



Published in final edited form as:

J Am Osteopath Assoc. 2010 November ; 110(11): 667–674.

Reliability of Bony Anatomic Landmark Asymmetry Assessment in the Lumbopelvic Region: Application to Osteopathic Medical Education

Bradley A. Stovall, OMS IV and Shrawan Kumar, PhD

Department of Osteopathic Manipulative Medicine at the University of North Texas Health Science Center—Texas College of Osteopathic Medicine in Fort Worth

Abstract

The objective of this review is to establish the current state of knowledge on the reliability of clinical assessment of asymmetry in the lumbar spine and pelvis. To search the literature, the authors consulted the databases of MEDLINE, CINAHL, AMED, MANTIS, Academic Search Complete, and Web of Knowledge using different combinations of the following keywords: *palpation, asymmetry, inter- or intraexaminer reliability, tissue texture, assessment, and anatomic landmark*. Of the 23 studies identified, 14 did not meet the inclusion criteria and were excluded. The quality and methods of studies investigating the reliability of bony anatomic landmark asymmetry assessment are variable. The κ statistic ranges without training for interexaminer reliability were as follows: anterior superior iliac spine (ASIS), -0.01 to 0.19 ; posterior superior iliac spine (PSIS), 0.04 to 0.15 ; inferior lateral angle, transverse plane (ILA-A/P), -0.03 to 0.11 ; inferior lateral angles, coronal plane (ILA-S/I), -0.01 to 0.08 ; sacral sulcus (SS), -0.4 to 0.37 ; lumbar spine transverse processes L1 through L5, 0.04 to 0.17 . The corresponding ranges for intraexaminer reliability were higher for all associated landmarks: ASIS, 0.19 to 0.4 ; PSIS, 0.13 to 0.49 ; ILA-A/P, 0.1 to 0.2 ; ILA-S/I, 0.03 to 0.21 ; SS, 0.24 to 0.28 ; lumbar spine transverse processes L1 through L5, not applicable. Further research is needed to better understand the reliability of asymmetry assessment methods in manipulative medicine.

Historically, manipulative medicine in the United States has been associated primarily with the osteopathic and chiropractic professions.^{1,2} Central to manipulative medicine in these professions is the importance of the musculoskeletal system in health and disease.^{1–3} To assess for musculoskeletal dysfunction, the clinician obtains a history, performs a physical examination, and, if indicated, conducts more specific evaluation by palpating for biomechanical abnormalities.^{2,3} In the spine and pelvis, a wide range of palpatory approaches has been indicated for detecting dysfunction suitable for manipulative intervention.^{2–6} In the osteopathic and chiropractic professions, bony anatomic landmark asymmetry is considered an important sign but not always the defining component of musculoskeletal dysfunction.^{2,3}

In osteopathic medicine, tenderness, asymmetry, restricted range of motion, and tissue texture change make up the commonly used mnemonic *TART* or *STAR*, in which the *S* refers

Address correspondence to Shrawan Kumar, PhD, University of North Texas Health Science Center—Texas College of Osteopathic Medicine, 3500 Camp Bowie Blvd, Fort Worth, TX 76107-2644. shrawan.kumar@unthsc.edu.

Publisher's Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the American Osteopathic Association, the National Center for Complementary and Alternative Medicine, or the National Institutes of Health.

Financial Disclosures: None reported.

to *symptom reproduction*.^{2,4,7-9} It has been suggested that bony asymmetry and these associated findings represent musculoskeletal dysfunction. The degree to which findings of tissue texture change or tenderness are associated with asymmetry is intermittently discussed in the leading osteopathic texts, with little consensus about specific findings.^{2,4,7-9} In the chiropractic profession, the mnemonic *PARTS* is similarly used to represent characteristics of joint dysfunction: pain or tenderness; asymmetry or alignment; range of motion abnormality; tone, texture, or temperature of soft tissues; and special tests.^{3,10} With these approaches, a positive finding of bony anatomic landmark asymmetry does not by itself necessarily identify musculoskeletal dysfunction. It is the collection of findings that guides identification of musculoskeletal dysfunction suitable for manipulative intervention.

Bony anatomic landmark asymmetry is hypothesized to give information on the relative positions of the structures in question. In this context, the role of asymmetry makes intuitive sense. Combining bony anatomic landmark asymmetry findings with motion test findings allows the clinical picture of sacroiliac joint, pelvic, and lumbar dysfunction to be assessed and more specific interventions to be implemented.^{1-4,6-9} This clinical picture, however, is based on the assumption that clinically significant displacement occurs and is detectable. Radiographic and biomechanical evidence has demonstrated that definite but small degrees of relative displacement occur at the sacroiliac joint.¹¹⁻¹⁵ Tullberg et al¹¹ evaluated the role of sacroiliac joint manipulation in patients with known sacroiliac joint pain and implanted metal markers. Pre- and posttreatment stereophotogrammetric radiography demonstrated no significant change in relative displacement at the joints. In another experiment, no clinically significant association was found between pelvic asymmetry and unilateral nonspecific low back pain.¹⁶ More recently, however, increased pelvic asymmetry has been associated with statistically significant differences of coupled lumbar motion asymmetry in lateral flexion and axial rotation. This motion asymmetry enabled differentiation between asymptomatic and symptomatic patients with low back pain.^{17,18}

Importance of Reliable Palpation

For the results of a specific palpatory test to provide useful information that can be communicated between clinicians, that test must have acceptable interexaminer reliability. However, although such reliability is important, it does not necessarily define the clinical usefulness of a test or method of assessment.¹⁹ For instance, although cardiac auscultation has not been clearly demonstrated to have adequate interexaminer reliability,^{20,21} it is still used in clinical practice along with history taking, physical examination, and other diagnostic testing to assess clinical presentation in a patient encounter.¹⁹ As pointed out by Joshua et al,²² there is a need for improved understanding in many forms of clinical examination.²² From this perspective, palpation for bony anatomic landmark asymmetry is similar to cardiac auscultation, highlighting the need for critical analysis of methods used in clinical examination.

Most methods of palpatory examination in manipulative medicine lack a reference standard, which further increases the importance of reliability.¹⁹ In this context, if two examiners can be demonstrated experimentally to consistently agree about diagnostic findings, there is a strong likelihood that what is being tested for exists.¹⁹ Therefore, demonstrating the reliability of palpatory methods in manipulative medicine is of primary importance.

Evaluating Reliability Research in Manipulative Medicine

A variety of statistical methods were used in early reliability experiments in manipulative medicine.^{19,23,24} The current standard of reliability analysis is the κ statistic, which has gained widespread acceptance as a measure of reliability because it accounts for the role of chance in dichotomous agreement.^{23,24} As pointed out in a recent literature review²⁵ of

manual examination in the spine and updated reliability criteria, there are limitations to the use of this statistic.^{15,25} Prevalence index and selection of patient population are two important elements that must be carefully addressed when making a qualitative interpretation of κ values, because they can lead to falsely low or high κ values.¹⁵

Two methods currently exist to assess the quality of reliability studies in manipulative medicine.^{15,25} Stochkendahl et al²⁵ devised a 6-point system based on clinical research guidelines to assess the method quality of reliability studies. These criteria were used in our review and can be found in Figure 1. More recently, the Scientific Committee of the International Academy for Manual/Musculoskeletal Medicine (IAMMM)¹⁵ developed a 78-point system for reviewing reliability studies. This protocol evaluates four primary domains in analysis of reliability studies: study design, study population, diagnostic procedures, and data analysis and presentation.¹⁵ We did not use this protocol in the present study because it does not clearly define analysis of asymmetry assessment. Although to our knowledge this protocol has not yet been used in literature analysis, it provides a useful guide for developing and assessing future reliability studies.

Attempts to experimentally evaluate the reliability of musculoskeletal assessment have explored various assessment methods, such as motion testing and pain provocation testing of the sacroiliac joint. For motion testing and pain provocation testing, current literature reviews exist.²⁶⁻²⁹ However, research into assessment of anatomic landmark asymmetry as defined in manipulative medicine is relatively new, and investigation into this area has evolved substantially in the past 10 years. Our initial literature search identified two systematic literature reviews^{25,26} assessing the reliability of a variety of palpatory methods, including static palpation. While completing the present review, we identified a comprehensive literature review specific to assessment of static anatomic landmark asymmetry,³⁰ which provides a broad overview of research investigating the reliability of static segmental identification and assessment of bilateral anatomic landmark asymmetry in the entire spine and pelvis. Our review incorporates two more recent studies and serves as a more focused complement to this recently published review. Thus, our goal was to analyze and assess the quality of current research investigating the reliability of assessments for bony anatomic landmark asymmetry in the lumbar spine and pelvis. We also sought to discuss these findings in the context of osteopathic medical education.

Methods

We searched the MEDLINE, CINAHL, AMED, MANTIS, Academic Search Complete, and Web of Knowledge literature databases by using multiple combinations of the following keywords: *palpation*, *asymmetry*, *inter- or intraexaminer reliability*, *tissue texture*, *assessment*, and *anatomic landmark*. Inclusion criteria were as follows: studies that assessed reliability of palpatory tests for static asymmetry of L1 through L5 transverse processes or the pelvic or sacral anatomic landmarks, experimental studies in which the design was appropriate to answer the raised question(s), and studies published in peer-reviewed scientific journals.

Excluded were studies that did not experimentally evaluate the reliability of assessments for bony anatomic landmark asymmetry, evaluated only the reliability of segmental identification of spinal levels, evaluated only iliac crest asymmetry assessment, or provided opinions without an experimental framework.

Search and Identification of Articles Reviewed

The initial search identified 40 articles of interest. Initial screening for subject relevance reduced the list to 18 articles. Of that group, 6 articles did not meet inclusion criteria

because of factors identified above. Reviewing the references of all articles after the initial screen revealed 5 further articles of interest that did not meet inclusion criteria. Thus, 23 studies were identified, 9 of which met inclusion criteria (Table).

Assessment Techniques

The posterior superior iliac spine (PSIS), inferior lateral angles (ILAs), and sacral sulcus (SS) were the posterior pelvic landmarks assessed in the reviewed studies. The ILAs are assessed for asymmetry in the transverse (ILA-A/P [anterior/posterior]) and coronal (ILA-S/I [superior/inferior]) planes. Some authors differentiate the SS and the sacral base in palpatory assessment, but it is our opinion that there is no clinically palpable difference between the landmarks.^{1,2,40,41} For the purpose of this review, “SS” will be used to describe the sacral base, as in some of the reviewed articles. The SS is evaluated by assessing depth medial to the PSIS through palpation rather than visual assessment. This “landmark” was included because historically it has been assumed to give information regarding sacral position. This assumption has since been challenged, and the finding seems to correspond more directly to the multifidus or soft-tissue density than to sacral position.⁴¹

The only landmark evaluated in the supine position was the anterior superior iliac spine (ASIS). In muscle energy technique, the ASIS is assessed for asymmetry in the transverse, coronal, and sagittal planes.^{4,42,43} Only research examining ASIS symmetry in the coronal plane (superior/inferior displacement) was found.^{31,33,34–39} Lumbar transverse processes were evaluated by palpation for asymmetry in various planes and in variations of flexion, extension, and neutral position.^{35,36}

Quality Assessment of Selected Studies

The 6-point quality assessment scale was implemented in the analysis (Figure 1). The first author performed assessments, and the second author reviewed the analysis. The modified 4-point intraexaminer reliability scores were not included in the Table, because the emphasis of this study was to assess the overall quality of reliability studies.

Results

Results of quality analysis using the 6-point scoring system ranged from 2 (33%) to 6 (100%). The 6-point scores are shown with κ values in the Table. The ranges for κ statistics without training for interexaminer reliability were as follows: ASIS, -0.01 to 0.19 ; PSIS, 0.04 to 0.15 ; ILA-A/P, -0.03 to 0.11 ; ILA-S/I, -0.01 to 0.08 ; SS, -0.4 to 0.37 ; and lumbar spine transverse processes L1 through L5, 0.04 to 0.17 . Ranges for intraexaminer reliability were higher for all associated landmarks: ASIS, 0.19 to 0.4 ; PSIS, 0.13 to 0.49 ; ILA-A/P, 0.1 to 0.2 ; ILA-S/I, 0.03 to 0.21 ; SS, 0.24 to 0.28 ; and lumbar spine transverse processes L1 through L5, not applicable (none of the reviewed studies reported intraexaminer reliability of lumbar spine transverse process asymmetry assessment). Results are summarized in the Table.

Comment

Criteria for reliability studies of clinical tests suggest using both symptomatic and asymptomatic patient populations.^{44,45} The goal is to provide a more accurate spread of the data in question. This is thought to be important, because the prevalence of positive findings is assumed to be higher in a symptomatic patient population. However, as Kmita and Lucas³¹ pointed out, the prevalence of bony anatomic landmark asymmetry in these populations has not been defined. To address the need for a balanced patient population, those authors selected a population that included symptomatic and asymptomatic patients

with low back pain. To our knowledge, this was the only study identified that evaluated a mixed population. Also supporting the use of symptomatic and asymptomatic populations was the assumption that examiners could be biased if they were aware that the population was only symptomatic or only asymptomatic. Other experiments using only symptomatic patient populations, however, have not demonstrated increased reliability for clinical tests of static asymmetry.^{32,33,39}

In addition to subject selection criteria, another important component of reliability studies is inclusion of an adequate number of subjects and examiners. Current guidelines suggest that the optimum number of subjects for reliability experiments is 25 to 40 subjects and that the optimum number of examiners is 2. The reviewed experiments demonstrated a wide range in numbers of examiners and subjects; the number of subjects ranged from 9 to 77, and the number of examiners from 2 to 10. Only two studies^{32,38} met the criteria for subject number, and three studies^{32,33,38} met the criteria for examiner number. All reviewed experiments used examiners with substantial training and varying levels of clinical experience (Table). Given the variation in subject and examiner number in the reviewed experiments, the ability to compare studies directly is limited.

The number of tests evaluated in a reliability study is also important in study design. Recently updated guidelines recommend studying only one test at a time for experiments designed to assess the reliability of a specific test. O'Haire and Gibbons³⁷ performed an experiment in which posterior pelvic and sacral landmarks were evaluated by 10 fifth-year Australian osteopathy students. One of the challenges noted was the 1200 examinations required of each examiner, which increased the chance for examiner fatigue. This challenge, coupled with the amount of time subjects were required to lie prone on the examination table, was a major experimental confounder. In the clinical setting, asymmetry assessment in the pelvis requires multiple assessments to determine the side of dysfunction and final diagnosis.^{2-4,7,9} Although a single assessment of asymmetry is not typical of clinical models of asymmetry assessment, only the study by Spring et al³⁶ and a study of medial malleolus asymmetry⁴⁶ have investigated isolated evaluation of one landmark for asymmetry.

Along with limiting the number of tests and having adequate numbers of subjects and examiners, the calibration of specific techniques is another important element of reliability research. Two of the reviewed studies predetermined what constitutes a positive finding.^{31,34} The experiment by Fryer et al³⁴ initially had the examiners reach an agreement on how much asymmetry constituted a positive finding, "approximately" 3 mm. The other experiment set 3 mm as a positive finding.³¹ Implementing such a threshold, however, introduces another confounder, because the method of doing so was not adequately described or discussed in any of the studies. There is no known ability to accurately determine amounts of asymmetry in vivo, and it is unlikely that the standardization sessions provided a mechanism for substantially improving the reliability of assessments. This finding was further confirmed by the results of these two experiments.

For the evaluation of low κ values in reliability research, the prevalence index has been suggested to be of primary importance.¹⁵ The prevalence index is defined as the frequency of positive judged tests in the study population.¹⁵ In applying the prevalence index in reliability analysis, only one palpatory test with dichotomous findings can be used.¹⁵ In assessments for asymmetry of bony anatomic landmarks in the pelvis, however, three options are often given to examiners: right side greater than left side, both sides equal, and left side greater than right side. Six studies offered three options in assessment.^{31-34,37,39} Examiners were given two to six options for assessing the lumbar spine but only one to evaluate for dichotomous asymmetry in the pelvis.^{35,36,38} Thus, the role of the prevalence index in the assessment of asymmetry is unclear.

When the prevalence index is incorporated into studies, a calibration session is needed to attain more than 85% agreement before conducting the experiment. This allows the prevalence index to be within optimum range to interpret the κ statistic. Therefore, qualitative comparison of κ values is limited for studies that do not provide information about the prevalence index of a given test. Readers are referred to the IAMMM guidelines for reliability studies for further discussion of the prevalence index.¹⁵ The prevalence of findings was determined in two studies, but no further mention of associated data could be found in these studies. The absence of this measure in reliability research highlights an important area for improvement.

Training

Two experiments examined the role of formal training programs with the goal of improving reliability.^{34,35} Two methods of training were implemented. One experiment used two groups of five examiners, with one group receiving training over two 1-hour training sessions with an experienced clinician.³⁴ During these training sessions, technique was standardized, and examiners evaluated and compared assessments. Interestingly, examiners demonstrated substantial improvements in reliability after training for some landmarks but not for others.³⁴ Assessments of the ASIS have consistently had higher reliability than assessments of other landmarks in the pelvic girdle, and they improved after training.³⁴ This improvement was not enough to reach statistical significance, but the change suggests that training changed something about how asymmetry was assessed in the trained group.

Another form of training introduced by Degenhardt et al³⁵ consisted of “consensus training,” which was focused on training examiners to perform assessments in as similar a manner as possible. This method of training was developed to better understand the known low interexaminer reliability of palpatory assessment. When results differed between examiners, the researchers compared evaluation techniques and attempted to understand the reason for observed differences.³⁵ The results of this experiment demonstrated statistically significant increases in interexaminer reliability after this method of training.

These results, however, must be interpreted with caution because of methodologic inconsistencies and nonstandardized changes. In the final two phases of the experiment, patients did not change positions during assessment until all examiners had evaluated them. This was different from phase 1 in that all assessments, which included active and passive motion testing, were conducted before another examiner made assessments. Although these limitations affect interpretation, having examiners with a high consensus as to what constitutes a positive finding is probably significant in musculoskeletal palpatory assessment.

In recent literature reviews of palpatory methods in manipulative medicine, intraexaminer reliability has consistently been higher than interexaminer reliability.^{25,26,28–30} In the studies we reviewed, this remained true except in one experiment.³⁶ Examiners intuitively can be expected to agree with themselves more than with other examiners, but the reasons for this commonly observed phenomenon are not well understood. The accuracy and consistency of finger placement as well as visual cues have been suggested to play a role in low interexaminer reliability.^{34,37} Although standardized training sessions were used in two experiments, the accuracy of finger placement for individual examiners was not specifically addressed.^{31,34}

Limitations

To date, minimal research has been conducted into methods of bilateral asymmetry assessment in manipulative medicine. The varied methods in the literature limit the

comparison of reliability studies. The quality assessment method that was used did not take prevalence of findings into account. Our review seeks to clarify the role of static asymmetry assessment of lumbar spine and pelvic landmarks, but owing to the varying sizes of these landmarks and differences in methods, the ability to make direct comparisons will remain limited until future studies explore this method of asymmetry.

Critical Thinking and Osteopathic Medical Education

Not all osteopathic medical students begin their education with a desire solely to practice manipulative medicine. However, most students approach their initial osteopathic manipulative medicine (OMM) courses with a genuine intellectual curiosity about OMM techniques. These students learn about Fryette's laws, pelvic and sacral mechanics, and concepts of palpatory diagnosis and treatment throughout their first 2 years of education. These concepts are often taught as basic facts with little discussion of validity or reliability, but there is growing evidence that some concepts are invalid or at least incompletely understood.⁴⁷⁻⁵² This fact marks one of the single greatest challenges facing osteopathic medical education today: how do educators best teach traditional, clinical models of palpatory diagnosis that the growing body of evidence suggests to be unreliable or invalid?

As previously discussed by Fryer,⁴⁷ various approaches can be used to address this issue in the OMT lecture hall or laboratory; one can ignore the challenge, hoping it is not brought up by inquisitive students, or address it directly by incorporating the role of critical thinking into the OMT curriculum.⁴⁷ It is our opinion that allowing students to develop hands-on skills from highly trained instructors while at the same time thinking critically of concepts presented will prove beneficial for the profession. Our review and previous analysis of palpatory methods can serve as a starting point for critical discussion of the palpatory methods found in manipulative medical professions.^{19,47,51} This approach can help cultivate respect for the history of diagnostic approaches in osteopathic medicine while allowing growth and adaptation of current methods and student skills. Meeting the challenges of teaching today's osteopathic medical student with explanations that address current scientific knowledge in the most uniquely osteopathic component of osteopathic medical education can only foster further understanding of and interest in the art and science of osteopathic medicine. Research continues to grow and establish itself in the osteopathic profession as never before, and we in the profession have the opportunity to equip future osteopathic physicians with both a high level of manipulative skill and an appreciation for the balance of clinical and evidence-based knowledge.

Future Directions for Research

Overall improvement in the number of high-quality reliability studies with similar methods is paramount. With the increasing need for an evidence base in medical professions, the objective understanding of palpatory assessment methods is especially important for those professions with an emphasis on manipulative medicine. When evaluating a specific test, studies should follow the IAMMM protocol to allow standardization of experimental methods. Furthermore, the manner in which expert practitioners of manipulative medicine developed a high degree of diagnostic skill must begin to be experimentally established. Another emerging goal is to establish detection thresholds for pelvic landmark positional assessment to aid in training.⁵³ The use of highly controllable models should be explored to understand the degree of accuracy in human perception. The extent of bony anatomic landmark asymmetry that may be normal and asymptomatic needs to be quantified. Through such advancements in knowledge, the profession can strengthen the scientific base needed to better explain the diagnosis of somatic dysfunction.

Conclusion

The key findings of the present review article are presented in Figure 2. The quality and methods of studies investigating the reliability of assessments for bony anatomic landmark asymmetry are variable. Further research is needed to understand the reliability of evaluation methods in manipulative medicine. Assessment of bony anatomic landmark asymmetry is only one component of the diagnostic paradigm in manipulative medicine, but an improved understanding of its reliability would be of substantial value to the osteopathic medical profession, as well as other professions in manipulative medicine.

Acknowledgments

This work was supported by a fellowship grant from the American Osteopathic Association and also by an award from the National Center for Complementary and Alternative Medicine (Award No. 5T35AT004388-02).

References

1. Isaacs, ER.; Bookhout, MR.; Bourdillon, JF. Bourdillon's Spinal Manipulation. 6th ed.. Woburn, MA: Butterworth-Heinemann; 2002.
2. Greenman, PE. Principles of Manual Medicine. 3rd ed.. Philadelphia, PA: Lippincott Williams & Wilkins; 2003.
3. Bergmann, TF.; Peterson, DH.; Lawrence, DJ. Chiropractic Technique: Principles and Procedures. New York, NY: Churchill Livingstone; 1993. p. 803
4. Mitchell, FL.; Mitchell, PKG. The Muscle Energy Manual. East Lansing, MI: MET Press; 1995.
5. Jones, LH.; Kusunose, RS.; Goering, EK. Strain-Counterstrain. Boise, ID: Jones Strain-CounterStrain; 1995. p. 163
6. Walton, WJ. Textbook of Osteopathic Diagnosis and Technique Procedures. 2nd ed.. St Louis, MO: American Academy of Osteopathy, Distributed by Matthews Book Co; 1972. p. 547
7. DiGiovanna, EL.; Schiowitz, S.; Dowling, DJ. An Osteopathic Approach to Diagnosis and Treatment. 3rd ed.. Philadelphia, PA: Lippincott Williams & Wilkins; 2005.
8. Gibbons, P.; Tehan, P. Manipulation of the Spine, Thorax and Pelvis: An Osteopathic Perspective. 2nd ed.. Edinburgh, Scotland: Churchill Livingstone Elsevier; 2006.
9. Kuchera, WA.; Kuchera, ML. Osteopathic Principles in Practice. 2nd rev ed.. Kirksville, MO: Kirksville College of Osteopathic Medicine; 1992.
10. Gatterman, MI. Foundations of Chiropractic: Subluxation. 2nd ed.. St Louis, MO: Elsevier Mosby; 2005. p. 590
11. Tullberg T, Blomberg S, Branth B, Johnsson R. Manipulation does not alter the position of the sacroiliac joint: a roentgen stereophotogrammetric analysis. *Spine (Phila Pa 1976)*. 1998; 23(10): 1124–1128. [PubMed: 9615363]
12. Smidt GL, McQuade K, Wei SH, Barakatt E. Sacroiliac kinematics for reciprocal straddle positions. *Spine (Phila Pa 1976)*. 1995; 20(9):1047–1054. [PubMed: 7631234]
13. Freburger JK. Using published evidence to guide the examination of the sacroiliac joint region. *Phys Ther*. 2001; 81(5):1135–1143. [PubMed: 11319939]
14. Sturesson B, Uden A, Vleeming A. A radiostereometric analysis of the movements of the sacroiliac joints in the reciprocal straddle position. *Spine (Phila Pa 1976)*. 2000; 25(2):214–217. [PubMed: 10685486]
15. Patjin, J.; Remvig, L. Protocol Formats for Diagnostic Procedures in Manual/Musculoskeletal Medicine. 2nd rev ed.. International Academy of Manual/Musculoskeletal Medicine; 2010. http://www.iammm.net/?download=REPRODUCIBILITYANDVALIDITY2010_part2.pdf
16. Levangie PK. The association between static pelvic asymmetry and low back pain. *Spine (Phila Pa 1976)*. 1999; 24(12):1234–1242. [PubMed: 10382251]
17. Al-Eisa E, Egan D, Deluzio K, Wassersug R. Effects of pelvic skeletal asymmetry on trunk movement: three-dimensional analysis in healthy individuals versus patients with mechanical low back pain. *Spine (Phila Pa 1976)*. 2006; 31(3):E71–E79. [PubMed: 16449891]

18. Al-Eisa E, Egan D, Deluzio K, Wassersug R. Effects of pelvic asymmetry and low back pain on trunk kinematics during sitting: a comparison with standing. *Spine (Phila Pa 1976)*. 2006; 31(5):E135–E143. [PubMed: 16508537]
19. Seffinger, MA. Palpation reliability and validity. In: Chaitow, L.; Chaitow, S., editors. *Palpation and Assessment Skills: Assessment Through Touch*. 3rd ed.. Edinburgh, Scotland: Churchill Livingstone; 2010. p. 330
20. Mangione S, Nieman LZ. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. *JAMA*. 1997; 278(9):717–722. [PubMed: 9286830]
21. Raferty E, Holland W. Examination of the heart: an investigation into variation. *Am J Epidemiol*. 1967; 85(3):438–444. [PubMed: 4225873]
22. Joshua AM, Celermajer DS, Stockler MR. Beauty is in the eye of the examiner: reaching agreement about physical signs and their value. *Intern Med J*. 2005; 35(3):178–187. [PubMed: 15737139]
23. Haas M. The reliability of reliability. *J Manipulative Physiol Ther*. 1991; 14(3):199–208. [PubMed: 2045731]
24. Haas M. Statistical methodology for reliability studies. *J Manipulative Physiol Ther*. 1991; 14(2): 119–132. [PubMed: 2019821]
25. Stochkendahl MJ, Christensen HW, Hartvigsen J, et al. Manual examination of the spine: a systematic critical literature review of reproducibility. *J Manipulative Physiol Ther*. 2006; 29(6): 475–475. [PubMed: 16904495]
26. Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson V, Reinsch S. Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. *Spine (Phila Pa 1976)*. 2004; 29(19):E413–E425. [PubMed: 15454722]
27. Szadek KM, van der Wurff P, van Tulder MW, Zuurmond WW, Perez RSGM. Diagnostic validity of criteria for sacroiliac joint pain: a systematic review. *J Pain*. 2009; 10(4):354–368. [PubMed: 19101212]
28. Hestock L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. *J Manipulative Physiol Ther*. 2000; 23(4):258–275. [PubMed: 10820299]
29. van der Wurff P, Hagmeijer RHM, Meyne W. Clinical tests of the sacroiliac joint: a systematic methodological review, I: reliability. *Man Ther*. 2000; 5(1):30–36. [PubMed: 10688957]
30. Haneline MT, Young M. A review of intraexaminer and interexaminer reliability of static spinal palpation: a literature synthesis. *J Manipulative Physiol Ther*. 2009; 32(5):379–386. [PubMed: 19539121]
31. Kmita A, Lucas NP. Reliability of physical examination to assess asymmetry of anatomical landmarks indicative of pelvic somatic dysfunction in subjects with and without low back pain. *Int J Osteopath Med*. 2008; 11(1):16–25.
32. Holmgren U, Waling K. Inter-examiner reliability of four static palpation tests used for assessing pelvic dysfunction. *Man Ther*. 2008; 13(1):50–56. [PubMed: 17210266]
33. Tong HC, Heyman OG, Lado DA, Isser MM. Interexaminer reliability of three methods of combining test results to determine side of sacral restriction, sacral base position, and innominate bone position. *J Am Osteopath Assoc*. 2006; 106(8):464–468. [PubMed: 16943516]
34. Fryer G, McPherson HC, O’Keefe P. The effect of training on the inter-examiner and intra-examiner reliability of the seated flexion test and assessment of pelvic anatomical landmarks with palpation. *Int J Osteopath Med*. 2005; 8(4):131–138.
35. Degenhardt BF, Snider KT, Snider EJ, Johnson JC. Interobserver reliability of osteopathic palpatory diagnostic tests of the lumbar spine: improvements from consensus training. *J Am Osteopath Assoc*. 2005; 105(10):465–473. [PubMed: 16314679]
36. Spring F, Gibbons P, Tehan P. Intra-examiner and inter-examiner reliability of a positional diagnostic screen for the lumbar spine. *J Osteopath Med*. 2001; 4(2):47–55.
37. O’Haire C, Gibbons P. Inter-examiner and intra-examiner agreement for assessing sacroiliac anatomical landmarks using palpation and observation: pilot study. *Man Ther*. 2000; 5(1):13–20. [PubMed: 10688955]

38. Paydar D, Thiel H, Gemmell H. Intra- and interexaminer reliability of certain pelvic palpatory procedures and the sitting flexion test for sacroiliac joint mobility and dysfunction. *J Neuromusculoskel Syst.* 1994; 2(2):65–69.
39. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. *Phys Ther.* 1985; 65(11):1671–1675. [PubMed: 2932746]
40. Jordan TR. Conceptual and treatment models in osteopathy. II. Sacroiliac mechanics revisited. *Am Acad Osteopath J.* 2006; 16(2):11–17.
41. McGrath MC. Palpation of the sacroiliac joint: an anatomical and sensory challenge. *Int J Osteopath Med.* 2006; 9(3):103–107.
42. Mitchell, FL.; Moran, PS.; Pruzzo, NA. *An Evaluation and Treatment of Osteopathic Muscle Energy Procedures.* Valley Park, MO: Mitchell, Moran and Pruzzo Associates; 1979.
43. Mitchell FL. Structural pelvic function. *AAO Yearbook.* 1958:71–90.
44. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003; 3:25. [PubMed: 14606960]
45. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Radiol.* 2003; 58(8):575–580. [PubMed: 12887949]
46. Fryer G. Factors affecting the intra-examiner and inter-examiner reliability of palpation for supine medial malleoli asymmetry. *Int J Osteopath Med.* 2006; 9(2):58–65.
47. Fryer G. Teaching critical thinking in osteopathy: integrating craft knowledge and evidence-informed approaches. *Int J Osteopath Med.* 2008; 11(2):56–61.
48. Fryer G, Morris T, Gibbons P. Paraspinal muscles and intervertebral dysfunction: part two. *J Manipulative Physiol Ther.* 2004; 27(5):348–357. [PubMed: 15195042]
49. Gibbons P, Tehan P. Patient positioning and spinal locking for lumbar spine rotation manipulation. *Man Ther.* 2001; 6(3):130–138. [PubMed: 11527452]
50. Fryer G, Morris T, Gibbons P. Paraspinal muscles and intervertebral dysfunction: part one. *J Manipulative Physiol Ther.* 2004; 27(4):267–274. [PubMed: 15148466]
51. Fryer G. Muscle energy concepts: a need for change. *J Osteopath Med.* 2000; 3(2):54–59.
52. Gibbons P, Tehan P. Muscle energy concepts and coupled motion of the spine. *Man Ther.* 1998; 3(2):95–101.
53. Fossum C, Snider E, Fryer G, Gillette N, Degenhardt B. The introduction of a novel approach to the teaching and assessment of osteopathic manipulative medicine assessment skills. *Int J Osteopath Med.* 2008; 11(4):165–165.

1. Order of observers conducting the test(s) randomized (1 point)
2. Case mix (1 point)
 - Both symptomatic and asymptomatic subjects. For studies with an intentional case mix, the case mix must represent a natural clinic population, that is, a population consisting of a variety of patients with regard to site, age, and problems in different part of the spine. If the purpose is to look at subgroups, such as asymptomatic patients only, credit will be given.
3. Observers blind to other observers' findings (1 point)
4. Observers blind to confounding information (1 point)
5. Subjects blind to observer findings (1 point for true or complete blinding, 0.5 point for no blinding but with subjects unable to affect results)
6. Kappa statistics (κ , κ_{11} , κ_{22} , κ_{33}) or ICC (1-way ICC, 2-way ICC, or generalizability coefficients for single rating by each observer) used for analysis (1 point)

Figure 1. Six-point system for assessing the method quality of reliability studies. Criteria 2, 4, 5, and 6 apply to intraexaminer comparisons. The 6-point assessment of each study appears in the Table.

Abbreviation: ICC, interclass correlation coefficient.²⁵

- Assessment of bony anatomic landmark asymmetry has not been demonstrated to be a reliable method of palpatory assessment.
- Steps should be taken in the osteopathic medical educational setting to address the known low reliability of such assessment and to foster critical thinking on the subject.
- Future research should focus on improving the quality of reliability studies and investigate reasons for this low reliability through increasing the understanding of factors involved in human perception.

Figure 2.
Key findings of the present review article.

Table
Inter- and Intraexaminer Reliability of Bony Anatomic Landmark Asymmetry Assessment

Study	Methods	No. of Participants and Examiners	Landmarks Evaluated, Reliability (κ Coefficient)*	Quality Score (%) With Breakdown†	Author Conclusions
Kmita and Lucas, 2008 ³¹	Double-blind assessment performed twice	5 symptomatic and 4 asymptomatic patients; 4 examiners (2 clinicians, 2 students)	PSIS, 0.04/0.13 SS, -0.40/0.283 ILA-S/L, -0.01/0.058 ILA-A/P, -0.03/0.095 ASIS, 0.13/0.403	6 (100) 1, 1, 1, 1, 1, 1	Alternatives to static asymmetry assessment are recommended for assessment of low back pain and/or pelvic dysfunction.
Holmgren and Waling, 2008 ³²	Independent examination performed once; only interexaminer reliability was assessed	25 symptomatic patients; 2 examiners (experienced clinicians)	L5 transverse processes, 0.17 SS, 0.11 ILA-A/P, 0.11	3.5 (58) 0, 1, 1, 0, 0.5, 1	Interexaminer reliability observed was only slightly better than expected by chance; low interexaminer reliability was attributed to differences in palpation technique.
Tong et al., 2006 ³³	2 rounds of evaluation; 3 methods for analyzing results; only interexaminer reliability was assessed	24 symptomatic patients; 2 examiners (training level unknown)	SS [‡] in trunk flexion, 0.37 SS [‡] in trunk extension, 0.05 ASIS, 0.15	3.5 (58) 0, 1, 1, 0, 0.5, 1	Maximum interexaminer reliability occurs when the most reliable test is used to evaluate SIJ dysfunction; this method is suggested in clinical decision making.
Fryer et al., 2005 ³⁴	Trained group of examiners had 2 1-h training sessions; each landmark examined 3 times	10 asymptomatic patients; 2 groups of 5 examiners (trained and untrained fifth-year students)	Untrained: PSIS, 0.15/0.49 ILA-S/L, -0.01/0.03 ILA-A/P, -0.01/0.2 ASIS, -0.01/0.19 Trained: PSIS, 0.08/0.54 ILA-S/L, 0.04/0.2 ILA-A/P, 0.04/0.07 ASIS, 0.24/0.65	6 (100) 1, 1, 1, 1, 1, 1	Osteopathic physicians should reconsider these tests in evaluation of the SIJ. Training inconclusively improved assessment of anatomic landmark asymmetry; an improved understanding of these evaluation procedures is recommended.
Degenhardt et al., 2005 ³⁵	3 phases of experiment: phase 1, multiple tests; phase 2, consensus training over 4 mo for most reliable tests from phase 1; and phase 3, examinations with trained assessments	42 symptomatic patients evaluated before training, 77 after training; 3 examiners (trained in manual medicine)	L1-L4 transverse processes, [§] 0.17 (untrained) and 0.34 (trained)	5 (83) 0, 1, 1, 1, 1, 1	Consensus training can significantly improve interexaminer agreement for palpatory examinations.
Spring et al., 2001 ³⁶	Fifth-year students; 3-part positional screen in neutral,	10 asymptomatic patients; 10 examiners (fifth-	L4, 0.04/0.037	6 (100) 1, 1, 1, 1, 1, 1	No significant agreement above chance was found for inter- or intraexaminer

Study	Methods	No. of Participants and Examiners	Landmarks Evaluated, Reliability (κ Coefficient) *	Quality Score (%) With Breakdown †	Author Conclusions
	hyperflexed, and extended positions; 1 h of training before examination; total of 3 examinations	year students			reliability. Poor reliability may be attributed to anatomy of lumbar spine; caution is suggested in using static asymmetry for lumbar spine assessment.
O'Haire and Gibbons, 2000 ³⁷	4 assessments per examiner; 1-h training session to standardize methods	10 asymptomatic patients; 10 examiners (fifth-year students)	PSIS, 0.04/0.326 SS, 0.07/0.24 ILA-S/I, 0.08/0.211	6 (100) 1,1,1,1,1,1,1	Further studies are needed to better understand the low reliability of anatomic landmark assessment of the SIJ.
Paydar et al, 1994 ^{38//}	Standing and sitting landmarks assessed; 2 evaluations with second 3 h after the first	32 asymptomatic patients; 2 examiners (student interns with ≥ 1 year of clinical experience)	PSIS, 0.150/0.248	3 (50) 1, 1, 0, 0, 1	Palpatory findings should not be the primary factor in clinical decision making; the patient's response to the treatment is probably the only indication that the diagnosis was correct.
Potter and Rothstein, 1985 ^{39//}	Clinicians; 13 common tests assessed	17 symptomatic patients; 8 examiners (clinicians)	Standing PSIS, § 35.29% agreement sitting PSIS, § 35.29% agreement standing ASIS, § 37.50% agreement; χ^2 value calculated for goodness of fit with 90% and 70% agreement expected	2 (33) 0, 1, 1, 0, 0, 0	The poor reliability observed suggests that new operational definitions for SIJ evaluation are needed; given that clinicians in the same profession evaluated the patients, this study raises issues of continuity of care.

* Except where otherwise explained, values represent interexaminer reliability (single κ values) or interexaminer/intraexaminer reliability. The term *clinicians* may include osteopathic physicians, osteopaths, chiropractors, and others.

† Breakdown of quality score is listed in consecutive order, as found in the *Figure*.

‡ Patients evaluated in seated flexed and sphinx positions.

§ Other assessment methods were used during the study.

// Patients were seated.

Abbreviations: ASIS, anterior superior iliac spine; ILA-S/I, inferior lateral angle of sacrum, superior/inferior assessment; ILA-A/P, inferior lateral angle of sacrum, anterior/posterior assessment; PSIS, posterior superior iliac spine; SIJ, sacroiliac joint; SS, sacral sulcus.