Manual Therapy (2000) **5(1),** 30–36

DOI: 10.1054/math.1999.0228, available online at http://www.idealibrary.com on IDEAL®

Original article

Clinical tests of the sacroiliac joint

A systematic methodological review. Part 1: Reliability

P. van der Wurff*, R. H. M. Hagmeijer*, W. Meyne*†

**Manual Therapist, Department of Physiotherapy, Military Rehabilitation Centre 'Aardenburg', Doorn,*
†Physiotherapist in Private Practice, Bussum, The Netherlands

SUMMARY. **In the literature concerning the sacroiliac joint (SIJ) there are numerous specific tests used to detect joint mobility or pain provocation. In this article the authors have reviewed 11 studies which investigated the reliability of these tests. The methodological quality of the studies was tested by a list of criteria developed by the authors. This list consisted of three categories: (1) study population, (2) test procedures and (3) test results. To each criterion a weighting was attached. The methodological score for nine out of the 11 studies was found to be acceptable. The results of this review, however, could not demonstrate reliable outcomes and therefore no evidence on which to base acceptance of mobility tests of the SIJ into daily clinical practice. There are no indications that 'upgrading' of methodological quality would have improved the final conclusions. With respect to pain provocation tests, the findings did not show the same trend. Two studies demonstrated reliable results using the Gaenslen test and the Thigh thrust test. One study showed acceptable reliability for five other pain provocation tests; however, since other authors have described contradictory results, there is a necessity for further research in this area with an emphasis on multiple test scores and pain provocation tests of the SIJ.** © **2000 Harcourt Publishers Ltd**

INTRODUCTION

Many authors share the opinion that the sacroiliac joint (SIJ) may be a source of low back and buttock pain (Potter & Rothstein 1985; Sturesson et al. 1989; Shaw 1992). There is still, however, controversy surrounding the existence of motion, pain, dysfunction and clinical diagnostic procedures of the SIJ (Beal 1982; Walker 1992). It is now generally accepted that a small amount of motion exists in the SIJ (Vleeming 1990; Alderink 1991). Dysfunction of the SIJ is defined as a state of relative hypomobility within a portion of the joint's range of motion with subsequent altered structural (positional) relationships between the sacrum and the ilium (Dreyfuss et al. 1994). The prevalence of SIJ dysfunction in patients with low back pain was investigated by Schmid in 1985 among 1344 patients with low back pain and SIJ involvement was presumed in 467 cases (35%), if seven out of 14 SIJ dysfunction predictors were found to be positive.

Schwarzer et al. (1995) found in 100 patients with low back pain, a prevalence of at least 13% and perhaps as high as 30% of SIJ pain detected by a combination of SIJ anaesthetic blocks and pain provocation tests. The authors further argued that, until now, the study of the SIJ has been hindered by the fact that there has not been a satisfactory standard of criteria by which its prevalence can be measured and against which various clinical examinations can be validated.

For the clinician it is important to diagnose low back pain properly, and SIJ dysfunction in particular, in order to treat the problem in an appropriate way (Kirkaldy-Willis & Hill 1979). There is a wide variety of SIJ tests available to detect dysfunction, but none of them appears to be superior to others. In general there are three types of tests used to examine the SIJ: (1) motion palpation tests to assess movement; (2) pain provocation tests to stress SIJ structures and (3) tests for pelvic position. In this article, only the first and second test types will be discussed because they

Peter van der Wurff, Ruud H. M. Hagmeijer, Department of Physiotherapy, Military Rehabilitation Centre 'Aardenburg', Postbus 185, 3941 AD Doorn, **Wilco Meyne**, Physiotherapist in Private Practice, Bussum, The Netherlands.
Correspondence to: PvdW.

are the most frequently studied SIJ tests. To accept SIJ tests for clinical use it is essential that the tests used are reproducible so that they are capable of yielding meaningful results. A test may give illusory information without any reliability, although demonstration of reliability alone is insufficient to determine if a test is valid and valuable (Potter & Rothstein 1985). Since a single test might be not sufficient in diagnosing back pain, Haas (1991a) suggests using a complete battery of tests, the so-called 'multiple-test score'. This regime makes it possible for two observers to be in total disagreement about the exact number of positive tests yet be in complete agreement with the final conclusion when accepting that three out of five tests give positive results.

Interpreting SIJ tests is difficult, because the variations in anatomy and motion in the SIJ are accompanied by movements of the lumbar spine and hip (Paris 1992).

Nelson et al. (1979), found in their study of reliability and reproducibility of clinical findings in low back pain, an observer error of approximately 30%. Walker (1992) doubted the capability of clinicians particularly to detect movement in the order of 1–3° or 1–3 mm in the SIJ. The term 'play' would be more suitable to use instead of 'motion' even though there is no similarity with other synovial joints. The possible positive SIJ tests do not allow one to directly implicate the sacroiliac joint to the exclusion of the other adjacent structures (Maigne et al. 1996).

During the last decade two articles have been published which critically review the literature on tests of the SIJ (Oldreive 1995; Pescioli & Kool 1997). Neither of these articles include a systematic methodological review of the topic. In this article, the authors present a systematic methodological review of reliability studies concerning SIJ tests for motion palpation and pain provocation. The authors have also reviewed the results of validity studies on SIJ tests. This is the subject of a second article (part 2) to be published in the next issue of this journal.

The aim of this review was to determine reliable tests for the SIJ that are relevant in daily clinical practice. In addition the authors wished to make suggestions for further study of this subject.

METHODS

Study selection

For this systematic methodological review, the authors included studies that met the following conditions:

- The results must have been published as a full report before February 1999 (Abstracts, letters and unpublished studies were not selected)
- All relevant clinical tests of the SIJ were included
- The study had to be of either an inter- or intra-examiner reliability design. Validity studies were

Table 1 Criteria list for methodological assessment of reliability trials for SIJ-dysfunction

Criteria	Weighting
Study population	
A 1. Description of study population i.e. volunteers or patients, age, gender, etc.	8
2. Description of inclusion and exclusion criteria	7
B Drop-outs described, information from which group and with reason for withdrawal	5
C Number of subjects	
<25 subjects	0
>25 subjects	3
>50 subjects	6
>75 subjects	10
Test procedure	
D Standardization of test procedure	
1. Position of subject	3
2. Position of examiner	2
3. Description of palpation technique (position hands of examiner)	3
4. Description of neutralizing simple exercises for low back and pelvis before or during the test procedure	2
5. Information given to the subject about the test procedure	2
6. Standardization according to the original description of the test in the literature (referenced)	4
E Selection of examiner	
1. Description of the choice for experienced examiners	3
2. Description of less-experienced examiner	2
3. Description of a consensus procedure	9
Test results	
F Standardized measurement of test outcome	5
G Test/re-test procedure, description of time interval	2
H Procedure of blinding	
1. Attempt to blinding the examiner	2
2. Subject not informed of outcome	1
3. Results sealed, the examiners could not see each others' findings	5
I Descriptive statistics: frequencies and total agreement	10
J Inferential statistics: Cohen's Kappa or ICC	15

excluded because methodological review on both topics in one session was, in the authors' opinions, not justifiable
- The study to be written in English, German, French or Dutch.

A Medline, Embase and CINAHL literature search was carried out from the period between January 1980 and February 1999. The keywords that were used were; sacroiliac joint, physical examination, palpation, evaluation studies, reproducibility, reliability and assessment. In addition, the references found in relevant identified studies were also examined.

Method of assessment of studies

The studies were identified by one of the authors (PW). From the potential relevant publications, a total of eight were included for this study by computer search. Furthermore, three additional studies were found by reference to the already identified studies. The selected publications were blinded for author(s), source of publication, results and conclusions in order to minimize potential reviewer bias. RHMH and WM independently scored each publication according to a standardized set of 20 methodological criteria (Table 1).

The authors developed a criteria list according to the guidelines for meta-analysis evaluating diagnostic tests by Irwig et al. (1994) and the method guidelines for systematic reviews by Van Tulder et al. (1997). Items that seemed to be irrelevant for reliability studies were dropped and more appropriate items were added.

They consisted of three categories: (1) study population, (2) test procedure and (3) test results. Each criterion was given a weighting. The maximum score was set at 100 points for each study. In a subsequent meeting, the two reviewers tried to reach a consensus on each criterion where they had disagreed. Where disagreement persisted, a third reviewer (PW) made the decision. The assessment results are listed hierarchically in which a higher score indicates studies with a better methodology (Table 2).

RESULTS

Eleven articles met the inclusion criteria (Wiles 1980; Potter & Rothstein 1985; Carmichael 1987; Herzog et al. 1989; McCombe et al. 1989; Deursen et al. 1990; Laslett & Williams 1994; Dreyfuss et al. 1996; Van der Wurff et al. 1996; Strender et al. 1997; Meyne et al. 1999).

These studies are presented in Table 2 in a hierarchical order according to their methodological score. Initially, the two reviewers did not agree on the criterion being judged in 40 out of 220 times (18%). In the majority of the cases it appeared to be an error in reading. The disagreement between the two reviewers mostly concerned the description of the study population, description inclusion and exclusion criteria and the selection of examiners. After the consensus meeting between the two reviewers, the number of disagreements was completely reduced and it was not necessary for the third reviewer to make a final decision.

The inter-examiner reliability between the two reviewers was $\kappa = 0.63$, which is 'substantial' agreement according to the classification of Landis and Koch (1977). A comparison of the results by the reviewers with similar studies shows relatively equal results (Koes et al. 1996).

In this review, only two studies (Wiles 1980; Herzog et al. 1989) had a methodological score of less than 50 points and there were nine with more than 50 points. In general this indicates an acceptable quality in the majority of the studies.

Table 2 shows that the most prevalent methodological problems concerned:

- the description of the inclusion and exclusion criteria (A2)
- description of drop-outs (B)
- description of simple exercises before and/ ordering of the test procedure (D4)

Table 2 Overall results of pain provocation tests and mobility tests for the SIJ in order of method score

First author	Year	A1 8	A2 7	B 5	C 10	D1 3	D2 2	D3 3	D4 2
Van der Wurff	1996	8	7	5	6	3	2	3	0
Meyne	1999	8	7	5	3	3	2	3	2
Carmichael	1987	8	0	0	6	3	2	3	2
Strender	1997	8	0	0	6	3	2	3	0
Potter	1985	8	7	0	0	3	2	3	0
Laslett	1994	0	7	0	6	3	2	3	0
McCombe	1989	8	7	0	10	3	2	3	0
Dreyfuss	1996	8	0	5	10	3	0	0	0
Deursen	1990	8	0	0	10	3	2	3	0
Herzog	1989	0	0	0	3	3	0	3	2
Wiles	1980	8	0	0	0	0	0	3	0

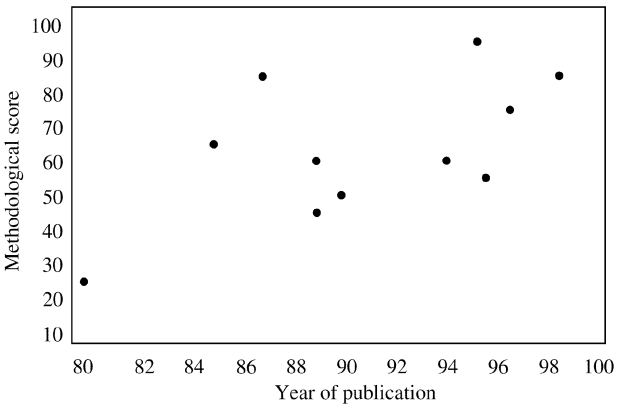

Fig. 1—Relationship between methodological score of trials and year of publication.

- information available to the subjects (D5)
- test/re-test procedure (G)
- results sealed (H3).

Figure 1 plots the methodological score against the year of publication to assess whether the quality of the published studies has increased during the past two decades. It appears that the quality of the reliability studies on the SIJ included in this review have improved since 1980; $r = 0.687$ ($P < 0.05$). The highest methodological scores (more than 60 points) were attained during the last 6 years. All studies examined inter-examiner reliability and four studies additionally examined intra-examiner reliability (Carmichael 1987; Herzog et al. 1989; Van der Wurff et al. 1996; Meyne et al. 1999).

The majority (6) of the studies reported negative results for reliability of SIJ tests for joint mobility and pain provocation tests. In two studies (Wiles 1980; Herzog et al. 1989), reliable results are described. Unfortunately, the methodological scores for these studies were the lowest in this review. Three other studies (McComb et al. 1989; Laslett & Williams 1994; Dreyfuss et al. 1996) report partly reliable results (5 out of 7; 2 out of 4 and 3 out of 5 respectively).

There was a tendency for 'positive' conclusions about the reliability of the reviewed studies to be inversely proportional to the methodological score.

Pain provocation tests

Table 2 shows the overall results of the pain provocation tests and mobility tests for the SIJ in order of methodological score. The majority of the studies included a detailed description of the study population (item A1); only two failed (Herzog et al. 1989; Laslett & Williams 1994). The drop-outs were described in only two studies (Van der Wurff et al. 1996; Meyne et al. 1999). This item was not mentioned in the other studies. With respect to the interpretation of outcome measures, the study by Laslett and Williams (1994) was the only study to include unclear information. For item J, the use of Cohen's κ or the interclass correlation coefficient (ICC), only three studies showed shortcomings (Wiles 1980; Potter & Rothstein 1985; Herzog et al. 1989). The ICC was not used in any of the reviewed studies.

Table 3 reflects the reliability of the pain provocation tests for the SIJ. Only Laslett and Williams (1994) and Dreyfuss et al. (1996) report acceptable reliability for the Gaenslen test and the Thigh thrust test. These positive conclusions about reliability (Laslett & Williams 1994; Dreyfuss et al. 1996) cannot be compared with other authors' work because there is no data available.

The Gapping or distraction test and Compression test were the most frequent tests studied, respectively three and four times. Laslett and Williams (1994) also showed positive reliability for the Gapping or distraction test while Potter and Rothstein (1985) and McCombe et al. (1989) found no acceptable reliability. The methodological score of Laslett and Williams (1994) was similar to that of Potter and Rothstein (1985) and McCombe et al. (1989) as far as the Gapping or distraction test was concerned.

Laslett and Williams (1994) concluded that the Compression test is reliable with substantial Kappa-scores, while Potter and Rothstein (1985), McCombe et al. (1989) and Strender et al. (1997) found no reliability.

D5	D6	E1	E2	E3	F	G	H-1	H-2	H-3	I	J	Method
2	4	3	2	9	5	2	2	1	5	10	15	score
0	4	3	2	9	5	2	2	1	5	10	15	92
2	4	0	2	9	5	2	2	1	5	10	15	90
0	4	3	2	9	5	0	2	1	5	10	15	80
2	0	3	2	9	5	0	2	1	5	10	15	76
0	4	3	2	9	5	0	2	0	5	10	0	63
0	0	0	0	9	5	0	2	1	0	10	15	63
0	4	0	0	0	5	0	0	0	0	0	15	57
0	0	0	0	0	5	0	0	0	0	10	15	56
0	4	0	2	0	5	0	0	0	0	0	15	52
0	0	3	2	9	5	2	2	1	0	10	0	45
0	0	0	2	0	5	0	0	0	0	10	0	28

Table 3 Overview of the reliability by author for the individual pain provocation tests

Test	First author	Agreement %	Max κ	Method score	Authors' conclusion
Gapping or distraction test	Potter	94	κ = na	63	Unreliable
	Laslett	91	κ = 0.69	63	Reliable
	McCombe	na	κ = 0.36	57	Unreliable
Compression test	Strender	79	κ = 0.26	76	Unreliable
	Potter	76	κ = na	63	Unreliable
	Laslett	91	κ = 0.77	63	Reliable
	McCombe	na	κ = 0.16	57	Unreliable
Gaenslen test	Laslett	86*	κ = 0.72*	63	Reliable
	Dreyfuss	82	κ = 0.61	56	Reliable
Sacral thrust	Laslett	73	κ = 0.32	63	Unreliable
	Dreyfuss	66	κ = 0.30	56	Unreliable
Thigh thrust	Laslett	91	κ = 0.82	63	Reliable
	Dreyfuss	82	κ = 0.64	56	Reliable
Cranial shear test	Laslett	73	κ = 0.31	63	Unreliable
Patrick's sign	Strender	96	κ = na	76	Unreliable
	Dreyfuss	85	κ = 0.62	56	Reliable
	Deursen	na	κ = 0.38	52	Unreliable
Flexion-adduction hip	Deursen	na	κ = 0.13	52	Unreliable

* = Average score; na = not available.

Table 4 Overview of the reliability by author for the individual mobility tests

Test	First author	Agreement %	Max κ	Method score	Authors' conclusion
Overtake phenomenon	Wurff	74	κ = 0.29	92	Unreliable
	Potter	44	κ = na	63	Unreliable
	Deursen	na	κ = 0.11	52	Unreliable
Spine test	Wurff	68	κ = 0.19	92	Unreliable
	McCombe	na	κ = 0.42	57	Reliable
	Deursen	na	κ = 0.10	52	Unreliable
Lateroflexion test	Wurff	67	κ = 0.12	92	Unreliable
	Deursen	na	κ = 0.13	52	Unreliable
Gillet test	Meyne	80	κ = 0.08	90	Unreliable
	Carmichael	85	κ = 0.02	80	Unreliable
	Potter	47	κ = na	63	Unreliable
	Dreyfuss	54	κ = 0.22	56	Unreliable
	Herzog	79	κ = na	45	Reliable
	Wiles	64	κ = na	28	Reliable
Sitting flexion test	Potter	50	κ = na	63	Unreliable
Long sitting test	Potter	40	κ = na	63	Unreliable
Translation SIJ	Deursen	na	κ = 0.14	52	Unreliable
Prone knee flexion test	Potter	24	κ = na	63	Unreliable
Maitland test	McCombe	na	κ = 0.38	57	Unreliable

na = not available.

For Patrick's sign, again ambiguous conclusions have been drawn. Dreyfuss et al. (1996) presented acceptable reliability while Deursen et al. (1990) and Strender et al. (1997) found no acceptable reliability. The reliability of the Flexion-adduction test, Cranial shear test and Sacral thrust test has also proved negative.

Mobility tests

Table 4 shows the reliability of the mobility tests for the SIJ. There are only three reliable outcomes stated by McCombe et al. (1989, Spine test) and for the Gillet-Liekens test by Wiles (1980) and Herzog et al. (1989). The Spine test has been studied three times: McCombe et al. (1989) found a positive reliability, Van Deursen et al. (1990) and Van der Wurff et al. (1996) found no reliability. The Gillet-Liekens test was the most frequently studied test (six occasions). The two authors who drew positive reliability (Wiles 1980; Herzog et al. 1989) reached a low method score compared with the other four authors (Potter & Rothstein 1985; Carmichael 1987; Dreyfuss et al. 1996; Meyne et al. 1999).

DISCUSSION

This article contains a review of 11 studies in which the authors assessed reliability of the SIJ for mobility tests and pain provocation tests. It is still possible that the authors might have missed unpublished studies where results might differ from those in this review (publication bias) despite our effort to obtain all available studies.

The items developed for the criteria list and the weightings point system used were arbitrary. The main reason for using a weighting is that some criteria seem to be more important than others. Van der Heijden et al. (1995) showed that using a weighting system was not as important as giving each item equal weighting and then ranking them.

In this article the authors have tried to evaluate the methodological quality of reliability on sacroiliac tests for mobility and pain provocation. The majority of the studies showed acceptable scores; nine out of 11 scored over 50 points. The authors' overall negative conclusions about reliability cannot be attributed to an inappropriate design of the reviewed studies, even though most authors stated that a lack of precision in the design of their trial and imprecise definitions of positive results might influence the results (Oldreive 1995). This present methodological review demonstrates the reverse position.

For the purposes of this study, the authors reviewed trials which dealt with both inter- and intra-examiner reliability. It has been shown that the results of intra-examiner reliability are better than those of inter-examiner reliability. Haas (1991b) stated that a greater incidence of intra-examiner reliability might be the result of influence of conscious and unconscious remembrance and by a systematic error of the examiner.

The results of the studies for mobility tests and pain provocation tests were disappointing. Laslett and Williams (1994) claim reliability for the Gapping or distraction test and the Compression test while other authors do not confirm these positive test conclusions.

The Gaenslen test and Thigh thrust test are concluded as being reliable in the studies of Laslett and Williams (1994) and Dreyfuss et al. (1996) but are not confirmed by other investigators. The level of the methodological score of these studies is moderate and therefore it is not appropriate to accept them a priori as reliable tests.

The review of the Laslett and Williams (1994) study raises some issues. In their study, one permanent examiner was compared with a pool of five other examiners who assessed 51 patients. For part of this trial there was a selection of a series of 22 patients who were examined by a pair of two permanent examiners. This part of Laslett and Williams' (1994) analysis was used in our review. The results of this section of their work contained less positive conclusions about reliability than the overall conclusions suggested.

Pescioli and Kool (1997) later discussed the study by Laslett and Williams (1994). Pescioli and Kool (1997) suggested a multi-test score (MTS) as a possible solution to solve the problem of inappropriate evaluation of the SIJ using four out of seven tests investigated by Laslett and Williams (1994). In the authors' opinion, this is questionable for various reasons. Firstly, the fact that Laslett and Williams (1994) used the Gaenslen test for both SIJs of each patient whether they suffered from problems in one or both SIJs. This means that for patients with complaints in only one single SIJ, a Gaenslen test at this joint would be satisfactory. Secondly, the conclusions stated by Laslett and Williams (1994) are not convincing for the tests where they used applied force. The use of applied force during the cranial shear and sacral shear test can influence the final conclusion. In a recent study, Leven et al. (1998) concluded that the applied force during three pain provocation tests (distraction, compression and sacral thrust test) showed acceptable intra-examiner reliability but insufficient inter-examiner reliability.

The results of the mobility tests were even more disappointing. The Gillet-Liekens test is recommended and accepted as a reliable test on opinion-based arguments by some authors (Bernard 1994; Lee 1998). An important point of criticism about the studies of Wiles (1980) and Herzog et al. (1989) concerns the statistics they used, respectively χ^2 and Pearson's r. According to Haas (1991b), no conclusion concerning reliability can be supported by χ^2 and Pearson's r for categorical data because they are used to test examiners' performance for significant differences from chance alone.

McCombe et al. (1989) reported acceptable reliability for the Spine test, although the κ-value was 'moderate' by the classification of Landis and Koch (1977) adopting a 'cut-off point' of $\kappa = 0.40$. The results of our review do not deviate from the final conclusions stated by Oldreive (1995) and Pescioli and Kool (1997) as far as the mobility tests are concerned. Based on the present review, the authors disagree with the general conclusion of Pescioli and Kool (1997) where they state that the pain provocation tests are a reliable and a valid method. It seems more likely that MTS, as stated by Haas (1991a), particularly for the pain provocation tests of the SIJ, are more appropriate. Therefore, it is necessary to develop a series of tests which prove to be reliable. This should be followed by validity studies to investigate sensitivity and specificity.

CONCLUSION

This review has identified that the methodological quality of reliability studies for SIJ mobility tests as

well as pain provocation tests was sufficient to confirm the more or less negative conclusions drawn by the original authors. The overall impression is that these tests are not reliable and it is difficult to presume that 'methodological upgrading' of these tests would have improved the results. An exception is made for the Gaenslen test and the Thigh thrust test, which do seem to be reliable.

References

Alderink GJ 1991 The sacroiliac joint: review of anatomy, mechanics, and function. Journal of Orthopaedic and Sports Physical Therapy 2: 71–85

Beal MC 1982 The sacroiliac problem: review of anatomy, mechanics and diagnosis. Journal American Osteopathic Association 81: 73–85

Bernard TN, Cassidy JD 1991 The sacroiliac joint syndrome: pathophysiology, diagnosis, and management. In: Frymoyer JW (ed) The Adult Spine: Principles and Practice. Raven Press, New York, pp 2107–2130

Carmichael JP 1987 Inter- and intra-examiner reliability of palpation for sacroiliac joint dysfunction. Journal Manipulative Physiological Therapeutics 10(4): 164–171

Deursen van LLJM, Patijn J, Ockhuysen AL et al. 1990 The value of some clinical tests of the sacroiliac joint. Journal Manual Medicine 5: 96–99

Dreyfuss P, Dreyer S, Griffin J, Hoffman J, Walsh N 1994 Positive sacroiliac screening tests in aymptomatic adults. Spine 19: 1138–1143

Dreyfuss P, Michaelsen M, Pauza K, McLarty J, Bogduk N 1996 The value of medical history and physical examination in diagnosing sacroiliac joint pain. Spine 21(22): 2594–2602

Haas M 1991a The inter-examiner reliability for multiple diagnostic test regimens. Journal of Manipulative Physiological Therapeutics 14(2): 95–103

Haas M 1991b The reliability of reliability. Journal of Manipulative Physiological Therapeutics 14(3): 199–208

Herzog W, Read LJ, Conway PJW et al. 1989 Reliability of motion palpation procedures to detect sacroiliac joint fixations. Journal Manipulative Physiological Therapeutics 12(2): 86–92

Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F 1994 Guidelines for meta-analysis evaluating diagnostic tests. Annals International Medicine 120(8): 667–676

Kirkaldy-Willis WH, Hill RJ 1979 A more precise diagnosis for low back pain. Spine 4(2): 102–109

Koes BW, Assendelft JJ, van der Heijden GJMG, Bouter LM 1996 Spinal manipulation for low back pain. An updated systematic review of randomized clinical trials. Spine 21(24): 2860–2873

Landis RJ, Koch GG 1977 The measurement of observer agreement for catagoral data. Biometrics 33: 159–174

Laslett M, Williams M 1994 The reliability of selected pain provocation tests for sacroiliac joint pathology. Spine 19(11): 1243–1249

Lee DG 1998 The Pelvic Girdle, 2nd edn. Churchill Livingstone, Edinburgh

Levin U, Nillson-Wikmar L, Stenström CH 1998 Reproducibility of manual pressure force on provocation of the sacroiliac joint. Physiotherapy Research International 3(1): 1–14

Maigne J-Y, Aivaliklis A, Pfefer F 1996 Results of sacroiliac double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. Spine 21(16): 1889–1892

McCombe PF, Fairbank JCT, Cockersole BC, Pynsent PB 1989 Reproducibility of physical signs in low-back pain. Spine 14(9): 908–917

Meyne W, Neerbos K, Aufdemkampe G, van der Wurff P 1999 Intra- and inter-examiner reliability of the Gillet test. Journal Manipulative Physiological Therapeutics 1: 4–10

Nelson MA, Allen P, Clamp SE, de Dombal FT 1979 Reliability and reproducibility of clinical findings in low back pain. Spine 4(2): 97–101

Oldreive WL 1995 A critical review of the literature on tests of the sacroiliac joint. Journal of Manual and Manipulative Therapy 4: 157–161

Paris SV 1992 Differential diagnosis of sacroiliac joints from lumbar spine dysfunction. In: Vleeming A et al. (eds) Course proceedings for the first interdisciplinary World Congress on low back pain and its relation to the sacroiliac joint. San Diego, California, pp 313–329

Pescioli A, Kool J 1997 Die Zuverlassigkeit Klinischer Iliosakralgelenktests. Manuelle Therapie 1(1): 11–17

Potter NA, Rothstein JM 1985 Intertester reliability for selected clinical tests of the sacroiliac joint. Physical Therapy 65(11): 1671–1675

Schmid HJA 1985 Iliosacrale Diagnose und Behandlung 1978–1982. Manuelle Medizin 23: 101–108

Schwarzer AC, Aprill CN, Bogduk N 1995 The sacroiliac joint in chronic low back pain. Spine 20(1): 31–37

Shaw JL 1992 The role of the sacroiliac joint as a cause of low back pain and dysfunction. In: Vleeming A et al. (eds) Course proceedings for the first interdisciplinary World Congress on low back pain and its relation to the sacroiliac joint. San Diego, California, pp 67–80

Strender L-E, Sjöblom A, Sundell K, Ludwig R, Taube A 1997 Interexaminer reliability in physical examination of patients with low back pain. Spine 22(7): 814–820

Sturesson B, Selvik G, Uden A 1989 Movements of the sacroiliac joints: a rontgen stereophotogrammetric analysis. Spine 14(2): 162–165

van der Heijden GJMG, Beurskens AJHM, Koes BW, Assendelft WJJJ, de Vet HCW de, Bouter LM 1995 Traction for back and neck pain: a blinded review. Physical Therapy 75(2): 93–104

Van der Wurff P, Hagmeyer RHM, Kuhlmann P 1996 Het sacroiliacaal gewricht getest. Nederlands Tijdschrift voor Fysiotherapie 106(6): 178–184

Van Tulder MW, Assendelft WJJ, Koes BW, Bouter LM and the Editorial Board of the Cochrane Collaboration Back Review Group 1997 Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. Spine 22(20): 2323–2330

Vleeming A 1990 The sacroiliac joint. Thesis, Rotterdam

Walker JM 1992 The sacroiliac joint: A critical review. Physical Therapy 72(12): 903–916

Wiles MR 1980 Reproducibility and inter-examiner correlation of motion palpation findings of the sacroiliac joints. The Journal of the Canadian Chiropractic Association 24(2): 59–69