Manual Therapy (2000) **5(1),** 13–20

DOI: 10.1054/math.1999.0203, available online at http://www.idealibrary.com on IDEAL®

Original article

Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study

C. O'Haire, P. Gibbons

School of Health Sciences, Faculty of Human Development, Victoria University, Australia

SUMMARY. **Despite the paucity of research into the reliability of static palpation, it is still employed extensively as a diagnostic tool by manual medicine practitioners. This study tested the inter- and intra-examiner agreement of ten senior osteopathic students using static palpation on ten asymptomatic subjects. Four assessments of the posterior superior iliac spine (PSIS), sacral sulcus (SS), and the sacral inferior lateral angle (SILA) on every subject by all examiners resulted in 1200 assessments in total. Kappa (Kg) yielded intra-examiner agreement that ranged between less-than-chance to substantial for the SILA (Kg = − 0.05 to 0.69; mean Kg = 0.21), and slight to moderate for the PSIS (Kg = 0.07 to 0.58; mean Kg = 0.33) and the SS (Kg = 0.02 to Kg = 0.60; mean Kg = 0.24), with 50% significant beyond the 0.05 level. Inter-examiner agreement was slight (PSIS Kg = 0.04; SILA Kg = 0.08; SS Kg = 0.07) and significant at the 0.01 level. Intra-examiner agreement was greater than inter-examiner agreement, which was consistent with existing palpation reliability studies. The poor reliability of clinical tests involving palpation may be partially explained by error in landmark location.**

INTRODUCTION

In order to diagnose and treat a patient with a musculoskeletal complaint, a manual medicine practitioner relies upon clinical skills. In particular the ability to take a history of the patient's health and to undertake a thorough physical examination of the patient. Physical examination includes observation, palpation, motion testing, and neuro-vascular assessment (Ward 1996). A manual medicine practitioner uses palpatory analysis of a patient to identify somatic problems, to treat the problems found and to assess the result of treatment (Dvorak & Dvorak 1990; DiGiovanna & Schiowitz 1991; Basmajian & Nyberg 1993; Greenman 1996; Ward 1996). Despite this reliance upon palpation as a diagnostic tool, the reliability of palpation remains to be proven.

Systematic and orderly observation and palpation of a patient's bony structures provide a method of evaluating a patient's anatomy (Bourdillon 1970; Hoppenfeld 1976; Dvorak & Dvorak 1990; DiGiovanna & Schiowitz 1991; Basmajian & Nyberg 1993; Greenman 1996; Ward 1996). The limitation of static palpation as a diagnostic tool is that the presence of perceived osseous asymmetry may not be related to abnormal mechanical function (Bowen & Cassidy 1981; Russell 1983; Diek et al. 1985; Vleeming et al. 1989). This is why manual medicine texts advise diagnosis be based upon collections of clinical findings (Bourdillon 1970; Dvorak & Dvorak 1990; DiGiovanna & Schiowitz 1991; Basmajian & Nyberg 1993; Greenman 1996; Ward 1996). This may also explain why so few studies have concentrated purely on static palpation. In an extensive review of chiropractic literature, Haas (1991b) found six studies that investigated static palpation, only two of the six used suitable statistics to support their conclusions, but neither of these two assessed independent palpatory procedures. He concluded that there were no studies to prove or refute the reliability of static palpation.

McConnell et al. (1980) compared diagnostic agreement among six osteopathic physicians free to use their own customary neuro-musculoskeletal examination procedures to identify spinal segmental dysfunction in patients with acute spinal pain. The research protocol imposed certain restraints upon the

Catherine O'Haire, BSc, MHSc(Ost), School of Health Sciences,
Peter Gibbons, MB, BS, DO, DM-SMed,
Associate Professor, Osteopathic Medicine Program; Head, School of Health Sciences, Faculty of Human Development, City Campus, Victoria University, PO Box 14428, MC, Melbourne, VIC 8001, Australia.
Correspondence to PG.

examiners that may have altered their diagnostic precision. For example, only one of the six physicians was privy to the patient's medical history. The remaining physicians diagnosed the patient's condition based only upon the results of physical examination of the patient. Under these circumstances any agreement would be remarkable. They reported low inter-examiner agreement and stated that this disagreement arose primarily due to differences in examination technique, and 'filtering of perceptions based upon conceptual orientation'. In order to improve reliability in future studies, they recommended that examiners agree upon the areas to be examined, the test procedure to be used, the method for quantifying severity, and the method of recording.

An examiner's level of experience may affect their clinical skills. It would be easy to presume that more experience equals better skills, but this view has not been supported in the literature. Some authors suggest that experience enables an examiner to filter out insignificant observations and recognize common palpatory patterns (Kappler 1980; Wiles 1980; Mann et al. 1984), but these studies were not rigorous in their definition of examiners clinical experience. Moir et al. (1990) clearly defined the experience of their examiner groups and found that the examiner level of experience could not explain the inter-examiner differences in accuracy of palpation of sacroiliac motion. However, they reported that intra-examiner agreement was improved for experienced examiners compared to less experienced examiners. Other researchers have been unable to support the hypothesis that experience plays a significant role in improving inter-examiner or intra-examiner reliability (Koran 1975a; 1975b; McConnell et al. 1980). The role experience plays in clinical accuracy remains to be established.

Static palpation has been used as an integral part of the methodology in studies whose primary focus was not the reliability of static palpation (Taylor & Slinger 1980; Montgomery et al. 1995). Taylor and Slinger (1980) used a variety of anthropometric measurements that incorporated location of osseous landmarks of the leg, pelvis and spine. They reported that these measurement techniques were reliable based upon a re-measurement study within their experimental protocol. The methodology and statistical basis for their claim of reliability was not reported, limiting the usefulness of their claim and preventing a repeat study. Montgomery et al. (1995) reported that static palpation of the sacral inferior lateral angle was almost as accurate an indicator (78%) of an anatomical short leg as a proven radiographic procedure. Unfortunately, their statistical analysis did not correct for chance agreement between tests, and only one examiner carried out the palpatory examinations with no re-test measurement for testing intra-examiner reliability. It is impossible to determine if the high accuracy reported was a function of this particular examiner's skill, or the skill of palpation itself.

The inter-examiner agreement of palpatory pain, static palpation for misalignment and visual observation of the lumbar spine was investigated by Keating et al. (1990). They found good agreement for production of pain over spinous processes, fair agreement for visual observation, and reported that static palpation was not reliable. However, examiners moved each subject from a prone to a seated position during the testing procedure, which may explain why static palpation was reported to be unreliable. It remains to be established whether static palpation reliability may be improved if the subject remains as motionless as possible during the experimental procedure. Granted, this is not reflective of clinical practice, but it permits greater chance that examiners are palpating landmarks that are in the same relationship.

Examiners' ability to agree on anatomical points in the lumbar spine is complicated because the examiner must define and name spinal segments and this may be a source of inter-examiner error (Gonnella et al. 1982; Keating et al. 1990). Breen (1992) suggested that sacroiliac osseous landmark palpation is less problematic because there is not the same need for segmental definition. Palpation reliability studies using invisible ink to mark the skin of subjects confirm this opinion, but fail to prove the reliability of location of sacroiliac landmarks (Burton et al. 1990; Byfield et al. 1992; Simmonds & Kumar 1993).

Studies into the reliability of sacroiliac motion palpation have produced conflicting results; reporting inter-examiner agreements as high (Wiles 1980), fair (Bowman & Gribble 1995), and slight (Carmichael 1987; Herzog et al. 1989). Problems with the methodology of some of these studies mean that their results must be viewed with caution. For instance, only one repeat observation was made to measure intra-examiner reliability (Carmichael 1987), a control group ($n = 1$) was used for comparison with the experimental group ($n = 11$) (Herzog et al. 1989), statistics not published to support conclusions drawn (Wiles 1980; Carmichael 1987; Herzog et al. 1989), and finally the high agreement reported by Wiles (1980) was not substantiated by the results reported. In addition, the statistical methods used by Wiles (1980), Carmichael (1987) and Herzog et al. (1989) have been called into question as not being valid measures of reliability (Haas 1991a). Researchers trying to establish the reliability of sacroiliac motion have not discussed the possibility that their disappointing results may be due to examiners failing to palpate the same structures. If the basis for motion palpation is assumed to be static palpation, it may be helpful to establish whether examiners can reliably locate static anatomical landmarks before it is possible to validate motion testing associated with use of those landmarks.

A study by Mann et al. (1984) was designed to test intra-examiner and inter-examiner ability to palpate and observe iliac crest heights in the standing subject. Eleven examiners were instructed in the method of palpation and observation then randomly assigned to one of ten subjects to assess that subject's iliac crest heights on two occasions. Their data was summarized descriptively, and without applying further statistical analysis they concluded that comparing iliac crest heights using observation and palpation was an unreliable test. The reliability of examiners may have been reduced because the subjects were covered from the waist up and this reduced the visual clues available to the examiner, in addition, reliability may have been affected due to overcrowding and rushing examiners and subjects during data collection. The examiners were not instructed to use their dominant eye to sight the crest heights, and if they alternated eyes between tests this may reduce reliability (Greenman 1996; Ward 1996). The subjects were required to stand for extended periods of time and fatigue may have caused them to shift their weight from one foot to the other, possibly producing alteration in hip adduction–abduction with a resultant change in pelvic tilt and therefore change in crest heights. Addressing these issues may improve reliability in future studies.

Despite the paucity of research into the reliability of static palpation, it is still employed extensively as a diagnostic tool by manual medicine practitioners. Manual medicine practitioners' ability to accurately identify bony anatomical landmarks needs to be validated as motion testing is often predicated upon the assumption that this is the case. The aim of this study was to test the ability of examiners to agree within themselves (intra-examiner) and with each other (inter-examiner) when observing and palpating sacroiliac anatomical landmarks.

METHODS

Subjects

The Victoria University of Technology ethics review board approved this study. Volunteers were screened and excluded if they experienced discomfort lying prone (e.g. pain, stiffness, shortness of breath), if they felt they could not lie still (e.g. cough, itchy rash, pregnancy), or if they had identifying characteristics (e.g. tattoo, body piercing, birthmark) that may have aided examiners' ability to identify subjects between examinations. Ten asymptomatic female volunteers with a mean age of 24 years (range, 18–30 years) were recruited. The subjects were provided with identical elastic waisted shorts to wear for the duration of the experiment, further limiting the examiners' ability to identify subjects between examinations.

Examiners

Since the role of experience in clinical accuracy remains to be established, examiners with comparable levels of experience were used. The examiners consisted of ten 5th year osteopathic students, who were required to attend a 1 hour training session to standardize the method of observation, palpation, and recording of results.

Method of observation and palpation

The examiners stood at the side of the plinth that corresponded with their dominant eye, thus a right eye dominant examiner stood at the subject's right hand side during all examinations. Stereognosis was enhanced by placing the whole hand on the surface of the subject's body and gently sliding the skin around over the bone. Once the landmark had been identified, the examiner's palpating thumbs or fingers were then carefully placed on the specific landmark and the dominant eye was aligned over the subjects mid-sagittal line central to the anatomical landmark being assessed. The static palpation of the posterior superior iliac spine (PSIS), sacral sulcus (SS), and the sacral inferior lateral angle (SILA) were all examined with the subjects prone.

To examine the PSIS, the examiner placed their hands over the subject's iliac crests, slid their whole hand posteriorly along the iliac crest until they could place their thumb on the inferior surface of the PSIS to inspect the relative heights of their thumb in regard to the horizontal plane (Greenman 1996; Ward 1996). A decision as to which of the three following findings best described the relative position of the subject's PSIS's was made; (1) Right higher than left ($R > L$); (2) Left higher than right ($L > R$); or (3) Right equal to left ($R = L$).

The SS was examined by locating the PSIS as described above, then the examiner slid their thumbs medially until they could no longer palpate the PSIS and then cephalward to locate the sacral sulcus where they gently depressed the tissue beneath their thumbs (DiGiovanna & Schiowitz 1991; Greenman 1996; Ward 1996). The thumb that depressed further into the sacral sulcus indicated the side of the deep sacral base. A decision as to which of the three following findings best described the relative position of the subject's SS's was made; (1) Right deeper than left ($R > L$); (2) Left deeper than right ($L > R$); or (3) Right equal to left ($R = L$).

To examine the SILA the examiner placed their hand over the sacrum with fingers pointing caudad to palpate the end of the sacral crest, and then the sacral hiatus. They placed each of their thumbs lateral to this central landmark and rolled their thumbs inferiorly to contact the inferior aspect of the SILA to assess the relative heights of their thumbs in regard to the horizontal plane (DiGiovanna & Schiowitz 1991; Greenman 1996; Ward 1996). A decision as to which of the three following findings best described the relative position of the subject's SILAs was made; (1) Right higher than left (R > L); (2) Left higher than right (L > R); or (3) Right equal to left (R = L).

Method for recording results

A set of result cards was made for each subject upon which examiners could record their findings. A single card was used to record the results of one examiner palpating one pair of anatomical landmarks on one subject (Fig. 1), thus a set of ten cards provided findings of one landmark by all ten examiners for analysis of inter-examiner reliability. To allow analysis of intra-examiner reliability, each examiner palpated each site on each subject a total of four times. For this reason each subject had four sets of ten recording cards for each site.

Site: **PSIS**
Examiner: ....
Result: R > L
L > R
R = L

Fig. 1—A result card for the PSIS of one subject. This subject would have four sets of cards for each landmark, with ten identical cards in each set.

Experimental procedure

A large room was set up with ten plinths placed longitudinally such that they formed the circumference of a circle–there was at least 1 metre between the top of one plinth and the bottom of the next. Subjects lay prone on a plinth with a sheet placed over their shoulders and head. The subjects were encouraged to lie as still as possible for the duration of the experiment, just over 2.5 hours. A small table was placed at the top of each plinth. On this table was a box with a slit in the top for posting of result cards, a pen for recording results, and the result cards.

Prior to the examiners entering the room, each subject had a group of ten recording cards randomly drawn and placed on the table. All ten examiners then entered the examination room and were randomly assigned to an examination plinth. The examiner had up to 50 seconds to read the card, to palpate and observe the designated anatomical landmark, circle the appropriate finding on the card, initial the card and finally post it in the subject's box. Once all examiners had posted their results they were asked to move in a clockwise direction to the next subject. Once all ten examiners had palpated all ten subjects, all the examiners left the room. Outside the examination room the examiners were given memory tasks to distract them and reduce their ability to remember subject results. Meanwhile, inside the examination room, another set of cards was randomly drawn for each subject before examiners re-entered the room and were randomly assigned a new starting position. This sequence of events continued until all the data had been collected.

Statistical methods

Reliability was determined using the Generalised Kappa (Kg) statistic which allows assessment of observer agreement for more than two examiners and multiple examinations (Fleiss 1971; Haas 1991a). $Kg = (Po - Pe)/(1 - Pe)$ where Po is the proportion of observed agreement, Pe is the proportion of agreement expected by chance alone, Kg is the proportion of observed agreement above chance divided by the maximum possible proportion of agreement above chance (Fleiss 1971; Haas 1991a). Kg = 1 for perfect agreement, Kg = 0 for chance agreement, and Kg is negative when observed agreement is less than agreement expected by chance. To interpret kappa values between 0 and 1 the guidelines proposed by Landis and Koch (1977) were used, i.e. $0.0 - 0.2 =$ slight, $0.21 - 0.4 =$ fair, $0.41 - 0.6 =$ moderate, $0.61 - 0.8 =$ substantial, and $0.81 - 0.99 =$ almost perfect. These guidelines have been widely used in clinical research probably because they are appealing in their simplicity and aid the informal evaluation of kappa, but these scales are arbitrary (Kramer & Feinstein 1981; Lantz 1997). In this study, Kg was tested for significant difference from chance agreement using the standard error (SEo) statistic (Fleiss 1971; Haas 1991a). A small sample ($n < \kappa$, where k = number of categories i.e. 3, and n = sample size i.e.: 10) may risk error in judging the significance of kappa values (Kramer & Feinstein 1981; Haas 1991a), thus it must be emphasized that the actual magnitude of Kg is of greater importance than the significance level in interpreting reliability in this pilot study.

RESULTS

The findings of all examiners for each landmark on all subjects are presented in Tables 1, 2 and 3.

Intra-examiner agreement for four independent palpations of each landmark pair on all ten subjects was calculated for each examiner, the results of which are presented in Table 4. Static palpation of the PSIS

Table 1. Findings of all examiners on all subjects' posterior superior iliac spine (PSIS)

	Subjects									
Examiners	1	2	3	4	5	6	7	8	9	10
1	LLLR	EEEE	REEE	RRRR	LEEE	EEEE	LEEE	REEE	LLRE	EEEE
2	RRRE	EEEE	REEE	REEE	REEE	EEEE	EEEE	RRRE	EEEE	EEEE
3	LREE	REEE	LRRE	LLEE	EEEE	EEEE	LREE	LRRE	LLEE	LRRR
4	RREE	RRRE	RRRR	RRRR	RRRR	EEEE	RREE	RRRE	RREE	RREE
5	LLLE	EEEE	REEE	EEEE	EEEE	EEEE	LEEE	EEEE	EEEE	EEEE
6	EEEE	LLLE	RRRR	RRRE	LRRE	LREE	LLLE	RRRE	RRRE	LRRE
7	REEE	LLLE	EEEE	EEEE	EEEE	EEEE	LLLL	LLRE	EEEE	LEEE
8	EEEE	EEEE	EEEE	EEEE	EEEE	EEEE	EEEE	EEEE	RRRE	RRRR
9	REEE	LREE	RRRR	EEEE	LEEE	REEE	RRRR	LRRE	LEEE	RRRR
10	EEEE	EEEE	REEE	EEEE	EEEE	EEEE	REEE	RREE	RRRR	EEEE

L = left higher than right, R = right higher than left, E = right and left sides equal.

Table 2. Findings of all examiners on all subjects' sacral inferior lateral angle (SILA)

	Subjects									
Examiners	1	2	3	4	5	6	7	8	9	10
1	LEEE	EEEE	LLRE	LLRE	EEEE	REEE	REEE	RRRE	REEE	RRRR
2	RRRE	RRRE	REEE	REEE	REEE	REEE	RRRE	RRRR	RREE	RRRR
3	LLRE	LLEE	LLEE	LLLE	EEEE	LLLE	LLRE	LLLR	REEE	LLLE
4	RRRR	RREE	EEEE	RREE	RREE	REEE	RRRE	RRRR	RRRE	RRRR
5	RREE	EEEE	LEEE	LLEE	EEEE	LLLL	RRRE	RLLL	REEE	RREE
6	RRRE	EEEE	RREE	LREE	LREE	LLEE	LEEE	RLLL	EEEE	LREE
7	RREE	EEEE	EEEE	RRRE	LEEE	REEE	RRRE	EEEE	REEE	REEE
8	RRRR	RRRR	RRRE	RRRR	REEE	EEEE	RRRE	EEEE	EEEE	RRRR
9	LEEE	REEE	LEEE	LEEE	EEEE	EEEE	REEE	REEE	REEE	RRRR
10	RRRE	RRRR	RRRE	RRRE	LRRR	LLEE	RRRR	RREE	RRRE	RRRR

L = left higher than right, R = right higher than left, E = right and left sides equal.

Table 3. Findings of all examiners on all subjects' sacral sulcus (SS)

	Subjects									
Examiners	1	2	3	4	5	6	7	8	9	10
1	EEEE	EEEE	EEEE	REEE	LLEE	LRRE	LREE	REEE	LLRE	LLLE
2	RRRR	EEEE	REEE	EEEE	EEEE	LLLE	EEEE	EEEE	EEEE	RREE
3	RRRE	REEE	LLEE	LRRE	LEEE	LLLL	LLEE	LLLE	LLLE	RREE
4	RREE	REEE	LLLL	EEEE	REEE	REEE	LRRE	LRRE	REEE	LREE
5	LRRR	EEEE	LEEE	LLEE	EEEE	EEEE	LLEE	LLLL	LEEE	EEEE
6	LREE	LEEE	RRRR	LLRR	LLLR	LRRE	LLLR	LLLL	LLRE	LLLL
7	LRRR	EEEE	REEE	LLLL	LREE	LLLR	EEEE	EEEE	LEEE	EEEE
8	RRRR	LEEE	LEEE	LRRL	LREE	LEEE	EEEE	REEE	EEEE	REEE
9	EEEE	REEE	LEEE	LRRR	LREE	LRRE	LLLE	LLLE	LEEE	LREE
10	LREE	LLRR	LLRR	RRRR	LLLR	LRRR	LLLR	LLLR	LRRR	LLRR

L = left deeper than right, R = right deeper than left, E = right and left sides equal.

yielded slight to moderate agreement (Kg = 0.07 to 0.58; mean Kg = 0.33), six of these 10 Kappa scores were significantly different from chance agreement at the 0.01 level. The SILA yielded less than chance to substantial agreement (Kg = −0.05 to 0.69; mean Kg = 0.21), five of these 10 Kg tested significant, two at the 0.05 level and three at the 0.01 level. The SS yielded slight to moderate agreement (Kg = 0.02 to Kg = 0.60; mean Kg = 0.24), four of these 10 Kappa scores were significant at the 0.01 level.

Inter-examiner agreement was assessed for each sacroiliac landmark for all examinations by all examiners, the results are shown in Table 5. All three landmarks yielded slight agreement (PSIS Kg = 0.04; SILA Kg = 0.08; SS Kg = 0.07) significant at the 0.01 level.

Table 4. Intra-examiner reliability using four independent assessments of three sacro-iliac anatomical landmarks (PSIS, SILA, SS) on 10 subjects

	PSIS			SILA			SS		
Examiner	Po	Pe	Kg	Po	Pe	Kg	Po	Pe	Kg
1	0.67	0.46	$0.38^{\square}$	0.58	0.44	$0.26^{\square}$	0.53	0.46	0.13
2	0.75	0.71	0.13	0.63	0.52	0.24^{+}	0.85	0.62	$0.60^{\square}$
3	0.43	0.39	0.07	0.40	0.43	−0.05	0.47	0.36	$0.17^{\square}$
4	0.63	0.55	0.20	0.65	0.53	0.25	0.48	0.41	0.13
5	0.85	0.77	$0.35^{\square}$	0.60	0.44	$0.28^{\square}$	0.72	0.50	$0.43^{\square}$
6	0.50	0.36	0.22	0.46	0.46	0.0	0.48	0.40	0.13
7	0.77	0.55	$0.48^{\square}$	0.63	0.57	0.15	0.72	0.45	$0.49^{\square}$
8	0.94	0.86	$0.58^{\square}$	0.85	0.51	$0.69^{\square}$	0.58	0.49	0.19
9	0.63	0.41	$0.37^{\square}$	0.65	0.57	0.18	0.45	0.38	0.11
10	0.83	0.68	$0.48^{\square}$	0.62	0.57	0.11	0.47	0.46	0.02

$^{+}P<0.05$, $^{\square}P<0.01$, Po = proportion of observed agreement, Pe = proportion of chance agreement, Kg = agreement beyond chance agreement.

Table 5. Inter-examiner reliability of assessment of three sacro-iliac anatomical landmarks (PSIS, SILA, SS) all examinations (1200) by all examiners (10)

	Number of findings for each site								
	PSIS			SILA			SS		
Subjects	L > R	R > L	R=L	L > R	R > L	R=L	L > R	R > L	R=L
1	7	7	26	4	22	14	4	20	15
2	7	5	28	2	15	23	4	5	31
3	1	18	21	6	10	24	11	8	21
4	2	12	26	9	15	16	11	15	14
5	3	7	30	4	8	28	12	6	22
6	1	2	37	11	4	25	15	11	14
7	10	8	22	3	22	15	15	5	20
8	4	17	19	3	17	20	18	5	17
9	5	13	22	0	13	27	11	6	23
10	3	11	26	4	28	8	11	8	21
Total	43	100	257	46	154	200	112	90	198
Po		0.51			0.46			0.39	
Pe		0.49			0.41			0.34	
Kg		0.04			0.08			0.07	
Seo		0.005			0.009			0.008	
P		<0.01			<0.01			<0.01	

Po=proportion of observed agreement, Pe=proportion of chance agreement, Kg=agreement beyond chance agreement, SEo=null standard error, *P*=significance beyond chance agreement.

DISCUSSION

Intra-examiner agreement revealed a range of reliability. The PSIS and SS yielded slight to moderate agreement, and the SILA yielded less-than-chance to substantial agreement. Examiners appear to have some measure of self-consistency that does not extend to agreement with each other, and this is consistent with previous palpation reliability studies (Mann et al., 1984; Keating et al. 1990; Boline et al. 1993). The fact that inter-examiner agreement was significant at the 0.01 level is inconsequential since there was no better than slight agreement beyond chance for all three landmarks. Conclusions regarding reliability should be weighted more heavily upon inter-examiner agreement (Haas 1991a). The results of this study fail to prove the reliability of static palpation of the PSIS, SILA or SS, as inter-examiner agreement beyond chance was very low and was reflected in kappa values that did not exceed 0.08. The lack of reliability locating sacroiliac landmarks may partially explain the poor reliability found in motion palpation studies of the sacroiliac joint (Potter & Rothstein 1985; Carmichael 1987; Hertzog et al. 1989; Bowman & Gribble 1995).

The greater agreement within examiners than between examiners may be due to systematic intra-examiner error. The source of this error can not be deduced from this study but it is possible that each examiner used their own reference point to identify the osseous anatomical landmarks despite training prior to testing. The training procedure was not intensive, and did not include any formal or informal comparison of results between examiners.

Inter-examiner reliability may be improved by training that includes discussion about findings, followed by retesting until greater agreement between examiners could be decided upon. This is relevant because inter-examiner agreement is less likely to suffer from the effect of systematic intra-examiner error (Haas 1991a).

Assessment of the SS involved a decision about tissue compliance based on the degree of indentation of the tissue over the site. The spatial pressure distribution over a contact region depends not only on the tissue compliance but also upon the force applied (Srinivasan & LaMotte 1995), and the degree of force influences the perceived amplitude of a movement thereby resulting in kinaesthetic errors known as force-movement illusions (Jones 1988; Jones & Hunter 1990). Subjects in this study reported that examiners used highly variable amounts of force during palpation. It is possible that different degrees of palpatory pressure used by examiners could contribute to systematic intra-examiner error and a reduction in inter-examiner agreement. The training procedure did not include instruction regarding palpation pressure beyond urging the examiners to be as gentle as possible. Obviously this instruction invites a highly variable and unpredictable response by the examiners. There is some evidence that palpation pressure biofeedback techniques may improve reliability (Jull & Bullock 1987; Bendtsen et al. 1995).

The variable range of intra-examiner agreement may reflect differing levels of palpatory ability among the group of 5th year osteopathic students. Two-point discrimination is one of the many physical components of the fingers that enables discrimination of tissue texture and shape during palpation. Two point discrimination of the fingers ranges from about 1.0 to 5.0 mm and varies depending on the number of specialized tactile receptors in any given region (Guyton 1991). Examiners may have varying sensory limits, or use different surfaces of their fingers for palpation and this could contribute to variability between examiners further investigation would be needed to establish this.

All subjects were asymptomatic and knowledge of this may have led to observer bias. Fifty-five percent of the findings were that the right and left sides were equal. Intra-examiner kappa statistics were low for some examiners despite having high agreement. For example, examiner 2 when examining the PSIS had an observed agreement of $Po = 0.75$, yet kappa was only 0.13. This was because most of the agreement fell into the right-equals-left category and Kappa becomes unstable when there is limited variability (Kramer & Feinstein 1981; Haas 1991a; Lantz & Nebenzahl 1996; Lantz 1997). Perhaps if a symptomatic subject cohort were used there would not be such over-representation of one category and Kappa would be more stable.

The subjects were required to lay prone for 2.5 hours. They were observed over this time and although they appeared to remain remarkably still there was little doubt that they would adjust their position. Small movements while prone may lead to changes in alignment of anatomical landmarks. Conclusions about the stability of palpatory cues can not be made from this study, but instability of palpatory cues could contribute to reduced intra- and inter-examiner agreement.

CONCLUSION

Intra-examiner reliability was greater than inter-examiner reliability, which is consistent with findings of previous palpation reliability studies. Static palpation of the PSIS, SILA, and SS yielded intra-examiner reliability ranging from moderate to less-than chance, and only slight inter-examiner reliability. As palpatory findings play a significant role in diagnosis, treatment, and analysis of the effect of treatment the variable range of intra-examiner agreement and the slight inter-examiner agreement beyond chance should be a cause for concern. The lack of reliability locating sacroiliac landmarks beyond chance may partially explain the poor reliability found in motion palpation studies of the sacroiliac joint. The reliability of palpation of anatomical landmarks needs to be validated. Motion palpation and the palpatory assessment of the effectiveness of treatment applied to alter motion restriction are predicated upon accurate localisation of anatomical landmarks. Information derived from palpation should be consistent within a practitioner and interpreted in a form that is transmissible to other practitioners in the field who can then identify comparable findings. As palpation is considered an important tool within manual medicine, further studies are required to determine why agreement on both static and motion palpatory findings remain poor and an effort should be made to identify means of significantly improving levels of agreement.

References

Basmajian JV, Nyberg R 1993 Rational Manual Therapies. Williams & Wilkins, Baltimore

Bendtsen L, Jensen R, Jensen NK, Olsen J 1995 Pressure-controlled palpation: a new technique which increases the reliability of manual palpation. Cephalgia 15: 205–210

Boline PD, Haas M, Meyer JJ, Kassak K, Nelson C, Keating JC 1993 Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality: Part II. Journal of Manipulative Physiological Therapies 166, 363–374

Bourdillon JF, 1970 Spinal Manipulation. Heinemann Medical Books Ltd, London

Bowen V, Cassidy D 1981 Macroscopic and microscopic anatomy of the sacroilliac joint from embryonic life until the eighth decade. Spine 66: 620–628

Bowman C, Gribble R 1995 The value of the forward flexion test and three tests of leg length changes in the clinical assessment of movement of the sacroiliac joint. The Journal of Orthopaedic Medicine 172: 66–67
Breen A 1992 The reliability of palpation and other diagnostic methods. Journal of Manipulative Physiological Therapies 15: 54–56
Burton AK, Edwards VA, Sykes DA 1990 Invisible skin marking for testing palpatory reliability. Journal of Manual Medicine 5: 27–29
Byfield DC, Mathiasen J, Sangren C 1992 The reliability of osseous landmark palpation in the lumbar spine and pelvis. European Journal of Chiropractic 40: 83–88
Carmichael JP 1987 Inter- and intra-examiner reliability of palpation for sacroiliac joint dysfunction. Journal of Manipulative Physiological Therapies 104: 164–171
Diek GS, Kelsey JL, Goel VK, Panjabi MM, Tech D, Walter SD, Laprade MH 1985 An epidemiologic study of the relationship between postural asymmetry in the teen years and subsequent back and neck pain. Spine 10: 872–877
DiGiovanna EL, Schiowitz S (Eds) 1991 An Osteopathic approach to Diagnosis and Treatment. JB Lippincott, Philadelphia
Dvorak J, Dvorak V 1990 Manual Medicine Diagnostics 2nd Edition. Thieme Medical Publishers, New York
Fleiss JL, 1971 Measuring nominal agreement among many raters. Psychological Bulletin 76: 378–382
Gonnella C, Paris SV, Kutner M 1982 Reliability in evaluating passive intervertebral motion. Physical Therapy 624: 436–444
Greenman PE 1996 Principles of Manual Medicine, 2nd edn. Williams & Wilkins, Baltimore
Guyton AC 1991 Textbook of Medical Physiology, 8th edn. WB Saunders, Philadelphia
Haas M 1991a Statistical methodology for reliability studies. Journal of Manipulative Physiological Therapies 14: 119–132
Haas M 1991b The reliability of reliability. Journal of Manipulative Physiological Therapies 14: 199–208
Herzog W, Read LJ, Conway PJW, Shaw LD, McEwen MC 1989 Reliability of motion palpation procedures to detect sacroiliac joint fixations. Journal of Manipulative Physiological Therapies 12: 86–92
Hoppenfeld S 1976 Physical Examination of the Spine and Extremities. Appleton & Lange, California
Jones LA 1988 Motor illusions: what do they reveal about proprioception? Psychological Bulletin 103: 72–86
Jones LA Hunter IW 1990 A perceptual analysis of stiffness. Experimental Brain Research 79: 150–156
Jull G, Bullock MA 1987 A motion profile of the lumbar spine in an ageing population assessed by manual examination. Physiotherapy 10: 1–4
Kappler RE 1980 A comparison of structural examination findings obtained by experienced physician examiners and student examiners on hospital patients. Journal of the American Osteopathic Association 79: 97–100
Keating JC, Bergmann TF, Jacobs GE, Finer BA, Larson K 1990 Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. Journal of Manipulative Physiological Therapies 13: 463–470
Koran LM 1975a The reliability of clinical methods, data and judgements. The New England Journal of Medicine: Part 1 293: 642–646
Koran LM 1975b The reliability of clinical methods, data and judgements. The New England Journal of Medicine: Part 2 293: 695–701
Kramer MS, Feinstein AR 1981 The biostatistics of concordance. Clinical Pharmacological Therapy 29: 111–123
Landis JR, Koch G 1977 The measurement of observer agreement for categorical data. Biometrics 33: 159–174
Lantz CA 1997 Application and evaluation of the kappa statistic in the design and interpretation of chiropractic clinical research. Journal of Manipulative Physiological Therapies 20: 521–528
Lantz CA, Nebenzahl E 1996 Behaviour and interpretation of the *K* statistic: resolution of the two paradoxes. Journal of Clinical Epidemiology 49: 431–434
Mann M, Glasheen–Wray M, Nyberg R 1984 Therapist agreement for palpation and observation of iliac crest heights. Physical Therapy 64: 334–338
McConnell DG, Beal MC, Dinnar U 1980 Low agreement of findings in neuromusculoskeletal examinations by a group of osteopathic physicians using their own procedures. Journal of the American Osteopathic Association 79: 59–68
Moir SA, McGregor M, Schut B 1990 The role of experience in clinical accuracy. Journal of Manipulative Physiological Therapies 13: 68–71
Montgomery DM, Egan IF, Pollard HP 1995 Palpable unilateral sacral prominence as a clinical sign of lower limb anisomelia: a pilot study. Journal of Manipulative Physiological Therapies 18: 353–356
Potter NA, Rothstein JM 1985 Intertester reliability for selected clinical tests of the sacroiliac joint. Physical Therapy 65: 1671–1675
Russell R 1983 Diagnostic palpation of the spine: a review of procedures and assessment of their reliability. Journal of Manipulative Physiological Therapies 6: 181–183
Simmonds MJ, Kumar S 1993 Health care ergonomics. Part 2: Location of body structures by palpation—a reliability study. International Journal of Industrial Ergonomics 11: 145–151
Srinivasan MA, LaMotte RH 1995 Tactual discrimination of softness. Journal of Neurophysiology 73: 88–101
Taylor JR, Slinger BS 1980 Scoliosis screening and growth in western australian students. The Medical Journal of Australia 1: 475–478
Vleeming A, Stockart R, Volkers ACW, Snijders CJ 1989 Relation between form and function in the sacroiliac joint. Part 1: Clinical anatomical aspects. Spine 15: 130–131
Ward RC (Ed) 1996 Foundations for Osteopathic Medicine. Williams & Wilkins, Baltimore
Wiles MR 1980 Reproducibility and inter-examiner correlation of motion palpation findings of the sacroiliac joints. The Journal of the Canadian Chiropractic Association 24: 59–69

