0008-3194/2007/30–41/$2.00/©JCCA 2007

Specificity, sensitivity, and predictive values of clinical tests of the sacroiliac joint: a systematic review of the literature

Kent Jason Stuber, BSc, DC*

Objective: *To determine which physical examination tests have the highest sensitivity, specificity, and predictive values for determining the presence of sacroiliac joint injuries and/or dysfunction when compared with the gold standard of a sacroiliac joint block.*

Data sources: *A systematic search of the literature was conducted for articles that evaluated clinical sacroiliac joint tests for sensitivity, specificity, and predictive value when compared to sacroiliac joint block. The search was conducted using several online databases: Medline, Embase, Cinahl, AMED, and the Index to Chiropractic Literature. Reference and journal searching and contact with several experts in the area was also employed.*

Data extraction: *Studies selected for inclusion were evaluated with a data extraction sheet and assessed for methodological quality using an assessment tool based on accepted principles of evaluation.*

Data synthesis: *Article results were compared, no attempt to formally combine the results into a meta-analysis was made.*

Results: *Seven papers were identified for inclusion in the review, two of which dealt with the same study, thus six studies were to be assessed although one paper could not be obtained. The most recently published article had the highest methodological quality. Study designs rarely incorporated randomized, placebo controlled, double blinded study designs or confirmatory sacroiliac joint blocks. There was considerable inconsistency between studies in design and outcome measurement, making comparison difficult. Five tests were found to have sensitivity and specificity over 60% each in at least one*

Objectif : *Déterminer quels tests d'examen physique ont la plus haute sensibilité, spécificité et valeur prédictive pour déterminer la présence de blessures à articulation sacro-iliaque ou le dysfonctionnement quand on le compare à l'exemple idéal du blocage de l'articulation sacro-iliaque.*

Source de données : Une recherche systématique a été effectuée de la documentation médicale portant sur les articles traitant de l'évaluation des tests cliniques de l'articulation sacro-iliaque quant aux aspects de la sensibilité, de la spécificité et de la valeur prédictive par comparaison au blocage de l'articulation sacro-iliaque. La recherche a été menée en utilisant plusieurs bases de données en ligne : Medline, Embase, Cinahl, AMED et l'Index to Chiropractic Literature. Nous avons également eu recours à des références, des recherches dans des magazines spécialisés et des communications avec des spécialistes.

Extraction des données : *Les études retenues ont été évaluées avec une grille d'extraction des données et évaluées pour la qualité de leur méthodologie en utilisant un outil d'évaluation fondé sur des principes acceptés d'évaluation.*

Synthèse des données : *Les conclusions des articles ont été comparées, aucune tentative n'a été faite de rassembler les résultats dans une méta-analyse.*

Résultats : *Sept études ont été retenues pour analyse, dont deux portaient sur le même sujet, donc six études devaient être évaluées bien qu'une d'entre elles n'a pu être obtenue. Le plus récent article publié possédait la plus haute qualité méthodologique. La méthodologie inclut rarement la technique aléatoire double aveugle*

* Graduate student (MSc in Health and Social Care Research Program, University of Sheffield School of Health and Related Research, Private practice, West Springs Chiropractic & Health Centre – 19–8 Weston Drive SW, Calgary, Alberta, Canada; 403-685-5252. Post-graduate faculty, University of Bridgeport Chiropractic College.

study with at least moderately high methodological quality. Using several tests and requiring a minimum number to be positive yielded adequate sensitivity and specificity for identifying sacroiliac joint injury when compared with sacroiliac joint block.

Conclusion: *Practitioners may consider using the distraction test, compression test, thigh thrust/posterior shear, sacral thrust, and resisted hip abduction as these were the only tests to have specificity and sensitivity greater than 60% in at least one study. Further research using improved methodology is required to determine the optimal tests and combinations of tests to identify sacroiliac joint injuries.*
(JCCA 2007; 51(1):30–41)

KEY WORDS: sacroiliac, joint, examination.

randomisée avec placebo ou blocage de joint sacro-iliaque confirmatoire. Il y avait une incompatibilité considérable entre les études dans la méthodologie et les mesures de résultat, ce qui a rendu les comparaisons difficiles. Dans cinq études, on a trouvé que la sensibilité et la spécificité supérieure à 60% dans au moins une étude avec au moins une qualité méthodologique modérément élevée. Le recours à plusieurs tests et à un nombre minimum de résultats positifs nous a permis d'obtenir une sensibilité et une spécificité adéquates pour identifier une blessure à l'articulation sacro-iliaque en comparaison à un blocage de l'articulation sacro-iliaque.

Conclusion : *Les spécialistes de la santé peuvent envisager d'utiliser le test d'Apley à la distension ligamentaire, le test de compression, poussée de la cuisse/cisaillement au postérieur, poussée sacro-iliaque et résistance à l'abduction de la hanche qui se sont avérés les seuls tests ayant une sensitivité et une spécificité supérieures à 60% dans au moins une étude. Il faut d'autres recherches en faisant usage d'une méthodologie améliorée pour déterminer les tests optimaux et les conjugaisons de tests pour identifier les lésions articulaires sacro-iliaques.*
(JACC 2007; 51(1):30–41)

MOTS CLÉS : sacro-iliaque, articulation, examen.

Introduction

The sacroiliac joint (SI joint) is a frequent source of low back and referred leg pain, being an indicated pain source in approximately 30% of patients with chronic low back pain.[1,2,3] Unfortunately as with many sources of low back pain, it can be difficult to determine which physical examination tests to use to identify the presence of sacroiliac joint pain and/or dysfunction. The anesthetic sacroiliac joint block (hereafter referred to as the SI joint block) has been identified as a controversial gold standard in the identification of pain originating from the SI joint.[4] Unfortunately, this type of testing is not feasible for practitioners who lack training in joint injections and is not as cost-effective as a physical examination maneuver (or maneuvers).[4] Many physical examination tests have been developed for eliciting SI joint pain, but it is desirable to know which tests have a higher probability of accurately ruling in or ruling out this joint as the source of pain for patients with this condition.

Objective

The objective of this review was to determine which physical examination tests have the highest sensitivity, specificity, and predictive values for determining the presence of SI joint injuries and/or dysfunction when compared with the gold standard of an SI joint block.

Methods

A systematic review of the literature was conducted to determine the sensitivity, specificity, and predictive values of any clinical SI joint tests when compared to the gold standard of a SI joint block (injection).

Table 1 Initial search strategy

Concept 1	Sacroiliac joint injur* OR Sacroiliac joint sprain OR Sacroiliac joint syndrome
	AND
Concept 2	Provocation test* OR maneuver* OR Assessment OR Palpation OR Palpatory OR Examination OR Physical Examination
	AND
Concept 3	Sacroiliac joint block OR Joint block
	AND
Concept 4	Sensitivity OR Specificity OR Sensitivity and Specificity OR Accurate OR Accuracy OR predictive value OR Predictive value of tests

Table 2 Modified search strategy

Concept 1	Sacroiliac joint injur* OR Sacroiliac joint sprain OR Sacroiliac joint syndrome
	AND
Concept 2	Provocation test* OR maneuver* OR Assessment OR Palpation OR Palpatory OR Examination OR Physical Examination
	AND
Concept 3	Sensitivity OR Specificity OR Sensitivity and Specificity OR Accurate OR Accuracy OR predictive value OR Predictive value of tests

Search strategy

A multi-component search strategy was employed. A search was conducted, without restrictions on language, date or article type, in the following online databases: Medline (1966–Oct Week 3 2005), Embase (1980–2005), Cinahl (1982–Oct Week 3 2005), AMED (1985–Oct 2005), and the Index to Chiropractic Literature (1985–2005). The search terms and strategy employed can be seen in tables one and two. The initial search strategy yielded very few results (only two), so the search strategy was modified to retrieve more articles.

In order to retrieve articles from the "grey" literature, the author attempted to contact experts in the field of low back examinations to identify previously unpublished data, conference abstracts, or text-book chapters. The author also hand searched several journals for unidentified articles including *Spine*, *Journal of Manipulative & Physiological Therapeutics*, *Journal of Manual and Manipulative Therapy*, *Journal of Spinal Disorders*, *Archives of Physical Medicine and Rehabilitation*, and *Manual Therapy*. These journals were searched from their publication start date to December 2005, including any special editions. Finally, the references in all retrieved articles were searched for relevant papers.

Study selection

Studies were selected for inclusion in this review if they assessed the accuracy and specifically the sensitivity, specificity, and/or predictive values of any physical examination test of the SI joint in comparison with a gold standard (an anaesthetic SI joint block or injection). Articles were excluded if they assessed the validity of clinical SI joint tests but did not compare them with the gold standard (SI joint block). All languages, dates and journals of publication, professions and names of authors were included with the hope of minimizing bias.

Table 3 Criteria descriptions (based on van derWurff, Meyne, and Hagmeijer (4)) and scoring for included studies.

Criterion description	*Laslett 2003, 2005 (6, 7)*	*Dreyfuss 1996 (12)*	*Broadhurst 1998 (9)*	*Maigne 1996 (11)*	*Slipman 1998 (10)*	*Fortin 1997 (8)*
Study population described (age, gender, duration of symptoms, time off work – each worth 2.5 points apiece)	10	2.5, 2.5, 2.5, 0	2.5, 2.5, 0, 0	2.5, 2.5, 2.5, 0	2.5, 2.5, 2.5, 0	/
Study inclusion and exclusion criteria clearly described (5 points apiece)	10	5, 0	10	10	10	/
Number of subjects described and number of drop-outs, subjects excluded described (5 points apiece)	10	10	5,0	10	5, 0	/
Clinical test description (10 for a complete and detailed description, 5 for a partial description)	10	10	5	10	0	/
Examiner described (out of 10)	10	10	0	0	0	/
Description of outcome measure of test and gold standard (5 points apiece)	10	10	10	10	10	/
Description of gold standard and examiner conducting the gold standard testing (5 points apiece)	10	5, 0	5,0	5, 0	10	/
Blinding of subject and examiner to group allocation (2.5 points apiece) and blinding of gold standard examiner to presence or absence of positive clinical test signs (5 points)	0, 0, 5	0, 0, 0	10	0, 0, 0	0, 0, 0	/
Randomization of subjects into an active gold standard group or placebo group (5 points), use of confirmatory diagnostic blocks (5 points).	0, 5	0, 0	5, 0	0, 5	0, 0	/
Data reporting frequency of positive/ negative tests (either individual tests or groups of tests (5 points), sensitivity, specificity, or predictive values (5 points)	10	10	10	0, 0	10	/
Total score (out of 100)	90	67.5	65	57.5	52.5	/

The population of interest had to be patients with mechanical low back pain of likely SI joint origin. Reliability studies for clinical SI joint tests were excluded unless they included data on the above mentioned measures. Studies that examined pregnant patients or those with degenerative joint disease, inflammatory arthropathies, malignancy, or other systemic diseases were excluded.

Table 4 Predictive values of individual tests (all values converted to percentages)

First author, year	*Test*	*Positive predictive value*	*Negative predictive value*
Laslett, 2005	Distraction	60%	81%
	Compression	52%	82%
	Thigh thrust	58%	92%
	Gaenslen's (right)	47%	76%
	Gaenslen's (left)	50%	77%
	Sacral thrust	56%	80%

Assessment of studies

A data extraction tool was formulated (based upon a previously devised form[5]) and revised throughout the article assessment process. This tool allowed for the sensitivity, specificity, and predictive value of the different tests in each study to be calculated or recorded. In addition the extraction tool allowed for documentation and critical appraisal of the methods used in each study.

A ten item study methodology assessment sheet was devised by the author to assess studies for methodological quality (see table three). Using this sheet each study was assessed subjectively by the author. This sheet was not formally tested for validity or reliability but was based on principles described in the review by van der Wurff, Meyne, and Hagmeijer[4] and appeared to have face validity (see table 3). Use of this sheet led to a score out of 100 for each study and allowed for method quality comparison between the included studies.

Data were reported in the form of tables illustrating the specificity, sensitivity, and predictive values of different tests or composites of positive tests, as well as a table illustrating the scores of the different studies on the methodology assessment sheet (included as part of table three).

Results

Search results

An initial yield of six papers for analysis was garnered from the database searches. One further paper by Fortin and Falco[8] was located through reference searching. Two of the papers actually pertained to the same study,[6,7] the results from these two were combined during data extraction and analysis, thus six studies were to be analyzed. One paper could not be obtained[8] and thus five studies (in six papers) were included in the analysis. One systematic review was identified[4] via reference searching. Contacting several experts in the field did not yield any further materials, nor did hand-searching of journals.

Methodological quality

The methodological quality of the included studies can be seen in table three. Each study suffered from at least one methodological flaw. The study by Laslett et al.[6,7] had the highest methodological quality score, whereas that by Slipman et al.[10] had the lowest. Only one study employed any form of randomization into placebo and active injection groups and employed double blinding of study staff;[9] despite this it had merely the third highest methodology score. The study by Maigne et al.[11] did not include calculations of specificity, sensitivity, or predictive values; however along with Laslett et al.[6,7] it did employ confirmatory blocks. Only Maigne et al.[11] and Laslett et al.[6,7] blinded the doctor performing the injections.

Predictive value

The positive predictive value of a test is how frequently those who have a positive test will actually have the condition, whilst the negative predictive value of a test refers to how frequently those with a negative test do not have the condition.[4] As can be seen from table four, reporting on Laslett et al.,[6] none of the tests exhibited a positive predictive value of greater than 60%, whereas the negative predictive values for these tests were considerably higher.

Table 5 Predictive values of composites of tests (all values converted to percentages)

First author, year	*Tests*	*Positive predictive value*	*Negative predictive value*
Slipman, 1998	History and clinical examination (at least three of Patick's test*, pain with pressure application to sacroiliac ligaments at sacral sulcus*, Shear test, standing extension, Gaenslen's test, Yeoman's test)	60%	N/A
Laslett, 2005	Two out of four tests positive (distraction, thigh thrust, compression, and sacral thrust)	67%	93%
Laslett, 2005	1 or more positive tests	47%	100%
	2 or more positive tests	58%	96%
	3 or more positive tests	68%	96%
	4 or more positive tests	60%	81%
	5 or more positive tests	50%	72%

* – test had to be positive for inclusion

Table five illustrates the predictive values of using a group of tests, or of having one or more positive tests. This concept was evaluated by Laslett et al.,[6] as well as by Slipman et al.[10] Slipman et al. showed that a history and clinical examination with at least three positive tests has a positive predictive value of 60%,[10] whereas Laslett et al. found that having two out of four SI joint tests positive led to positive and negative predictive values of 67% and 93% respectively.[6] Interestingly Laslet et al. found that with an increasing number of positive SI joint tests (more than three), positive and negative predictive values actually began to decrease.[6]

Sensitivity

Sensitivity is the proportion of people with a positive test result who have the target disorder, essentially true positives.[4,13] Table six depicts the sensitivity of individuals tests evaluated in the various studies. Several tests appeared in more than one paper, each with different results. Such differences may be at least partially attributed to variations in pain relief standards required for the injection considered to be positive (90% or greater in Dreyfuss et al.,[12] ≥70% in Broadhurst and Bond,[9] and ≥80% in Laslett et al.[6]). Other variations in methodology could also explain these differences. For example, Broadhurst and Bond[9] used the clinical tests before and after the patient received the injection and looked at the change in VAS results on the clinical tests to determine positive or negative findings[9] as opposed to Dreyfuss et al.[12] and Laslett et al.[6] who performed the SI joint tests initially and then gave subjects injections and looked for positive or negative responses to the injection, they did not re-perform the clinical tests.[6,7,12]

The thigh thrust test in Dreyfuss et al.[12] and Laslett et al.,[6,7] is the same as the posterior shear test noted in Broadhurst and Bond.[9] This test had sensitivity of 36%,[12] 80%[9] and 88%[6] in the different studies and 69% when Broadhurst and Bond[9] applied a 90% pain relief criteria. The Patrick's test (from Dreyfuss et al.[12]) and the FABER (Flexion ABduction External Rotation)[9] test are one and the same. In one study this test had a sensitivity of 69%[12] as opposed to 77% in another[9] with a 70% pain relief criteria or 50% using a 90% pain relief criteria. The Gaenslen's test had results of 71%[12] and 51.5% (on average)[6] in two different studies. The sacral thrust had sensitivity of 53%[12] and 63%[6] in the same studies respectively. Overall the test with the lowest sensitivity was the thigh thrust as per Dreyfuss et al.,[12] and the test with the highest sensitivity was the sacral sulcus test conducted by Dreyfuss et al.[12]

Table 6 Sensitivity of individual tests (all values converted to percentages)

First author, year	*Test*	*Sensitivity*
Dreyfuss, 1996	Gillet test	43%
	Thigh thrust	36%
	Patrick's	69%
	Gaenslen's	71%
	Midsacral thrust	53%
	Spring	75%
	Sacral sulcus tenderness	95%
Broadhurst, 1998	≥70%* FABER**	77%
	≥70%* POSH**	80%
	≥70%* REAB**	87%
Broadhurst, 1998	≥90%* FABER**	50%
	≥90%* POSH**	69%
	≥90%* REAB**	65%
Laslett, 2005	Distraction	60%
	Compression	69%
	Thigh thrust	88%
	Gaenslen's (right)	53%
	Gaenslen's (left)	50%
	Sacral thrust	63%

* – 70% or 90% pain relief on VAS
** – FABER = Flexion Abduction External Rotation test; POSH = Posterior Shear test; REAB = Resisted Abduction of the hip test

The sensitivity of having numerous positive results decreased in the studies by Dreyfuss et al.[12] and Laslett et al.[6] as the number of tests required to be positive increased, as seen in table seven. Laslett et al. evaluated the sensitivity of using at least two positive out of four tests and at least three positive out of five tests and this was found to have sufficiently high sensitivity (88% and 91%, respectively).[6,7]

Specificity

Specificity is the proportion of people with a negative test result who do not have the target disorder, essentially true negatives.[4,13] Table eight depicts the specificity of individual tests evaluated in the various studies. Again the results of several tests can be compared from one study to another. The thigh thrust/posterior shear test had specificity of 50%,[12] 100% (with both 70% and 90% pain relief criteria),[9] and 69%[6] in three different studies. The Gaenslen's test had results of 26%[12] and 74% (on average)[6] in two studies. The sacral thrust had specificity of 29%[12] and 75%[6] in the same respective studies. Again differences in results between studies may be explained by different pain relief standards required for the injection to be considered positive and by other methodological differences.

The test with the lowest specificity was the sacral sulcus tenderness test at 9%[12] and the highest specificity was for the REsisted hip ABduction (REAB) test, POste-

Table 7 Sensitivity of composites of tests (all values converted to percentages)

First author, year	*Test*	*Sensitivity*
Dreyfuss, 1996	6 positive tests	57%
	7	53%
	8	29%
	9	29%
	10	0%
	11	0%
Laslett, 2003	Three or more positive tests out of five (Distraction, Gaenslen's, Thigh Thrust, Compression, Sacral Thrust)	91%
Laslett, 2005	Two or more out of four tests positive (distraction, thigh thrust, compression, and sacral thrust))	88%
Laslett, 2005	1 or more positive tests	100%
	2 or more positive tests	93%
	3 or more positive tests	94%
	4 or more positive tests	60%
	5 or more positive tests	27%

rior SHear test (POSH), and FABER's test used by Broadhurst and Bond,[9] each with specificities of 100% regardless of pain relief criteria, followed by the distraction test with a specificity of 81%.[6,7]

As table nine depicts, the specificity of numerous positive results generally increased as the number of positive tests increased.[6,7,12] Laslett et al. examined the specificity of using at least two positive out of four tests and at least three positive out of five tests and this was found to have sufficiently high specificity (78% and 87%, respectively).[6,7]

Discussion

From the results of this review, it appears that two trends can be discerned. First, employing a group of tests with a requisite number of positive tests (such as two out of four or three out of five) in order to diagnose a SI joint injury may be desirable[4] as Laslett et al.[6,7] demonstrate that employing this approach results in adequate sensitivity, specificity and predictive values, and Slipman et al. demonstrated an acceptable positive predictive value when three out of five tests were positive.[10] Second, certain tests appear to have higher sensitivity and specificity than others. It would seem prudent to include tests with higher sensitivity and specificity in an examination where a SI joint injury is among the differential diagnoses, as van der Wurff et al.[4] note that a test should have higher values of both in order to be of use, in addition to being reliable.

Only five tests had both sensitivity and specificity greater than 60% in at least one study (a value of 50% or lower for a test has been deemed unacceptable by some authors[4]): distraction, compression, thigh thrust/poster shear, sacral thrust, and resisted abduction of the hip. Most of these tests were supported by studies with either very high (Laslett et al.[6,7]) or moderately high methodology scores (Broadhurst and Bond[9]). Unfortunately the positive predictive values of the distraction, compression, thigh thrust/posterior shear, and sacral thrust were not extremely high, although their negative predictive values were at least 80% in the study by Laslett et al.[6,7]

Methodological quality

The methodological quality of the studies included in this trial was variable and each study had at least one significant methodological flaw. The results from the studies

Table 8 Specificity of individual tests (all values converted to percentages)

First author, year	*Test*	*Specificity*
Dreyfuss, 1996	Gillet test	68%
	Thigh thrust	50%
	Patrick's	16%
	Gaenslen's	26%
	Midsacral thrust	29%
	Spring	35%
	Sacral sulcus tenderness	9%
Broadhurst, 1998	≥70%* FABER**	100%
	≥70%* POSH**	100%
	≥70%* REAB**	100%
Broadhurst, 1998	≥90%* FABER**	100%
	≥90%* POSH**	100%
	≥90%* REAB**	100%
Laslett, 2005	Distraction	81%
	Compression	69%
	Thigh thrust	69%
	Gaenslen's (right)	71%
	Gaenslen's (left)	77%
	Sacral thrust	75%

* – >70% or >90% pain relief on VAS
** – FABER = Flexion Abduction External Rotation test; POSH = Posterior Shear test; REAB = Resisted Abduction of the hip test

with the lowest methodological quality (in particular Maigne et al.[11] and Slipman et al.[10]) should be cautiously considered before incorporating the findings into practice. The quality of the study by Laslett et al. was considerably higher than the remainder of the studies.[6,7] Broadhurst and Bond's[9] study was significantly different from the others in that a positive or negative response to the injections was determined by the amount of pain relieved when re-performing the physical tests. As this protocol deviated significantly from the others, used the lowest pain relief criteria on injection (≥70%, although results were also calculated using ≥90%) and found 100% specificity of all of the tests that it examined, the results of this study may require cautious consideration.

It is encouraging that the studies evaluated in this review that have been previously evaluated[4] had the same order of decreasing quality in both reviews. This paper is the first to examine the studies by Laslett et al.[6,7] for methodological quality.

Shortcomings in the literature

Aside from the low methodological quality of some of the included studies, one of the glaring shortcomings from most of the studies was the lack of single or double-blinding, confirmatory blocks, and randomized placebo controlled trials.[4] Inconsistency in the definition of a positive response to the joint block (≥70% to ≥90% pain relief on VAS in the different studies) was also a detriment

Table 9 Specificity of composites of tests (all values converted to percentages)

First author, year	*Test*	*Specificity*
Dreyfuss, 1996	6 positive tests	42%
	7	55%
	8	52%
	9	68%
	10	87%
	11	83%
Laslett, 2003	Three or more positive tests out of five (Distraction, Gaenslen's, Thigh Thrust, Compression, Sacral Thrust)	87%
Laslett, 2005	Two out of four tests positive (distraction, thigh thrust, compression, and sacral thrust)	78%
Laslett, 2005	1 or more positive tests	44%
	2 or more positive tests	66%
	3 or more positive tests	78%
	4 or more positive tests	81%
	5 or more positive tests	88%

as a whole, as were inconsistencies in the type and concentrations of anesthetic employed.[4] A standardized value of minimum pain relief on VAS should be identified and future studies should employ this value to allow for easier comparison. A minimum of 80% pain relief on VAS is recommended here for determining a positive response to joint injection, as this value was employed in two out of the five studies evaluated in this review, and is a reasonable compromise between the two extreme values of $\geq$70% and $\geq$90%. It would seem unlikely that relieving at least 80% of a patient's pain on injection would be due to chance. In addition, using a standardized injection approach, and type and concentration of anesthetic would make studies more comparable.

Another noticeable issue is that of selecting the clinical tests evaluated in each study. The protocol for test selection employed by Dreyfuss et al.[12] seemed the most reasonable as they assembled a panel of experts and determined the tests that the experts felt would be most reliable out of a list of twenty, and conducted their testing on the twelve deemed most reliable by the panel.[12] Future studies could consider employing a similar test selection process as it would hopefully minimize the potential bias of authors selecting the tests that they are passionate (or perhaps dispassionate) about for evaluation.

This review had several potential shortcomings, the first being that one study (Fortin and Falco) that was anticipated for inclusion in the analysis could not be obtained. However, van der Wurff et al. reviewed this article and found it to have very low methodological quality (18 out of 100),[4] indicating that it might not have had a tremendous impact on the results herein. Next, the data extraction and methodological assessments of each study were only performed by one person, eliminating the possibility of establishing consensus with multiple reviewers or evaluating agreement/disagreement between reviewers, and the paper may accordingly suffer from the inherent potential biases of any subjective evaluation and rating of the literature.

The methodology assessment criteria employed in this review were not evaluated for validity or reliability. They were loosely based on the criteria employed by van der

Wurff et al.,[4] employed a reasonable and easy-to-use scoring system, and appeared to have face validity. This review did not include an assessment of reliability of the tests, and this is an essential quality for diagnostic tests to possess in order for them to be useful clinically.[4] A systematic review that assesses the reliability of clinical tests for the sacroiliac joint has been previously conducted.[14]

Conclusions

There is a general paucity of evidence to support the use of most clinical SI joint tests, yet the studies conducted thus far have produced some useful results. No test has proven to be overwhelmingly superior to the others, but combining several tests with higher sensitivity and specificity may allow for more accurate results, although this notion still requires formal evaluation.

Implications for practice

While definitive conclusions cannot be made from the literature to date, it appears that clinicians may want to consider using an array of provocative tests with a minimum number of positive tests (two out of four or three out of five, etc) needed for the diagnosis of SI joint dysfunction/sprain/injury/etc.[4] The distraction test, compression test, resisted hip abduction, thigh thrust/posterior shear, and sacral thrust may be likely candidates to include in such an examination as they have each produced specificity and sensitivity greater than 60% in at least one study.

Implications for research

As there have only been six studies in this area of assessment of SI joint tests, there is still a need for more research to be done. There is a need to confirm much of the work that has previously been done with improved and consistent study methods. A large scale double-blinded randomized trial comparing placebo and active joint blocks and using a minimum of 80% pain relief criteria for a positive on injection, with the use of a confirmatory block should be undertaken. Having a control group for comparison's sake would also be beneficial.[4] Such a study should assess as many SI joint tests as reasonably possible (as agreed to by a consensus panel). This would hopefully determine the most specific and sensitive SI joint tests, or combinations of tests to aid with diagnosis of this condition. Further assessments of these tests and protocols for reliability are also in need.

Potential conflicts of interest

None identified.

Sources of support and acknowledgment

The author would like to thank the University of Sheffield School of Health and Related Research for their support via a University of Sheffield School of Health and Related Research Training in Health bursary.

References

1 Schwarzer AC, Aprill CN, Bogduk N. The sacroiliac joint in chronic low back pain. Spine 1995; 20(1):31–37.
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(13):1475–1482.
3 Slipman CW, Jackson HB, Lipetz JS, Chan KT, Lenrow D, Vresilovic EJ. Sacroiliac joint pain referral zones. Arch Phys Med Rehabil 2000; 81:334–338.
4 van der Wurff P, Meyne W, Hagmeijer RHM. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 2: validity. Man Ther 2000; 5(2):89–96.
5 NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness. CRD's guidance for those carrying out or commissioning reviews. CRD Report Number 4 (2nd Ed) 2001. Appendix 3 – example of data extraction sheets. www.york.ac.uk/inst/crd/pdf/crd4_app3.pdf
6 Laslett M, Aprill CN, McDonald B, Young SB. Diagnosis of sacroiliac joint pain: validity of individual provocation tests and composites of tests. Man Ther 2005; 10:207–218.
7 Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of McKenzie evaluation and sacroiliac provocation tests. Aust J Physiother 2003; 49:89–97.
8 Fortin JD, Falco FJ. The Fortin finger test: an indicator of sacroiliac pain. Am J Orthop 1997; 26(7):477–480.
9 Broadhurst NA, Bond MJ. Pain provocation tests for the assessment of sacroiliac joint dysfunction. J Spine Disorders 1998; 11(4):341–345.
10 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.
11 Maigne JY, Aivailiklis 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(16):1889–1892.
12 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(22):2594–2602.

13 Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test: B. What are the results and will they help me in caring for my patients? 1994; 271(5):389–391.

14 van der Wurff P, Hagmeijer RHM, Meyne W. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: reliability. Man Ther 2000; 5(1):30–36.