Inter-examiner reliability of four static palpation tests used for assessing pelvic dysfunction

Ulrika Holmgren*, Kerstin Waling

Department of Community Medicine and Rehabilitation, Physiotherapy Umeå University, Umeå, Sweden

Accepted 30 August 2006

Abstract

In muscle energy technique (MET), palpation is an important tool aimed at detecting asymmetry and selecting interventions. The aim of this study was to test the inter-examiner reliability of static palpation of the transverse processes of L5 (L5), sacral sulci (SS), inferior lateral angles of the sacrum (ILA), and the medial malleoli (MM) in a clinical setting. Twenty-five participants, aged 18–78 years, with low back pain and/or sacroiliac pain with or without radiating pain to the knee, were independently examined by two experienced physiotherapists. For L5, SS, ILA, the proportion of observed agreement was 40–44% and the κ coefficient 0.11 (SE = 0.12) to 0.17 (SE = 0.10). For MM, the observed agreement was 52% and the κ coefficient 0.28 (SE = 0.15). Differences in palpation technique seem to be the most likely source of the low inter-examiner reliability in this study. For clinical practise, continued use of these tests as methods for detecting asymmetry and selecting interventions is of doubtful utility.

Keywords: Inter-tester reliability; Physical therapy; Reproducibility; Sacroiliac joint

1. Introduction

In the last two decades, researchers have concluded that the sacroiliac joint may be one cause of low back pain (Walker, 1992; Daum, 1995; Schwarzer et al., 1995; Dreyfuss et al., 1996; Maigne et al., 1996; Slipman et al., 1996; Laslett et al., 2003; Laslett et al., 2005; van der Wurff et al., 2006). The prevalence of the sacroiliac joint as a source of low back pain is reported to be between 13% and 48% in different studies (Daum, 1995; Schwarzer et al., 1995; Maigne et al., 1996; Touissant et al., 1999a, b).

Manual examination is a standard method of examining the sacroiliac joint (Touissant et al., 1999a, b). Some manual tests are designed to assess the location of and relative symmetry between the left and right side of the joint (static palpation), or to assess movement of bony landmarks associated with the sacroiliac joint (motion palpation). Application of force to the sacroiliac joint or related structures in an attempt to reproduce pain is another kind of test (pain provocation) (Walker, 1992; Simmonds and Kumar, 1993; Laslett et al., 2003, 2005; Meijne et al., 1999; van der Wurff et al., 2000, 2006; Freburger and Riddle, 2001). Tests that fall into these three categories are typically performed without assistance of instrumentation (van der Wurff et al., 2000; Freburger and Riddle, 2001). Attempts to use X-ray, bone scan or laboratory tests in investigating the sacroiliac joint have not been successful (Dreyfuss et al., 1996; Slipman et al., 1996; Tullberg et al., 1998).

Since tests of the sacroiliac joint are often based on palpation, accurate and reliable palpation skills are a prerequisite for them to be considered potentially valuable. Both static palpation tests (Potter and Rothstein, 1985; Lindsay et al., 1995; O’Haire and Gibbons, 2000) and motion palpation tests (Potter and Rothstein, 1985; Meijne et al., 1999; Vincent-Smith and Gibbons, 1999) in the...
pelvic region have to date failed to demonstrate acceptable levels of both inter- and intra-examiner reliability.

In muscle energy technique (MET) palpation is an important diagnostic tool (Chaitow, 2003). Bilateral palpation of the transverse processes of L5 (L5), the sacral sulci (SS), and the inferior lateral angles of the sacrum (ILA) are tests used to detect asymmetry and dysfunction in the pelvic region. Studies examining the reliability of pelvic tests used in MET are scarce, and exclusively performed on asymptomatic subjects (O’Haire and Gibbons, 2000). O’Haire and Gibbons (2000) evaluated the inter- and intra-examiner reliability in static palpation of the posterior superior iliac spine (PSIS), the SS, and the ILA. For all three landmarks, the results indicated an inter-examiner agreement only slightly better than what was expected by chance. The intra-examiner reliability was somewhat higher than the inter-examiner reliability, a result consistent with other studies (Meijne et al., 1999; Vincent-Smith and Gibbons, 1999). The authors suggested that difficulties in locating anatomical landmarks and the use of asymptomatic subjects might have reduced the reliability (O’Haire and Gibbons, 2000). In a study by Downey et al. (1999), the ability to palpate lumbar spinous processes was evaluated. Three pairs of manipulative physiotherapists presented almost perfect inter-examiner reliability in locating the nominated spinal levels. Studies evaluating the inter-examiner agreement in bilateral palpation of the transverse processes of L5 have to our knowledge not been published.

As functional leg length discrepancy is believed to appear due to for example pelvic distortions (Walker, 1992; Chaitow, 2003), leg length evaluation often accompanies the pelvic examination in MET (Chaitow, 2003). Leg length can be assessed by comparing the positions of the medial malleoli (MM) in a prone position. To our knowledge, no study has been presented assessing the inter-examiner reliability of static palpation of the MM in a prone position.

Since the results of the four mentioned tests (L5, SS, ILA, MM) are important in the selection of intervention techniques (Weiselfish-Giammatteo and Giammatteo, 2003), we felt it important to study their reliability in a clinical setting. Consequently, the aim of this study was to test the inter-examiner reliability of three static palpation tests used to detect asymmetry in the pelvic region (L5, SS, ILA), and one static palpation test to evaluate leg length (MM) in a symptomatic population.

2. Material and method

2.1. Participants

For 4 months, 27 patients with symptoms of low back pain and/or sacroiliac pain at two primary health care clinics were consecutively recruited to participate in the study. To be included they should not have radiating pain below the knee or have been subject to lumbar surgery during the last year. They had to be 18 years or older, and be able to lay prone for 10 min.

Two potential participants (a woman at one clinic and a man at the other clinic) failed to come at the time appointed leaving totally 25 patients to the study. Of these 18 were women and seven men. The mean age was 46 years (range 18–78 years), mean height 170 cm (range 156–187 cm) and mean weight 76 kg (range 56–99 kg). During the 12 months preceding the study, 14 of the participants had experienced daily pain. Nine had not been able to work full time or fulfil their domestic duties, and 13 had reduced their spare time activities. Twenty participants reported pain during the last 7 days. On the day of the assessment, the participants were asked to rate pain intensity on a visual analogue scale (VAS) ranging from 0 = no pain to 100 = worst possible pain (Huskisson, 1974; Carlsson, 1983). The ratings ranged between 0 and 70 with a group average of 29.

The participants were given both verbal and written information about the study and gave their written consent to participate. All participants were informed that they were free to withdraw from the study at any time. The managers of the two primary health care clinics where the study took place and the Ethics Committee of the Medical Faculty of Umeå University (Ref. no. 03–351) approved the study.

2.2. Examiners

The examiners consisted of three female physiotherapists (PT) (A, B and C), who were working in primary health care. All three PTs had around 15 years of experience of patients with low back pain and sacroiliac joint dysfunction. They were all trained in MET and had practised the technique for more than 10 years.

2.3. Procedure

The test order, the performance of the tests, and the position of patients and examiners were determined prior to the study. The tests were performed in the following order: L5, SS, ILA, MM.

PT A, also one of the authors (UH), was one of the examiners during all test sessions. Depending on which of the two clinics the patient belonged to, one of the other two PTs (B or C) was the second examiner. Tossing a coin randomized the order of which PT first examined each new participant.

The test procedure was performed in a true clinical setting on the participant’s initial visit to the physiotherapy clinic. Each participant underwent two examination sessions with a different examiner each time. The participants were instructed to lie in a prone position...
with lowered arms, looking down through a hole in the examination table (Fig. 1). They were told not to move until both PTs had completed their examinations. In addition, they were instructed not to inform the second examiner of the results of the first examination. The first PT palpated and observed the stated anatomical landmarks, filled in an assessment form, put the form in an envelope, sealed the envelope and left the examination room. The second PT entered the room within 2 min and performed the same procedure as the first PT. The two envelopes remained sealed until all participants were examined.

2.4. Assessments

During all examinations the PTs were instructed to stand on the right-hand side of the examination table while performing static palpation of L5, SS, and ILA. While performing static palpation of MM, the PTs stood at the end of the examination table. The PTs were instructed to palpate gently in order to feel the beginning of a sense of resistance in the structure, and to assess if the anatomical landmarks were symmetrical or not (Chaitow, 2003; Weiselfish-Giammatteo and Giammatteo, 2003). If necessary, they used their dominant eye to evaluate the results (Weiselfish-Giammatteo and Giammatteo, 2003).

To palpate the transverse processes of the L5, the PTs first located the L5 spinous process. From the L5 spinous process the thumbs were moved laterally over the area of the transverse processes to assess if one side appeared to be more posterior (Figs. 2 and 3). The SS were examined by placing the pads of the thumbs on both PSIS, followed by moving the thumbs medially into the SS to assess any differences in depth of the right and left side (Fig. 3). To examine the ILA, the pads of the thumbs were placed laterally to the sacral hiatus on the posterior aspects of the ILAs to assess whether one side appeared more posterior compared to the other side (Weiselfish-Giammatteo and Giammatteo, 2003) (Fig. 3). The leg length was examined by placing the participant’s legs side by side, followed by the examiner placing their thumbs at the medial border of the MM. The PTs compared visually if one leg appeared to be longer (Fig. 4).

The possible findings for each test are summarized in Table 1.

2.5. Statistical methods

The Statistical Package for the Social Sciences (SPSS) version 9.0 was used to calculate the percentage agreement between therapists (%), the Kappa agreement coefficient (\( \kappa \)), its standard error (SE), its 95% confidence interval (CI), and its significance (\( p \)). The observed proportion of positive agreement (\( P_{pos} \)) and the observed proportion of negative agreement (\( P_{neg} \)) between the therapists were calculated using Excel 6.0.

The \( \kappa \) coefficient evaluates inter-examiner agreement of categorical data and determines the proportion of agreements, which is expected by chance contrary to the percentage agreement, which is not chance corrected. The \( \kappa \) coefficient ranges in value from \(-1\) to \(+1\) where positive values signify agreement better than chance, a
value 0 denotes agreement no better than chance and a negative value signifies agreement worse than chance. According to the guidelines proposed by Landis and Koch (1977), the strength of agreement between 0 and 1 can be interpreted as follows: 0.00–0.20 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, 0.81–1.00 almost perfect agreement. A $\kappa$ value that reaches 0.40 or better is considered to be acceptable for clinical use.

3. Results

The inter-examiner reliability of examination of the four static palpation tests in all 25 participants is presented in Table 2. The proportion of observed
agreement for L5, SS, and ILA, varied from 40% to 44%, and the κ coefficients for the same tests varied from 0.11 (SE = 0.12) to 0.17 (SE = 0.10) indicating an agreement slightly better than what was expected by chance. For the MM, the proportion of observed agreement was somewhat higher (52%), and the κ coefficient was also higher (0.28, SE = 0.15) indicating fair agreement.

The proportion of positive agreements varied from 0% to 34%, and the proportion of negative agreements varied from 8% to 36%, indicating that the examiners disagreed frequently on both positive and negative findings.

### 4. Discussion

The results indicated poor reliability for all four tests, a result consistent with other static palpation studies (Potter and Rothstein, 1985; Lindsay et al., 1995; O’Haire and Gibbons, 2000). In the study we included all subjects that for a period of 4 months, consulted the PTs at two primary health clinics for low back pain and/or sacroiliac pain, with or without radiating pain to the knee. A broad inclusion criterion of reported low back and/or sacroiliac pain was chosen to make sure that the tests were evaluated on participants on whom the tests normally would be conducted routinely as part of their...
clinical evaluation (Domholdt, 2000; Fritz and Wainner, 2001). The majority of women only reflect the present gender distribution in primary health care. The study was performed in a clinical setting and designed to mirror the clinical situation where a physiotherapist at times uses colleagues for a second opinion or where different PTs meet the patient from one occasion to another.

This study showed that the inter-examiner reliability in a symptomatic population was just as low as in studies of asymptomatic populations (Lindsay et al., 1995; Vincent-Smith and Gibbons, 1999; O’Haire and Gibbons, 2000). The proportion of percentage agreement was in most cases less than 50%, and the $\kappa$ values for L5, SS and ILA indicated a result only slightly better than what was expected by chance. For the MM, the result was somewhat better but still not acceptable for clinical use. Regarding inter-examiner reliability the result of the present study must be considered unacceptable, especially in the light of the $P_{\text{pos}}$ and $P_{\text{neg}}$ values.

Standardization of measurement techniques is claimed to be the best way to improve reliability (Domholdt, 2000; Fritz and Wainner, 2001). In the present study, efforts were made to have a highly standardized but at the same time clinically practical and realistic test situation. Thereby the risk of bias due to the testing conditions was minimized. The distinct description of the palpation technique in MET possibly contributed to improved reliability in contrast to other palpation studies where therapists have been permitted to use their own methods of palpation (Lindsay et al., 1995, O’Haire and Gibbons, 2000). However, during the examination some of the participants spontaneously reported that the force used differed between therapists. Variation in examination technique, such as applied force and timing, has been shown to occur both within and between therapists irrespective of experience (Levin et al., 2005), and could influence both the test outcome and the reliability. Since all three PTs had similar experiences of the method MET and we wanted to resemble a clinical situation, no attempt to synchronize the examiners by training sessions was performed. According to Simmonds and Kumar (1993), not synchronizing the examiners may result in poorer reliability, but perhaps it gives a more clear indication of clinical practise. Keating et al. (1993), have shown that it is possible to train therapists to recognize and communicate the amount of force during palpation, but studies utilizing training programmes to synchronize examiners prior to the investigation have reported both high (Laslett and Williams, 1994; Kokmeyer et al., 2002) and low levels of reliability (Vincent-Smith and Gibbons, 1999; O’Haire and Gibbons, 2000; Riddle and Freburger, 2002).

Previously it has been suggested that findings of pelvic tests could be more stable in symptomatic participants compared to asymptomatic ones (Vincent-Smith and Gibbons, 1999). The time between the two examinations in our study was less than 2 min, which was considered to be sufficient to prevent the participants from moving. However, the second examiner in our study more often reported positive findings. Small movements when examining the MM, and/or increasing pain due to the prone position, or from the first examination, may have led to changes in the soft tissue texture and contributed to different assessments by the examiners. The low reliability could also be a result of a systematic difference between examiners in their evaluation, but perhaps the most likely explanation is that the structures palpated lay too far below the surface to be palpated with any kind of accuracy.

It is important to note that the results of our study are not applicable to other conditions than those presented here, and that the results should be interpreted with caution. In this study, the inter-examiner reliability is evaluated, and reliability is only a measure of reproducibility, not a gauge of accuracy. However, the results of this study clearly show that the four studied tests performed by experienced examiners and in a clinical setting have poor or limited reliability. Consequently, continued use of these tests as methods for detecting asymmetry and selecting interventions is doubtful and should be reconsidered, especially since a selection of pain provocation tests (distraction test, thigh thrust test, Gaenslen’s test, compression test and sacral thrust test) has been able to show both acceptable reliability and validity and can also be used to identify the sacroiliac joint as a source of low back pain (Laslett and Williams, 1994; Kokmeyer et al., 2002; Laslett et al., 2003; Young et al., 2003, Laslett et al., 2005; van der Wurff et al., 2006).

5. Conclusion

The static palpation of L5, SS, ILA, and MM performed by experienced examiners in symptomatic subjects demonstrated an inter-examiner reliability only slightly better than what was expected by chance. Differences in palpation technique seem to be the most likely source of the low inter-examiner reliability in this study. For clinical practise, continued use of these tests as methods for detecting asymmetry and selecting interventions is of doubtful utility.

Acknowledgements

We thank the men and women who consented to participate in the study and the physiotherapists Berith Joelsson and Anette Persson for their valuable help with the study.
The study was approved by the Ethical Committee of the Medical Faculty of Umeå University, Umeå, Sweden (§437/03, ref. no. 03-351).

References


Downey BJ, Taylor NF, Niere KR. Manipulative physiotherapists can reliably palpate nominated lumbar spine levels. Man Ther 1999;4:151–6.


Laslett M, Aprill CN, McDonald B, Young SB. Diagnosis of sacroiliac joint pain: validity of individual provocation tests and composites of tests. Man Ther 2005;10:207–18.


