Journal of Bodywork & Movement Therapies (2013) 17, 143—150

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/jbmt

RELIABILITY STUDY

The effects of experience on the inter-reliability of osteopaths to detect changes in posterior superior iliac spine levels using a hidden heel wedge

Claire Sutton, M.Ost, BSc (Hons), D.O, N.D [a], Lazarus Nono, BSc (Hons) Ost [a], Ross G. Johnston, MSc, BSc (Hons) Ost Med, D.O [b], Oliver P. Thomson, MSc, BSc (Hons) Ost Med, D.O [a,c,*]

[a] *The British College of Osteopathic Medicine, Research Department, London, United Kingdom*
[b] *Swansea University, Department of Inter-professional Studies, Wales, United Kingdom*
[c] *Oxford Brookes University, The Department of Sport and Health Sciences, Oxford, United Kingdom*

Received 27 March 2012; received in revised form 16 July 2012; accepted 17 July 2012

KEYWORDS
Reliability;
Validity;
Diagnosis;
Osteopathy;
Manual therapy;
Palpation;
Sacroiliac

Summary *Background:* The use of palpation to diagnose musculoskeletal dysfunction is commonly taught within osteopathy and other manual therapies. However the clinical tests used to detect sacroiliac joint dysfunction have not shown good reliability.
Objectives: To investigate the inter-examiner reliability of osteopaths to detect asymmetries of the posterior superior iliac spine (PSIS), and to determine if inter-examiner reliability was affected by the level of practitioner experience.
Methods: Fifteen final year osteopathic students ($n = 15$), fifteen third year osteopathic students ($n = 15$) and ten experienced osteopaths ($n = 10$) manually palpated the levels of the PSIS in one model nine consecutive times. A hidden 5 mm heel wedge was used to alter the height of the PSIS which was hidden from the examiners. Scores were analysed using Fleiss Kappa ($F\kappa$) statistics and one way analysis of variance on ranks (ANOVA).
Results: All three groups produced $F\kappa$ results below 0.4 (0.025—0.065), indicating poor inter-examiner reliability. $F\kappa$ values less than 0.4 are considered to be clinically unreliable. ANOVA testing did not show any significant difference between groups.
Conclusion: This study showed 'poor' inter-examiner reliability in detecting asymmetries of the PSIS. This is in accordance with other studies in the field. It is suggested that the inclusion of this osteopathic model within osteopathic education should be reviewed.

* Corresponding author. Oxford Brookes University, The Department of Sport and Health Sciences, Oxford, United Kingdom. Tel.: +44 01865482600.
E-mail address: OThomson@brookes.ac.uk (O.P. Thomson).

1360-8592/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jbmt.2012.07.005

Introduction

Low back pain (LBP) accounts for nearly 50% of presenting complaints made by new patients to osteopaths (GOsC, 2001). The sacroiliac joints (SIJ) are believed to be the 'pain generators' in 20–30% of cases of non-specific LBP (Schwarzer et al., 1995; Bogduk, 2004). Furthermore, the SIJ's have taken a central role in a large number of practice models adopted by the manual therapy professions, and there is an array of manual diagnostic tests and treatment techniques aimed at identifying and restoring normal SIJ movement (Ward and Hruby, 2002; Greenman, 2003; Haldeman, 2004; Lee, 2004). SIJ tests are often divided into pain provocation tests and motion tests, with the former claiming greater clinical value (Laslett et al., 2005). The SIJ examination procedure investigated in this study is considered as the latter and involves the palpation and of bony landmarks, which, once identified, may be assessed for their static or dynamic symmetry. It is thought that a manipulable 'lesion' may be identified by the findings of asymmetric bony landmarks, altered quality and quantity of movement, and altered tissue texture (Kuchera and Kuchera, 1992; DiGiovanna and Schiowitz, 1997; Greenman, 2003). Thus, palpation has been considered a vital component in thediagnosis of sacroiliac joint dysfunction (SIJD).

Over the last three decades, palpation based diagnostic methods have been tested for their reliability to assess and detect SIJ dysfunction (Gonnella et al., 1982; Carmichael, 1987; Burton et al., 1990; Breen, 1992). So far, this research suggests that the vast majority of palpatory tests regularly used by osteopaths and manual therapists to identify SIJD, are not reliable or valid (Hestboek and Leboeuf-Yde, 2000; van der Wurff et al., 2000a; van der Wurff et al., 2000b; Freburger and Riddle, 2001; Tong et al., 2006). Despite this, manual procedures to assess pelvic motion are still commonly used by osteopaths both in the UK (Fryer et al., 2010) and the United States (Fryer et al., 2009). *Validity* refers to the extent to which the test measures what it is intended to; alternatively, how it compares to the gold standard, of which there is none for LBP and pelvic assessment (Lucas and Bogduk, 2011). *Reliability* refers to the reproducibility and consistency of the test outcome between repeated examinations and different examiners (Kmita and Lucas, 2008). For a test to be clinically useful, it must be both reliable and valid (Fleiss et al., 2003; Lucas and Bogduk, 2011).

Research investigating the inter- and intra-examiner reliability of manually assessing pelvic motion, tissue texture and asymmetry of anatomical regions has generally shown low to fair reliability, regardless of the clinical test or technique used (Maher et al., 1998; French et al., 2000; van der Wurff et al., 2000a; Fryer et al., 2005; Hungerford et al., 2007; Moriguchi et al., 2009). *Inter*-examiner reliability refers to the agreement between two or more examiners assessing the same entity, while *intra*-examiner reliability refers to the agreement of an examiner repeatedly assessing the same benchmark (Lucas and Bogduk, 2011).

In the last twenty years there has been a drive for clinical practice to be increasingly informed by research evidence rather than opinion and experience alone (Sackett, 2000) and the osteopathic profession has not escaped the debate (Licciardone, 2007; Fryer, 2008). With an ongoing movement towards evidence-based practice (EBP), osteopaths are required to take a more reflective stance towards their practice and integrate research evidence into their clinical reasoning processes (Thomson et al., 2011). The lack of reproducibility demonstrated around the detection of somatic dysfunction (SD) and low inter-examiner reliability displayed in clinical tests is a continuing challenge for the profession (Fryer, 2003).

The aims of this study were:

- To investigate the inter-examiner reliability of osteopaths to detect asymmetries of a pelvic landmark, the posterior superior iliac spine (PSIS), created with a 5 mm hidden heel wedge;
- To determine if the level of osteopathic experience had any effect on inter-examiner reliability of PSIS asymmetry detection.

Method

Participants

Forty participants were recruited from the British College of Osteopathic Medicine (BCOM) on a volunteer basis after being emailed and approached by the researcher. All participants provided written informed consent prior to taking part in the study. Ethical approval for the study was granted by the British College of Osteopathic Medicine ethics committee.

Model

One participant ($n = 1$) was selected to be used as the pelvis model throughout the experiment. The model was a 26 year old female with no history of LBP, with a normal body mass index (20.2), and was independent of the research. A model with this BMI was selected in attempt to reduce the variable of lumbopelvic soft tissue thickness, which has been shown to affect palpation accuracy (Harlick et al., 2007). Throughout the testing period (one week), the model refrained from any osteopathic or manual therapy treatment and strenuous physical exercise, to minimize any possible changes in the pelvis.

Examiners

Forty examiners ($n = 40$) were divided into three groups depending upon their level of experience.

Group 1 – Student Osteopaths with 3 years training ($n = 15$)
Group 2 – Student Osteopaths with 4 years training ($n = 15$)
Group 3 – Experienced Osteopaths with greater than 5 years in clinical practice (range: 5–25 years; mean 14.5 years) and involved in osteopathic education in either a clinical supervisory role, or teachers of osteopathic manipulative technique ($n = 10$)

Exclusion criteria included previous massage or other manual therapy qualifications prior to training to become an osteopath.

Training

In order to standardize the palpation procedure, all participants took part in a 10 min consensus training session immediately prior to the test procedure, which has been shown to increase reliability (Vincent-Smith and Gibbons, 1999; Degenhardt et al., 2005; Fryer et al., 2005). Participants read detailed descriptions of the test procedure, with pictures demonstrating how they were expected to assess the PSIS levels on the model (Fig. 1). All participants were shown the test procedure technique demonstrated on a model skeleton and given opportunity to practice prior to commencing the test on the real model.

Assessment of posterior superior iliac spine

The examination procedure took place in a clinic room with only the model, researcher and participant present. Examination of the PSIS involved the examiner sitting on a stool behind the model whilst placing one's hands on the iliac crests and placing their thumbs on the inferior aspect of the PSIS to assess the level in the horizontal plane (Mitchell and Mitchell, 1995; Greenman, 2003), as shown in Fig. 2.

Before each PSIS assessment the examiner wore a blindfold. When the blindfold was removed by the chief researcher (CS), the examiner assessed the PSIS levels and recorded the results. In order to reduce potential bias, results were self-recorded by participants on a form with three outcome possibilities: Left PSIS higher, PSIS equal level, or Right PSIS higher. The heel wedge was inserted in a randomized order three times on the right, three time on the left, and three times it was not on either side (i.e., was absent), creating a total of nine readings. After all nine readings had been taken, the examiner folded the results sheet and placed it into a box, a format similar to previous studies (O'Haire and Gibbons, 2000).

Hidden foot wedge

Unknown to the examiner, a 5 mm heel wedge (trial posting wedge set, available from Canonbury Healthcare) (Fig. 3) was placed under the model's heel during the testing, three times on the left, three times on the right. The randomized order of heel wedge insertions was: 1 — Left, 2 — None, 3 — Left, 4 — Right, 5 — Right, 6 — None, 7 — None, 8 — Left, 9 — Right. The heel wedge and entire lower limb of the model was concealed from the tester using black cloth.

The model stood barefoot upon a foot plate with feet parallel and hip width apart, ensuring that the feet were in the same position consistently (Fig. 3). The model was instructed to stand straight with weight equally between both feet. The positioning of the model was in accordance with previous studies (Vincent-Smith and Gibbons, 1999).

Figure 2 Photo of PSIS assessment.

Statistical analysis

Examiner results were used to calculate the inter-examiner reliability of detecting PSIS positional asymmetries using Fleiss' Kappa ($F\kappa$) and agreement strength interpreted using the $F\kappa$ benchmark scale as shown in Table 1. Kappa is calculated as $\kappa = (P_o - P_e)/(1 - P_e)$, where P_o is the

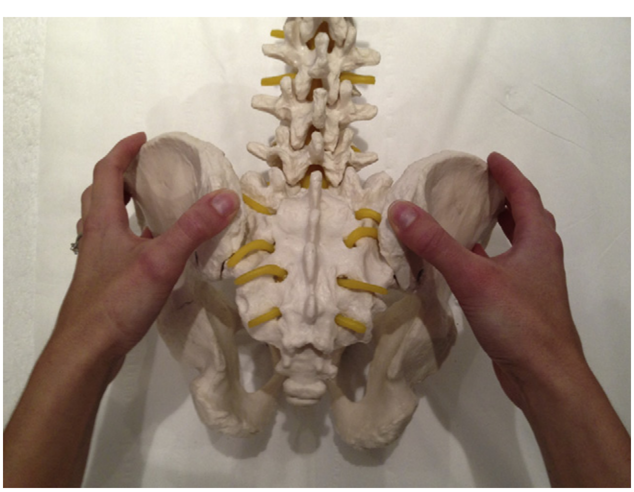

Figure 1 Assessment procedure of PSIS on an anatomical model shown during examiner training.

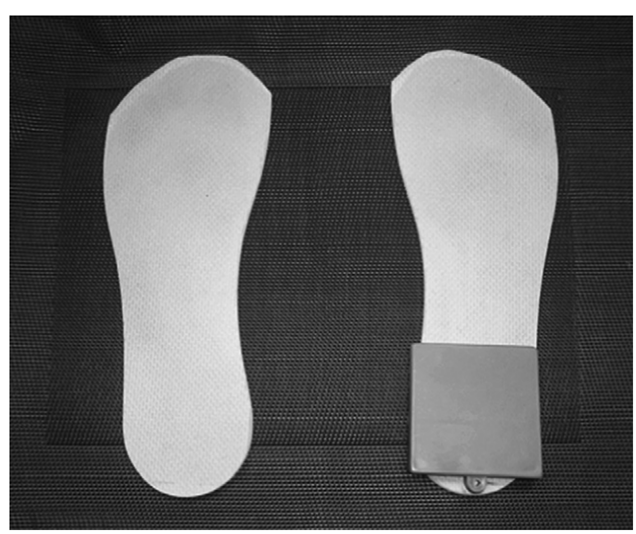

Figure 3 Photograph of foot plate with heel wedge inserted on the right foot.

Table 1 Fleiss' Kappa (Fκ) benchmark scale (SAS II).

Kappa statistic (Fκ)	Strength of agreement
<0.40	Poor
0.40–0.75	Intermediate to good
>0.75	Excellent

proportion of observed agreement between examiners, and P_e is the proportion of agreement achieved by chance alone (SAS II; Landis and Koch, 1977; Haas, 1991; Fleiss et al., 2003).

The scores for the correct determination of the pelvic asymmetry were analysed using one-way repeated measures analysis of variance (ANOVA) to determine if there were any differences between the accuracy of the three sample groups, using Sigma Plot 11.0.

Results

Fleiss Kappa Statistic

The inter-reliability of determining PSIS positional asymmetry was shown to have 'poor' levels of inter-rate agreement using the Fκ statistic. As shown in Table 2, all Fκ values were below 0.4 which are considered to have poor clinical reliability (Sim and Wright, 2005).

Detection of heel wedge using PSIS level assessment score

Data were analysed using the Kruskal–Wallis One Way ANOVA. This non-parametric test was used as the data did not conform to assumptions required for the parametric equivalent. There was no statistically significant difference between the scores of the three groups ($H = 0.435$ with 2 degrees of freedom ($P = 0.805$). Fig. 4 illustrates that all three groups displayed similar median values for the scores out of nine, with the experienced osteopaths having slightly more accuracy than the two student groups. The horizontal reference bar indicates the score that should have been obtained even if examiners were guessing.

Fig. 5 displays the distribution of data for each of the three groups. This illustrates that the experienced osteopaths displayed a smaller range in results, consistent with the smaller error bars seen in Fig. 4. The experienced osteopath's scores were closer to the midline and there were no outliers at the lower end of the score range. The third year students showed the largest amount of variance,

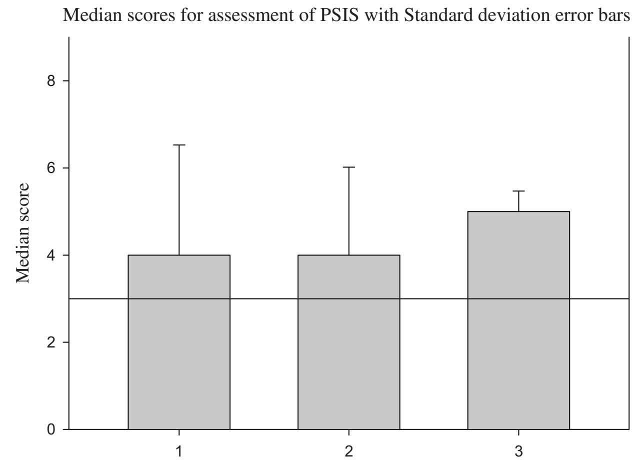

Figure 4 Median scores of all three groups: 1 — Third year students; 2 — Fourth year students; 3 — Experienced osteopaths.

with outliers at both end of the spectrum. The fourth year group had a smaller range than the third years but still shows outliers at both ends of the spectrum.

Differences between detection of 'left', 'right' or 'equal' PSIS levels

There was no statistically significant difference found between the scores for left, right and equal PSIS positions in the third year group ($H = 1.018$ with 2 degrees of freedom, $P = 0.601$), or the fourth year group ($H = 0.738$ with 2 degrees of freedom, $P = 0.692$).

Fig. 6 shows the median scores for the third year students, with scores out of three for, 'left', 'right' and 'equal', with standard deviation error bars. The median score is no better than would be expected by chance, as shown by the reference line at $y = 1$.

Fig. 7 shows the median scores for the fourth year students, with scores out of three for, 'left', 'right' and 'equal', with standard deviation error bars. The median

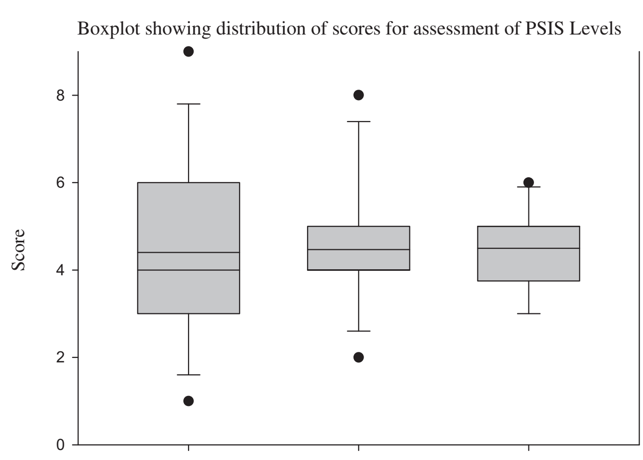

Figure 5 Box plot showing distribution of scores. 1 — Third year students; 2 — Fourth year students; 3 — Experienced osteopaths.

Table 2 Inter-reliability of determining PSIS positional asymmetry.

Group	Fleiss Kappa statistic	Fleiss Kappa benchmark
Third years	0.025	Poor
Fourth years	0.065	Poor
Tutors	0.058	Poor
All combined	0.063	Poor

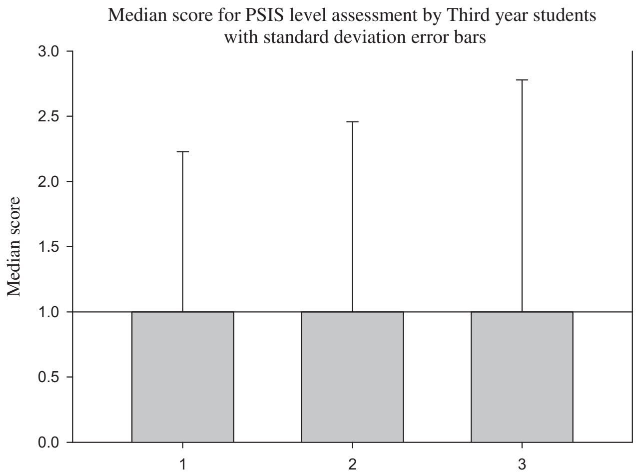

Figure 6 Median scores by third year students. 1 — Left; 2 — Equal; 3 — Right.

score is better than would be expected by chance, as shown by the reference line at $y = 1$.

There was a statistically significant difference found between the scores for left, right and equal PSIS positions detected by experienced osteopaths ($H = 9.348$ with 2 degrees of freedom, $P = 0.009$). The difference between detection of the right side lesions was statistically more accurate than both the left and equal positions as determined by the Pairwise Multiple Comparison Procedures (Tukey Test) ($P < 0.05$).

Fig. 8 shows the median scores out of three for 'left', 'right' and 'equal' as detected by the experienced osteopaths. The right 'lesions' show a higher median than the left and equal. The left and equal scores are only as good as would be expected by chance, as shown by the reference line at $y = 1$.

Discussion

This investigation found that the inter-reliability of osteopaths examining the levels of the PSIS during a static

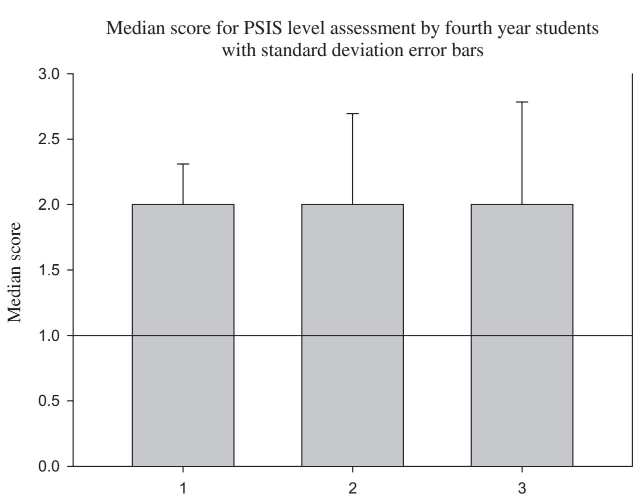

Figure 7 Median scores by fourth year students. 1 — Left; 2 — Equal; 3 — Right.

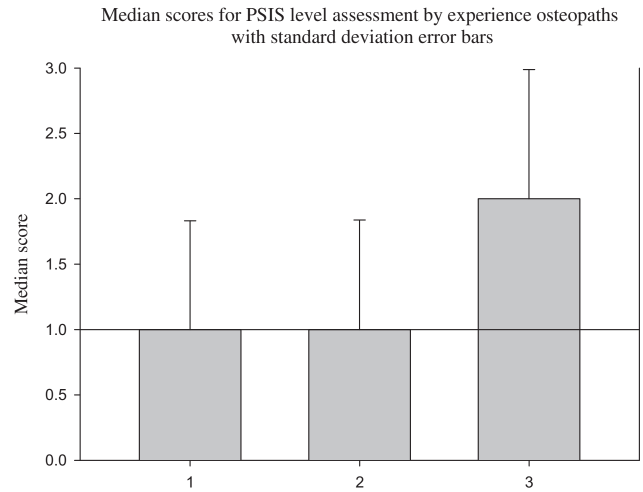

Figure 8 Median scores for experienced osteopaths. 1 — Left; 2 — Equal; 3 — Right.

palpation test was below the level required for a test to be clinically reliable (κ all < 0.4) (Sim and Wright, 2005). The inter-reliability was poor for all levels of experience assessed: i.e., students with three years experience, students with four years training and experienced osteopaths. There is no consensus on what constitutes expertise in osteopathy or other manual therapies (Thomson et al., 2011). For example, years in clinical practice and advanced professional certification are not able to differentiate between 'expert' and 'average' practitioners when using clinical outcomes to identify expert physical therapists (Resnik and Jensen, 2003). This current study used practitioners with at least five years post qualification clinical experience with a mean of 14.5 years which is consistent with models of expertise frequently used in health professions research (Chase and Simon, 1973; Doody and McAteer, 2002). However, it is likely that expertise encompasses a wide range of clinical attributes (Jensen et al., 2000) and future research should explore variables such as hours in clinic, numbers of patients, and self-characterized 'styles' of practice (for example 'cranial', 'visceral' or 'structural' approaches).

There was no significant difference between the three groups to accurately detect the changes in PSIS levels as analysed using ANOVA. The median scores showed all groups were better than would be expected by chance. The median score for third and fourth year students was 4 out of 9 with score ranges of 1—9 and 2—8 respectively. The median score for experienced osteopaths was 5 out of 9 with a smaller range (3—6). Thus although ANOVA shows no significant difference, there is a trend that the most experienced group were slightly more accurate and were the only group to have no scores below the level which would be achieved by pure guess work.

The heel wedge was placed three times on the left heel, three times on the right, and three times was not used. This allowed investigators to determine if examiners were better at detecting changes on one side or the other. ANOVA testing showed that for the third and fourth year students there was no difference in the accuracy of detection between, left, right and equal. However, in the experienced osteopath group ANOVA testing showed they were more accurate at detecting the heel wedge on the right than on the left or equal. It is postulated that this could be due to an effect of the dominant hand; however, in all three groups the vast majority of examiners were right handed. Thus, the trend would have been expected to occur in all groups. Of all forty examiners used, only two were left handed.

The outcomes of this study are in accordance with existing research indicating that clinical experience or training does not influence palpatory precision (Maher and Adams, 1994; Potter and Rothstein, 1985). Fryer et al. (2005) investigated inter- and intra-examiner reliability of assessment of pelvic anatomical landmarks with palpation. In this study, the PSIS assessment kappa values were poor, ranging from 0.08 to 0.15. Kmita and Lucas (2008) found that assessment of asymmetry of pelvic anatomical landmarks was generally low. PSIS kappa values ranged from −0.038 to 0.35. O'Haire and Gibbons (2000) also found low inter-reliability for assessing pelvic anatomical landmarks, with kappa scores of 0.07–0.58 for PSIS assessment. More recently, Rajendran and Gallagher (2010) investigated pelvic landmark assessment using palpation within a group of undergraduate osteopathic students and recorded values of −0.0476 to 0.0330, which are below an acceptable level to be clinically reliable ($F\kappa < 0.4$) (Sim and Wright, 2005).

Limitations of our study include small sample sizes and the use of only one model. Future studies should incorporate a larger sample size to give greater statistical power, and the use of more than one model. In this study a 26 year old female of normal BMI was used to reduce the influence that soft tissue thickness may have had on palpatory accuracy (Harlick et al., 2007). Follow up studies should investigate whether different morphologies give different results. It could be that certain palpation techniques are valid for certain subsets of patients, and not others, depending upon for example their gender or BMI (Harlick et al., 2007).

Another limitation was that the degree of asymmetry was not defined. Only one size of heel wedge was used, the choice of the 5 mm wedge was based upon a small pilot study which found it to be a mid-sized, detectable wedge. Due to use of heel wedges to detect changes in PSIS levels being a novel method there were no other previous studies to base this choice on. While a unilateral heel wedge inducing 5° of calcaneal eversion can affect the positions of the hip, pelvis and thorax in all three dimensions (Tateuchi et al., 2011), it may be that a 5 mm wedge was not large enough to produce what examiners considered a clinically significant amount of pelvic asymmetry. Future studies should incorporate different wedge sizes, so that sensitivity of palpation could be graded according to what size wedge is detectable. Experimentally at least, the larger the foot wedge the greater the change in pelvic position (Tateuchi et al., 2011). This was also found by Fryer (2006) who reported that the level of asymmetry was a strong factor in the inter-reliability of medial malleoli asymmetry. Fryer (2006) found that inter-reliability was almost perfect ($\kappa = 0.94$) in a group selected to have medial malleoli asymmetry greater than 4 mm but less than 10 mm. In the group that were not pre-selected to have asymmetries of medial malleoli, the inter-reliability was fair ($\kappa = 0.22$).

Studies investigating inter-reliability have proposed fatigue as a possible cause for low reliability, notably when multiple landmark assessments and multiple models have been used (Fryer et al., 2005; Paulet and Fryer, 2009). Reliability studies investigating manual and palpatory assessment procedures are faced with a number of methodological challenges (see Stochkendahl et al., 2006). For example, it is possible that after the repeated application of load during testing, the range of motion is likely to change between the first and last examiner due to the tissue mechanics and viscoelastic properties of ligament and connective tissue in the lumbar and pelvic regions (Vleeming et al., 1992; Bussey et al., 2009). Likewise, the act of palpating of soft tissue structures in itself to detect tissue changes may have an effect on the surrounding myofascial structures (Ercole et al., 2010), thus altering what the examiners are assessing throughout the course of the investigation. Therefore, it may be unsurprising that reliability of these types of tests is frequently found to be low (Maher et al., 1998; Meijne et al., 1999; Vincent-Smith and Gibbons, 1999; Riddle and Freburger, 2002; Fryer et al., 2005; Schneider et al., 2008; Paulet and Fryer, 2009). However, reliability tests using anatomical landmark assessment avoids these problems.

This study contributes to the growing body of research suggesting that static palpation of the SIJ is unreliable (Meijne et al., 1999; Vincent-Smith and Gibbons, 1999; O'Haire and Gibbons, 2000; van der Wurff et al., 2000a; van der Wurff et al., 2000b; Riddle and Freburger, 2002; Tong et al., 2006; Mitchell et al., 2007; Robinson et al., 2007; Kmita and Lucas, 2008; Rajendran and Gallagher, 2010). If basic surface palpation of anatomical landmarks is unreliable, it is unsurprising that motion testing and clinical test reliability is also shown to have low reliability, as these invariably and fundamentally rely upon the correct identification of landmarks (Meijne et al., 1999; Vincent-Smith and Gibbons, 1999; Fryer et al., 2005).

A profession is required to possess a discrete body of knowledge to be used within its teaching and practice, which is vital for academic credibility and legitimacy (Richardson et al., 2004). With the landscape of manual therapy becoming increasingly competitive, the osteopathic profession faces challenging questions with regard to its professional values and identity (Tyreman, 2008; Thomson et al., in press), epistemology of practice (Tyreman, 2008; Lucas and Moran, 2007), and the process of clinical reasoning (Thomson et al., 2011). Members of the osteopathic profession should not fear relegating elements of practice with such shaky foundations, fearing that it is as 'unosteopathic'. Rather, osteopathy needs to reflectively examine its models of practice which it has embraced so tightly, in order that it continues to grow and move forward as a profession.

Conclusion

This study demonstrated that inter-reliability of palpation to locate the PSIS and assess their levels was poor in both students and experienced osteopaths. This is consistent with other studies, and as such it is suggested that any diagnostic tests based upon techniques which rely upon palpation alone have low clinical utility. In taking an evidence informed approach to osteopathy, it is difficult to ignore the results of studies such as this and others. Taking a reflective stance towards all aspects of practice will help ensure that osteopathy is a progressive profession. The triangulation of the results from this study and others suggest that the role of such tests in osteopathic curriculum should be reconsidered.

References

Bogduk, N., 2004. Management of chronic low back pain. Medical Journal of Australia 181, 55.

Breen, A., 1992. The reliability of palpation and other diagnostic methods. Journal of Manipulative and Physiological Therapeutics 15 (1), 54–56.

Burton, A., Edwards, V., Sykes, D., 1990. Invisible skin marking for testing palpatory reliability. Journal of Manual Medicine 5, 27–29.

Bussey, M.D., Bell, M.L., Milosavljevic, S., 2009. The influence of hip abduction and external rotation on sacroiliac motion. Manual Therapy 14 (5), 520–525.

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

Chase, W., Simon, H.A., 1973. Perception in chess. Cognitive Psychology 4, 55–81.

Degenhardt, B.F., Snider, K.T., Snider, E.J., Johnson, J.C., 2005. Interobserver reliability of osteopathic palpatory diagnostic tests of the lumbar spine: improvements from consensus training. JAOA: The Journal of the American Osteopathic Association 105 (10), 465–473.

DiGiovanna, E.L., Schiowitz, S., 1997. An Osteopathic Approach to Diagnosis and Treatment. Lippincott–Raven, Philadelphia.

Doody, C., McAteer, M., 2002. Clinical reasoning of expert and novice physiotherapists in an outpatient orthopaedic setting. Physiotherapy 88 (5), 258–268.

Ercole, B., Antonio, S., Julie Ann, D., Stecco, C., 2010. How much time is required to modify a fascial fibrosis? Journal of Bodywork and Movement Therapies 14 (4), 318–325.

Fleiss, J.L., Levin, B.A., Paik, M.C., 2003. Statistical Methods for Rates and Proportions. J. Wiley, Oxford, UK.

Freburger, J.K., Riddle, D.L., 2001. Using published evidence to guide the examination of the sacroiliac joint region. Physical Therapy 81 (5), 1135–1143.

French, S.D., Green, S., Forbes, A., 2000. Reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low-back pain. Journal of Manipulative and Physiological Therapeutics 23 (4), 231–238.

Fryer, G., 2003. Intervertebral dysfunction: a discussion of the manipulable spinal lesion. Journal of Osteopathic Medicine 6 (2), 64–73.

Fryer, G., 2006. Factors affecting the intra-examiner and inter-examiner reliability of palpation for supine medial malleoli asymmetry. International Journal of Osteopathic Medicine 9 (2), 58–65.

Fryer, G., 2008. Teaching critical thinking in osteopathy – integrating craft knowledge and evidence-informed approaches. International Journal of Osteopathic Medicine 11 (2), 56–61.

Fryer, G., Johnson, J.C., Fossum, C., 2010. The use of spinal and sacroiliac joint procedures within the British osteopathic profession. Part 1: assessment. International Journal of Osteopathic Medicine 13 (4), 143–151.

Fryer, G., McPherson, H.C., O'Keefe, P., 2005. 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 8 (4), 131–138.

Fryer, G., Morse, C.M., Johnson, J.C., 2009. Spinal and sacroiliac assessment and treatment techniques used by osteopathic physicians in the United States. Osteopathic Medicine and Primary Care 3 (1), 1–11.

Gonnella, C., Paris, S.V., Kutner, M., 1982. Reliability in evaluating passive intervertebral motion. Physical Therapy 62 (4), 436–444.

GOsC, 2001. General Osteopathic Council Survey and Statistics. From: http://www.osteopathy.org.uk/uploads/survey2snapshot_survery_results_2001.pdf.

Greenman, P.E., 2003. Principles of Manual Medicine. Lippincott Williams & Wilkins, Philadelphia, USA.

Haas, M., 1991. Statistical methodology for reliability studies. Journal of Manipulative and Physiological Therapeutics 14 (2), 119.

Haldeman, S., 2004. Principles and Practice of Chiropractic. McGraw-Hill, Medical Pub. Division, New York.

Harlick, J.C., Milosavljevic, S., Milburn, P.D., 2007. Palpation identification of spinous processes in the lumbar spine. Manual Therapy 12 (1), 56–62.

Hestboek, L., Leboeuf-Yde, C., 2000. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. Journal of Manipulative and Physiological Therapeutics 23 (4), 258–275.

Hungerford, B.A., Gilleard, W., Moran, M., Emmerson, C., 2007. Evaluation of the ability of physical therapists to palpate intrapelvic motion with the Stork test on the support side. Physical Therapy 87 (7), 879–887.

Jensen, G.M., Gwyer, J., Shepard, K.F., 2000. Expert practice in physical therapy. Physical Therapy 80 (1), 28–43. discussion 44–52.

Kmita, A., Lucas, N.P., 2008. 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 11 (1), 16–25.

Kuchera, W.A., Kuchera, M.L., 1992. Osteopathic Principles in Practice. Greyden Press, Original Works Books, Dayton, Ohio, USA.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Laslett, M., Aprill, C.N., McDonald, B., Young, S.B., 2005. Diagnosis of sacroiliac joint pain: validity of individual provocation tests and composites of tests. Manual Therapy 10 (3), 207–218.

Lee, D., 2004. The Pelvic Girdle: an Approach to the Examination and Treatment of the Lumbopelvic–Hip Region. Churchill Livingstone, London.

Licciardone, J.C., 2007. Responding to the challenge of clinically relevant osteopathic research: efficacy and beyond. International Journal of Osteopathic Medicine 10 (1), 3–10.

Lucas, N., Bogduk, N., 2011. Diagnostic reliability in osteopathic medicine. International Journal of Osteopathic Medicine 14 (2), 43–47.

Lucas, N.P., Moran, R.W., 2007. Is there a place for science in the definition of osteopathy? International Journal of Osteopathic Medicine 10 (4), 85–87.

Maher, C., Adams, R., 1994. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. Physical Therapy 74 (9), 801–811.

Maher, C.G., Latimer, J., Adams, R., 1998. An investigation of the reliability and validity of posteroanterior spinal stiffness judgments made using a reference-based protocol. Physical Therapy 78 (8), 829–837.

Meijne, W., van Neerbos, K., Aufdemkampe, G., van der Wurff, P., 1999. Intraexaminer and interexaminer reliability of the Gillet test. Journal of Manipulative and Physiological Therapeutics 22 (1), 4—9.

Mitchell, F.L., Mitchell, P.K.G., 1995. The Muscle Energy Manual, vol. 1. MET Press, East Lansing, MI, USA.

Mitchell, T.D., Urli, K.E., Breitenbach, J., Yelverton, C., 2007. The predictive value of the sacral base pressure test in detecting specific types of sacroiliac dysfunction. Journal of Chiropractic Medicine 6 (2), 45—55.

Moriguchi, C.S., Carnaz, L., Silva, L.C.C.B., Salasar, L.E.B., Carregaro, R.L., Sato, T.O., Coury, H.J.C.G., 2009. Reliability of intra- and inter-rater palpation discrepancy and estimation of its effects on joint angle measurements. Manual Therapy 14 (3), 299—305.

O'Haire, C., Gibbons, P., 2000. Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study. Manual Therapy 5 (1), 13—20.

Paulet, T., Fryer, G., 2009. Inter-examiner reliability of palpation for tissue texture abnormality in the thoracic paraspinal region. International Journal of Osteopathic Medicine 12 (3), 92—96.

Potter, N.A., Rothstein, J.M., 1985. Intertester reliability for selected clinical tests of the sacroiliac joint. Physical Therapy 65 (11), 1671—1675.

Rajendran, D., Gallagher, D., 2010. The assessment of pelvic landmarks using palpation: a reliability study of undergraduate students. International Journal of Osteopathic Medicine 14 (2), 57—60.

Resnik, L., Jensen, G.M., 2003. Using clinical outcomes to explore the theory of expert practice in physical therapy. Physical Therapy 83 (12), 1090—1106.

Richardson, B., Higgs, J., Dahlgren, M.A., 2004. Recognising practice epistemology in the health professions. In: Higgs, J., Richardson, B., Dahlgren, M.A. (Eds.), Developing Practice Knowledge for Health Professionals. Butterworth Heinemann, Edinburgh, pp. 1—14.

Riddle, D.L., Freburger, J.K., 2002. Evaluation of the presence of sacroiliac joint region dysfunction using a combination of tests: a multicenter intertester reliability study. Physical Therapy 82 (8), 772—781.

Robinson, H.S., Brox, J.I., Robinson, R., Bjelland, E., Solem, S., Telje, T., 2007. The reliability of selected motion- and pain provocation tests for the sacroiliac joint. Manual Therapy 12 (1), 72—79.

Sackett, D.L., 2000. Evidence-Based Medicine: How to Practice and Teach EBM. Churchill Livingstone, Edinburgh.

SAS II. Compute Estimates and Tests of Agreement Among multiple raters. From: http://support.sas.com/kb/25/006.html.

Schneider, M., Erhard, R., Brach, J., Tellin, W., Imbarlina, F., Delitto, A., 2008. Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability study. Journal of Manipulative and Physiological Therapeutics 31 (6), 465—473.

Schwarzer, A.C., Aprill, C.N., Bogduk, N., 1995. The sacroiliac joint in chronic low back pain. Spine 20 (1), 31—37.

Sim, J., Wright, C.C., 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical Therapy 85 (3), 257—268.

Stochkendahl, M.J., Christensen, H.W., Hartvigsen, J., Vach, W., Haas, M., Hestbaek, L., Adams, A., Bronfort, G., 2006. Manual examination of the spine: a systematic critical literature review of reproducibility. Journal of Manipulative and Physiological Therapeutics 29 (6), 475—485. e410.

Tateuchi, H., Wada, O., Ichihashi, N., 2011. Effects of calcaneal eversion on three-dimensional kinematics of the hip, pelvis and thorax in unilateral weight bearing. Human Movement Science 30 (3), 566—573.

Thomson, O.P., Petty, N.J., Moore, A.P., 2011. Clinical reasoning in osteopathy — more than just principles? International Journal of Osteopathic Medicine 14 (2), 71—76.

Thomson, O.P., Petty, N.J., Moore, A.P., in press. Reconsidering the patient-centeredness of osteopathy. International Journal of Osteopathic Medicine. doi:10.1016/j.ijosm.2012.03.001.

Tong, H.C., Heyman, O.G., Lado, D.A., Isser, M.M., 2006. Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position. JAOA: The Journal of the American Osteopathic Association 106 (8), 464—468.

Tyreman, S., 2008. Valuing osteopathy: what are (our) professional values and how do we teach them? International Journal of Osteopathic Medicine 11 (3), 90—95.

van der Wurff, P., Hagmeijer, R., Meyne, W., 2000a. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: reliability. Manual Therapy 5 (1), 30—36.

van der Wurff, P., Meyne, W., Hagmeijer, R.H.M., 2000b. Clinical tests of the sacroiliac joint: a systematic methodological review. Part 2: validity. Manual Therapy 5 (2), 89—96.

Vincent-Smith, B., Gibbons, P., 1999. Inter-examiner and intra-examiner reliability of the standing flexion test. Manual Therapy 4 (2), 87—93.

Vleeming, A., Van Wingerden, J., Dijkstra, P., Stoeckart, R., Snijders, C., Stijnen, T., 1992. Mobility in the sacroiliac joints in the elderly: a kinematic and radiological study. Clinical Biomechanics 7 (3), 170—176.

Ward, R.C., Hruby, R.J., 2002. Foundations for Osteopathic Medicine. Lippincott Williams & Wilkins, Philadelphia, USA.