This is the peer reviewed author accepted manuscript (post print) version of a published work that appeared in final form in:

Mine, Koya, Ono, Kenta, Tanpo, Nobuhito 2021 'The reliability of palpatory examinations for pelvic landmarks to determine pelvic asymmetry: a systematic review and meta-analysis', Physical Therapy Reviews, online, pp. 1-10

This un-copyedited output may not exactly replicate the final published authoritative version for which the publisher owns copyright. It is not the copy of record. This output may be used for non-commercial purposes.

The final definitive published version (version of record) is available at:
https://doi.org/10.1080/10833196.2021.2019368

Persistent link to the UniSA Research Outputs Repository:
https://researchoutputs.unisa.edu.au/11541.2/27996

General Rights

Copyright and moral rights for the publications made accessible in the UniSA Research Outputs Repository are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the persistent link identifying the publication in the UniSA Research Outputs Repository

If you believe that this document breaches copyright, please contact us and provide details. We will remove access to the work immediately and investigate your claim.

The reliability of palpatory examinations for pelvic landmarks to determine pelvic asymmetry: a systematic review and meta-analysis.

Koya Mine[a], Kenta Ono[b], Nobuhito Tanpo[c]

[a]Allied Health & Human Performance, University of South Australia, Adelaide, South Australia, Australia

[b]Department of Rehabilitation, Kuwano Kyoritsu Hospital, Fukushima, Japan

[c]Department of Rehabilitation, Takeda General Hospital, Fukushima, Japan

Corresponding author: Koya Mine

Email: Koya.Mine@unisa.edu.au

Tel: +61 468377539

ORCID iD: https://orcid.org/0000-0003-4321-0537

Twitter handle: @mine_koya

Address: UniSA City East Campus, Corner of North Terrace and Frome Road, Adelaide, SA, 5000, Australia

Word count: 2,837 words

ABSTRACT

Background: Assessments and treatments focused on pelvic asymmetry are common amongst health professionals. However, there is no clear evidence for the reliability of palpatory examinations for pelvic landmarks.

Objectives*:* This review aimed to investigate the reliability of palpatory examinations for pelvic landmarks.

Methods: A systematic search was undertaken using eight databases from inceptions to 2021 January 8th, to identify relevant primary studies. Methodological quality of included studies was evaluated using quality appraisal tool for studies of diagnostic reliability (QAREL). Data on kappa statistics were synthesized quantitatively.

Results: 10 papers were included. Six papers had moderate methodological quality (50-70% in QAREL). Four pelvic landmarks, including anterior superior iliac spine, posterior superior iliac spine, sacral sulcus or inferior lateral angle were examined. For both inter-rater and intra-rater reliability, kappa statistics were consistently lower than 0.60 regardless of examiners' qualifications or experience levels, or standardisation trainings. As a result of meta-analyses, pooled kappa statics were consistently lower than 0.06 for all landmarks.

Conclusions: Consistent evidence was found to suggest that palpatory examinations of pelvic landmarks to detect pelvic asymmetry do not have an acceptable reliability. Therefore, the current evidence does not support the use of these tests in clinical or educational settings.

(199 words)

Keywords

Pelvic landmarks; Pelvic asymmetry; Palpatory examinations; Reliability; Systematic review

Introduction

When one of the most successful golfers in recent decades suffered from low back pain in 2014, he was reported to have made the following comment in the press (1). 'My sacrum went out, it pinched the nerve and hence the spasms. Once the bone was put back, it was all good.' This comment quickly caught a great deal of attention and led to discussions and controversy among health professionals (2). This social phenomenon reflects persistent fascination and uncertainty regarding pelvic alignments and its relationship with pain and physical functions, among health professionals and the public.

Traditionally, palpatory examinations for pelvic bony landmarks to determine pelvic asymmetry have been often taught and used by various health professions, such as osteopaths, chiropractors and physiotherapists (3-6). Pelvic asymmetry has been described in a variety of terms, such as anterior rotation of ilium or sacral torsion (7,8). The bony landmarks to be palpated typically include anterior superior iliac spine (ASIS), posterior superior iliac spine (PSIS), inferior lateral angle (ILA) and sacral sulcus (SS). Some health professionals believe that the existence of pelvic asymmetry can contribute to pain disorders and physical impairments, and thus should be corrected by the use of specific manual techniques, such as joint manipulation or muscle energy technique (7,9). To the best of the authors' knowledge, however, the validity of the concept of 'pelvic malalignment' as a contributing factor to pain disorders has not been established.

One review which systematically investigated the reliability of palpatory examinations for pelvic landmarks (10). This study was published in 2010 and concluded that the examinations were not proven to be reliable. However, this review did not consider risks of biases of the included studies. A systematic review is supposed to provide evidence regarding a clearly formulated question, by critically appraising and analysing included studies (11). A critical

appraisal enables authors to evaluate methodological quality in individual studies and perform accurate assessments for strength of overall evidence. A more recent systematic review published in 2016 concluded that the current palpating methods for PSIS asymmetry do not have acceptable levels of inter-rater reliability (12). Although this conclusion was based on sound methodology with critical appraisal, reliability of palpatory examination for other pelvic landmarks was not available. In addition, those two reviews did not perform meta-analysis due to a technical difficulty or no intention to perform meta-analysis (10,12). The lack of quantitative analysis might have biased the findings, failing to address potentially low statistical power in individual study and provide a precise estimate of the reliability (13). The paucity of clear evidence regarding the reliability of pelvic alignment examinations can make it difficult for clinicians to monitor patients' conditions accurately. This uncertainty can also mislead clinicians to inappropriate decision making. Therefore, it was considered to be timely and useful to update systematic reviews regarding the reliability of palpatory examinations for pelvic asymmetry, utilising critical appraisal and meta-analysis .

The objective of this study is to systematically review and synthesize findings from all available primary studies which investigated the reliability of palpatory examinations for pelvic asymmetry.

Methods

Study protocol and registration

This systematic review was written in line with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement (11). This review was registered in the international prospective register of systematic reviews PROSPERO network (CRD42017073220) before the commencement of a systematic search.

Inclusion and exclusion criteria

Eligibility criteria are shown in Table 1. This systematic review included any primary studies investigating the intra-rater and/or inter-rater reliability of static palpatory examinations for pelvic landmarks to determine pelvic symmetry or asymmetry. Studies exploring the reliability of dynamic examinations using patients' movements or any examinations using mechanical tools were excluded. Studies using cadavers or skeleton models were also excluded. Anatomical landmarks of interests in this review were ASIS, PSIS, ILA and SS. Participants could be healthy subjects or patients with low back or pelvic girdle pain. Limitations were not set on professional disciplines or training levels of assessors. Only studies using kappa statistics as a scale of reliability were included. Only papers published in peer-reviewed journals in English or Japanese language were considered for eligibility.

Databases and search strategy

A systematic search was performed using six English-language databases (Cochrane Library, Embase, MEDLINE, Ovid Emcare, Scopus and Web of Science). Two Japanese-language databases, including CiNii and Ichushi were also utilized to decrease the effects of language bias (14,15). The reference lists of included studies were also hand searched. One reviewer (KM) conducted searches.

All databases were searched from the earliest available dates to 8th January 2021. Search terms were established to maximize sensitivity and specificity of the systematic search (Table 2). In Japanese-language databases, Japanese translations for these terms were used. After the removal of duplicates, titles and abstracts of all articles identified in the search were screened. Lastly, full texts of potentially relevant papers were assessed for inclusion. This screening process was performed by three authors (KM, KO and NT) independently, using a citation management software EndNote X7 (Clarivate Analytics, USA). All three reviewers were

fluent in both English and Japanese languages. Disagreements in study selections were resolved by consensus-based discussion.

Critical appraisal of included studies

An assessment of methodological quality was performed using the quality appraisal tool for studies of diagnostic reliability (QAREL) (16). This specific appraisal tool has been demonstrated to be reliable to evaluate studies of diagnostic reliability (16). Three physiotherapists (KM, KO and NT) independently reviewed all included articles. After individual assessments, disagreements were mediated through discussions between the three reviewers. In this study, study quality was graded as 'high' (>70% in QAREL), 'moderate' (50-70% in QAREL) or 'low' (<50% in QAREL) to enable a nominal categorisation for convenience.

Data extraction and synthesis

Data were extracted on country of origin, sample characteristics (sample size, age and gender), assessors' characteristics (professions, experience levels, and whether there were training sessions before experiments or not), methods of palpatory examinations, types and numbers of kappa statistics, and conclusions. This procedure was undertaken collaboratively by three authors (KM, KO and NT) in a predesigned manner. When the reporting of data was incomplete, corresponding authors were contacted via e-mails to confirm necessary information.

Meta-analysis was undertaken to calculate summary estimates and 95% confidence intervals (95% CI) for the kappa statistics, given clinical homogeneity among two or more studies and sufficient data sets. Meta-analysis was carried out using generic inverse variance model and random effects models in Review Manager 5.4 (Cochrane Collaboration, UK). Magnitude of

statistical heterogeneity across studies was estimated using I^2 statistics. P values < 0.05 were considered to be statistically significant in heterogeneity tests. Subgroup analyses on the basis of the proficiency levels of assessors or the presence of symptoms among subjects were planned if sufficient numbers of studies were available. In order to minimize the risk of statistical errors, subgroup analysis was expected to be performed only when there are at least five data sets in each subgroup (17). In this systematic review, an acceptable level of reliability was defined as ≥ 0.60 of kappa score (18).

Results

Study selection

A detailed systematic literature search process in the eight electronic databases is presented in Figure 1. The initial search yielded a total of 1,213 papers. Subsequently, 475 duplicate papers were excluded. After screening titles and abstracts of 738 remaining articles, 13 potentially relevant articles were chosen for full-text assessment. Six corresponding authors were contacted by e-mails and the requested information was provided by one author (19). Two papers were excluded because kappa statistics was not available (20,21). Another paper was excluded because it examined only medial malleoli of tibia (22). As a result, 10 English-language articles were included in this systematic review and data sets from six papers were used for meta-analysis.

Study characteristics

A summary of included studies is described in Table 3. Studies came from the UK (n = 4), Australia (n = 3), the US (n = 2) and Sweden (n = 1).

Participants: Participants' profiles were varied across studies. Sample sizes for subjects ranged from 1 to 65. A total sample size across the 10 articles was 188. Out of the total

number of subjects, 119 people had low back pain (23-26), 61 people were asymptomatic (19,27-30), and one study recruited eight skiers without disclosing their demographic information (31).

Assessors: Osteopaths and/or osteopathic students were recruited in five studies (19,24,27,29,30), physiotherapists in three studies (23,25,31), chiropractors or chiropractic students in two studies (26,28), and a medical doctor in one study (26). Professional experiences in their fields were varied, ranging from none to 15 years. Standardising training sessions were specifically administered before experiments in five studies (19,24,27,28,30).

Examination Methods: Examined pelvic landmarks included the following; ASIS in three studies (24,26,31), PSIS in seven studies (19,23-31), ILA in five studies (19,23,24,27,29), and SS in four studies (23,24,27,29). The location of ASIS was assessed in either craniocaudal dimension (19,24,26,31) or mediolateral dimension (31). ILA was evaluated in either craniocaudal dimension(19,24,27) or anteroposterior dimension (19,23,24,29).

Methodological quality

The critical appraisal scores are presented in Table 4. The methodological quality scores were varied (33.3% to 90.0%). Six studies had moderate quality (19,23,26-29), while the other four studies had either low or high quality scores (refer to Table 3) (24,25,30,31). Common methodological flaws were; no consideration to stability of the tests (100%), no blinding of raters to additional cues in the testing procedure (60%), no blinding of raters to participants' clinical information (50%), and failure to recruit representative participants (50%). The three reviewers (KM, KO and NT) showed almost perfect inter-rater reliability in determining methodological quality of included papers (overall kappa of 0.91).

Reliability of palpatory examinations for ASIS

ASIS was examined in a supine position in two high-quality studies (24,31) and one moderate-quality study (see Table 3 and 4) (26). The inter-rater reliability was consistently lower than 0.60, regardless of professions, standardisation trainings or raters' experience levels. Contradictory findings were seen for intra-rater reliability. One high-quality study showed that intra-rater reliability for ASIS was 0.40 (24), whereas another reported a reliability of 0.65 (19). Both groups in the two studies had standardisation training sessions. There was an overall tendency for intra-rater reliability to be higher than inter-rater reliability. As a result of meta-analyses, pooled kappa scores for the intra-/inter-rater reliability of ASIS tests were 0.42 (95% CI 0.05, 0.79) and 0.12 (95% CI -0.10, 0.33) respectively (Figure 2). Statistical heterogeneity was significant for both intra- and inter-rater reliability.

Reliability of palpatory examinations for PSIS

PSIS was investigated in two high-quality studies (24,31), three moderate-quality studies (19,27,28), and two low-quality studies (see Table 3 and 4) (25,30). All papers consistently reported poor intra-rater and inter-rater reliability, with kappa scores being lower than 0.60, irrespective of professions, standardisation trainings, or examiners' experience levels. Overall, intra-rater reliability tended to be higher than inter-rater reliability across the studies. Meta-analyses showed that pooled kappa scores for the intra-/inter-rater reliability of PSIS tests were 0.38 (95% CI 0.29, 0.48) and 0.09 (95% CI 0.03, 0.15) respectively (Figure 3). Statistical heterogeneity was significant for both intra- and inter-rater reliability.

Reliability of palpatory examinations for ILA

One high-quality study (24) and four moderate-quality studies (19,23,27,29) examined palpatory examinations for ILA (Table 3 and 4). All these papers found that kappa scores of intra-rater and inter-rater reliability were consistently lower than 0.60. Types of professions, the presence of standardisation training sessions, or examiner experience did not appear to

have influences on the reliability. Intra-rater reliability was generally higher than inter-rater reliability across the studies. As a result of meta-analytic synthesis of the intra-/inter-rater reliability for ILA-CC tests, pooled kappa statistics were 0.15 (95% CI 0.05-0.25) and 0.07 (95% CI -0.03-0.17) respectively (Figure 4). For ILA-AP tests, the values were 0.15 (95% CI 0.06, 0.23) and 0.01 (95% CI -0.03, 0.06) respectively (Figure 5). Statistical heterogeneity was significant for intra-rater reliability.

Reliability of palpatory examinations for SS

SS was examined in one high-quality study (24) and three moderate-quality studies (Table 3 and 4) (23,27,29). Kappa scores for intra-rater and inter-rater reliability were constantly lower than 0.60, again regardless of types of professions, the presence of standardisation training sessions, or examiners' experience levels. Overall, intra-rater reliability was higher than inter-rater reliability across studies. Meta-analyses showed that pooled kappa scores for the intra-/inter-rater reliability of SS tests were -0.13 (95% CI -0.59, 0.32) and 0.24 (95% CI 0.14, 0.35) respectively (Figure 6). Statistical heterogeneity was significant for intra-rater reliability ($p = 0.008$).

Discussion

The purpose of this systematic review was to investigate the reliability of palpatory examinations for pelvic bony landmarks. Methodological quality was moderate in the majority of the included studies, which implied moderate overall study quality. Examined pelvic landmarks included the ASIS, PSIS, ILA and SS. The intra- and inter-rater reliability was consistently lower than the acceptable level (kappa = 0.60) for all the landmarks across the included studies. Overall, intra-rater reliability tended to be higher than inter-rater reliability. This phenomenon might be explained by systematic errors within individual raters, leading to internal consistencies. However, the lack of gold standard test makes it difficult for

us to validate this hypothesis. Meta-analyses consistently showed that pooled kappa statistics were below 0.60 for the landmarks (ASIS, PSIS, ILA and SS). In summary, findings from the current evidence consistently negate the acceptable level of reliability for palpatory examinations to identify pelvic asymmetry.

The fact that it is extremely difficult for clinicians to agree upon pelvic asymmetry might be related to minimal movements that occur at the sacroiliac joint (SIJ). One study using roentgen stereophotogrammetry identified very little movements (0.7-mm translation and 2.5-degree rotation on average) at the SIJ, which may results in minimum actual asymmetry in pelvis (32). Considering this finding, it seems to be understandable that the reliability to detect pelvic asymmetry is lower than the adequate level for clinical utility. Furthermore, one cadaveric study found considerable amount of variations in pelvic morphology within and between individuals (33). This finding challenges the content validity of palpatory examinations for pelvic asymmetry, since what raters regard as asymmetry might derive from normal morphological side-to-side differences in individuals.

Although the notion of pelvic malalignment as a contributing factors to pain disorders or physical impairments is prevalent amongst health professionals (7,34), this view has not been validated by research evidence, and some authors challenge this belief system (2,35). In fact, one cross-sectional study reported that there was no significant difference in mobility between symptomatic and asymptomatic sacroiliac joints (32). Another roentgen stereophotogrammetric study investigated whether sacroiliac joint manipulation could change the relative position between the ilium and the sacrum (9). Although subjective findings in positional tests changed from positive to negative after the manipulation, the objective roentgen measurements did not find any changes. This research implies that examiners were biased, assuming that manipulation could alter pelvic alignment. This mode of thinking might

be confounding and reinforce clinicians' beliefs regarding the validity of palpatory tests for determining the relationship between pelvic alignment and symptoms (36,37).

Furthermore, there may be potentially negative influences of using assessment techniques to determine pelvic alignment, and how these findings are conveyed to patients. When verbal messages like "upslip", "torsion" or "out of place" are conveyed in clinical settings, they can increase patients' perceived vulnerability, compromising their long-term recoveries and fuelling the chronicity of pain (38,39). Therefore, it is questionable whether clinicians should continue to advocate these assessments and treatments which focus on pelvic asymmetry.

Several limitations have to be acknowledged in this review. Although we attempted to reduce language bias and publication bias by utilising Japanese-language electronic databases, no relevant Japanese-language article was identified. This may indicate that palpatory examinations for pelvic landmarks are not used in Japan as often as in western countries (the UK, Australia, the US and Sweden), where identified studies were conducted. It should be noted that language and publication bias still exists in this review since we did not consider papers written in other languages. Future review should explore relevant studies written in other languages. Although conducting meta-analyses was a strength and a novel element of this review compared to previous reviews (10,12), there were only two to five studies in each data sets. The small numbers of studies might have led to high statistical heterogeneity and less accurate summary estimates of the reliability. Further studies with higher methodological quality are required to contribute to more precise evidence.

Conclusions

This systematic review found consistent evidence suggesting that palpatory examinations for pelvic landmarks do not have an acceptable level of intra-rater and inter-rater reliability for clinical utility. Considering its low reliability and potentially negative influences on patients'

perceptions of pain, clinicians are not recommended to use these assessment techniques. In addition, health professionals involved in educational fields should confront the current evidence and reconsider the value of teaching these physical examinations to their students. Since the small numbers of studies might have led to less accurate summary estimates of the reliability, further studies with higher methodological quality are required to contribute to stronger evidence.

Conflict of interest

The authors declare no conflict of interests.

Funding

The authors received no financial support for the research.

References

1. Khan K. 'Sacrum went out'…what went wrong? Tiger Woods, media doctors, and collective responsibility: BMJ Blog; 2014 [cited 2015 November 24]. Available from: http://blogs.bmj.com/bjsm/2014/08/08/sacrum-went-out-what-went-wrong-tiger-woods-media-doctors-and-collective-responsiblity/
2. O'Sullivan P. Common misconceptions about back pain in sport: Tiger Woods' case brings five fundamental questions into sharp focus. British journal of sports medicine. 2015:bjsports-2014-094542.
3. Kilby J, Heneghan NR, Maybury M. Manual palpation of lumbo-pelvic landmarks: a validity study. Manual therapy. 2012;17(3):259-262.
4. Fryer G, Johnson JC, Fossum C. The use of spinal and sacroiliac joint procedures within the British osteopathic profession. Part 1: Assessment. International Journal of Osteopathic Medicine. 2010;13(4):143-151.
5. Byfield D, Mathiasen J, Sangren C. The reliability of osseous landmark palpation in the lumbar spine and pelvis. European Journal of Chiropractic. 1992;40:83-88.
6. Barnes MF. Efficacy study of the effect of a myofascial release treatment technique on obtaining pelvic symmetry. Journal of Bodywork and Movement Therapies. 1997;1(5):289-296.
7. Chaitow L, Crenshaw K. Muscle energy techniques. Elsevier Health Sciences; 2006.
8. Fryer G. Muscle energy technique: An evidence-informed approach. International Journal of Osteopathic Medicine. 2011;14(1):3-9.
9. Tullberg T, Blomberg S, Branth B, et al. Manipulation does not alter the position of the sacroiliac joint: a roentgen stereophotogrammetric analysis. Spine. 1998;23(10):1124-1128.
10. Stovall BA, Kumar S. Anatomical landmark asymmetry assessment in the lumbar spine and pelvis: a review of reliability. PM&R. 2010;2(1):48-56.
11. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009;151(4):264-269.

12. Cooperstein R, Hickey M. The reliability of palpating the posterior superior iliac spine: a systematic review. The Journal of the Canadian Chiropractic Association. 2016;60(1):36.
13. Harrison F. Getting started with meta-analysis. Methods in Ecology and Evolution. 2011;2(1):1-10.
14. Egger M, Zellweger-Zähner T, Schneider M, et al. Language bias in randomised controlled trials published in English and German. The Lancet. 1997;350(9074):326-329.
15. Mine K, Nakayama T, Milanese S, et al. Acute effects of stretching on maximal muscle strength and functional performance: A systematic review of Japanese-language randomised controlled trials. Manual therapy. 2016;21:54-62.
16. Lucas N, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). BMC medical research methodology. 2013;13(1):111.
17. Spineli LM, Pandis N. Problems and pitfalls in subgroup analysis and meta-regression. American journal of orthodontics and dentofacial orthopedics. 2020;158(6):901-904.
18. McHugh ML. Interrater reliability: the kappa statistic. Biochemia medica. 2012;22(3):276-282.
19. Fryer G, McPherson HC, O'Keefe P. The effect of training on the inter-examiner and intra-examiner reliability of the seated flexion test and assessment of pelvic anatomical landmarks with palpation. International Journal of Osteopathic Medicine. 2005;8(4):131-138.
20. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. Physical therapy. 1985;65(11):1671-1675.
21. Suwanasri C, Sakullertphasuk W, Tosiriphattana M, et al. Inter-and intra-rater reliability of postural assessment for scoliosis. Journal of the Medical Association of Thailand= Chotmaihet thangphaet. 2014;97:S1-5.
22. Fryer G. Factors affecting the intra-examiner and inter-examiner reliability of palpation for supine medial malleoli asymmetry. International Journal of Osteopathic Medicine. 2006;9(2):58-65.
23. Holmgren U, Waling K. Inter-examiner reliability of four static palpation tests used for assessing pelvic dysfunction. Manual therapy. 2008;13(1):50-56.

24. Kmita A, Lucas NP. Reliability of physical examination to assess asymmetry of anatomical landmarks indicative of pelvic somatic dysfunction in subjects with and without low back pain. International Journal of Osteopathic Medicine. 2008;11(1):16-25.
25. Riddle DL, Freburger JK. Evaluation of the presence of sacroiliac joint region dysfunction using a combination of tests: a multicenter intertester reliability study. Physical Therapy. 2002;82(8):772-781.
26. Tong HC, Heyman OG, Lado DA, et al. Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position. JOURNAL-AMERICAN OSTEOPATHIC ASSOCIATION. 2006;106(8):464.
27. O'Haire C, Gibbons P. Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study. Manual therapy. 2000;5(1):13-20.
28. Paydar D, Thiel H, Gemmell H. Intra-and interexaminer reliability of certain pelvic palpatory procedures and the sitting flexion test for sacroiliac joint mobility and dysfunction. J Neuromusculoskel Sys. 1994;2(2):65-69.
29. Rajendran D, Gallagher D. The assessment of pelvic landmarks using palpation: a reliability study of undergraduate students. International Journal of Osteopathic Medicine. 2011;14(2):57-60.
30. Sutton C, Nono L, Johnston RG, et al. The effects of experience on the inter-reliability of osteopaths to detect changes in posterior superior iliac spine levels using a hidden heel wedge. Journal of bodywork and movement therapies. 2013;17(2):143-150.
31. Lindsay DM, Meeuwisse WH, Mooney ME, et al. Interrater reliability of manual therapy assessment techniques. Physiotherapy Canada. 1995;47(3):173-180.
32. Sturesson B, Selvik G, UdÉn A. Movements of the sacroiliac joints. A roentgen stereophotogrammetric analysis. Spine. 1989;14(2):162-165.

33. Preece SJ, Willan P, Nester CJ, et al. Variation in pelvic morphology may prevent the identification of anterior pelvic tilt. Journal of Manual & Manipulative Therapy. 2008;16(2):113-117.
34. Mitchell FL, Mitchell PKG. The muscle energy manual. Met Press; 1999.
35. Meakins A, Lehman G. Talking About Unstable Pelvises 2019 [cited 2021 14th January 2021]. Available from: https://ptpodcast.com/the-naf-physio-podcast-episode-37-talking-about-unstable-pelvises/
36. Alonso-Perez JL, Lopez-Lopez A, La Touche R, et al. Hypoalgesic effects of three different manual therapy techniques on cervical spine and psychological interaction: A randomized clinical trial. Journal of bodywork and movement therapies. 2017;21(4):798-803.
37. Pickar JG. Neurophysiological effects of spinal manipulation. The Spine Journal. 2002;2(5):357-371.
38. Jones SL, Jones PK, Katz J. Compliance for Low-Back Pain Patients in the Emergency Department: A Randomized Trial. Spine. 1988;13(5):553-556.
39. O'sullivan P. It's time for change with the management of non-specific chronic low back pain. British Association of Sport and Excercise Medicine; 2011.

Table 1. Eligibility criteria

Eligibility category	**Description**
Target population	Healthy subjects or patients with low back or pelvic girdle pain
Index tests	Palpatory examinations for pelvic landmarks (ASIS, PSIS, ILA and SS)
Reference tests	Not applicable
Outcome measures	Kappa statistics for intra-rater and/or inter-rater reliability

Table 2. Search terms

#	**Search terms**
1	(pelvis OR pelvic OR sacroiliac OR (anterior superior iliac crest) OR (posterior superior iliac crest) OR (inferior lateral angle) OR (sacral sulcus)) AND (alignment OR symmetry OR asymmetry OR landmark)
2	test* OR exam* OR assess* OR palpation
3	reliability OR reliable
4	1 AND 2 AND 3

Table 3. Summary of included studies

Study (Country)	Assessors	Subjects	Examined landmarks and decision criteria	Outcome measures	Results
Sutton et al., 2013 (UK)	· 3rd-year OPS (n = 15) · 4th-year OPS (n = 15) · Experienced OP (n = 10) A one-hour standardising training was given.	Healthy female adult (n = 1) Age: 26 years	PSIS in standing position · Left side higher · Equal height · Right side higher	Fleiss' kappa	Inter-rater reliability · 3rd-year OPS: 0.03 · 4th-year OPS: 0.07 · Experienced OP: 0.06 · All combined: 0.06
Rajendran et al., 2011 (UK)	3rd-year OPS (n = 20) No standardising training was given.	Healthy adults (n = 4) Mean age: not specified (described as "third-year students") Gender: not specified	SS and ILA-P in prone · Left side deeper · Equal · Right side deeper · Unclear	Fleiss' kappa	Inter-rater reliability · SS: 0.03 · ILA-AP: -0.05
Holmgren et al., 2008 (Sweden)	Experienced PT (n = 3) No standardising training was given.	Patients with LBP (n = 25) Mean age: 46 years (range 18-78 years) Gender: 18 women and 7 men	SS and ILA-P in prone · Left side deeper · Equal · Right side deeper	Cohen's kappa	Inter-rater reliability · SS: 0.11 · ILA-AP: 0.11
Kmita et al., 2008 (Australia)	Final-year OPS (n = 2) and experienced OP (n = 2) A one-hour standardising training was given.	Patients with LBP (n = 5) and healthy adults (n = 4) Age range: 21-78 Gender: 5 women and 4 men	PSIS, SS, ILA-P and ILA-I in prone, and ASIS in supine · Left side higher/deeper · Equal · Left side higher/deeper	Cohen's kappa	Inter-/Intra-rater reliability · PSIS: 0.04 (-0.38 to 0.35)/0.13 (-0.29 to 0.39) · ILA-AP: -0.03 (-0.18 to 0.51) /0.10 (-0.29 to 0.44) · ILA-CC: -0.01 (-0.13 to 0.36)/0.06 (-0.29 to 0.34) · SS: -0.40 (-0.34 to 0.26)/0.28 (-0.28 to 0.83) · ASIS-CC: 0.13 (-0.13 to 0.50)/0.40 (0.25 to 0.63)

Tong et al., 2006 (US)	MD (n = 1) and CP (n = 2) No standardising training was given.	Patients with LBP (n = 24) Mean age: 68.3 ± 13.2 years Gender: 15 women and 9 men	ASIS in supine · Left side higher · Equal · Right side higher	Fleiss' kappa	Inter-rater reliabil;ity · ASIS-CC: 0.15
Fryer et al., 2005 (Australia)	· Trained OP (n = 5) · Untrained OP (n = 5) Two one-hour standardisation trainings were given for the trained OP group.	Healthy female adults (n = 10) Age range: 22-24 years	PSIS, ILA-P and ILA-I in prone, ASIS in supine · Left side higher/deeper · Equal · Right side higher/deeper	Cohen's kappa	Inter-/Intra-rater reliability in the trained group · PSIS: 0.08/0.54 (0.12 to 0.83) · ILA-AP: 0.04/0.07 (-0.04 to 0.39) · ILA-CC: 0.04/0.20 (-0.10 to 0.64) · ASIS-CC: 0.24/0.65 (0.54 to 0.80) Inter-/Intra-rater reliability in the untrained group · PSIS: 0.15/0.49 (0.22 to 0.76) · ILA-AP: -0.01/0.20 (-0.08 to 0.64) · ILA-CC: -0.01/0.03 (-0.17 to 0.28) · ASIS-CC: -0.01/0.19 (-0.01 to 0.40)
Riddle et al., 2002 (US)	Experienced PT (n = 34) Detailed instructions regarding the examinations were given beforehand and assessors were encouraged to use them in clinical practice.	Patients with LBP (n = 65) Mean age: 47.4 ± 14.0 years (range 18-81 years) Gender: 42 women and 23 men	PSIS in sitting · Left side higher · Equal · Right side higher	Cohen's kappa	Inter-rater reliability · PSIS: 0.37
O'Haire et al., 2000 (Australia)	OPS (n = 10) A one-hour standardising training was given.	Healthy female adults (n = 10) Mean age: 24 years (range 18-30 years)	PSIS, SS and ILA-I in prone · Left side higher · Equal · Right side higher	Fleiss' kappa	Inter-/Intra-rater reliability · PSIS: 0.04/0.33 (0.13 to 0.58) · ILA-CC: 0.08/0.21 (-0.05 to 0.69) · SS: 0.07/0.24 (0.11 to 0.60)
Lindsay et al., 1995 (UK)	Experienced PT (n = 2) No standardising training was given.	Skiers (n = 8) Age, gender and symptom status were unclear.	PSIS and ASIS Positions and decision criteria were not specifically described.	Kappa (type not specified)	Inter-rater reliability · PSIS: -0.10 · ASIS-CC: 0.10 · ASIS-ML: 0.00
Payder et al., 1994	CPS (n = 2)	Healthy adults (n = 32) Mean age: 24.0 ± 7.8 years	PSIS in sitting · Left side higher	Cohen's kappa	Inter-/Intra-rater reliability · PSIS: 0.15/0.25

(UK)	A standardising training was given.	(range 18-50 years) Gender: 15 women and 17 men	· Equal · Right side higher	

OPS = osteopathy students, OP = osteopaths, PT = physiotherapists, MD = medical doctor, CP = chiropractors, CPS = chiropractic students
PSIS = posterosuperior iliac spine, ASIS = anterosuperior iliac spine, SS = sacral salcus, ILA = inferior lateral angle, CC = craniocaudal directions, AP = anteroposterior directions

Table 4. Critical appraisal of included studies

Authors	1	2	3	4	5	6	7	8	9	10	11	Total
Sutton et al., 2013	N	Y	U	N/A	N/A	U	U	N	N	Y	Y	33.3%
Rajendran et al., 2011	N	Y	Y	N/A	N/A	Y	U	Y	N	Y	Y	66.7%
Holmgren et al., 2008	Y	Y	Y	N/A	N/A	U	U	Y	N	Y	Y	66.7%
Kmita et al., 2008	Y	Y	Y	Y	N/A	Y	Y	Y	N	Y	Y	90.0%
Tong et al., 2006	Y	Y	Y	N/A	N/A	Y	U	N	N	Y	Y	66.7%
Fryer et al., 2005	N	Y	Y	Y	N/A	U	Y	Y	N	Y	Y	70.0%
Riddle et al., 2002	Y	Y	U	N/A	N/A	N	U	N	N	Y	Y	44.4%
O'Haire et al., 2000	N	Y	Y	Y	N/A	U	Y	Y	N	Y	Y	70.0%
Lindsay et al., 1995	N	Y	Y	N/A	N/A	Y	Y	Y	N	Y	Y	77.8%
Payder et al., 1994	Y	Y	U	U	N/A	U	U	Y	N	Y	Y	50.0%

QAREL checklist:

1. Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended to the results to be applied?
2. Was the test performed by raters who were representative of those to whom the authors intended the results to be applied?
3. Were raters blinded to the findings of other raters during the study?
4. Were raters blinded to their own prior findings of the test under evaluation?
5. Were raters blinded to the results of the accepted reference standard or disease status for the target disorder (or variable) being evaluated?
6. Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design?
7. Were raters blinded to additional cues that were not the part of the test?
8. Was the order of examination valied?
9. Was the stability (theoretical stability) of the variable being measure taken into account when determining the suitability of the time-interval between repeated measures?
10. Was the test applied correctly and interpreted appropriately?
11. Were appropriate statistical measures of agreement used?

Y = Yes, N = No, U = Unclear, N/A = Not applicable

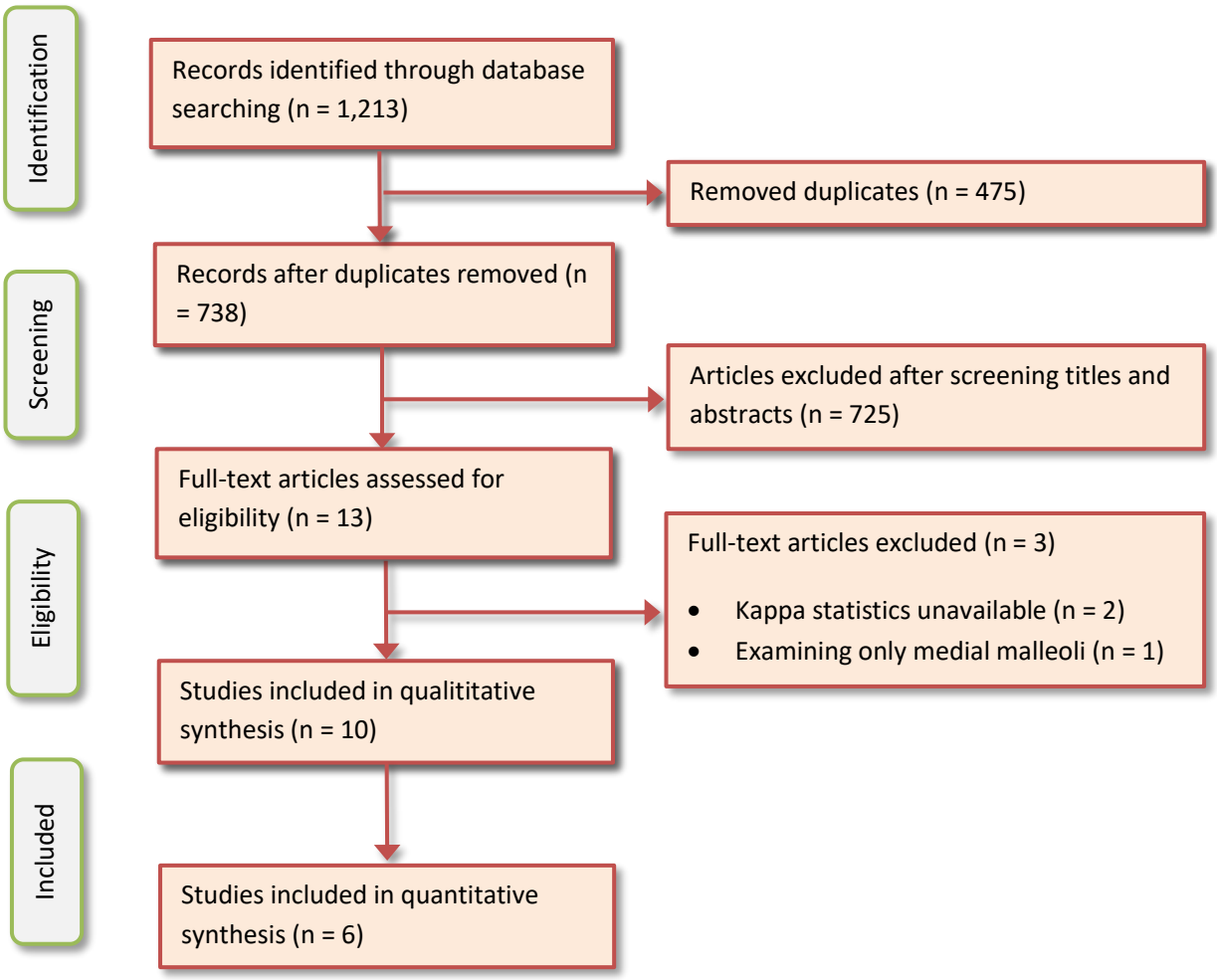


Figure 1. Flowchart of the study selection.

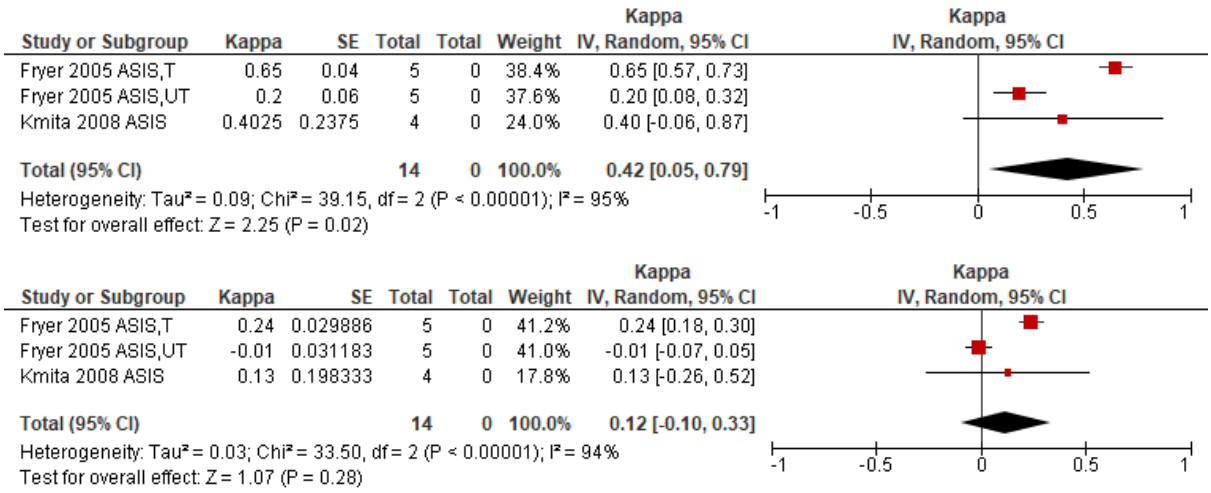


Figure 2. Forest plots for ASIS test reliability (up: intra-rater, down: inter-rater)

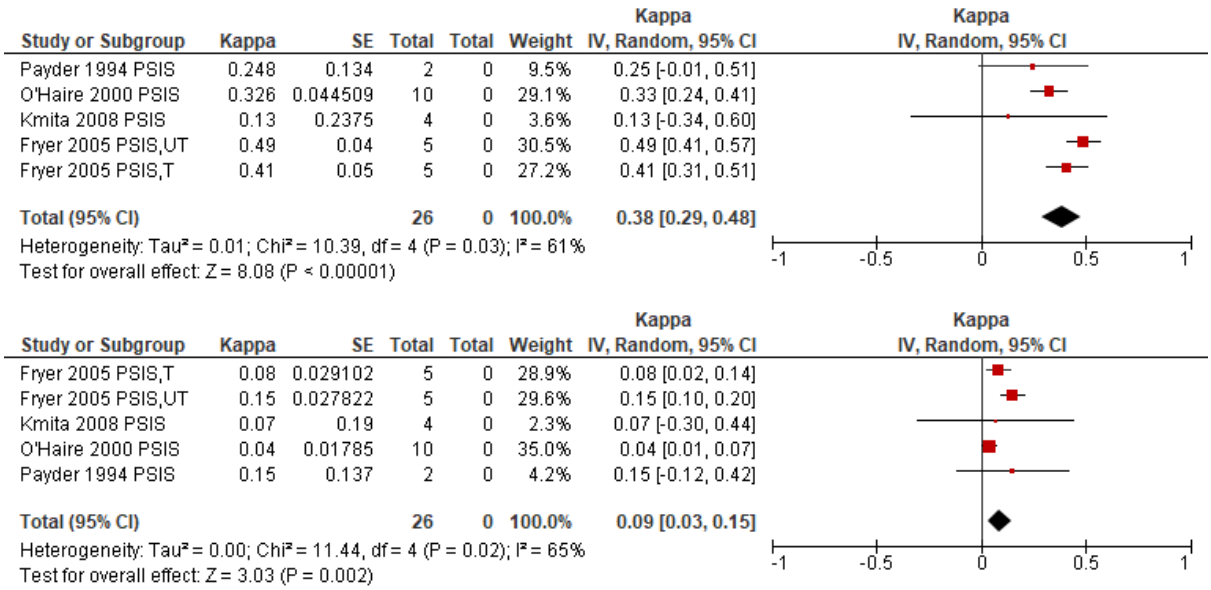


Figure 3. Forest plots for PSIS test reliability (up: intra-rater, down: inter-rater)

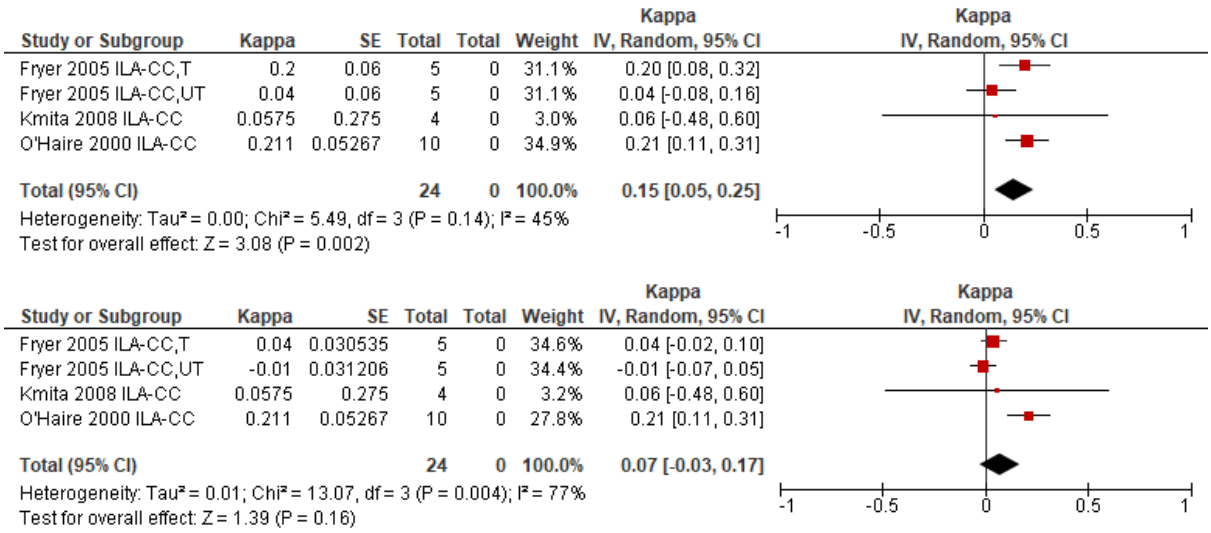


Figure 4. Forest plots for ILA-CC test reliability (up: intra-rater, down: inter-rater)

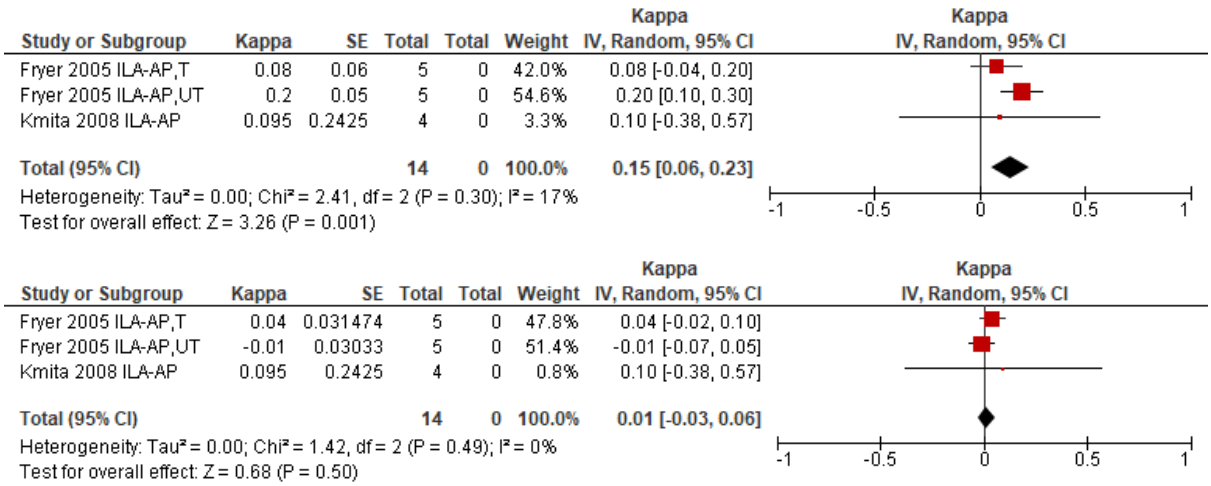


Figure 5. Forest plots for ILA-AP test reliability (up: intra-rater, down: inter-rater)

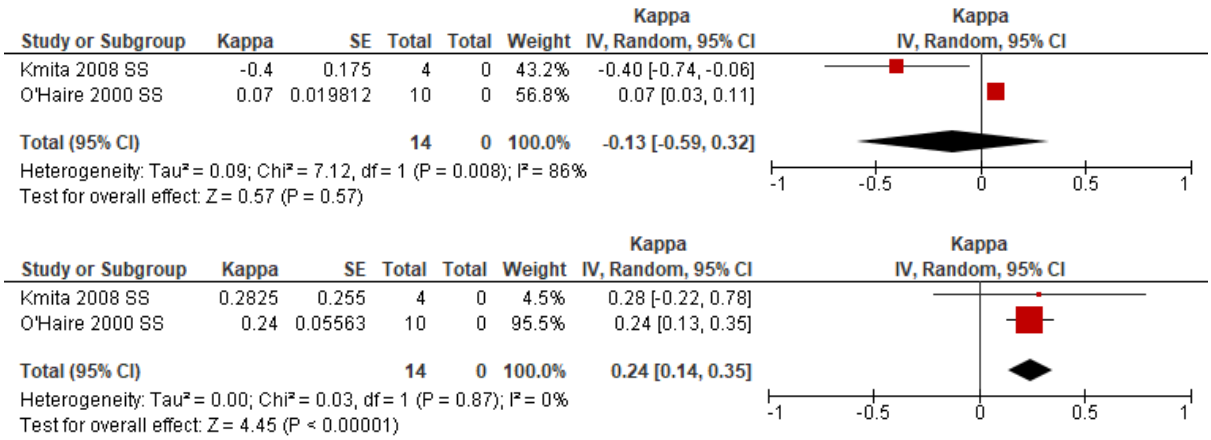


Figure 6. Forest plots for ILA-AP test reliability (up: intra-rater, down: inter-rater)