 95% CI. The odds ratio and 95% CI for the standing hip flexion test was 4.57 (1.51 - 13.86); for the standing flexion test 0.77 (0.42 - 1.42); for the sitting flexion test 1.52 (0.63-3.64); and for the supine-to-sit test 1.23 (0.75 - 2.02). The author concluded that only the standing hip flexion test was associated with LBP and suggested that the standing hip flexion test might asses SIJ hypomobility as a cause of LBP. She also noted that the standing flexion and hip flexion tests did not appear to be responsive to the same phenomena.

Albert et al[47] also studied inter-rater reliability and construct validity of the sitting flexion test. The rating scale was dichotomous. Raters, subjects, and statistical measures were as discussed. Inter-rater agreement was 88%; κ was only reported as >0. Sensitivity for detecting pain in the all three pelvic joints, the symphysis, one SIJ, or both was reported as 0.14; 0.00; 0.69; and 0.21; specificity was 0.98.

Sturesson et al[57] studied concurrent criterion-related validity of the standing hip flexion test. The raters were an orthopaedic surgeon, a chiropractor, and two physical therapists. The subjects were 22 patients with SIJD diagnosed by all four raters by way of physical examination including SIJ motion palpation and provocation tests. The gold standard test was RSA. No statistical analysis was performed. RSA showed that movements during the standing hip flexion tests were too minute to be detected with manual methods and that in addition motions, when they occurred, were similar in both joints. The authors concluded that the standing hip flexion test could not be recommended for evaluation of SIJ motion.

In the study discussed earlier, Riddle et al[48] also studied the inter-rater reliability of three motion palpation tests: the standing flexion, prone knee flexion, and supine-to-sit tests. The rating scale consisted of three points for the standing flexion test (right positive, left positive, or negative) and of five points for the other two tests indicating absence or

presence of both side and type of dysfunction. The raters, subjects, and statistical measures were as discussed above. Inter-rater agreement for the standing flexion, prone knee flexion, and supine-to-sit tests was 55.4%, 60.0% and 44.6%, respectively. The κ-values with standard errors were 0.32 (0.09); 0.26 (0.10); and 0.19 (0.09), respectively. The respective κ/κ$_{max}$-values were 40.5; 28.6; and 21.1. The authors concluded that the κ-values of the individual tests were too low to justify clinical use of these tests.

Provocation Tests

In the study mentioned above, Potter and Rothstein[41] also examined inter-rater reliability of two SIJ provocation tests, the compression (Figure 6) and distraction tests (Figure 7). The three-point rating scale allowed for a choice of left painful, right painful, or no pain. Raters, subjects, and statistical measures were as discussed earlier. The inter-rater agreement for the compression test was 76.47% and for the distraction test 94.12%. The tests achieved statistical significance at $P < 0.05$ for the χ^2 goodness-of-fit tests for 70% and 90% agreement, respectively. The authors concluded that their study only showed that these two tests, which relied on patient response, were somewhat reliable.

Laslett and Williams[58] studied inter-rater reliability of SIJ provocation tests. The tests were the distraction, compression, thigh thrust (Figure 8), pelvic torsion (Figure 9), sacral thrust (Figure 10), and cranial sacral shear test (Figure 11). The rating scale was dichotomous: symptom reproduction or not. The raters were six pairs of physical therapists; two training sessions were provided for standardization. The subjects were 51 patients with unilateral LBP or buttock pain, with or without radiation below the knee. The statistical measures used were

percentage agreement, κ, and a modified κ called κn. Table 6 provides the results for the tests studied. The authors concluded that the distraction, compression, thigh thrust, and pelvic torsion tests had substantial inter-rater reliability, whereas the sacral thrust and shear tests were found to be potentially reliable tests.

In the study discussed above, Maigne et al[5] also studied concurrent criterion-related validity of SIJ provocation tests. The tests studied were the distraction, compression, sacral thrust, pelvic torsion, flexion-abduction-external rotation (FABER) (Figure 12), resisted external rotation of the hip, and pubic symphysis pressure. The rating scale was dichotomous. Raters, subjects, statistical measures, and gold standard tests were as noted above. There was no statistically significant association of any pain provocation test and the gold standard test (P = 0.09 - 0.67). The authors concluded SIJ provocation tests were not useful predictors of SIJ-related pain.

In the study discussed earlier, Dreyfuss et al[6] also studied inter-rater reliability and concurrent criterion-related validity of SIJ provocation tests. The tests were the thigh thrust, FABER, pelvic torsion, and sacral thrust tests. The raters, subjects, and gold standard tests were as discussed earlier. The statistical measures used for the reliability study were percentage agreement and κ-values; the validity portion of the study used sensitivity, specificity, and likelihood ratio. Inter-rater reliability yielded 82% 82% 85% and 66% agreement for the thigh thrust, FABER, pelvic torsion, and sacral thrust tests, respectively; respective κ-values were 0.64; 0.62; 0.61; and 0.30. Table 7 contains data on sensitivity, specificity, and likelihood ratio for these tests.

Strender et al[59] studied inter-rater reliability of the FABER

Figure 6. Compression.

Figure 7. Distraction.

and compression test. The rating scale was dichotomous: normal or pathologic. The raters were two physical therapists and two physicians; a session was held prior to the study for standardization. The subjects were 50 patients with LBP examined by the therapists and 21 examined by the physicians; pregnant, post-operative, obese, and adolescent subjects were excluded. The statistical measures used were the percentage agreement and κ-values. The therapists achieved 96% agreement on the FABER test, the physicians 88%. The values for the compression test were 79% (κ=0.26) and 74% (κ=0.26), respectively. The authors concluded these tests were insufficiently reliable.

Broadhurst and Bond[7] studied concurrent criterion-related validity of three SIJ provocation tests. The tests were the FABER, thigh thrust, and resisted abduction test. The rating scale for these tests was dichotomous: reproduction of pain or not. The raters were physicians. The subjects were 40 patients with suspected SIJD. The gold standard test was 70 or 90% reduction of pain after a fluoroscopically guided double blind SIJ block. The statistical measures used were an analysis of variance, sensitivity, and specificity. At the 70% criterion, sensitivity and specificity for the FABER test were 77% and 100%; at 90%, they were 50 and 100%, respectively. The sensitivity and specificity of the thigh thrust test were 80% and 100% at the 70% criterion; at 90%, they were 69% and 100%, respectively. Sensitivity and specificity of the resisted abduction test were 87% and 100% at the 70% criterion and 65% and 100% at the 90% criterion. The ANOVA was significant for all three tests indicating a significantly greater post-test pain reduction in treated versus control subjects. The authors concluded that the three tests studied in combination with the pain referral pattern established by Fortin et al[2,3,39] would add to the clinician's diagnostic capabilities.

Mens et al[60] studied the construct validity of the active straight leg raise test (ASLR) exploring the correlation between this test and pelvic joint instability. The rating scale

Figure 8. Thigh thrust.

Figure 9. Pelvic torsion.

Figure 10. Sacral thrust.

was a four-point scale going from no restriction to inability to raise the leg. The rater for the ASLR test was a physical therapist; the raters for the radiograph were physicians. The subjects were 21 non-pregnant women with mainly asymmetric peri-partum pelvic pain and impaired ASLR. Patients with a history of neoplasm, fracture, surgery or signs of radiculopathy were excluded. The patients were tested with the ASLR test, the same test after application of a pelvic belt fastened around the pelvic girdle, and a radiograph as described by Chamberlain. The statistical measure used was a binomial two-tailed test for statistical significance. Application of a pelvic belt reduced impairment in 20 patients (significant at P=0.0000). Of 21 patients, 17 had a greater step when standing on the reference side than on the symptomatic side on radiograph and four had an equal step (significant at P=0.01). The authors suggested that the step visible on a radiograph was the result of an anterior innominate rotation on the symptomatic side. They concluded that the results showed a clear correlation between impaired ASLR and mobility of the pelvic joints in patients with peri-partum pelvic girdle pain and suggested further research into diagnostic accuracy and responsiveness.

In the study mentioned above, Albert et al[47] also studied inter-rater reliability and construct validity of the thigh thrust, FABER, compression, and distraction tests. The rating scale for these tests was dichotomous: reproduction of pain over the SIJ or not. Raters, subjects, and statistical measures were as discussed earlier. Inter-rater agreement for the thigh thrust test was 91% (κ=0.70); for the FABER test 88% (κ=0.54); for the compression test 97% (κ=0.79); and for the distraction test 97% (κ=0.84). Table 8 provides data on sensitivity and specificity.

Kokmeyer et al[61] studied the inter-rater agreement of the distraction, compression, pelvic torsion, FABER, and thigh thrust tests. The rating scale for these tests was dichotomous: ipsilateral pain in the gluteal region under L5 was defined as positive. The raters were two physical therapy students. The raters completed training sessions prior to the study to standardize the force applied. The subjects were 59 patients with LBP and 19 asymptomatic subjects. Statistical measures used were percent agreement, κ, and 95% CI of κ, and variants of κ adjusted for bias and both prevalence and bias. Table 9 reports reliability measures for the individual tests.

Damen et al[62] studied the predictive validity of the ASLR and thigh thrust tests for post-partum pregnancy related pelvic pain (PRPP). The subjects were 55 women with PRPP at 36 weeks of gestational age; exclusion criteria were low back or pelvic pain prior to pregnancy, pain below the knee, or known rheumatological or congenital abnormalities. The statistical measures used were sensitivity, specificity, predictive values, and relative risk. The ASLR test yielded a sensitivity of 76.9%, a specificity of

Figure 11. Cranial shear.

Figure 12. FABER

55.2%, a positive predictive value of 60.6%, a negative predictive value of 66.7%, and a relative risk of 2.4; the values for the thigh thrust test were 61.5%; 72.4%; 66.7%; 67.7%; and 2.1, respectively. The authors related PRPP during and, to some extent, after pregnancy to asymmetric SIJ laxity established by way of Doppler imaging of vibrations over the joints, thus also lending support to the construct linking SIJD to an underlying instability.

Levin and Stenstrom[63] studied concurrent criterion-related validity of the distraction test. The test was performed from both sides of the patient because an earlier study[64] showed lower forces in the SIJ closest to the examiner. The rating scale was dichotomous; at least two of three positive tests were required for a test to be rated positive. The raters were three physical therapists. The subjects were seven subjects with ankylosing spondylitis, four with undifferentiated spondylarthropathy and 11 asymptomatic subjects. Ankylosis, neurological

involvement, and a history of surgery, fracture, or neoplasm were reason for exclusion. The gold standard test was verification of sacroiliitis on radiograph or MRI. The statistical measures used were sensitivity, specificity, and positive and negative predictive values. Sensitivity and negative predictive value of the test performed from the right was 0.55 and 0.69; from the left, these values were 0.55 and 0.67, respectively. The specificity and positive predictive values were 1.0.

Multiple Test Regimens

Considering the lack of reliability of the individual special tests meant to detect SIJD, some authors[43,52] have suggested the use of multi-test regimens to diagnose SIJD. We will review studies that researched reliability and validity of multiple test regimens. The regimens studied have consisted of various combinations of the individual positional palpation, motion palpation, and provocation tests reviewed above.

Cibulka et al[65] studied inter-rater reliability of a cluster of four SIJ tests. The tests included the standing flexion, the sitting PSIS palpation, the supine-to-sit, and the prone knee flexion test. The rating scale for the individual tests was dichotomous: positive or negative for SIJD. The overall rating scale was also dichotomous: three of four tests positive were needed for a diagnosis of SIJD. The raters were two physical therapists. The subjects were 26 patients with non-specific LBP or buttock pain; patients with a neurological deficit, pain below the knee, ankylosing spondylitis, and symptom magnification were excluded. The statistical measure used was the κ-value. Inter-rater agreement yielded a κ-value of 0.88. The authors concluded that the combination of tests studied was reliable for diagnosing SIJD as defined in this study and also suggested that the additional training on standardization of test performance might have improved reliability.

In the study discussed earlier, Dreyfuss et al[6] also studied concurrent criterion-related validity of a combination of all history items and physical tests discussed separately earlier in this article. The raters, subjects, and gold standard tests were as discussed earlier. The statistical measures used were sensitivity, specificity, and likelihood ratio. Sensitivity for six to 11 positive tests was 0.57; 0.53; 0.29; 0.29; 0.0; and 0.0, respectively. Specificity values were 0.42; 0.55; 0.52; 0.68; 0.87; and 0.83, respectively. Likelihood ratios were 1.0; 1.2; 0.6; 0.9; 0.0; and 0.0, respectively. Specific variable

Table 6. Inter-rater agreement provocation tests in study by Laslett and Williams[58].

Test	Percentage agreement	κ	κn
Distraction	88.2%	0.69	0.76
Compression	88.2%	0.73	0.76
Thigh thrust	94.1%	0.88	0.88
Pelvic torsion right	88.2%	0.75	0.76
Pelvic torsion left	88.2%	0.72	0.76
Sacral thrust	78.0%	0.52	0.56
Cranial shear	84.3%	0.61	0.69

Table 7. Validity individual SIJ provocation tests in study by Dreyfuss et al[6].

Test	Sensitivity	Specificity	Likelihood ratio
Thigh thrust	0.36	0.50	0.7
FABER	0.69	0.16	0.8
Pelvic torsion	0.71	0.26	1.0
Sacral thrust	0.53	0.29	0.8

Table 8. Construct validity provocation tests in study by Albert et al[47].

| Test | Sensitivity | | | | Specificity |
	All pelvic joints	Sympysis	One-sided SIJD ratio	Double-sided SIJD	
Thigh thrust	0.90	0.17	0.84	0.93	0.98
FABER	0.70	0.40	0.42	0.40	0.99
Distraction	0.40	0.13	0.04	0.14	1.00
Compression	0.70	0.13	0.25	0.38	1.00

combinations of sacral tenderness, the Fortin finger test, and groin pain yielded likelihood ratios from 0.4 to 1.2.

Slipman et al[8] studied concurrent criterion-related validity of a cluster of SIJ provocation tests. These tests always included the FABER test and pain with pressure to the SIJ ligaments at the sacral sulcus with the patient prone; other tests could consist of a shear, standing extension, pelvic torsion, or prone hip extension test (Figure 13). The rating scale for the individual tests was dichotomous, as was the rating scale for the test cluster: a positive response to at least three tests was considered indicative of SIJD. The raters were physicians. The subjects were 50 patients with sub-acute and chronic LBP. Patients with symptoms of spondylarthropathy or neurological signs were excluded. The gold standard test was at least 80% reduction of pain after a fluoroscopically guided SIJ block. The statistical measure used was the positive predictive value. The positive predictive value of the cluster of tests was 60%. The authors suggested that the cluster of tests might play a role in a clinical algorithm culminating in diagnostic SIJ blocks.

Cibulka and Koldehoff[66] researched construct validity of a cluster of four SIJ tests exploring the association between LBP and this cluster of tests. The tests again consisted of the standing flexion, the sitting PSIS palpation, the supine-to-sit, and the prone knee flexion test. The rating scale was similar to the one used in the earlier study. The raters were two physical therapists. The subjects were 219 patients: 105 with (sub) acute LBP and 114 without LBP. Subjects with signs of nerve root involvement were excluded. The statistical measures used were sensitivity, specificity, positive and negative predictive values, and prevalence. Sensitivity of the cluster of tests was 0.82; specificity was 0.88, and prevalence 0.48. The positive predictive value of the cluster was 0.86 and the negative predictive value 0.84. The authors concluded that the cluster of tests appeared to be clinically useful to detect SIJD in patients with LBP, but noted that usefulness was not determined for diskogenic patients.

In the study reported above, Kokmeyer et al[61] also studied the inter-rater reliability of a cluster of provocation tests. Using the same tests, rating scale, raters, and subjects, they reported 83.33% inter-rater agreement and a κ of 0.63 (95% CI: 0.47 - 0.83) for diagnosing SIJD based on 1 positive test. Two positive tests yielded 92.31% and κ of 0.74 (0.54 - 0.94); three yielded 93.59% and a κ-value of 0.70 (0.45 - 0.95); four yielded 96.15% and a κ of 0.71 (0.38 - 1.03). Finally, agreeing on five tests to diagnose SIJD produced 98.72% agreement and a κ-value of 0.66 (0.00 - 1.32). The authors suggested using a regimen requiring three positive tests out of five tests to decrease chance agreement as well as false negative decisions.

In the study discussed earlier, Riddle et al[48] also studied the inter-rater reliability of the cluster of SIJ tests consisting of the standing flexion, prone knee flexion, supine-to-sit, and seated PSIS palpation tests. Three of four tests needed to be positive for the diagnosis of SIJD. Raters, subjects, and statistical measures were as discussed above. The authors used three rating scales for this cluster of tests. With the dichotomous rating scale (SIJD present or absent) agreement was 61.5 % with κ (SE) = 0.18 (0.12). When using a three-point rating scale (right positive, left positive, or negative), agreement was 60.0% with κ(SE) = 0.11 (0.11). A five-point rating scale indicating both side and type of innominate positional fault yielded 69.2% agreement with κ(SE) = 0.23 (0.12). The κ/κmax-values were 20.2; 12.2; and 27.1, respectively. The authors suggested using an alternative approach to identifying patients suspected of SIJD due to the poor reliability of the cluster of tests in identifying SIJD irrespective of the rating scale used.

Comprehensive examination

In the clinical situation, a diagnosis of SIJD is not made based on the results of an isolated history item or AROM test; also the result of an isolated special test or even the results of a cluster of special tests in isolation is not used to establish a diagnosis of SIJD. Instead, clinically a diagnosis of SIJD will be the result of a comprehensive examination consisting of a history, AROM tests, and special tests within the framework of a clinical reasoning process. We will review the one study done on the validity of a comprehensive examination in the diagnosis of SIJD.

Laslett et al[11] studied concurrent criterion-related validity of a comprehensive examination consisting of a McKenzie evaluation combined with a cluster of SIJ provocation tests. The tests used were the distraction, compression, thigh thrust, pelvic torsion, and sacral thrust tests. The rating scale for the individual tests was dichotomous; the subjects were diagnosed with SIJD when three or more tests were positive after exclusion of diskogenic complaints with a McKenzie

Figure 13. Prone Hip Extension.

evaluation. The raters were physical therapists. The subjects were 48 patients with buttock pain with or without lumbar or leg symptoms. Patients with only midline or symmetrical LBP above L5 or signs of nerve root involvement were excluded. The gold standard test was a fluoroscopically guided double SIJ block with at least 80% pain reduction. Statistical measures used were sensitivity, specificity, and positive and negative likelihood ratios, all with 95% CI. Sixteen subjects had a positive response to the first SIJ block; five subjects had significant relief from the first block and did not receive a second block. Eleven patients responded positively to the second block. Excluding these five patients, this subset of 43 patients yielded a sensitivity of 0.91 (95% CI: 0.62 - 0.98); specificity of 0.87 (0.68 - 0.96); positive likelihood ratio 6.97 (2.70 - 20.27); and a negative likelihood ratio of 0.11 (0.02 - 0.44). Excluding the diskogenic patients produced a second subset of 34 subjects. This subset yielded a sensitivity of 0.91 (95% CI: 0.62 - 0.98); specificity of 0.78 (0.61 - 0.89); positive likelihood ratio 4.16 (2.16 - 8.39); and a negative likelihood ratio of 0.12 (0.02 - 0.49). The authors concluded that SIJ provocation tests within the context of a specific clinical reasoning process allow the clinician to differentiate between a symptomatic and asymptomatic SIJ.

Discussion

When interpreting the studies reviewed in this article, we need to address the research validity of these studies. Domholdt[67] defined research validity as the extent to which the conclusions of a study are believable and useful. Three aspects of research validity are relevant when interpreting the studies reviewed: statistical conclusion validity, external validity, and construct validity[28].

Using inappropriate statistical tools for data analysis is a threat to statistical conclusion validity[67]. The reliability studies reviewed have used a multitude of statistical measures. Some have used descriptive statistics[31]. Other studies have used measures of agreement:
- Percentage agreement[6,41,47,48,50,51,52,55,56,58,59,61]
- A κ-value [3,6,43,47,84,51,53,55,56,58,59,61,65]
- Mean κ-value [51,53,56]
- Generalized κ[46]
- Modified κ, $κ_n$, to allow for unrestricted distribution of judgments made by the raters[58]
- Maximal κ, $κ_{max}$, allowing for quantification of the effect of a limited upper margin for κ-values by calculating $κ/κ_{max}$[48]
- Bias-adjusted κ[61]
- Bias- and prevalence-adjusted κ[61]

Some of these studies also supplied a 95% CI with these statistics[43,61]. Table 10 provides bench mark κ-values for evaluating reliability studies[28]. Some studies have used measures of correlation, the Pearson product moment correlation coefficient[50]. Some have used statistics to establish statistical significance[41,52].

The validity studies reviewed, similarly, have used a multitude of statistical measures. Some studies only used descriptive statistics[2,3,9,10,42,44,54,57]. Some studies have reported

measures of diagnostic accuracy, such as:
- Sensitivity[6,7,11,32,47,62,63,66]
- Specificity[6,7,11,32,47,62,63,66]
- Positive and negative predictive values[8,11,32,62,63,66]

Table 10. κ Benchmark values[28].

< 40%	Poor to fair agreement
40-60%	Moderate agreement
60-80%	Substantial agreement
>80%	Excellent agreement
100%	Perfect agreement

- Prevalence[66]

Some studies have combined these measures and provided likelihood ratios[6]. Other studies provided odds ratios[32,45] or calculations of relative risk[62]. Some studies provided a 95% CI with these statistics[11,32,45]. Some studies reported a measure of correlation, the Pearson product moment correlation coefficient[49]. Other studies have used measures of significance to establish statistical significance of between-group differences[4-7,9,54,60] or to accept or reject hypotheses related to sensitivity and specificity[50].

In general, one needs to review study methodology to appreciate whether a specific statistical measure was appropriate. Statistical analysis with the appropriate tools is preferred over a descriptive presentation of data. Generally, for reliability studies variations of the κ-statistic are preferred over percentage agreement values, as the latter do not correct for chance agreement[28]. Limited variation in the data set analyzed (e.g., due to a study population which is highly homogenous on the variable of interest) may result in high percentage agreement values, but low κ-values giving a false impression of deficient reliability. Interpretation of κ-values is facilitated if the study presents data on prevalence or even the complete original data set[28]. Combining κ-statistics into a mean κ-value is only allowed if (reported) standard errors of the individual κ-values are similar in magnitude. A generalized κ is the weighted average of pair-wise κ's: assignment of weights needs to be clarified[28]. Measures of correlation are inappropriate statistics for reliability studies: they express covariance rather than agreement[28]. Determining statistical significance of agreement values is similarly inappropriate due to sample size effects on significance[28]. As discussed, measures of diagnostic accuracy, likelihood ratios, odds ratios, and relative risk are appropriate measures for validity studies. Establishing statistical significance of between-group differences when a gold standard test is used for group assignment seems equally appropriate. Providing a 95% CI allows the reader to identify whether the possible values for a statistical measure include those similar to chance agreement or those results irrelevant to changing pre-test probability: the fact that Dreyfuss et al[6] provided no 95% CI for the only history item with a likelihood ratio significantly higher than 1.0 (pain relief with standing in patients

diagnosed with SIJ-related pain; LR = 3.9) makes the value of this finding unclear[68]. The reader needs to review the study results and conclusions presented in the light of this information.

External validity deals with the degree to which study result can be generalized to different subjects, settings, and times[67]. Similarity in subjects, raters, operational definition of history or examination item, rating scale, and setting allow for a greater degree of generalization of the studies discussed to the reader's setting[28]. Physicians, physical therapists, chiropractors, and athletic trainers are not necessarily trained similarly: inter-professional differences in operational definitions of tests and rater bias based on theoretical constructs underlying these different professions may affect study outcomes. Information from studies using asymptomatic subjects[2,31,42,46,49-51,56], cannot simply be generalized to a clinical population. Similarly, results from studies involving a highly specific population of patients referred to the specialist physician for diagnostic studies[3-11] cannot necessarily be generalized to the population seen by the average primary care provider. Pictures of the tests studied have been provided in this article. However, these pictures only serve as a general mnemonic for the techniques involved. More specific operational definitions have usually been provided in the articles and need to be reviewed before adopting the study results to justify one's own clinical practice. Rating scales studied in the articles reviewed are generally dichotomous, but involve up to five points. If the reader intended to use the tests described to identify only the painful side, a dichotomous scale reporting reproduction or absence of symptoms may be sufficient.

However, if the tests were used to, e.g., draw conclusions regarding side and direction for the application of a manipulative thrust, then a study with a five-point rating scale indicating side and type of innominate positional fault would provide the more appropriate information. A (somewhat insidious) example of this issue is provided by the studies by Cibulka et al[65] and Cibulka and Koldehoff[66]. The rating scale used for these studies, as discussed, is dichotomous. SIJD is considered present if at least three of four tests are positive. However, the findings of the individual tests need not be similar: raters might arrive at a completely different diagnosis of side and type of positional fault present[68]. In defense of these studies, the intervention proposed is less dependent on a precise diagnosis of side and type of positional fault, thereby justifying this specific methodological approach. Studies set in an actual primary care clinic may provide more relevant information to the primary care provider than studies set in a strictly controlled research environment or studies set in a specialist office. Laslett[69] addressed additional issues concerning external validity: he discussed the risk of false negative findings when insufficient force was applied during SIJ provocation tests. Levin and Stenstroem[64] agreed showing that lower forces were applied to the SIJ closest to the clinician during the distraction test and warned that inter-rater force variability measured in their study could negatively affect reliability and sensitivity. The time a provocative force is applied also seems to play a role: Levin et al[63] found symptom reproduction with the distraction tests after as

much as 20 seconds. Laslett[69] also suggested more attention be paid to whether a test reproduced symptoms thought related to the SIJ rather than to unrelated (hip or back) complaints. Description of a physical test seems to require extensive operational definition including level of force applied and duration of application. Again, readers are urged to review aspects of external validity when interpreting the studies presented.

Construct validity within the framework of research validity is somewhat different from construct validity as described earlier in the framework of construct validity studies on SIJD. The main threat to construct validity in reliability and validity research is the discrepancy between the construct as labeled and the construct as implemented[67]. For reliability studies, adding training sessions to standardize techniques and rating scales may inadvertently change the construct as implemented from studying test reliability to the effect of rater training on test reliability. Inadvertent manipulation of the SIJ during repeated motion palpation or provocation tests during a reliability study changes the construct as implemented to the effect of repeated mobilizing stress on SIJ mobility and pain response as measured by the tests studied[28]. The greatest threat to construct validity in the concurrent criterion-related studies reviewed is related to the gold standard test used. We discussed above how the constructs of SIJD as a painful joint dysfunction relate abnormal articular as well as peri-articular mechanical stresses to the pain associated with SIJD. Maigne et al[5] acknowledged that a major part of symptomatic SIJ pathology may be related to the irritation of peri-articular tissues. Consequently, a fluoroscopically guided intra-articular anaesthetic infiltration might serve as the reference test for intra-articular pathology, but probably should not serve as the gold standard test for peri-articular pathology thought to be part of the patho-mechanical diagnosis of SIJD[69]. This consideration changes the construct as labeled for the infiltration validity studies to validity of history items and physical tests for the diagnosis of intra-articular SIJ pathology rather than SIJD, an important point to consider when interpreting the studies reviewed.

Conclusion

The discussion above on research validity allows, to some degree, a summary of research findings, which can be used to guide evidence-based diagnosis of SIJD by way of history and physical examination:

History
- Referred pain from the SIJ is located mainly in the buttock, lower lumbar, and postero-lateral thigh region[9,10]. However, it may extend all the way down the leg into the foot[9].
- Predominant unilateral pain in an area just inferior to the PSIS is especially indicative of SIJ-related pain[2,3,6].
- Groin pain may or may not be a sensitive indicator of SIJ-related pain[4,6].
- Older patients with pain below the knee are more likely to be diagnosed with complaints other than SIJD[9].

- No aggravating or easing factors have been identified with diagnostic value for SIJ-related pain[4,6].

AROM Tests

- AROM tests, including trunk flexion, extension, bilateral rotation, bilateral side bending, and bilateral rotation combined with contralateral extension are not useful to discriminate between patients with or without SIJ-related pain[4,5].

Positional Palpation Tests

- When using a three-point rating scale, positional palpation tests in standing of the iliac crest levels, PSIS, and ASIS have insufficient inter-rater reliability[31,41]. Similarly, with a three-point rating scale positional palpation tests of the PSIS, ILA, and SS with the patient in prone position[46] and of the PSIS with the patient sitting have insufficient inter-rater reliability[48].
- A dichotomous rating scale (absence or presence of pelvic torsion) produces moderate inter-rater agreement for palpation of ASIS and PSIS palpation in sitting[43]; data on the standing test are equivocal ranging from poor to moderate[43,47].
- Innominate torsional asymmetry is not associated with positive findings on the standing hip flexion, standing flexion, sitting flexion, and supine-to-sit motion palpation tests, nor is it related to positive findings on a cluster of two or more of these tests[32].
- Innominate torsional asymmetry is not associated with LBP[45,49].
- Palpation in standing, prone, or supine of the iliac crests, PSIS, and ASIS, and of the ILA in supine is not a valid descriptor of SIJ position as confirmed by RSA[44].

Motion Palpation Tests

- Inter-rater agreement for the standing hip flexion test is poor to substantial when using a dichotomous rating scale[6,43,51]; with a dichotomous rating scale, inter-rater agreement is moderate for the sacral springing and sitting flexion tests[43].
- Inter-rater agreement using a three-point rating scale is poor for the standing hip flexion, standing flexion, sitting flexion, supine-to-sit, and prone knee flexion tests[41,48,55,56].
- Inter-rater agreement when using a five-point rating scale for the prone knee flexion and supine-to-sit tests is poor[48].
- Positive findings on the standing flexion, sitting flexion, and supine-to-sit tests are not associated with LBP; in contrast, a positive finding on the standing hip flexion test is associated with LBP[32].
- The standing hip flexion test and the sacral base springing test lack diagnostic accuracy for identifying patients with a positive SIJ block[6].
- The standing hip flexion test with a five or two-point rating scale was shown to be neither sensitive, nor specific for diagnosing SIJD[50].
- The standing hip flexion, sitting flexion, and standing flexion tests have a false positive rate of near 20%,

resulting in decreased specificity[54].
- The standing hip flexion test is not a valid indicator of SIJ motion as confirmed by RSA[57].

Provocation Tests

- Using a dichotomous rating scale, the compression test produced poor to substantial inter-rater agreement[47,58,59,61]. The distraction test yielded moderate to excellent agreement[6,47,58,61]. The FABER test yielded moderate to substantial agreement[6,47,58,61]. The pelvic torsion test yielded moderate to substantial inter-rater agreement[6,58,61] and the cranial sacral shear test yielded substantial inter-rater agreement[58]. The sacral thrust test showed poor to moderate agreement[6,58] and the thigh thrust test showed substantial to excellent agreement[6,47,58,61].
- A positive ASLR test is associated with ipsilateral increased SIJ mobility[60].
- The compression, sacral thrust, pelvic torsion, resisted hip external rotation, and pubic symphysis pressure tests individually lack diagnostic accuracy for identifying patients with a positive SIJ block[5,6]. Data on diagnostic accuracy of the distraction, FABER and thigh thrust tests are equivocal[5-7,63]. The resisted hip abduction test has acceptable diagnostic accuracy[7].
- The ASLR test and the thigh thrust test have good predictive validity for identifying patients with post-partum pelvic pain associated with asymmetric SIJ laxity[62].

Multiple Test Regimens

- Inter-rater reliability of a cluster of tests consisting of the standing flexion, sitting PSIS palpation, supine-to-sit, and prone knee flexion tests varied from poor to excellent when using a dichotomous rating scale[48,65]. Three- and five point rating scales produced poor inter-rater agreement[48].
- Using a criterion of three of five provocation tests (consisting of the distraction, compression, pelvic torsion, FABER, and thigh thrust tests) to diagnose SIJD produced substantial inter-rater agreement[61].
- A positive result on a dichotomous rating scale for the test cluster consisting of the standing flexion, sitting PSIS palpation, supine-to-sit, and prone knee flexion tests was associated with LBP[66].
- A test cluster consisting of the FABER test and tenderness to palpation at the SS with a variation of other tests (including the shear, standing extension, pelvic torsion, or prone hip extension tests) was useful in determining which patients might need a diagnostic SIJ block[8].

Comprehensive Examination

- A comprehensive examination consisting of a McKenzie evaluation to exclude patients with diskogenic complaints and a score of three or more positive tests out of a cluster of SIJ provocation tests (consisting of the compression, distraction, thigh thrust, pelvic torsion, and sacral thrust tests) allowed for excellent diagnostic accuracy in identifying

patients responding to a double SIJ block[11].

In summary, when considering evidence-based diagnosis of SIJD:

- Patient history provides little information other than a report of predominant unilateral pain just inferior to the PSIS and a decreased likelihood of SIJD in older patients with pain radiating below the knee
- Trunk AROM tests provide no information helpful for the diagnosis of SIJD.
- Positional palpation test have insufficient reliability on any rating scale that might be useful to guide (manual) physical therapy interventions. The SIJD construct linking positional abnormalities to hypomobility or even to LBP is not supported by research. Positional palpation tests are not a valid indicator of SIJ position.
- Motion palpation tests lack sufficient reliability on any rating scale that might be useful to guide (manual) physical therapy interventions. The construct linking hypomobility to LBP is not supported for the standing flexion, sitting flexion, and supine-to-sit tests. The standing hip flexion, sitting flexion, standing flexion and sacral springing tests lack diagnostic accuracy. The standing hip flexion test is not a valid indicator of SIJ motion.
- The individual provocation tests studied generally have shown sufficient inter-rater reliability for clinical use. The FABER, thigh thrust, and resisted hip abduction tests appear to have sufficient diagnostic accuracy. The ASLR and thigh thrust test have predictive validity for post-partum SIJ-related pain. The SIJD construct linking SIJ laxity to LBP appears supported by current research.
- Using a cluster of SIJ provocation tests increases inter-rater agreement to a clinically useful range and specific clusters may help in establishing a course for further specialist differential diagnosis.
- A comprehensive examination consisting of a McKenzie evaluation and a cluster of SIJ tests provides for excellent accuracy in the diagnosis of SIJ-related pain.

It is obvious that an appropriate history and physical examination as discussed in this article can be used to diagnose SIJ-related pain. However, currently research does not support a specific diagnosis in the sense of an SIJ positional fault or specific hypomobility as needed to guide manual medicine interventions of SIJD. Further research on the level of patient outcome with manual medicine would seem the only route open to validate claims of manual medicine diagnostic and therapeutic efficacy.

Correspondence to:
Dr. Peter Huijbregts, PT
Shelbourne Physiotherapy Clinic
100B-3200 Shelbourne Street, Victoria, BC V8P 5G8
CANADA
(250) 598-9828 (Phone); (250) 598-9588 (Fax)
shelbournephysio@telus.net (E-mail)

References

1. Sakamoto N, et al. An electrophysiologic study of mechanoreceptors in the sacroiliac joint and adjacent tissues. Spine 2001;26:E468-471.

2. Fortin JD, Dwyer AP, West S, Pier J. Sacroiliac joint: Pain referral maps upon applying a new injection/arthrography technique, part I: Asymptomatic volunteers. Spine 1994;19:1475-1482.

3. 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.

4. Schwarzer AC, Aprill CN, Bogduk N. The sacroiliac joint in chronic low back pain. Spine 1995;20:31-37.

5. Maigne JY, Aivaliklis A, 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.

6. Dreyfuss P, Michaelsen M, Pauza K, McLarty J, Bogduk N. The value of medical history and physical examination in diagnosing sacroiliac joint pain. Spine 1996;21:2594-2602.

7. Broadhurst NA, Bond MJ. Pain provocation tests for the assessment of sacroiliac joint dysfunction. J Spinal Disord 1998;11:341-345.

8. Slipman CW, Sterenfeld EB, Chou LH, Herzog R, Vresilovic E. The predictive value of provocative sacroiliac joint stress maneuvers in the diagnosis of sacroiliac joint syndrome. Arch Phys Med Rehabil 1998;79:288-292.

9. Slipman CW, Jackson HB, Lipetz JS, Chan KT, Lenrow D, Vresilovic EJ. Sacroilac joint pain referral zones. Arch Phys Med Rehabil 2000;81:334-338.

10. Fukui S, Nosaka S. Pain patterns originating from the sacroiliac joints. J Anesth 2002;16:245-247.

11. Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: A validity study of a McKenzie evaluation and sacroiliac provocation tests. Aust J Physiother 2003;49:89-97.

12. Bernard TN, Cassidy JD. The sacroiliac syndrome. In: Frymoyer JW, Ed. The Adult Spine: Principles and Practice. New York, NY: Raven Press, 1991: 2107-2130.

13. Chan K, et al. Pelvic instability after bone graft harvesting from superior iliac crest: Report of nine patients. Skeletal Radiol 2001;30:278-281.

14. Ribeiro S, Prato-Schmidt A, Wurff P van der. Sacroiliac dysfunction. Acta Ortop Bras 2003;11:118-125.

15. Braun J, Sieper J, Bollow M. Imaging of sacroiliitis. Clin Rheumatol 2000;19:51-57.

16. Kamradt T, Loreck D. Sacroiliitis-it's not all B27. Z Rheumatol 1999;58:213-217.

17. Payer M. Neurological manifestations of sacral tumors. Neurosurgical Focus 2003;15 (2):Article 1. Available at: http://www.neurosurgery.org/focus/aug03/15-2-1.pdf. Accessed December 13, 2003.

18. El Maghraoui A, Tabache F, Bezza A, et al. A controlled study of sacroiliitis in Behcet's disease. Clin Rheumatol 2001;20:189-191.

19. Battistone MJ, Manaster BJ, Reda DJ, Clegg DO. The prevalence of sacroiliitis in psoriatic arthritis: New perspectives from a large, multicenter cohort. Skeletal Radiol 1999;28:196-201.

20. Weyland BM, Gimenez MV, Mueller-Haberstock S, Rommens PM. Tuberkuloese Destruktion des Iliosakralgelenks. Unfallchirurg 2001;104:359-362.

21. Paris SV. Mobilization of the spine. Phys Ther 1979;49:988-995.

22. Wurff P van der. Welke testen zijn aan te bevelen bij problematiek

van het SI-gewricht? Stimulus 2003;2:172-184.

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

24. Najm WI, Seffinger MA, Mishra SI, et al. Content validity of manual spinal palpatory exams: A systematic review. BMC Complementary and Alternative Medicine 2003;3:1. Available at: http://www.biomed-central.com/1472-6882/3/1. Accessed December 11, 2003.

25. Guide to Physical Therapist Practice, 2nd ed. Phys Ther 2001;81:9-744.

26. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Norwalk, CT: Appleton & Lange, 1993.

27. Fritz JM, Wainner RS. Examining diagnostic tests: An evidence-based perspective. Phys Ther 2001;81:1546-1564. Available at: http://www.ptjournal.org/PTJournal/September2001/ad090101546p.pdf. Accessed December 28, 2003.

28. Huijbregts PA. Spinal motion palpation: A review of reliability studies. J Manual Manipulative Ther 2002;10:24-39

29. Cummings GS, Crowell RD. Sources of error in clinical assessment of innominate rotation. Phys Ther 1988;68:77-78.

30. Badii M, Shin S, Torregiani WC, et al. Pelvic bone asymmetry in 323 study participants receiving abdominal CT scans. Spine 2003;28:1335-1339.

31. Mann M, Glasheen-Wray M, Nyberg R. Therapist agreement for palpation and observation of iliac crest heights. Phys Ther 1984;64:334-338.

32. Levangie PK. Four clinical tests of sacroiliac joint dysfunction: The association of test results with innominate torsion among patients with and without low back pain. Phys Ther 1999;79:1043-1057.

33. Huijbregts PA. Lumbopelvic region: Aging, disease, examination, diagnosis, and treatment. In: Wadsworth C, Ed. Current Concepts of Orthopedic Physical therapy, Home Study Course 11.2. LaCrosse, WI: Orthopaedic Section, APTA, 2001.

34. Bland JM, Altman DG. The odds ratio. BMJ 2000;320:1468. Available at: www.bmj.com/cgi/content/full/320/7247/1468. Accessed December 28, 2003.

35. Davidson M. The interpretation of diagnostic tests: A primer for physiotherapists. Aust J Physiother 2002;48:227-233.

36. Simon SD. Understanding the odds ratio and relative risk. J Andrology 2001;22:533-536.

37. Slipman CW, Patel RK, Whyte WS, et al. Diagnosing and managing sacroiliac pain. J Musculoskel Med 2001;18:325-332.

38. Gatterman MI. Disorders of the pelvic ring. In: Gatterman MI, Ed. Chiropractic Management of Spine Related Disorders. Philadelphia, PA: Lippincott, Williams & Wilkins, 1990.

39. Fortin JD, Falco FJE. The Fortin finger test: An indicator of sacroiliac pain. Am J Orthop 1997;24:477-480.

40. Winkel D, Ed. Het Sacroiliacale Gewricht. Houten, The Netherlands: Bohn Stafleu Van Loghum, 1991.

41. Potter NA, Rothstein JM. Intertester reliability for selected tests of the sacroiliac joint. Phys Ther 1985;65:1671-1675.

42. Janos SC. Palpation of selected bony landmarks in the lumbo-pelvic region. In: Paris SV, Ed. Proceedings 5th International Conference IFOMT. Vail, CO: June 1-5, 1992: A 111.

43. Richter T, Lawall J. Zur Zuverlaessigkeit manualdiagnostischer Befunde. Man Med 1993;31:1-11.

44. Tullberg T, Blomberg S, Branth B, Johnsson R. Manipulation does not alter the position of the sacroiliac joint: A Roentgen stereophotogrammetric analysis. Spine 1998;23:1124-1128.

45. Levangie PK. The association between static pelvic asymmetry and low back pain. Spine 1999;24:1234-1242.

46. O'Haire C, Gibbons P. Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation. Man Ther 2000;5:13-20.

47. Albert H, Godskesen M, Westergaard J. Evaluation of clinical tests used in classification procedures in pregnancy-related pelvic joint pain. Eur Spine J 2000;9:161-166.

48. Riddle DL, Freburger JK, North American Orthopaedic Rehabilitation Research Network. Evaluation of the presence of sacroiliac region

dysfunction using a combination of tests: A multicenter intertester reliability study. Phys Ther 2002;82:772-781.

49. Krawiec CJ, Denegar CR, Hertel J, Salvaterra GF, Buckley WE. Static innominate asymmetry and leg length discrepancy in asymptomatic collegiate athletes. Man Ther 2003;8:207-213.

50. Wiles MR. Reproducibility and interexaminer correlation of motion plapation findings of the sacroiliac joints. J Can Chiropr Assoc 1980;24:59-69.

51. Carmichael JP. Inter- and intra-examiner reliability of palpation for sacroiliac joint dysfunction. J Manipulative Physiol Ther 1987;10:164-171.

52. Herzog W, Read LJ, Conway PJW, Shaw LD, McEwen MC. Reliability of motion palpation procedures to detect sacroiliac joint fixations. J Manipulative Physiol Ther 1989;12:86-92.

53. Mior SA, McGregor M, Schut B. The role of experience in clinical accuracy. J Manipulative Physiol Ther 1990;13:68-71.

54. Dreyfuss P, Dreyer S, Griffin J, Hoffman J, Walsh N. Positive sacroiliac screening tests in asymptomatic adults. Spine 1994;19:1138-1143.

55. Bowman C, Gribble R. The value of the forward flexion tests and three tests of leg length changes in the clinical assessment of movement of the sacroiliac joint. J Orthop Med 1995:17:66-67.

56. Vincent-Smith B, Gibbons P. Inter-examiner and intra-examiner reliability of the standing flexion test. Man Ther 1999;4:87-93.

57. 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.

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

59. Strender LE, Sjoeblom A, Sundell K, Ludwig R, Taube A. Interexaminer reliability in physical examination of patients with low back pain. Spine 1997;22:814-820.

60. Mens JMA, Vleeming A, Snijders CJ, Stam HJ, Ginai AZ. The active straight leg raise test and mobility of the pelvic joints. Eur Spine J 1999;8:468-473.

61. Kokmeyer DJ, Wurff P van der, Aufdemkampe G, Fickenscher TCM. The reliability of multitest regimens with sacroiliac pain provocation tests. J Manipulative Physiol Ther 2002;25:42-48.

62. Damen L, Buyruk HM, Gueler-Ysal F, Lotgering FK, Snijders CJ, Stam HJ. The prognostic value of asymmetric laxity of the sacroiliac joints in pregnancy-related pelvic pain. Spine 2002;27:2820-2824.

63. Levin U, Stenstroem CH. Force and time recording for validating the sacroiliac distraction test. Clin Biomech 2003;18:821-826.

64. Levin U, Nilsson-Wikmar L, Harms-Ringdahl K, Stenstroem CH. Variability of forces applied by experienced physiotherapists during provocation of the sacroiliac joint. Clin Biomech 2001;16:300-306.

65. 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.

66. Cibulka MT. Koldehoff R. Clinical usefulness of a cluster of sacroiliac joint tests in patients with and without low back pain. J Orthop Sports Phys Ther 1999;29:83-92.

67. Domholdt E. Physical Therapy Research: Principles and Applications. Philadelphia, PA: WB. Saunders Company, 1993.

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

69. Laslett M. Letter to the editor. Spine 1998;23:962-963.